المجلة التربوية

مجلة فصلية، تخصصية، محكّمة تعنى بدعم الفكر التربوي ودراسة
المشكلات التربوية واقتراح الحلول لها وتوفير مجال
النشر العلمي المحكّم للبحوث التربوية

توجه جميع المراسلات باسم رئيس التحرير على العنوان التالي:

المجلة التربوية – مجلس النشر العلمي
ص. ب: ١٣٤١١ كيفان – الرمز البريدي 71955 الكويت
هاتف: ٤٨٤٧٩٦١(٩٦٥) – ٤٩٨٤٤٠٩ – ٤٩٨٤٤٠٣
فاكس: ٤٨٣٧٧٩٤
http: www.pubcouncil.kuniv.edu.kw/joe
E-mail: TEJ@kuc01.kuniv.edu.kw.

جميع الآراء الواردة في المجلة تعبر عن وجهة نظر كاتبيها ولا تعبر بالضرورة عن وجهة نظر المجلة.

Issn: 1029-810

المجلة التربوية

ملحق العدد الثامن والسبعون - المجلد العشرون - مارس ٢٠٠٦م

رئيس التحرير

أ. د. صالح عبدالله جاسم

مديرة التحرير

أ. عبير حسن الكندري

هيئة التحرير

أ. د. كافيه رمضان

أ. د. وصفي عزيز بولس

د. بدر عمر العمر

د. معدي مهدي العجمي

د. علي جاسم الشهاب

حجم أثر المعالجات التجريبية ودلالة الدلالة الإحصائية

د. رجاء محمود أبو علام
أستاذ بمعهد الدراسات التربوية
جامعة القاهرة

الملخص

يتناول هذا الكتيب موضوعاً من أهم الموضوعات المعاصرة المرتبطة بتحليل البيانات الإحصائية للبحوث. هذا الموضوع هو حجم أثر المعالجات التجريبية. وحجم الأثر أصبح الآن جزءاً لا يتجزأ من تحليل البيانات حتى إن التحليل الإحصائي الذي لا يستخدم حجم الأثر أساساً لتفسير مستوى الدلالة الإحصائية يعتبر غير مستكمل.

ويتكون هذا الكتيب من ثلاثة فصول، يتناول الفصل الأول منها القضية الجدلية في علم النفس حول الاختبارات الإحصائية، والأحداث التي قادت إليها. وبخاصة فيما يتعلق بالتوجهات التي اتخذتها رابطة علم النفس الأمريكية وضمنتها الطبعة الخامسة من دليل النشر فيما يتعلق بالاختبار الإحصائي للفرض الصفري. واهتم هذا الفصل بإبراز مشكلة الدلالة الإحصائية وأثرها في إعاقة تطور علم النفس والبحوث النفسية، وبين أن المشكلة لا تتعلق بالدلالة الإحصائية في حد ذاتها، لأننا إذا قمنا بشكل روتيني بتحديد فروض معقولة، ودرسنا عينة عشوائية، وراجعنا المسلمات المتعلقة بالتوزيع العيني، وحددنا القوة الإحصائية، وفهمنا المعنى الحقيقي لقيم ل (مستوى الدلالة)، لا تكون هناك أية مشكلة في استخدام اختبارات الدلالة الإحصائية كوسيلة في الاستدلال الإحصائي.

والحل الأفضل حتى الآن لهذه المشكلة هو إضافة حجم الأثر إلى نتائج البحوث التجريبية. وقدم الفصل الأول عدة تعريفات لحجم الأثر تقوم على طرق حساب حجم الأثر. وأول معادلة معروفة لحساب حجم الأثر هي معادلة جلاس التي تقوم على قسمة

الفرق بين متوسطي المجموعة التجريبية والمجموعة الضابطة على الانحراف المعياري للمجموعة الضابطة على اعتبار أنها لم تتأثر بالمعالجة التجريبية.

وتناول الفصل الثاني حجم الأثر في الإحصاء المعلمي فناقش أحجام الأثر المتعلقة بالفروق بين المتوسطات. وأهم العمليات الإحصائية التي تناولها هذا الفصل هي معاملات الارتباط، واختبار "ت" للفروق بين المتوسطات، وتحليل التباين البسيط، وتحليل التباين العاملي. وتبين من تلك الأساليب الإحصائية أن هناك طرقاً متعددة لتقدير حجم الأثر، قد يتعارض بعضها مع البعض الآخر، ولذلك يجب أن يذكر الباحث في تقريره أي طرق استخدمها لتقدير حجم الأثر. وذلك لمساعدة القراء على عدم المزج بين طرق متعارضة لتقدير حجم الأثر.

ويمكن أن تختلف أحجام الأثر باختلاف التباين في أفراد العينة من دراسة لأخرى، لأن نوع المجتمع الذي ندرسه أو طبيعة العينة التي نختارها وما يرتبط بها من تباين أثر كبير على حجم الأثر. ولذلك يجب أن يكون الباحث حريصاً عند مقارنة أحجام الأثر التي يحصل عليها من دراسات مختلفة، لأن اختلاف التباين المحتمل بين أفرادها يمكن أن يكون السبب فيما قد يلاحظ من اختلاف بين أحجام الأثر في الدراسات المختلفة.

وتناول الفصل الثالث حجم الأثر في الإحصاء اللامعلمي والمتغيرات القطعية. وتقيس بعض مؤشرات حجم الأثر الاحتمالات النسبية عبر مجموعتين مختلفتين بالنسبة لنتيجة غير مرغوبة في مقابل نتيجة أخرى مرغوبة، وهو ما يطلق عليه المجازفة النسبية. ومؤشر المجازفة النسبية الذي يتميز بالخاصية السيكومترية الأفضل هو نسبة الأرجحية.

مقدمة

أصبح الاهتمام كبيراً بين الباحثين بذكر حجم الأثر* إلى جانب ذكرهم للدلالة الإحصائية، كما أن عدداً متزايداً من المجلات العلمية العالمية يتطلب ضمن شروط النشر أن يذكر الباحثون تقدير حجم الأثر ضمن نتائجهم.

* يستخدم الكثيرون مصطلح «حجم التأثير» بدلاً من «حجم الأثر»، ورغم أن هناك تقارباً كبيرا بين المصطلحين إلا أن الكاتب يرى أن الباحثين عادة ما يستخدمون مسطلح «أثر المتغير المستقل على المتغير التابع» في دراساتهم التجريبية، لأنهم يريدون معرفة نتيجة (أثر) معالجة المتغير المستقل. ولذلك فمن المنطقي، وحتى نستخدم نفس المصطلح في مناقشة نتائج البحوث التجريبية، استخدام مصطلح «حجم الأثر» الذي يشير إلى حجم أثر المعالجة التجريبية، أي أثر المتغير المستقل على المتغير التابع. كما أن استخدام مصطلح "أثر" ترجمة بمصطلح "effect" ومصطلح «تأثير» ترجمة لمصطلح "impact" بجعل التمييز بين المصطلحين سهلاً وواضحاً.

وبالنسبة لأنواع معينة من البحوث التطبيقية لم يعد مقبولاً أن يذكر الباحث فقط الدلالة الإحصائية. فالنتائج الدالة إحصائياً تشير إلى أن الباحث قد اكتشف أدلة على وجود فروق حقيقية بين معالم المجتمع، أو ارتباطاً حقيقياً بين المتغيرات، إلا أن حجم هذه المعالم أو حجم الارتباط غير معروف. وكثيراً ما تحتاج العبارات المتعلقة بالدلالة الإحصائية إلى إضافات تعطي تقديرات عن مدى الفروق بين متوسطات المجموعات التي يدرسونها، أو عن قوة الارتباط بين المتغيرات. وبالإضافة إلى هذا، كثيراً ما يحتاج أولئك الذين يطبقون نتائج البحوث إلى معرفة المزيد عن مدى الاختلاف بين طريقتي تدريس، أو أسلوبين في العلاج، والدرجة التي يمكن بها تفضيل مجموعة على مجموعة أخرى. وبمعنى آخر فإنهم يحتاجون إلى معرفة حجم أثر المعالجة التجريبية التي أعطت نتائج معينة.

وسوف نحاول في هذه الدراسة أن نعرف المقصود بحجم الأثر، ومناقشة الظروف التي يكون فيها لحجم الأثر أهمية خاصة بالنسبة للنتائج، وتحديد لماذا نحتاج أنواعاً مختلفة من حجم الأثر، هذا بالإضافة إلى ذكر العوامل التي دعت إلى استخدام حجم الأثر كطريقة مكملة للدلالة الإحصائية.

وتعكس المناقشة الجادة لتحليل نتائج البحوث سخطاً كبيراً على الاختبارات الإحصائية استمر عدداً كبيراً من السنوات. والواقع أن هذا الجدل قد ازداد من عقد لآخر ودار في مجالات متنوعة بالإضافة إلى علم النفس والتربية. وكاد يصل الأمر في أواخر التسعينيات من القرن الماضي إلى حدود بالغة الشدة. وبدأ عدد من المجلات يطلب إضافة حجم الأثر إلى اختبارات الدلالة الإحصائية. وكان الجدل حول استخدام الاختبارات الإحصائية القضية المنهجية الأولى في تلك الفترة، حتى إن البعض اعتقد أن التغيير قادم لا محالة (Kaufman, 1998).

وتهدف هذه الدراسة إلى مساعدة القراء في العالم العربي في الاطلاع على التغيرات الجارية في طرق تحليل البيانات وإعداد تقارير البحث في العلوم

السلوكية. وهؤلاء القراء يمكن أن يكونوا تربويين أو باحثين أو محكمين للبحوث المقدمة إلى المجلات العلمية أو طلاب الدراسات العليا في علم النفس والتربية. ويفترض الكاتب أن كثيراً من القراء قد دربوا على استخدام الاختبارات الإحصائية باعتبارها الطريقة الرئيسية وربما الوحيدة لاختبار الفروض. ولا شك في أن الكثيرين قد تساءلوا عما إذا كانت هناك طرق لتقويم الفروض، أو أنه لا بد من استخدام نتائج الاختبارات الإحصائية في تقارير البحوث (Kaufman, 1998).

وليس الغرض من هذه الدراسة الدخول في جدول حول جدوى تعديل الطريقة المستخدمة في الاختبارات الإحصائية، فالمعروف أن دور الاختبارات الإحصائية آخذ في التضاؤل. ولذلك فإن الهدف من هذه الدراسة مساعدة القراء على فهم:

أ - الجدل الدائر حول نواحي القصور في الاختبارات الإحصائية.

ب - نواحي القوة والضعف في بدائل الاختبارات الإحصائية.

ج - الطرق الأخرى المرتبطة بالإقلال من الاختبارات الإحصائية.

ويرتبط بهذه العناصر في المقام الأول تقدير حجم الأثر الذي تطلب كثير من المجلات العلمية استخدامه كما سبق أن ذكرنا، كما يرتبط بها تقدير حدود الثقة بناء على القيم التي نحصل عليها في حساب حجم الأثر.

وتتكون هذه الدراسة من ثلاثة فصول، يتناول الفصل الأول منها المفاهيم الأساسية المرتبطة بحجم الأثر كوسيلة مكملة لاختبارات الدلالة الإحصائية، كما يتناول المشكلات التي نجمت عن الإفراط في استخدام الاختبارات الإحصائية للفرض الصفري وكيفية معالجة هذا الأمر. ويتناول الفصل الثاني حجم الأثر في الإحصاء المعلمي، أما الفصل الثالث والأخير فيتناول حجم الأثر في الإحصاء اللامعلمي.

الفصل الأول
الفرض الصفري واختبار الدلالة الإحصائية

كان استخدام حجم الأثر كتقدير مكمل لاختبار الدلالة الإحصائية رد فعل لقضية أثارت الجدل حول الدلالة الإحصائية واختبار الفرض الصفري، حتى إن حجم الأثر ربما كان حلاً وسطاً لحسم هذه القضية التي استمرت أكثر من ثلاثة عقود متوالية في أواخر القرن العشرين، واحتدمت في العقد الأخير منه. وسنحاول في هذا الفصل مناقشة هذه القضية وكيف انتهى الأمر بها إلى استخدام حجم الأثر.

مخاطر اختبار الفروض:

تستخدم الغالبية العظمى من الباحثين في التربية وعلم النفس، بل وفي غالبية العلوم الأخرى، اختبار الفرض الصفري كأساس لاختبار الفروض المتعلقة بالمعالجات التجريبية، إلا أن هذا الأسلوب رغم قوته ليس خالياً من العيوب، فهو لا يضمن الوصول إلى نتائج دقيقة وصادقة بالنسبة للعلاقات بين المتغيرات. وتحدث كثير من الأخطاء الشائعة بين الباحثين لدرجة أنه في عام ١٩٩٦ قام فريق من المنهجيين النفسانيين بجهود قوية لإقناع رابطة علم النفس الأمريكية (APA) بمنع استخدام اختبار الفرض الصفري في البحوث المقدمة لمجلاتها (Shrout, 1997; Furlong, Lovelace, & Lovelace, 2000). إلا أن هذا اتجاه خطير وليس علاجاً للمشكلات التي تواجه عملية اختبار الفروض، وليس من المحتمل تطبيقها، ومع ذلك فقد أسهمت هذه القضية الجدلية في توجيه صيحة لإيقاظ الباحثين لتذكرهم بأن أية وسيلة لا تكون فعالة إلا إذا استخدمت بشكل سليم. أي أن اختبار الفرض الصفري لا يكون صادقاً إلا إذا كان تفسير النتائج صحيحاً.

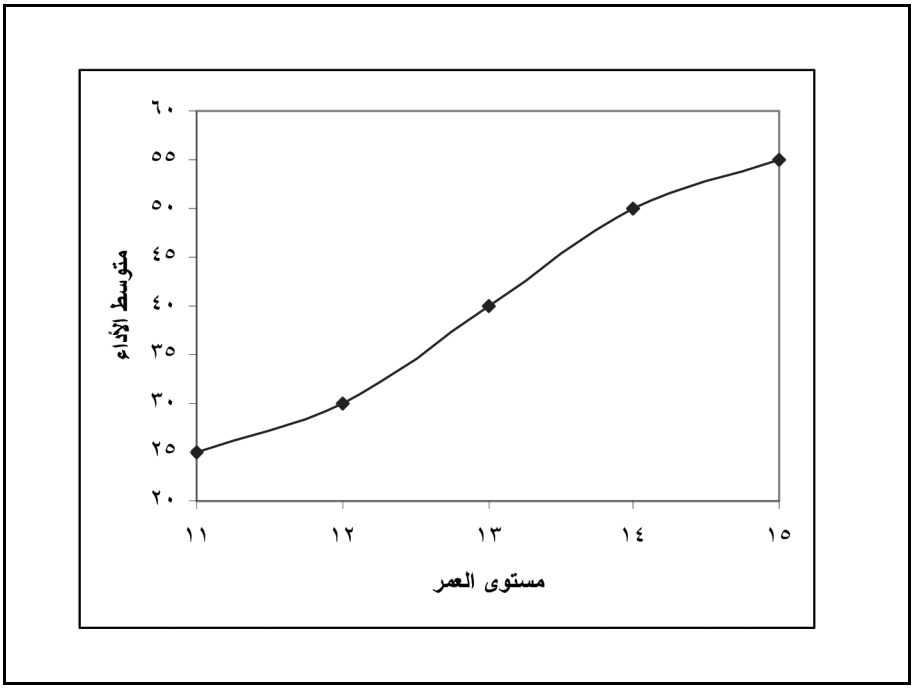
وفيما يلي مثال يسوقه روزنتال وزملاؤه حول الدلالة الإحصائية وكيف يمكن لها عدم إظهار الفروق بين المتوسطات (Rosenthal, Rosnow, & Rubin, 2000). لنفرض أن باحثاً مهتماً بدراسة نمو المهارات الحركية جعل ما مجموعه ٥٠ طفلاً

في خمسة مستويات عمرية هي (١١ و١٢ و١٣ و١٤ و١٥) يلعبون لعبة من ألعاب الفيديو. وقد وجه الباحث سؤالاً محدداً هو ما إذا كان العمر عاملاً محدداً في المهارة في هذه اللعبة. وكان متوسط أداء عشرة أطفال في الأعمار الخمسة المتتالية هو ٢٥، ٣٠، ٤٠، ٥٠ ،٥٥. وقد رسمت هذه القيم في الشكل (١-١)، كما أن الجدول (١-١) يعطي نتائج تحليل التباين لهذه الدرجات الخمس.

جدول (١-١)

تحليل التباين لمتوسط الأداء العام للأطفال

مصدر التباين	مجموع المربعات	د. ح.	متوسط المربعات	ف	ل
مستويات العمر	٦٥٠٠	٤	١٦٢٥	١,٠٣	٠,٤٠
داخل المجموعات	٧٠٨٧٥	٤٥	١٥٧٥		



شكل (١-١) العلاقة بين مستوى العمر ومتوسطات الأداء

هل يستطيع الباحث إذا حصل على قيمة ل = ٠,٤٠ لتحليل التباين أن يقبل الفرض الصفري بأن العمر ليس عاملا فعالاً في التنبؤ بالأداء؟ إذا اتخذ الباحث مثل هذا القرار فإنه يتجاهل ما هو واضح للعين المجردة من الخط البياني الموجود في شكل (١-١). فهذا الشكل يبين بوضوح أن العمر يزيد باستمرار مع زيادة درجة الكفاءة في الأداء من أقل درجة إلى أعلى درجة. والواقع أن معامل ارتباط بيرسون بين العمر ومتوسطات الأداء الخمسة يبلغ ٠,٩٩. ويطلق على معامل الارتباط بين المؤشر «مستوى المعالجة» (العمر) ومتوسط الدرجات (متوسط الأداء) ارتباط منذر (Alerting correlation)، لأنه ينذرنا بمستوى الأداء الكلي. وهذا الأداء الكلي تقدير ضعيف بين كل عمر من أعمار الأطفال ودرجات الأداء لأن معاملات الارتباط المحسوبة على البيانات الزمنية المجمعة (متوسطات المجموعات) يمكن أن تكون أكبر (أو أصغر) بشكل بالغ الشدة من الارتباطات المحسوبة على الدرجات الفردية. وفي البحوث السلوكية عادة ما تكون الارتباطات المنذرة أكبر بكثير من الارتباطات القائمة على الدرجات الفردية. ومع ذلك فإن هذا الارتباط ينذرنا بأنه رغم أن قيمة ف غير دالة إحصائياً عند المستوى التقليدي وهو ٠,٠٥ فإننا لا نستطيع أن نغفل الحقيقة بأن العمر منبئ فعال بالأداء. إذ يلاحظ أن النقاط التي تحدد متوسطات المجموعات تكاد تقع على الخط المستقيم الذي يحدد انحدار الدرجات على المتوسطات (ص = ٨ س – ٦٤). ومن الواضح طبقاً لهذا الشكل أنه يمكن التنبؤ بالأداء من العمر.

ويرجع استخدام ما هو معروف اليوم باختبار دلالة الفرض الصفري Null hypothesis significance testing والذي سوف نشير إليه اختصاراً بالاختبارات الإحصائية إلى عام ١٧٠٠، فقد ظهرت منذ ذلك الوقت بعض الدراسات العلمية التي استخدمت الاختبارات الإحصائية. إلا أن هذا الاستخدام لم يكن منظماً حتى عام ١٩٠٠. والواقع أن الاختبارات الإحصائية كما نستخدمها اليوم كانت مزيجاً من مدرستين فكريتين، إحداهما ظهرت في العشرينيات من القرن العشرين على يد فيشر، والأخرى ظهرت في الثلاثينيات

على يد نيمان وبيرسون ويعتبر النموذج الأخير امتداداً لنموذج فيشر، وقد تكوَّن هذا النموذج من الفرض الصفري وأطلق على الاحتمالات الناتجة عن الاختبارات الإحصائية قيم «ل» (p-values)، ولم يتضمن هذا النموذج فرضاً بديلاً للفرض الصفري. ويعزى إلى فيشر استخدام المستويين التقليديين للدلالة الإحصائية وهما ٠,٠٥ و٠,٠١ إلا أنه لم يطالب باستخدام هاتين القيمتين في جميع الدراسات. ونظراً لتركيز طريقة فيشر على قيم «ل» في ظل الفرض الصفري، أطلق عليه طريقة «قيمة ل».

ويرجع إلى نموذج نيمان وبيرسون إضافة الفرض البديل إلى نموذج فيشر الأساسي، وتحديد منطقتي رفض الفرض الصفري (المنطقة الحرجة)، باستخدام ذيل واحد أو ذيلين، وتطبيق مستويات ثابتة لألفا في جميع الدراسات. وكانت هذه الخاصية الأخيرة هي السبب غالباً في التطبيق الجامد لمستويي الدلالة ٠,٠٥ و٠,٠١ واللتان أصبحتا الممارسة الفعلية اليوم للدلالة الإحصائية. ولهذا السبب أطلق على نموذج نيمان وبيرسون «طريقة ل الثابتة» (Tht fixed-p approach). كما أدى تطبيق هذا النموذج أيضاً إلى إضافة المفاهيم المتعلقة بالقوة الإحصائية وما يرتبط بها من الخطأ من النوع الأول والخطأ من النوع الثاني، وهذه كلها مفاهيم أتت مع نموذج نيمان وبيرسون وليس نموذج فيشر.

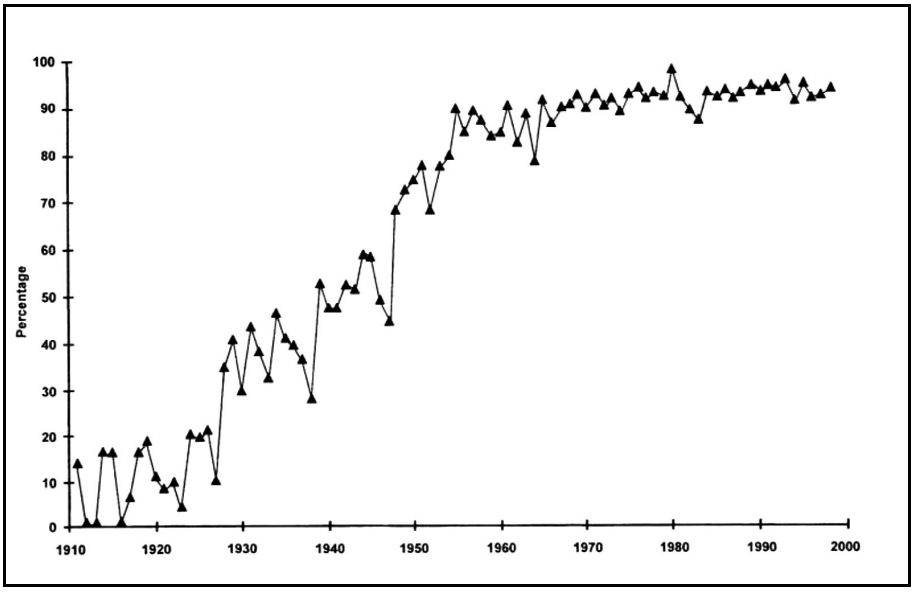
وكان استخدام الاختبارات الإحصائية في البحوث المنشورة في مجلات علم النفس محدوداً قبل عام ١٩٤٠، وكان التحليل الإحصائي في تلك الفترة قاصراً على بعض أساليب الإحصاء الوصفي وبعض جوانب محدودة من الاختبارات الإحصائية الأولية. إلا أن بين حوالي عامي ١٩٤٠ و١٩٦٠ حدث انقلاب في اساليب التحليل الإحصائي أطلق عليه الانقلاب الاستدلالي في علم النفس (inference revolution) ونتيجة له أدخل نظام (Intro Stats) في التحليل الإحصائي، وأخذ يسود في كتب الإحصاء الدراسية، ومناهج الإحصاء في الجامعات، وطالبت به مجلات علم النفس باعتباره النظام الذي يستخدم في

اختبار الفروض. وهناك عاملان ساعدا على حدوث هذا الانقلاب في علم النفس؛ أولهما التحول من دراسة الحالات الفردية في علم النفس التجريبي كما كان الحال في تجارب التعلم الإجرائي إلى دراسة المجموعات، والثاني ظهور الثورة الاحتمالية (probabilistic revolution) في العلوم والذي قام عليها مذهب اللاحتمية كمفهوم نظري أساسي في الدراسات العلمية الغرض منه زيادة فهم المادة الدراسية. وقد أدى استخدام هذا المذهب في علم النفس إلى تحريك العملية الاستدلالية باستخدام الاختبار الإحصائي للفرض الصفري (Kline, 2004, p. 8).

ولقد حدثت زيادة كبيرة في ذكر نتائج التحليل الإحصائي في المجلات العلمية في علم النفس والمجالات المرتبطة به. وهذا التوجه واضح من شكل (١-٢) الذي يبين ٨٠٠٠ بحث نشرت بين عامي ١٩١١ و١٩٩٨ في ١٢ مجلة مختلفة مختارة عشوائياً من المجلات الصادرة عن رابطة علم النفس الأمريكية (APA). ويلخص الشكل النسبة المئوية للمقالات التي استخدمت الاختبارات الإحصائية ضمن التحليل الإحصائي. وتبلغ هذه النسبة ١٧٪ في الفترة من ١٩١١ إلى ١٩٢٩. وقد زادت هذه النسبة إلى حوالي ٥٠٪ في عام ١٩٤٠، وإلى ٨٥٪ في عام ١٩٦٠ وفاقت ٩٠٪ ابتداء من عام ١٩٧٠. وقد بلغت أقصى زيادة في سرعة استخدام الاختبار الإحصائي للفرض الصفري في الفترة بين عامي ١٩٤٠ و١٩٦٠ وهي الفترة التي تواكب الثورة الاستدلالية في علم النفس. واستمر الجدل حوالي ٧٠ عاماً منذ عام ١٩٤٠ إلى الآن حول الاختبارات الإحصائية.

ويمكن تلخيص أهم جوانب نقد الاختبارات الإحصائية فيما يلي:

١ – هناك سوء فهم واسع الانتشار حول قيم «ل» الناتجة عن الاختبارات الإحصائية، ويشمل سوء الفهم هذا الاعتقاد بأنها تقيس احتمال خطأ المعاينة، وتكرار البحث وصحة الفرض الصفري أو الفرض البديل. وقد لا تكون هذه المعتقدات الزائفة خطأ مستخدمي الاختبارات الإحصائية، لأن الأسس المنطقية التي قام عليها الاختبار الإحصائي للفرض الصفري ليست كلها أسساً متجانسة.



شكل (١-٢) نسبة المقالات التي أعطت نتائج التحليل الإحصائي في ١٢ مجلة من المجلات الصادرة عن رابطة علم النفس الأمريكية (APA) في الفترة بين عامي ١٩١١ و١٩٩٨.
عن:

"The Historical Growth of Statistical Significance Testing in Psychology - And Its Future Prospects" by R. Hubbard and P. A. Ryan, 2000, *Educational and Psychological Measurement*, 60, p. 665.

٢ – أدت المعتقدات الخاطئة عن الاختبارات الإحصائية إلى قيام نوع من التشويش المعرفي في أخَّر تقدم البحث في العلوم السلوكية. وهذا واضح من فشل تنمية تقاليد أقوى لتكرار البحوث مقارنة بما هو حادث في العلوم الطبيعية، وفقدان كثير من البحوث للمصداقية، وضياع الجهد البحثي ومصادره. ولا يقول الناقدون أن الاختبارات الإحصائية هي السبب الوحيد، إلا أن الاستخدام المكثف لها يضخم هذه المشكلة.

٣ – من المحتمل أن قيم ل الناتجة من الاختبارات الإحصائية ليس لها معنى في كثير من البحوث السلوكية (إن لم يكن معظمها)، وبخاصة عند اختبار فروض غير معقولة، أو إذا كانت المسلمات المرتبطة بها غير صحيحة. فإذا كانت قيم «ل» مشكوكاً فيها فكذلك بالتالي تكون القرارات القائمة عليها.

٤ – المعلومات التي توفرها الاختبارات الإحصائية خاصة جداً، مما يجعلها غير قابلة للتعميم الذي يسعى وراءه الباحثون عادة.

٥ – الدلالة الإحصائية لا تذكر شيئاً يتعلق بحجم الأثر مباشرة، كما لا نستطيع الخروج منها بما يشير إلى مضمون نظري أو عملي أو تربوي.

٦ – ورغم جميع المشكلات الناجمة عن الاختبارات الإحصائية فليس لها بديل سحري (Cohen, 1994). فالمقترحات البديلة مثل تقدير حجم الأثر، وتقدير حدود الثقة بالنسبة للدراسات الفردية، واستخدام التحليل البعدي لتجميع نتائج البحوث وتركيبها لها هي الأخرى مشكلاتها الخاصة. ولذلك لا يجب تأييد بدائل الاختبارات الإحصائية بشكل مطلق.

فشل المقترحات الأولى لذكر حجم الأثر (١٩٩٤ إلى الآن):

من الطرق التي نعوض بها بعض أوجه القصور في الاختبارات الإحصائية إعطاء معلومات إضافية نتائج الاختبارات الإحصائية، مثل حجم الأثر. وقد ذكر كيرك (Kirk, 1996) أن فكرة تقدير حجم الأثر ليست فكرة جديدة. فقد استخدمها كارل بيرسون في مطلع القرن العشرين، ويعزى إلى فيشر واحد من أوسع أحجام الأثر استخداماً اليوم وهو مربع إيتا (η^2) والتي يطلق عليها أحياناً نسبة الارتباط (R^2). كما نادى بعض الناقدين المعاصرين للاختبارات الإحصائية بإعطاء حجم الأثر بجانب نتائج الدلالة الإحصائية.

ولقد شجعت الطبعة الرابعة من دليل النشر الصادر عن رابطة علم النفس الأمريكية (APA, 1994) الباحثين للمرة الأولى على إعطاء حجم الأثر إلى جانب نتائج الاختبارات الإحصائية، إلا أنها لم تشترط ذلك. إلا أن هذا التشجيع لم يلق صداه بعد عام ١٩٩٤ في أكثر من ٢٠ مجلة تابعة للرابطة، وظلت المقترحات باستخدام حجم الأثر دون تأثير يذكر. وكان ذلك مستغرباً في ضوء انتشار الحاسبات الآلية الرخيصة نسبياً، والتي وفرت للباحثين مختلف البرامج الإحصائية سهلة الاستخدام والتي لديها إمكانيات هائلة في تحديد شكل النتائج كما يريد المستخدم.

نشأة التحليل البعدي (١٩٧٦ إلى الآن):

أصبح التحليل البعدي منذ إدخاله في أواخر السبعينيات من القرن الماضي على يد جلاس (Glass, 1976) أداة هامة للبحث في كثير من المجالات. والتحليل البعدي في العلوم الاجتماعية في شكله العادي عبارة عن تحديد مقاييس النزعة المركزية والتشتت لحساب حجم الأثر المقنن عبر مجموعة من الدراسات تدرس نفس الظاهرة، مثال ذلك دراسة فاعلية معالجة المجموعة التجريبية في مقابل المجموعة الضابطة (رجاء أبو علام، ١٩٨٩). وهذا التركيز على حجم الأثر وليس اختبارات الدلالة الموجود في الدراسات الفردية شجع قارئ التحليل البعدي على التفكير خارج قيود الدلالة الإحصائية. وهناك أمثلة عديدة تُظهر فيها نتائج التحليل البعدي أن الحصول على خلاصات قائمة على رفض الفرض الصفري والحاصل في الدراسات المفردة كانت خاطئة.

وكانت للزيادة المطردة في استخدام التحليل البعدي أثرها في زيادة التفكير في هذا الأسلوب الإحصائي (Cumming & Finch, 2001, p. 555) والذي يمكن أن نعزو له الخصائص التالية:

١ – أصبح التقدير الدقيق لنتائج الدراسات السابقة شيئاً ضرورياً.

٢ – يجب أن ينظر الباحث إلى دراسته باعتبارها إضافة متواضعة إلى تراث البحوث السابقة.

٣ – يجب على الباحث أن يعد تقريره بحيث يمكن دمج نتائجه في دراسة مقبلة للتحليل البعدي.

٤ – إعادة تفسير النتائج الجديدة بعد جمعها عن طريق مقارنتها مباشرة مع أحجام الأثر السابقة.

ومن المحتمل أن يصبح التفكير البعدي سائداً أكثر مما هو اليوم، والمعروف أن التحليل البعدي لا يستخدم الاختبارات الإحصائية أداة رئيسية له.

توصيات دليل النشر (الطبعة الخامسة):

صدرت الطبعة الخامسة من دليل النشر لرابطة علم النفس الأمريكية (APA) في عام ٢٠٠١ (APA, 2001). وكان من بين توصياتها ما يلي:

١ – من الأفضل استخدام أقل تحليل إحصائي ممكن في البحث.

٢ – عدم إعطاء نتائج من مخرجات الحاسب الآلي دون دراية بمعناها.

٣ – توثيق المسلمات المتعلقة بأحجام الأثر في المجتمع، أو أحجام العينة، أو القياس التي تعتبر أساساً لاختبارات القوة الإحصائية للدراسة. ويجب استخدام حدود الثقة بالنسبة لنتائج الدراسة بدلاً من استخدام الاختبارات التتبعية (post hoc).

٤ – يجب إعطاء حجم الأثر المناسب لبيانات البحث مع النتائج الأولية أو عند استخدام قيم «ل». وهذا يحسن من البحث كما أنه يسهل القيام بالتحليل البعدي. وعدم ذكر حجم الأثر يعتبر عيباً في البحث.

٥ – التوصية بشدة باستخدام حدود الثقة لقيم حجم الأثر التي نحسبها.

٦ – إعطاء تأكيدات بدرجة معقولة أن البيانات تحقق المسلمات الإحصائية.

٧ – إعطاء نتائج كافية للتحليل الوصفي مثل المتوسط والتباين وحجم كل مجموعة مستخدمة والتباين والتغاير المجمع داخل المجموعات في الدراسات المقارنة، ومصفوفة الارتباط في دراسات تحليل الانحدار. وهذه المعلومات ضرورية في حالة القيام بدراسة للتحليل البعيد فيما بعد، أو رغبة باحثين آخرين إجراء بحوث ثانوية باستخدام بيانات البحث.

إلا أن رابطة علم النفس الأمريكية لم توص بمنع الاختبارات الإحصائية في مجلات علم النفس، ربما لأن مثل هذا المنع إجراء متطرف ليس هذا أوان إثارته. ورغم أن دليل النشر يعترف بالجدل الدائر حول الاختبارات الإحصائية، إلا أنه لا يرى أن مهمته أن يحل هذه المشكلة، بل يطالب في أن يستمر إعطاء

تقرير كامل لنتائج الاختبارات الإحصائية، والتي يجب أن تتضمن قيمة الاختبار الإحصائي، ودرجات الحرية، أو مستوى ألفا في جميع الاختبارات المستخدمة مثل ل < ٠٥, أو القيمة الفعلية لألفا من نتائج الحاسب الآلي، مثل ل = ٠,٠١٢.

ولعل أفضل الطرق لتعديل الأفكار السائدة حالياً حول الاختبارات الإحصائية تعديل مقررات الإحصاء في الجامعات، بحيث يضاف إليها مفاهيم جديدة تعدل الأخطاء المنتشرة بالنسبة للاختبارات الإحصائية. ومن هذه المفاهيم ما يلي:

١ – اختبار دلالة الفرض الصفري ليس هو الطريقة الوحيدة لتحليل البيانات، وليس بالضرورة أفضل طريقة لاختبار الفروض.

٢ – لا تعتبر النتيجة التي نحصل عليها هامة لمجرد أنها دالة إحصائياً.

٣ – إذا زدنا حجم العينة بشكل كاف فسوف تصبح كل نتيجة دالة إحصائياً غالباً. وهذا أمر مخيف.

٤ – الإقلال من أهمية البحوث التي تضع أهمية كبرى للدلالة الإحصائية.

٥ – لا يعني الحصول على دلالة إحصائية الحصول على دلالة عملية. ولذلك يجب إضافة حجم الأثر والقوة والمتوسط والانحراف المعياري إلى تقارير البحوث.

٦ – يجب أن يكون هناك فهم أفضل للاختبارات الإحصائية في العلوم الاجتماعية والسلوكية حتى يستطيع الباحثون القيام باختيار اختبارات أفضل وأكثر قيمة.

الدلالة الإحصائية في مقابل الدلالة النفسية:

طبقاً لإجراءات اختبار الفرض الصفري فإننا نرفض الفرض إذا كانت إحصاءة العينة (كالفرق بين المتوسطات، أو معامل الارتباط) أكبر أو أصغر مما يمكن توقعه طبقاً لعوامل الصدفة وحدها، ولذلك نستخلص أن هناك فرقاً «دالاً» أو علاقة «دالة» بين المتغيرات. إلا أن هناك خطأ شائعاً هو الخلط بين «الدلالة

الإحصائية» والفائدة العملية للنتائج. فالنتائج الدالة إحصائياً لا تنطوي بالضرورة على قيمة عملية أو نظرية. مثال ذلك أن الدراسات التي تتم على عينات كبيرة، وبالتالي يكون لها قوة كبيرة، كثيراً ما تعطينا نتائج دالة رغم أن هذه النتيجة قد تكون أقل من أن يعتد بها. لنفرض مثلاً أننا أجرينا اختباراً في الميول واختباراً آخر في الذكاء على عينة مكونة من ٢٠٠٠ طالب في الصف الأول الثانوي وحصلنا على معامل ارتباط قدره ٠,٤٤، هذه النتيجة دالة عند مستوى ٠,٠٥، ويمكن على أساسها أن نرفض الفرض الصفري ونعتبر أننا حصلنا على ارتباط دال إحصائياً بين الميول والذكاء. ولكن ما قيمة ارتباط يكاد يكون صفراً، إننا إذا قوَّمنا هذا المعامل من حيث معناه الإحصائي لوجدنا أنه يعني أن التباين المشترك بين المتغيرين (مربع معامل الارتباط) يبلغ حوالي ٠,٠٠٠٢، مما يعني أنه لا يوجد أي تباين مشترك بينهما لأن الارتباط بينهما هو في واقع الأمر صفر، أي أننا بمعنى آخر لا نستطيع أن نقبل القول بوجود ارتباط دال إحصائياً بين المتغيرين. وبالمثل إذا حصلنا في اختبار إحصائي للفرق بين متوسطين على قيمة ت = ١,٦٤٥ فإننا نجد أن هذه القيمة دالة عند مستوى ٠,٠٥ ويمكن على أساسها أن نرفض الفرض الصفري بعدم وجود فرق بين المتوسطين. ولكن إذا كان الفرق الفعلي بين المتوسطين يقل عن ٥, فلا يمكن اعتبار هذا فرقاً يعتد به، لأن دلالته ترجع إلى كبر حجم العينة التي أجري عليها الاختبار.

يتبين بوضوح من المثالين السابقين أن هناك فرقاً بين الدلالة الإحصائية التي قد نحصل عليها عند اختيار الفرض الصفري، والأهمية الفعلية للنتائج، فإذا كان حجم العينة كبيراً كما هو الحال في المثالين السابقين، قد لا يكون للنتائج أي قيمة عملية للمجتمع، أو نظرية للعلم.

ومن الأخطاء الشائعة أيضاً الخلط بين الدلالة الإحصائية والدلالة النفسية أو التربوية. إن الدلالة النفسية أو الدلالة التربوية تعني القدر الذي يمكن لنتيجة ما أن تضيفه للمعرفة (Furlong et al., 2000). وتتضمن الدلالة النفسية أو الدلالة التربوية ثلاثة عناصر:

* قيمة الفروض التي وضعها الباحث والأفكار النظرية التي استمدت منها هذه الفروض، وقدرتها على تفسير البيانات التي يحصل عليها الباحث.
* كفاية الدراسة كاختبار للفروض، بما في ذلك مدى جودة تصميمها، واستخدام أدوات حديثة صادقة في جمع البيانات.
* وضوح نتائج الدراسة.

فالنتيجة الدالة إحصائياً لا تضيف دائماً لفهمنا السلوك الإنساني، ومع ذلك فقد يكون لدى البعض نزعة لتركيز على الدلالة الإحصائية، رغم ما قد يكون بالنتائج من ضعف، لا يساعد على تفسير سليم له معناه لهذه النتائج.

قبول الفرض الصفري والخطأ المرتبط به:

يشعر معظم الباحثين بالاطمئنان عند استخدام فكرة أن مستوى الدلالة (قيمة ألفا (α) التي تحدد القيم الحرجة التي بواسطتها نتخذ قراراً برفض أو قبول الفرض الصفري) تعطينا الاحتمال بالوقوع العشوائي. ولسوء الحظ نجد بعض الباحثين يعتقدون أن ألفا هي «معدل الخطأ» بالنسبة للدراسة، ويظنون أننا عندما «نحدد أن ألفا تساوي ٠٥, فإن ذلك يعني أن احتمال الخطأ الذي قد يقع فيه الباحث يعادل ٥٪، وعلى هذا فإنه على ثقة بأن نتائجه صادقة بنسبة ٩٥٪». وقد أدى هذا بدوره إلى الاعتقاد بأن ٩٥٪ من البحوث تصل إلى نتائج دقيقة (Furlong et al., 2000). ويذكر جون هنتر (Hunter, 1997) بخطأ هذا التفكير إذ يقول إن الخطأ من النوع الأول لا ينطبق إلا على الحالات التي يكون فيها الفرض الصفري في الواقع **خطأ**، ولذلك لا تعتبر ألفا معدل الخطأ إلا في تلك الدراسات التي تنظر في علاقات **غير موجودة**، وعندما يكون الفرض الصفري خطأ في الواقع، لا يكون لألفا علاقة بالموضوع.

وإذا عرفنا «معدل الخطأ» بأنه احتمال الوصول إلى استنتاج خاطئ، فإننا في هذه الحالة نرتكب خطأ يطلق عليه الخطأ من النوع الثاني (β) أي بيتا، وليس ألفا (α). وتتوقف قيمة بيتا (وهي قيمة غير معروفة) على حجم الأثر أو قوة

العلاقة بين المتغيرات. فعندما تكون العلاقة ضعيفة جداً تكون قيمة بيتا كبيرة جداً. ويكون احتمال الخطأ (أي بعدم رفض الفرض الصفري) كبير جداً.

نقطة القطع الاعتبارية بين الأحداث «الشائعة» و«النادرة»:

يذكر جيفري لوتفس (Loftus, 1996) أن من التهديدات التي تواجه صدق عملية اختيار الفرض الصفري الطبيعة الاعتبارية لألفا (α) التي تحدد عند مستوى ٠٥, أو أقل بطريقة تقليدية. إذ يقول إن معظم الناس يتفقون على عدم وجود فروق مهمة بين النتيجة التي تحدث ٥٠ مرة من ١٠٠٠ (ل=٠٥٠,) والنتيجة التي تحدث ٥١ مرة من ١٠٠٠ (ل=٠٥١,) (ص١٦٣-١٦٤). ومع ذلك عندما نحدد مستوى ألفا بأنه ٠٥,، فإن النتيجة الأولى تكون دالة إحصائياً، في حين تكون النتيجة الثانية غير دالة. وتظهر المشكلة عندما يحاول الباحثون ترجمة النتيجة الإحصائية في علاقات مفهومة بين المتغيرات. إذ ينظر عادة للنتائج الدالة كبرهان على صدق النتائج بطريقة عملية، في حين أن النتائج غير الدالة كثيراً ما تفسر خطأ على أنها برهان على عدم وجود أثر. وعلى هذا فإن معرفتنا بمدى فاعلية النتائج تتوقف على أن الاختبار الإحصائي يعطي احتمالاً قدره ٠٥, في مقابل احتمال قدره ٠٥١,. ويرى لوفتس (Loftus, 1996) أن كثيراً من الباحثين لا يدركون أن التمييز بين «دال» و«غير دال» تمييز سطحي ويقوم على نقطة قطع اعتبارية. فالنتائج تميل إلى أن تقع في تقسيم ثنائي بين «وجود آثار فعلية» (ل $\leqslant$ ٠٥,)، و«عدم وجود أثر» (ل $>$ ٠٥,). والخطورة في هذا الموقف هو أن هذه الثنائية تصبح جزءاً من المعرفة التي نحصل عليها من البحث العلمي. وببساطة فإن لوفتس يعتقد أن النتائج المتشابهة (مثل ل = ٠٥, و ل = ٠٥١,) يجب أن تؤدي إلى نفس الاستنتاجات والخلاصات. ولذلك فهو يدين الفرض الصفري ويدعو إلى إلغائه من التحليل الإحصائي.

وقد أدى هذا الخط من التفكير إلى توجيه الانتباه نحو مشكلات الفرض الصفري والاختبارات الإحصائية. والتراث البحثي في هذا المجال واسع جداً. فقد وجد تومسون (Thompson, 2001) أكثر من ٤٠٠ مقال في المجالات المختلفة حول الاستخدام غير السليم للاختبار الإحصائي للفرض الصفري.

إساءة تفسير قيم (ل)

نناقش فيما يلي بعض الأخطاء في فهم الاحتمالات الناتجة عن الاختبارات الإحصائية وهي (قيم ل). وسنبدأ أولاً بعرض التفسيرات الصحيحة. ويلاحظ أن الغرض من الاختبارات الإحصائية قياس الفرق بين إحصاءة العينة ومعلم المجتمع المذكور في الفرض الصفري مع أخذ أخطاء المعاينة في الاعتبار، ويتم تحويل الاختبار الإحصائي الأمبيريقي إلى احتمال داخل الاختبار المركزي المناسب. وهذا الاحتمال احتمال مشروط للإحصاءة بالتسليم بصحة الفرض الصفري. وفيما يلي التفسيرات الأخرى الصحيحة للحالة الخاصة ل < ٠٥,٠:

١ – الأرجحية أقل من ١ إلى ١٩ في الحصول على نتيجة من عينة عشوائية أكثر تطرفاً من القيمة الملاحظة عندما يكون الفرض الصفري (فـ$_0$) صحيحاً.

٢ – أقل من ٥٪ من الاختبارات الإحصائية تبعد كثيراً عن متوسط التوزيع العيني بمقتضى الفرض الصفري (فـ$_0$) من المتوسط الذي نحصل عليه.

٣ – مع التسليم بصحة الفرض الصفري (فـ$_0$) وتكرار الدراسة مرات كثيرة فإن أقل من ٥٪ من هذه النتائج سيكون أقل اتساقاً مع الفرض الصفري من النتيجة التي حصلنا عليها.

ويمكن القول أن التفسيرات السابقة هي التفسيرات الصحيحة فقط، وأي تفسير صحيح آخر لا بد أن يكون تنويعاً منها، وهذه التفسيرات كما هو واضح محدودة للغاية. ويمكن بشكل عام الإشارة إلى أي تفسير صحيح باعتباره ل (ب|فـ$_0$) والتي تؤكد على الاحتمالات التي نحصل عليها من الاختبارات الإحصائية باعتبارها احتمالات مشروطة للبيانات (ب) في وجود الفرض الصفري.

والأخطاء التالية في التفسيرات أخطاء شائعة لحالة ل < ٠٥,٠ وبعضها يرجع إلى نسيان الاحتمالات المشروطة، أي إعطاء الوضع العكسي للحالتين الممثلتين بقيم (ل)، و(فـ$_0$). ومع كل تفسير خاطئ أعطينا سببه.

الخطأ الأول:

مضمون الخطأ الأول أن قيمة ل هي احتمال أن النتيجة ترجع إلى أخطاء المعاينة، ولذلك فإن ل < ٠٥, تعني أن هناك أرجحية تقل عن ٥٪ أن هذه االنتيجة حدثت بالصدفة. ويرجع الخطأ هنا إلى أن قيم ل تحسب مع التسليم بأن أخطاء المعاينة هي السبب في ابتعاد إحصاءات العينة عن الفرض الصفري. أن أن أرجحية أخطاء المعاينة اعتبرت بالفعل ١,٠٠ عند إجراء أي اختبار إحصائي. ولذلك فمن غير المنطقي النظر إلى قيم ل كمقياس لاحتمال خطأ المعاينة. وهذا الخطأ إلى جانب الأخطاء التالية قد تفسر الخطأ المرتبط بأن الاختبارات الإحصائية تفرز النتائج في فئات: تلك الناتجة عن الصدفة (قبول الفرض الصفري) وتلك الناجمة عن الآثار الحقيقية (رفض الفرض الصفري). والواقع أن الاختبارات الإحصائية المستخدمة في الدراسات المفردة لا يمكن لها أن تقوم بهذا التمييز. وذلك لأن أي قرار يتخذ بناء على نتائج الاختبار الإحصائي للفرض الصفري قد تكون خاطئة (أي الخطأ من النوع الأول أو الخطأ من النوع الثاني).

الخطأ الثاني:

أما الخطأ الثاني فمضمونه أن قيمة ل هي الاحتمال بأن الفرض الصفري صحيح بالنسبة للبيانات. ولذلك فإن ل < ٠٥, تعني أن ل (فـ$_0$|ب) < ٠٥, وهذا عكس خطأ الاحتمال (Cohen, 1994)، وهذا ناتج عن نسيان أن قيم ل هي احتمالات شرطية للبيانات، أي ل (ب|فـ$_0$) وليس بسبب الفرض الصفري، أو ل (فـ$_0$|ب). والأخير هو الاحتمال البعدي للفرض الصفري في ضوء البيانات، وهو غالباً ما يرغب الباحثون في معرفته. وتبين فكرة مبسطة من نظرية بايز (Bayes theorem) أن ل (ب|فـ$_0$) من اختبار إحصائي والاحتمال البعدي للفرض الصفري هما في الواقع

$$\text{ل (فـ}_0\text{|ب)} = \frac{\text{ل (فـ}_0\text{) ل (ب|فـ}_0\text{)}}{\text{ل (ب)}}$$ معادلة ١-١

وفي المعادلة رقم (١-١) ل(فـ$_0$) هي الاحتمال القبلي أن الفرض الصفري صحيح قبل جمع البيانات، ل (ب) هي الاحتمال البعدي للبيانات بغض النظر عن صحة الفرض الصفري. أي أنه بالحصول على قيمة ل من اختبار إحصائي مع تقديرات لقيم ل (فـ$_0$) و ل (ب) يمكننا أن نشتق بالمعادلة ل (فـ$_0$|ب) الاحتمال البعدي للفرض الصفري. ولكن الذين يستخدمون الاختبارات الإحصائية التقليدية لا يفكرون عادة في الاحتمال البعدي.

الخطأ الثالث:

الخطأ الثالث أنه إذا رفض الفرض الصفري فإن ل هي الاحتمال بخطأ هذا القرار. ولذلك إذا كانت ل< ٠٥, فهناك فرصة تقل عن ٥٪ بأن القرار برفض الفرض الصفري خطأ من النوع الأول. وهذا الخطأ نوع آخر من الاحتمال العكسي الخطأ والذي وصفه بولارد (Pollard, 1993) بالخلط بين الاحتمال الشرطي القبلي للخطأ من النوع الأول، أو:

$$\alpha = \text{ل (رفض فـ}_0\text{|فـ}_0\text{)}$$

والاحتمال الشرطي البعدي بأن الفرض الصفري رفض، أو:

$$\alpha = \text{ل (فـ}_0\text{| رفض فـ}_0\text{)}$$

ويستخدم بولارد نظرية بايز ليبين أنه من غير الممكن بشكل عام تقدير ل (فـ$_0$| رفض فـ$_0$) من α. ومن ناحية المستوى الحدسي فإن القرار برفض الفرض الصفري في دراسة مفردة قد يكون صحيحاً وقد يكون خاطئاً، ولذلك فليس هناك احتمال مرتبط به. ولا يمكن أن يتخذ الفرد قراراً خاصاً باعتبار رفض الفرض الصفري صحيحاً إلا بعد تكرار الدراسة عدداً كافياً من المرات.

الخطأ الرابع:

مضمون الخطأ الرابع أن القيمة المكملة لقيمة ل، أي (ل-١) هي احتمال أن الفرض البديل صحيح بالنسبة للبيانات المعطاة، أو ل (ف$_1$|ب). وعلى ذلك

ل< ٠٥, تعني أرجحية صحة ف١ أكثر من ٩٥٪. وهذه الفكرة الخاطئة يطلق عليها موليك (Mulaik et al., 1997) زيف الصدق validity fallacy وهذه القيمة المكملة لقيمة ل احتمال، ولكنها أرجحية الحصول على نتيجة وإن كانت أقل تطرفاً بمقتضى الفرض الصفري من القيمة التي نحصل عليها فعلاً. وعلى ذلك فإن مكملات ل ليس لها علاقة مباشرة بالاحتمال البعدي للفرض البديل.

الخطأ الخامس:

الخطأ الخامس هو أن القيمة المكملة لقيمة ل هي احتمال أن النتيجة سوف تتكرر تحت ظروف ثابتة، ومعنى هذا أن ل< ٠٥, تعني أن فرصة تكرار النتيجة تزيد على ٩٥٪. وهذا هو زيف القابلية للتكرار replicability fallacy (Carver, 1978). وهناك تنويع آخر لهذا الخطأ وهو أن ل< ٠٥, تعني أن التكرار له احتمال قدره ٩٥٪ بإعطاء نتيجة دالة إحصائياً، يفترض أنها في نفس اتجاه الدراسة الأصلية. ولو كان هذا الخطأ صحيحاً فإن معرفة احتمال الحصول على نفس النتيجة في تكرار مقبل للبحث يكون أمراً مفيداً جداً. إلا أن قيمة ل ليست إلا مجرد احتمال نتيجة معينة تحت فرض معين. ويقول كارفر أن التكرار أمر يتعلق بالتصميم التجريبي. أي أنه تساؤل أمبيريقي للدراسات المقبلة ولا يتعلق مباشرة بالاختبارات الإحصائية في دراسة واحدة.

الاستنتاجات الخاطئة بعد اتخاذ قرار بالنسبة للفرض الصفري

هناك أيضاً بعض الاستنتاجات الخاطئة التي تحدث بعد رفض الفرض الصفري أو قبوله بناء على قيم ل. ومعظمها لا يحتاج إلى شرح سبب خطئها.

الخطأ الأول:

أن قيم ل مؤشر بدرجة أثر معين، ولذلك فإن قيم ل الصغيرة تشير إلى آثار كبيرة. ويمكن أن نطلق على هذا الخطأ خطأ المرتبة. فقيم ل الصغيرة تدل على احتمالات شرطية أقل للبيانات، وذلك مع وجود المسلم الضروري بأن الفرض الصفري يصف المجتمع بدقة وهذا كل ما يمكن قوله دون أية أنواع

أخرى من التحليلات مثل حجم الأثر. ذلك أن الاختبارات الإحصائية وقيم ل الناتجة عنها تقيس حجم العينة وحجم الأثر، ولذلك فإن أثراً بسيطاً يحتاج فقط إلى عينة كبيرة الحجم جداً لتصبح قيمة ل الصغيرة دالة إحصائياً. وإذا كان حجم العينة كبيراً فعلاً، فإن قيم ل الصغيرة هي مجرد تأكيد لحجم العينة الكبير. كما أن النتائج ذات الحجم الكبير قد تكون قيم ل الناتجة عنها صغيرة. وبمعنى آخر فإننا لا نستطيع من مجرد معرفة قيم ل أن نحدد شكل النتائج.

الخطأ الثاني:

أن رفض الفرض الصفري يؤكد الفرض البديل وفرض البحث الذي يقف وراءه. هذا الخطأ الذي يبدو معقولاً في معناه يعكس خطأين في المفاهيم. الأول أن القرار برفض الفرض الصفري في دراسة واحدة لا يعني البرهنة على الفرض البديل. والثاني أنه حتى ولو كان الفرض الإحصائي البديل صحيحاً فإن هذا لا يعني أن الفرض الأساسي الذي يقف وراءه فرض صحيح هو الآخر.

والتمييز بين الفروض الإحصائية والفروض الأساسية أمر مهم فهما لا يختلفان فقط في مستوى التجريد (الإحصائي منخفض والعلمي مرتفع)، بل إن لهما مضامين مختلفة بعد رفض الفرض الصفري فإذا كان الفرض الصفري والفرض البديل يعكسان الفروض الإحصائية فقط، فليس هناك ما نفعله عند رفض الفرض الصفري سوى تكرار البحث. ولكن إذا كان الفرض البديل هو الفرض العلمي، فإن العمل يبدأ بعد رفض الفرض الصفري. لأن جزءاً من العلم يتطلب مقارنة فرض البحث بالفروض الأخرى الرئيسة والتي تتوافق أيضاً مع الفرض الإحصائي البديل. وإذا لم يكن من الممكن استبعاد تلك الفروض فإن ثقة الباحث في الفرض الأصلي يجب أن تقل. وقد يكون من الضروري أيضاً إجراء دراسات إضافية تحاول تخطئة النماذج المكافئة. وهذا ما يطلق عليه استراتيجية الاستدلال (Platt, 1964).

الخطأ الثالث:

أن الفشل في رفض الفرض الصفري معناه أن حجم الأثر في المجتمع صفر. وهذا ليس استدلالاً صادقاً لعدد من الأسباب. أحدها أن المعتقد الأساسي في العلم أن غياب القرينة ليس قرينة على عدم وجودها. كذلك إن الفشل في اتخاذ القرار برفض الفرض الصفري قد يكون راجعاً إلى الخطأ من النوع الثاني. مثال ذلك أنه قد يكون هناك أثر حقيقي، إلا أن الدراسة ينقصها قوة كافية لكشفها. وإذا أخذنا في الاعتبار الانخفاض النسبي العام للقوة في البحوث السلوكية، نجد أن من المحتمل أن هذه الحادثة ليست أمراً غير متكرر. فإن ضعف تصميم البحث أو استخدام أدوات غير سليمة قد يؤدي إلى الخطأ من النوع الثاني.

الخطأ الرابع:

أن الفشل في رفض الفرض الصفري بتساوي متوسطين يعني أن المجتمعين متكافئين. لنفرض أننا قارنا معالجة تمت البرهنة على فعاليتها بمعالجة جديدة تكاليفها أقل. فمن غير الصحيح استنتاج أن المعالجتين متكافئتين إذا لم نرفض الفرض الصفري. إن الاستدلال على التكافؤ يكون أيضاً غير صحيح إذا كان معاملا الثبات أو نسبتان في مجموعتين ليستا مختلفتين إحصائياً (Thompson, 2003). وإذا أردنا إعادة صياغة المعتقد السابق يمكن أن نقول إن غياب قرينة الاختلاف ليس قرينة على التكافؤ. وهناك طرق مناسبة لاختبار التكافؤ سوف نذكرها فيما بعد. وللأسف فإن هذا الخطأ منتشر انتشاراً عاماً في البحوث الاجتماعية والسلوكية في البلاد العربية حيث نجد أن حصول الباحث على مستوى دلالة يزيد على ٠٥, يتخذ قرينة على تكافؤ المجموعتين التجريبية والضابطة أو أي متغير آخر يرغب الباحث في تحقيق تكافئه في المجموعتين.

الخطأ الخامس:

أن رفض الفرض الصفري يؤكد جودة التصميم التجريبي. ذلك أن ضعف التصميم قد يوجد آثاراً زائفة تؤدي إلى رفض غير صحيح للفرض الصفري. كما أن خطأ المعاينة قد يقود إلى الخطأ من النوع الأول.

الخطأ السادس:

أنه إذا لم نرفض الفرض الصفري، فذلك يعني أن الدراسة فاشلة. وهذا الخطأ عكس الخطأ السابق. فرغم أن استخدام طرق غير سليمة أو ضعف القوة قد يؤدي إلى الخطأ من النوع الثاني، فإن الفشل في رفض الفرض الصفري يمكن أن يكون راجعاً أيضاً إلى علم جيد. مثال ذلك قد تكون بعض الادعاءات المبنية على دراسات أولية غير صحيحة، مما يعني أن تكرار الدراسة سوف يؤدي إلى نتائج سالبة. مثال ذلك إذا نتج عن بعض الدراسات نتيجة ما تعتبر مهمة، فقد لا يستطيع باحثون آخرون تحقيق نفس النتيجة وقد يخرجون برفض لهذه الفكرة.

الخطأ السابع:

أن رفض الفرض الصفري معناه معرفة الميكانزم المسبب. وهذا التفسير ينتمي للتفسيرات التي سقناها قبل قليل. فيجب أن يكون واضحاً الآن أن رفض الفرض الصفري من مرة واحدة لا يؤكد الأثر السببي المشار إليه في الفرض البديل.

الخطأ الثامن:

أن الفشل في تكرار نتائج البحث هو فشل باتخاذ نفس القرار حول الفرض الصفري عبر عدد من الدراسات. ومعنى هذا أن النتيجة لم تتكرر إذا رفض الفرض الصفري في الدراسة الأولى وليس في الدراسة الثانية. إلا أن هذه النظرة تتجاهل حجم العينة وحجم الأثر واتجاه الأثر عبر الدراستين. لنفرض أننا وجدنا فرقاً بين متوسطين في الدراسة الأولى، وتم رفض الفرض الصفري. ووجدنا نفس الفرق بالضبط في الدراسة الثانية، إلا أن الفرض الصفري لم يرفض لأن حجم العينة كان أصغر. وفي هذه الحالة يكون لدينا قرينة على تكرار النتيجة رغم اتخاذ قرارين مختلفين بالنسبة للفرض الصفري.

انتشار التفسيرات الخاطئة:

هناك شواهد على أن كثيراً من المعتقدات الخاطئة التي وصفناها قبل قليل معتقدات شائعة بين الباحثين النفسيين والتربويين، ومنهم أكاديميون يعملون في الجامعات. ولقد طلب أوكس (Oakes, 1986) من ٧٠ أكاديمياً من المختصين في علم النفس ذكر بعض التفسيرات التي يستخدمونها عادة لقيمة ل< ٠١, ويمكن للمستجيب أن يذكر أكثر من تفسير. ومن بين العبارات المبينة في جدول (رقم ١-٢) كان التفسير الأخير هو التفسير الصحيح، ولكن لم يذكره سوى ٨ من ٧٠ مستجيباً (حوالي ١١٪). وقد أيد حوالي ٥٠٪ من المستجيبين العبارتين ٢ و ٤ في الجدول بأن قيم ل تشير إلى الاحتمال الشرطي للفرض الصفري (خطأ الاحتمال العكسي) أو الفرض البديل (خطأ صدق فروض البحث) على التوالي. وقد أخطأ معظم المستجيبين في قولهم أن قيم ل احتمالات بعدية للخطأ من النوع الأول، كما ذكر حوالي الثلث أن مكملات قيم ل تشير إلى أرجحية تكرار النتائج (خطأ التكرار).

جدول ١-٢

التفسيرات التي يتبناها ٧٠ من الأكاديميين النفسيين للقيمة ل< ٠١,

	العبارة	التكرار	٪
١	تمت البرهنة على خطأ الفرض الصفري.	١	١,٤
٢	تم العثور على احتمال الفرض الصفري.	٣٢	٤٥,٧
٣	تمت البرهنة على الرفض التجريبي تماماً.	٢	٢,٩
٤	يمكن استنباط احتمال الفرض التجريبي.	٣٠	٤٢,٩
٥	أصبح احتمال خطأ القرار المتخذ معروفاً.	٤٨	٦٨,٦
٦	يبلغ احتمال دلالة تكرار النتيجة ٩٩, .	٢٤	٣٤,٣
٧	احتمال البيانات حيث الفرض الصفري معروف.	٨	١١,٣

Oakes, M. (1986). Statistical inference. New York: Wiley, (p. 81).

وقد يعتقد البعض أن نتائج أوكس خاصة بعينة غير ممثلة للنفسيين، إلا أن نتائج دراسات مسحية أخرى في العلوم الاجتماعية تشير إلى انتشار نفس المعتقدات الخاطئة انتشاراً واسعاً (Mittag & Thompson, 2000; Nelson, Rosenthal, & Rosnow, 1986). وقد وصف تفرسكي وكانمان (Tversy & Kahneman, 1971) نوعاً من التشويش المعرفي بين النفسيين يطلقون عليه «الاعتقاد في الأعداد الصغيرة». ويشير هذا الاعتقاد إلى:

١ – حتى العينات الصغيرة عينات ممثلة لمجتمعاتها.

٢ – يمكن أن نحصل على نتائج دالة إحصائياً من تكرار البحث على عينات يبلغ حجمها نصف حجم العينة الأصلية.

ويحتمل أن الاعتقاد في قانون الأعداد الصغيرة منتشر كذلك بين المختصين في العلوم الاجتماعية والسلوكية، وليس قاصراً على النفسيين (Kline, 2004).

الاختبار الإحصائي للفرض الصفري لا يعطينا المعلومات التي نسعى لمعرفتها:

تتعلق كثير من الأخطاء التي ذكرناها عن نتائج الاختبار الإحصائي للفرض الصفري والتي سبق ذكرها بأشياء يريد الباحثون فعلاً أن يصلوا إليها بما في ذلك احتمال صحة الفرض الصفري أو الفرض البديل، وأرجحية تكرار البحث، واحتمال خطأ القرار برفض الفرض الصفري، وذلك كله مع التسليم بالبيانات. وإذا استخدمنا «ت» لتكون رمزاً للتكرار فإن هذه الرغبات يمكن تلخيصها فيما يلي:

ل (فـ$_0$|ب) و ل (فـ$_1$|ب) و ل (ت|ب) و ل (فـ$_0$| رفض فـ$_0$)

ولسوء الحظ فإن الاختبارات الإحصائية لا تخبرنا إلا عن ل (ب|فـ$_0$). ويذكر كوهين (Cohen, 1994) أنه لا يوجد أسلوب إحصائي في الدراسات المفردة يمكنها أن تحقق تلك الرغبات المعبر عنها بالمعادلات السابقة. إلا أن هناك أسلوباً يمكن أن يخبرنا ما نريد أن نعرف حقيقة، إلا أنه ليس أسلوباً إحصائياً

بل إنه التكرار، وليس هذا الأسلوب هو الأسلوب الوحيد الذي يخبرنا فقط عن أخطاء المعاينة، ولكن التكرار معيار ذهبي في العلم. وفي الفقرة التالية تفاصيل أكثر حول هذه النقطة.

الفروض الصفرية خاطئة عادة:

الفروض الصفرية أكثر أنواع الاختبارات استخداماً في العلوم الاجتماعية والسلوكية. إلا أن معالم المجتمع نادراً ما تكون صفراً بالضبط، وبخاصة إذا كان الصفر يعني عدم وجود أثر أو ارتباط أو فرق بالمرة (Kirk, 1996). مثال ذلك يندر أن يكون معامل الارتباط في المجتمع بين أي متغيرين صفراً. ولذلك يكون واقعياً أكثر أن نسلم بأن الارتباطات أو الفروق بين المتغيرات في المجتمع ليست صفراً. ورغم أننا لا نعرف القيم الفعلية لهذه الارتباطات أو الفروق، إلا أن هذه القيم قد تبتعد كثيراً عن الصفر بحيث إننا إذا اخترنا عينات كبيرة الحجم فإننا سوف نحصل على ارتباطات أو فروق دالة إحصائياً وتقود بالتالي إلى رفض الفرض الصفري رغم أن القيم التي حصلنا عليها ليس لها قيمة أو دلالة عملية. وعلى ذلك يمكن القول:

١ – إن المعدل الفعال للخطأ من النوع الأول في كثير من الدراسات قد يكون صفراً بالضرورة.

٢ – إن النوع الوحيد من الخطأ الذي يمكن أن نقع فيه هو الخطأ من النوع الثاني.

وإذا أخذنا في اعتبارنا أن القوة في المتوسط تبلغ ٥٠, فإن الخطأ من النوع الثاني في الأحوال العادية سوف يكون هو الآخر ٥٠, ويقول شميث (Schimdt, 1996) في هذا المجال أن الطرق التي تتحكم في الخطأ من النوع الأول في التجارب التي نقوم بها مثل تصحيح بونفيروني Bonferroni correction قد يؤدي إلى خفض القوة إلى أقل من ٠,٥٠. وعلى هذا فإن أي نوع من الفروض سواء كان الفرض الصفري أو الفروض البحثية لن تكون صحيحة، ولهذا يكون من المنطقي أكثر اختبار فروض البحث ما دامت النتيجة واحدة في

الحالتين. لنفرض أننا اختبرنا الفرض الصفري ل = ٠,٣٠ فمن المحتمل جداً أن القيمة الفعلية في المجتمع تختلف عن ٠,٣٠ باعتبارها صفراً، ولذلك فإن اختبار الفرض غير الصفري يكون أكثر واقعية حيث إن هذا الاختبار سيؤدي إلى رفضه. وبذلك تكون فروض البحث فروضاً أكثر واقعية لاختبار قيم العينة. وربما كانت معظم قيم ل الواردة في تقارير البحوث في التراث البحثي في العلوم السلوكية والاجتماعية مرتبطة بفروض صفرية غير واقعية.

يسلم التوزيع العيني للاختبارات الإحصائية بالمعاينة العشوائية

وصف لنبورج (Lunneborg, 2001) هذه القضية بسوء التوافق بين التحليل الإحصائي وتصميم البحث. فإن تقدير قيم ل للاختبارات الإحصائية يتم بالنسبة للتوزيع العيني الذي يسلم بالمعاينة العشوائية من مجتمعات معروفة. وهذا هو نفس التوزيع الذي يُستخدم أساساً لحساب الخطأ المعياري التي تقدر على أساسها حدود الثقة للمجتمع الذي يتعلق بالصدق الخارجي لنتائج العينة. إلا أن معظم العينات في العلوم الاجتماعية لا تختار عشوائياً، بل هي عينات نحصل عليها بالطرق المتاحة.

لا تتحقق غالباً المسلمات الإحصائية المتعلقة بالاختبارات الإحصائية للفرض الصفري

تقوم الاختبارات الإحصائية على مسلمات معينة حول التوزيع. وبعض هذه المسلمات أكثر تشدداً من بعضها الآخر وبخاصة فيما يتعلق بتباين المتغير التابع. فإذا انتهكت المسلمات قد نحصل على قيم ل خاطئة. ولسوء الحظ لا يوفر لنا كثير من الباحثين بيانات عن مدى تحقق المسلمات المتعلقة بتوزيع المتغيرات. بل إن الباحثين غالباً ما يهملون الاختبارات المتعلقة بالمسلمات. ولذلك فإن لدينا فجوة بين ما يذكر في كتب الإحصاء عن هذه المسلمات وبين ما يحدث في الواقع. لأن قليلاً من البحوث وبخاصة التي تستخدم تحليل التباين هي التي تعطي نتائج صحيحة رغم انتهاك مسلمات استخدام الاختبارات الإحصائية.

اختبارات التكافؤ

يقصد باختبارات التوافق تلك الاختبارات التي تعالج مشكلة تكافؤ مجموعتين أو حالتين. واختبارات التكافؤ معروفة بشكل أكبر وأفضل خارج العلوم السلوكية والاجتماعية، فهي معروفة ومستخدمة بكثرة في البحوث الصيدلانية وبحوث البيئة والعلوم البيولوجية. مثال ذلك قد يرغب الباحث في معرفة هل تتكافأ المجموعة التجريبية والمجموعة الضابطة في المتغير المستقل قبل إجراء التجربة. وبالنسبة للاختبار الإحصائي التقليدي للفرض الصفري فإن قبول الفرض الصفري بعدم وجود فروق دالة إحصائياً لا يعني التكافؤ، فقد يزيد مستوى الدلالة قليلاً على ٥٪ وهذا غير كاف لاتخاذ قرار بالتكافؤ، ويرى الكاتب أن استخدام اختبار الدلالة الإحصائية لتحقيق التكافؤ يحتاج أن يكون مستوى الدلالة على الأقل ٢٥, أو أكثر، وكلما اقتربنا من ٥٠, كان أفضل. إلا أنه من الأفضل استخدام الاختبارات الخاصة بالتكافؤ بين المجموعات أو الحالات. وأحد صور التكافؤ يستبدل بالفرض الصفري ذو النقطة الواحدة فرضٌ يعبر عن مدى بين نقطتين (م١ – م٢) بحيث يعطي فرقاً واقعياً بين المتوسطين. ولتحقيق ذلك نستخدم زوجاً من الفروض وكل فرض منهما يتعلق بإحدى نقطتي المدى، وقد يصاغ الفرضان على النحو التالي:

$$\text{ف}_0 = \begin{cases} \text{ف}_{0\,1}: (\text{م}_1 - \text{م}_2) > -٥ \\ \text{ف}_{0\,2}: (\text{م}_1 - \text{م}_2) < ٥ \end{cases}$$

وهذان الفرضان الفرعيان يعنيان أنه لا يمكن اعتبار متوسطي المجموعتين متكافئين إذا كان الفرق بينهما ٥ درجات. والمدى بين هذين الفرضين يمكن صياغته على النحو التالي:

$$-٥ \geqslant (\text{م}_1 - \text{م}_2) \leqslant ٥$$

ويرفض هذا الفرض إذا كان الحد الأدنى لحدود ثقة ذات طرف واحد حول المتوسط الملاحظ يزيد على ٥, أو كان الحد الأعلى لحدود ثقة ذات طرف واحد أقل من –٥.

ما الطريقة المثلى للتغلب على هذه المشكلة؟

بعد مناقشة أوجه النقد للاختبارات الإحصائية يمكننا أن نختار واحداً من الطرق التالية:

١ – لا نفعل شيئاً، أي نستمر في استخدام الاختبارات الإحصائية كما تعودنا خلال الخمسين عاماً السابقة.

٢ – نتوقف عن استخدام الاختبارات الإحصائية كلية ونتوقف عن تدريسها في المقررات الجامعية. ونفرض حظراً على استخدامها في تقارير البحوث. وهذا هو نفس اقتراح هنتر (Hunter, 1997; Schmidt, 1996) والسابق الإشارة إليه في هذه الدراسة.

٣ – نرسم طريقاً وسطاً بين الإجراءين المتطرفين السابقين، بأن نستخدم درجات متباينة من الاختبارات الإحصائية التي تتوقف أهميتها على نوع الدراسة التي نقوم بها ولكن مع التشدد في استخدامها حتى لا تفلت الأمور وتصبح أسوأ مما كانت عليه من قبل.

والاختيار الأول ليس مقبولاً إذ يترتب عليه آثار سلبية وهذا ما لا يتفق مع التقدم العلمي في البحوث السلوكية، والتي يترتب عليها استبعاد فكرة عدم فعل شيء. كما إن منع الاختبارات الإحصائية من النشر في المجلات العلمية ليس أمراً منظوراً في المستقبل القريب على الأقل، خاصة وأن متطلب استخدام حجر الأثر إلى جانب الدلالة الإحصائية هو في واقع الأمر منع لاستخدام الاختبارات الإحصائية بمفردها. وعلى هذا لا بد من استبعاد الخيارين الأولين.

وقد ذكر كلاين (Kline, 2004) مجموعة من التوصيات تتعلق بالاختيار الثالث نوردها فيما يلي. والغرض منها وضع إطار بناء للتغيير. ومن المسلم به أن البعض قد لا يوافق على بعض عناصر هذه التوصيات، وهذا يرجع إلى صعوبة الاتفاق بشأن الجدل الدائر حول الاختبار الإحصائي للفرض الصفري، ولذلك من الصعب أن يوافق الجميع على توصيات بعينها.

والفكرة الأساسية لهذه التوصيات كما ذكرها كلاين يمكن تلخيصها على النحو التالي: إن استخدام الاختبار الإحصائي للفرض الصفري قد أعاننا كثيراً في علم النفس والتربية وغيرها من العلوم السلوكية والاجتماعية على استخدام الطرق المستخدمة في العلوم الطبيعية، ولكن العلوم السلوكية الآن تحتاج إلى انتهاج طرق أخرى تبعدها عن المراهقة العلمية، وربما كان فرض النمو الإجباري السبيل إلى تحقيق ذلك.

التوصيات:

١ – **يفضل استخدام التحليل الإحصائي للفرض الصفري في الدراسات الاستطلاعية** حيث إن هذا النوع من الدراسة عادة جديد ولا يعرف فيه إذا كان هناك أثر للمتغير المستقل على المتغير التابع. ذلك أن قدرة التحليل الإحصائي للفرض الصفري على معالجة السؤال الثنائي بوجود (أو عدم وجود) علاقات تزيد على المستوى المتوقع لأخطاء المعاينة قد تكون مفيدة في بعض أنواع البحوث الجديدة. وفي هذه الحالة يفيدنا الاختبار الإحصائي للفرض الصفري باتخاذ قرار يتعلق بهذا الموضوع.

٢ – **إذا استخدمت الاختبارات الإحصائية فيجب تحقيق ما يلي:**

أ – إعطاء معلومات عن القوة.

ب – أن يكون الفرض الصفري جديراً بالتصديق.

فمن المهم ذكر القوة وبخاصة إذا كانت معظم النتائج سالبة، أي أن عدداً قليلاً من الفروض الصفرية رفض، ذلك أن قراء الدراسة الأمبيريقية يمكنهم أن يعرفوا ما إذا كانت قوة الدراسة منخفضة جداً (أي أقل من ٠,٥٠). فمعرفتنا برفض الفرض الصفري مفيدة فقط عندما تكون الفروض معقولة وذات جدوى. كما أن واحداً ناقص القوة هي احتمال الخطأ من النوع الثاني، وهذا لا يحدث إلا إذا تم قبول الفرض الصفري عندما يكون هناك أثر حقيقي. وما زالت الإشارة إلى القوة محدودة جداً في معظم الدراسات، لأن الدراسات التي لا يرفض فيها الفرض الصفري غير مقبولة بشكل عام.

٣ – **لم يعد مقبولاً في العلوم السلوكية وصف نتائج التحليل الإحصائي للفرض الصفري فقط.** ويبرز هذا الاقتراح نواحي القصور التي أشرنا إليها سابقاً في الاختبار الإحصائي للفرض الصفري.

٤ – **لا يجب استخدام كلمة «دلالة» في التحليل الإحصائي على أن تستخدم فقط في الكتابة العادية لتدل على شيء له أهميته الفعلية.** ذلك أن استخدام كلمة «دلالة» لوصف الحدث ل $< \alpha$ تعبير ضعيف. ورغم أن الإحصائيين يعلمون أن كلمة دلالة في معرض الكلام على الاختبار الإحصائي للفرض الصفري لا أهمية له في الواقع، إلا أن استخدام هذه الكلمة قد يؤدي إلى تكوين معتقدات خاطئة لدى غير الإحصائيين. ولذلك يجب علينا الاستغناء عنها في العلوم السلوكية والاجتماعية، ونستخدم عوضاً عنها كلمة إحصائياً فقط عندما نرفض الفرض الصفري وهذا أمر كاف لذكر النتيجة التي حصلنا عليها ومعرفة نوع الفروق أو الارتباطات التي حصلنا عليها.

٥ – **يجب أن يكون الباحث مسؤولاً عن إعطاء حجم الأثر وحدود الثقة للنتائج الأولية وتفسيرهما كلما كان ذلك ممكناً.** كما أن إعطاء حدود الثقة لحجم الأثر أفضل وأكثر فائدة. فإن المدى بين الحدود لا يعرفنا فقط بمدى اتساع أخطاء المعاينة المرتبطة بالحدود التي يقع فيها حجم الأثر الملاحظ، بل إنها تعطي كذلك مدى من تقديرات أحجام الأثر في المجتمع والتي قد تكون مسؤولة عن النتيجة الملاحظة. على أنه ليس من السهل دائماً تقدير أحجام الأثر وبخاصة في التصميمات المعقدة، إلا أنه في أغلب الحالات يمكن إعطاء حجم الأثر وحدود الثقة المرتبطة به.

٦ – **إنها مسؤولية الباحث أن يبين الدلالة الجوهرية (نظرياً وعملياً) للنتائج. والاختبارات الإحصائية لا تصلح لهذا الغرض.** إن رفض الفرض الصفري لا يعني دلالة جوهرية. إذ يحتاج الباحثون إلى أطر مرجعية أخرى لتفسير كيف أن نتيجة ما هي نتيجة مهمة يعتد بها. والواقع أن هذه ليست مسألة سهلة، وأكثر صعوبة من استخدام قيمة ل في البحوث الاجتماعية والسلوكية. فإن تقديرات حجم الأثر تعتبر نقطة البداية لتحديد الدلالة

الجوهرية. ويجب أن يستعين الباحث بقراءاته وخبرته في هذا المجال كأن يشير إلى نتائج بحوث التحليل البعدي إذا كان لها وجود في مجاله.

٧ – **تكرار البحث خير وسيلة لمعالجة الخطأ المعياري.** ومنطق هذه التوصية واضح كذلك. فالأمر يكون مختلفاً تماماً إذا قام الباحث بتكرار بحثه، لأن هذا سوف يغير كثيراً من الأمور، وسوف يحسن مستوى البحوث المنشورة في المجلات العلمية. والتوصية بتكرار البحث تجعل نتائج البحث أكثر استمراراً وأكثر دواماً. إلا أن ذلك سوف يترتب عليه نقص شديد في عدد البحوث المنشورة لأن بحثاً واحداً قد يستغرق أطول من ضعف المدة التي يستغرقها البحث عادة.

٨ – **يجب أن تتغير مناهج الإحصاء بحيث يقل الاعتماد على التحليل الإحصائي للفرض الصفري**، وإنفاق وقت أكبر لتعليم الطلاب كيفية تحديد ما إذا كان لنتيجة ما دلالة جوهرية، وكذلك تعليمهم أسس تكرار البحوث.

٩ – **يحتاج الباحثون إلى عون أكبر من البرامج الإحصائية لحساب حجم الأثر وحدود الثقة.** وهذه آخذة في التحقق بالتدريج حيث إن كثيراً من الرزم الإحصائية تعطي أحجام الأثر في البحوث المعملية، وبقي أن تعطي تقديرات حجم الأثر في البحوث التي تستخدم الإحصاء اللامعلمي في تحليل بياناتها.

بدائل مقترحة لإجراء اختبار الفرض الصفري

اقترح جيفري لوفتس (Loftus, 1996) وجون هنتر (Hunter, 1997) استبدال أساليب إحصائية أقل تعرضاً لأخطاء تفسير البيانات بالإجراءات التقليدية لاختبار الفرض الصفري. وتدعو أكثر البدائل المقترحة إلى حساب حدود الثقة المستخدمة في تقدير معالم المجتمع، وذلك بتحديد أفضل «مدى للقيم». كما يوصي لوفتس باستخدام الباحثين للرسوم البيانية بدلاً من استخدام جداول البيانات، لأن الآثار يمكن تفسيرها بصرياً بشكل أفضل (رجاء أبو علام، ٢٠٠٦). لنفرض أننا قمنا بدراسة حول آثار قراءة القصص لأطفال ما قبل المدرسة (البنين والبنات) قبيل نومهم على نمو المفردات اللغوية لديهم في المرحلة العمرية

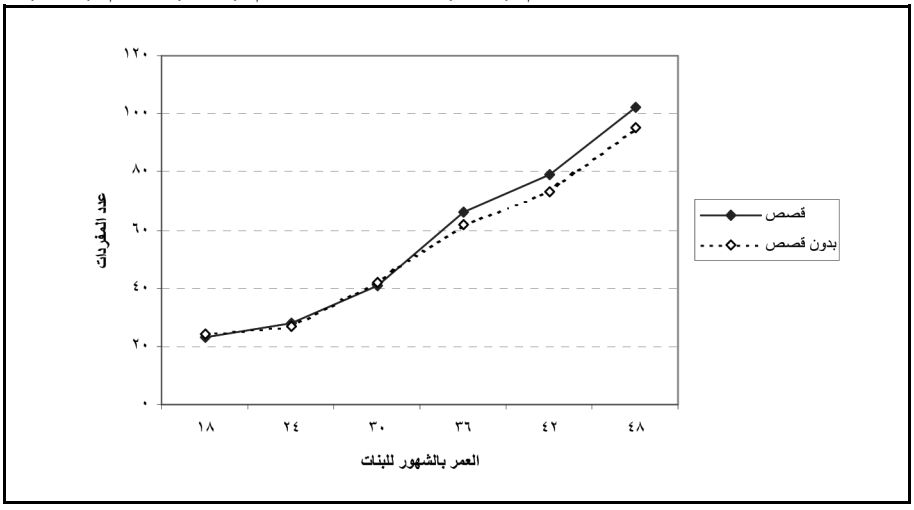
من ١٨ شهراً إلى ٤٨ شهراً. ويمكن تلخيص هذه النتائج الافتراضية كما يلي: «تشير البيانات إلى أن قراءة القصص لأطفال ما قبل المدرسة عند دخولهم السرير قبل النوم يزيد من مفرداتهم اللغوية زيادة دالة إحصائياً، وبخاصة بالنسبة للبنين الذي تزيد أعمارهم على ٣٦ شهراً». ويعتقد لوفتس أن معظم الناس يجدون تصوير البيانات بالرسم أسهل في تفسيرها. ولذلك فإننا يمكن أن نقدم البيانات التي نحصل عليها في جدول ونقرنه بالرسم البياني الذي يمثله.

جدول (١-٣)

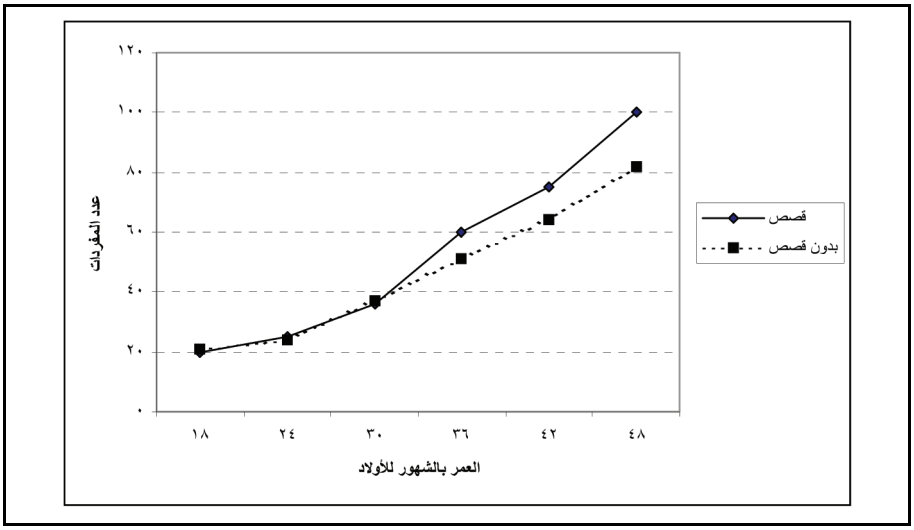
عدد المفردات في كل عمر (بالشهور) للبنين والبنات

الحالة	النوع	١٨	٢٤	٣٠	٣٦	٤٢	٤٨
قصص قبل النوم	بنين	٢٠	٢٥	٣٦	٦٠	٧٥	١٠٠
	بنات	٢٣	٢٨	٤١	٦٦	٧٩	١٠٢
بدون قصص	بنين	٢١	٢٤	٣٧	٥١	٦٤	٨٢
	بنات	٢٤	٢٧	٤٢	٦٢	٧٣	٧٣

ويمكن تصوير بيانات الجدول رقم (١-٣) في الشكلين رقم (١-٣) ورقم (١-٤).



شكل (١-٣): بيانات الجدول رقم (١-٣) الخاصة بالبنين



شكل (١-٤) بيانات الجدول رقم (١-٣) الخاصة بالبنات

استخدام حجم الأثر كوسيلة للتخلص من هذه المشكلة:

دعت نواحي القصور في إجراءات اختبار الفروض بعض المختصين في مناهج البحث بالدعوة إلى عدم استخدام الفرض الصفري. ورغم أن هذا أمر غير محتمل، على الأقل في الوقت الحاضر، إلا أن إثارة هذا الموضوع يلفت الأنظار إلى الأخطاء الشائعة التي يقع فيها كثير من الباحثين، حتى يحاولوا تجنب التفسيرات الخاطئة لاختبار الفرض الصفري. ومن المقترحات التي يمكن أن تكون إجراءات بديلة استخدام حدود الثقة والرسوم البيانية إلى جانب اختبار الفرض الصفري التقليدي. واستخدام أكثر من أسلوب إحصائي في نفس الوقت مع فهم عميق لما يرتبط به من معدلات الأخطاء لكل نتيجة إحصائية قد يساعد على تجنب التفسيرات غير السليمة ويزيد من فرص الوصول إلى استنتاجات سليمة من البيانات.

ومن الممارسات الهامة التي تساعد على تجنب تفسير نتائج اختبار الفرض الصفري تفسيراً خاطئاً استخدام حجم الأثر كجزء من نتائج الاختبار الإحصائي.

تعريف حجم الأثر

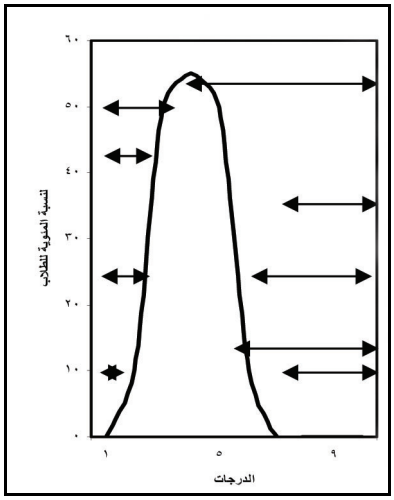
يحسن بنا قبل أن نتكلم عن معنى «حجم الأثر» أن نناقش مفهوماً أوسع وهو المقارنة بين متغيرين باستخدام معيار مقنن، وكثير من المعادلات الإحصائية ليست إلا مجرد مقارنة مقننة. فإذا أردنا أن نقارن بين ثراء شخصين فلا بد من البحث عن معيار، وأبسط معيار في هذه الحالة هو استخدام الجنيه. وبالمثل إذا أردنا أن نقارن بين متوسطي متغيرين فإننا نستخدم اختبار «ت»، ومعيارنا في هذه الحالة هو الانحراف المعياري. ومعظم العمليات الإحصائية تتبع هذا الخط من التفكير.

ويمكن اعتبار حجم الأثر فرقاً مقنناً، يمثل في أبسط صوره متوسط الفرق بين مجموعتين باستخدام درجة معيارية، هي في هذه الحالة نسبة الفرق بين المتوسطين إلى الانحراف المعياري. وهذا التعريف مستمد من مدرسة منهجية هي مدرسة «التحليل البعدي» الذي اقترحها جلاس Glass (رجاء أبو علام، ٢٠٠٦).

وحجم الأثر طريقة لتكميم فاعلية تدخل معين، أي تكميم مدى أثر المتغير المستقل (نتيجة لتدخلنا) على المتغير التابع (رجاء أبو علام، ١٩٨٩). ويتميز حجم الأثر بسهولة حسابه، وبساطة فهمه ويمكن تطبيقه على أي نتيجة استخدمنا فيها وسيلة من وسائل القياس في التربية أو العلوم الاجتماعية والسلوكية. ويسمح لنا حجم الأثر بالتحرك إلى أعد من مجرد معرفة هل يوجد فرق أو علاقة بين متغيرين. إذ يمكننا حجم الأثر من معرفة إلى أي مدى يمكن تطبيق النتائج في أكثر من مجال. كما أننا بالتأكيد على حجم الأثر نذهب إلى أبعد من معرفة هل توجد دلالة إحصائية أم لا، لأن حجم الأثر يطلعنا على أهم جوانب نتائج التجربة أو التدخل، وبذلك يوفر لنا إضافة أفضل للمعرفة. ولذلك فحجم الأثر وسيلة فعالة في تحليل النتائج وتفسيرها.

وهناك علاقة بين حجم الأثر وتحليل القوة Power Analysis. فإذا افترضنا مثلاً أنه لا يوجد لدينا سوى الفرض الصفري، وأننا رفضنا هذا الفرض، فإن النتيجة التي نذكرها هي غالباً «الفرق بين المتوسطين ليس

صفراً». ويوضح الشكل رقم (١-٥) أن هذا الفرق (ممثلاً في الأسهم) يمكن أن يكتسب أية قيمة، وقد تكون هذه القيمة كبيرة وقد تكون صغيرة. وبمعنى آخر فإن مستوى الدلالة مهما كانت قيمته ليس مؤشراً بمدى الفرق بين المتوسطين.



شكل (١-٥) - الفرق بين متوسطين ويمكن لهذا الفرق أن يكتسب أية قيمة كما هو موضح بالأسهم

ويجب ألا يكتفي الباحثون بمجرد ذكر أن الفرض الصفري خطأ أو صحيح، بل يجب عليهم كذلك الاهتمام بمدى هذا الخطأ، وبمعنى آخر إذا لم يكن الفرق صفراً، فما حجم الفرق الذي نتوقعه؟ وبتحديد حجم الأثر، وهو الحد الأدنى للفرق بين المتوسطين الذي يستحق اهتمام الباحث، يمكن للباحثين تصميم دراسة ذات قوة أعلى بدلاً من إضاعة الوقت على آثار تافهة. وتزيد قوة الاختبار الإحصائي بزيادة حجم الأثر (الفرق بين المتوسطات الصفرية والمتوسطات البديلة). وتستخدم قوة التحليل الإحصائي حجم الأثر في المجتمع. إلا أن الواقع العملي يفرض حساب حجم الأثر من بيانات العينة.

مما سبق يتبين أن حجم الأثر هو مدى الفرق بين مجموعتين من البيانات أي بين نقطتين لنفس المجتمع (أي درجة الخطأ في الفرض الصفري). ويقصد بحجم الأثر حجم الفروق الدالة إحصائياً أي الفرق في خاصية من

الخواص الرياضية (كثيراً ما تكون المتوسط) لتوزيع متغير تابع مرتبط بمستوى معين لمتغير مستقل ونفس الخاصية لتوزيع آخر يحدده مستوى آخر من نفس المتغير المستقل، وقبل حساب حجم الأثر يجب التأكد من أن هناك فرقاً دالاً إحصائياً.

ويعقد رشدي فام (١٩٩٧) مقارنة بين مستويات الدلالة الإحصائية ومقاييس حجم الأثر على النحو المبين في شكل (١-٦). ويتبين من هذه المقارنة أن كلا من اختبار الدلالة الإحصائية وحجم الأثر يكملان بعضهما البعض. فالدلالة الإحصائية تظهر وجود فرق أو ارتباط دون بيان حجم هذا الفرق أو درجة الارتباط، في حين أن حجم الأثر يبين مدى الفرق أو قوة الارتباط. ولذلك يرى رشدي فام أن حجم الأثر يكمل الدلالة الإحصائية ويفسرها، كما لو كانت الوجه الآخر للعملة.

مستويات الدلالة للاختبارات الإحصائية	مقاييس حجم الأثر
١ – تشير إلى حجم الثقة في النتيجة بصرف النظر عن حجم الفرق أو قوة الارتباط.	١ – تشير إلى حجم الفرق أو قوة الارتباط بصرف النظر عن مدى الثقة.
٢ – تتأثر مباشرة بحجم العينة.	٢ – لا تتأثر بحجم العينة.

شكل (١-٦) مقارنة بين مستويات الدلالة الإحصائية وحجم الأثر

مأخوذ بتصرف عن: رشدي فام: حجم التأثير: الوجه المكمل للدلالة الإحصائية. المجلة المصرية للدراسات النفسية، ١٦ (٧)، يونيو ١٩٩٧، ٥٧-٧٥.

وحجم الأثر في أبسط صوره هو الفرق المقنن لمتوسطي مجموعتين، أي الفرق بين متوسطين مقسوماً على الانحراف المعياري المشترك لهذين المتوسطين. ويشار عادة إلى هذا النوع من حجم الأثر بمعادلة كوهين (Cohen's d) التي يمكن صياغتها على النحو التالي:

$$d = \frac{م_٢ - م_١}{\sqrt{\frac{ع_١^٢ + ع_٢^٢}{٢}}}$$ معادلة ١-٢

حيث d حجم الأثر

$م_1$ متوسط المجموعة الأولى

$م_2$ متوسط المجموعة الثانية

$ع^2_1$ تباين المجموعة الأولى

$ع^2_2$ تباين المجموعة الثانية

وهناك طرق مختلفة لتفسير حجم الأثر ولكن أكثرها قبولاً التفسير الذي وضعه كوهين (Cohen, 1992) إذ يذكر أن ٠,٢ تعني حجم أثر ضعيف، و ٠,٥ تعني حجم أثر متوسط، و ٠,٨ تعني حجم أثر كبير. ولنفرض أننا في مثلنا السابق حول الفرق بين متوسطي الرجال والنساء حصلنا على القيم التالية:

- متوسط المجموعة الأولى (مجموعة الرجال) متوسط طول قدره ١٦٥ سم.
- متوسط المجموعة الثانية (مجموعة النساء) متوسط طول قدره ١٥٧ سم.
- انحراف معياري للمجموعة الأولى يبلغ ٦,٥.
- انحراف معياري للمجموعة الثانية قدره ٧,٢.

فإن حجم الأثر يبلغ ١,٢ وهو حجم أثر كبير ويمكن القول أن الفرق في طول النساء والرجال الذي حصلنا عليه فرق يعتد به*.

وإذا لم يكن واضحاً أي المجموعتين هي المجموعة التجريبية (أي المجموعة التي تعرضت للمعالجة) وأيها المجموعة الضابطة (التي لم تتعرض للمعالجة أو تعرضت للإجراءات العادية فقط) يمكن مع ذلك حساب حجم الأثر. وفي هذه الحالة يقيس حجم الأثر الفرق بين المجموعتين فقط، ولذلك فمن المهم أن يحدد الباحث في هذه الحالة كيف حسب حجم الأثر بالضبط محدداً موضع المجموعتين من المعادلة.

والانحراف المعياري هو بالطبع مقياس لتشتت مجموعة من الدرجات.

* توجد برامج إحصائية مجانية لحساب حجم الأثر يمكن تحميلها من شبكة الإنترنت، ومن أمثلتها: "http:\\en.wilipedia.org/wiki/Effect_size"

وفي حساب حجم الأثر يمكن استخدام الانحراف المعياري للمجموعة الضابطة، أو الانحراف المعياري المجمع من المجموعتين. وفي دراسة دوسون السابق الإشارة إليها كان الانحراف المعياري يساوي ٣,٣ ولذلك فإن حجم الأثر يبلغ:

$$(١٥,٢ - ١٧,٩) \div ٣,٣ = ٠,٨$$

ويعطي باحثون آخرون طرقاً مختلفة لحساب حجم الأثر. ومن أهم الطرق الأخرى طريقة جين جلاس وطريقة هنتر وشميت.

طريقة جلاس Glass's Approach:

يستخدم جلاس المعادلة التالية في حساب حجم الأثر الفرق بين متوسطي التجريبية والضابطة مقسوماً على الانحراف المعياري للمجموعة الضابطة (رجاء أبو علام ، ١٩٨٩) وذلك بالمعادلة التالية:

$$\Delta = \frac{م_{١} - م_{٢}}{ع^{٢}_{ض}}$$ معادلة ١-٣

حيث:

Δ حجم الأثر

$ع^{٢}_{ض}$ الانحراف المعياري للمجموعة الضابطة.

ويستخدم الانحراف المعياري للمجموعة الضابطة لأنه لا يتأثر بالمعالجة التجريبية (Glass, McGraw, & Smith, 1981). ولكن يشترط أن تكون المجموعة التجريبية كبيرة العدد فإن تقدير الانحراف المعياري للمجتمع يكون أقل دقة مما لو استخدمنا المجموعتين التجريبية والضابطة في حساب الانحراف المعياري. كما أنه في الدراسات التي لا يكون فيها مجموعة ضابطة حقيقية يصبح من الصعب تحديد أي انحراف معياري نستخدم لأن استخدام أي انحراف معياري سوف يؤثر على قيمة حجم الأثر.

ولهذه الأسباب من الأفضل حساب انحراف معياري «مجمع»، وهذا التقدير المجمع يعتبر بالضرورة متوسطاً للانحراف المعياري لكل من المجموعتين

التجريبية والضابطة. وتبين معادلة رقم (١-٤) كيفية حساب الانحراف المعياري المجمع:

$$ع_{مجمع} = \sqrt{\frac{(ن_ت - ١)ع^2_ت + (ن_ض - ١)ع^2_ض}{ن_ت + ن_ض - ٢}}$$

معادلة ١-٤

حيث:

$ن_ت$ عدد أفراد المجموعة التجريبية.

$ن_ض$ عدد أفراد المجموعة الضابطة.

$ع_ت$ تباين المجموعة التجريبية.

$ع_ض$ تباين المجموعة الضابطة.

ويتوقف استخدام الانحراف المعياري المجمع على مسلم أن الانحرافين المعياريين تقديران لنفس قيمة الانحراف المعياري في المجتمع، وأن اختلاف الانحراف المعياري للمجموعة الضابطة يرجع إلى أخطاء المعاينة. ولا يجب استخدام الانحراف المعياري المجمع إذا لم يكن من الممكن وضع هذا المسلم، أو إذا كانت القيم الفعلية للانحرافين المعياريين تختلفان اختلافاً كبيراً.

طريقة هنتر وشميت Hunter-Schmidts's Approach:

اقترح هنتر وشميت (Hunter & Schmidts, 1990) استخدام الانحراف المعياري المجمع من المجموعتين لأنه أقل تأثراً بأخطاء المعاينة من الانحراف المعياري للمجموعة الضابطة وذلك تحت شروط تساوي حجم العينة. وبالإضافة إلى ذلك يقوم هنتر وشميت بتصحيح حجم الأثر من خطأ القياس وذلك بقسمة حجم الأثر على الجذر التربيعي لمعامل الثبات.

$$\text{تصحيح أخطاء القياس} = \frac{\Delta}{\sqrt{ر_{11}}}$$ معادلة ١-٥

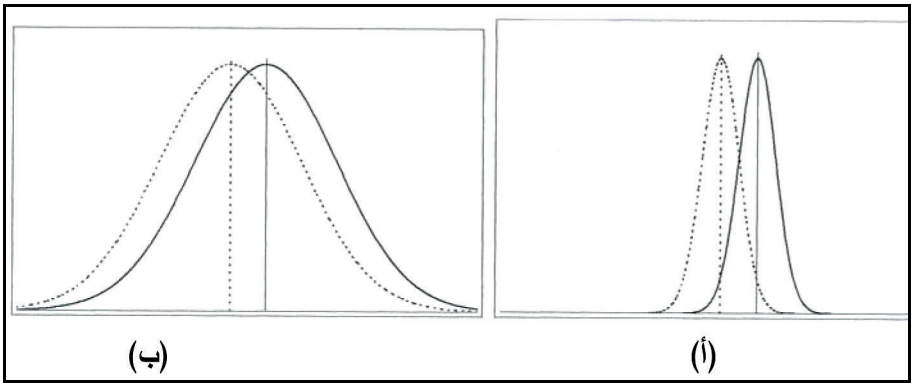
ويتوقف تفسير حجم الأثر على مجال تخصص الباحث، مثال ذلك أنه في البحوث الطبية قد يعتبر حجم أثر قيمته ٠٥, حجم أثر كبير إذا كان الدواء موضوع البحث ينقذ أرواحاً أكثر.

ونظراً لأن التخصصات المختلفة لها أحجام أثر مختلفة قال ولكوويتز وإيوان وكوهين (Welkowitz, Ewen, & Cohen, 1982) صراحة أنه لا يجب استخدام القيم التقليدية لحجم الأثر إذا تمكن الباحث من تحديد حجم الأثر المناسب لمشكلته الخاصة. ويقترح ويلكنسون وتاسك فورس (Wilkinson, & Task Force, 1999) أنه نظراً لأن حسابات القوة تكون ذا فائدة أكبر إذا حسبت قبل جمع البيانات وتحليلها، فمن المهم أن يبين الباحث أن قيم حجم الأثر قد حددت بالرجوع إلى البحوث والنظريات السابقة لاستبعاد أي شك حول تحديد هذه القيم من بيانات البحث، أو أن الباحث يستخدم حجم الأثر لتبرير النتائج التي حصل عليها من عينته.

لماذا نحتاج حجم الأثر؟

في تجربة أجراها فال دوسون لاستقصاء آثار الوقت من اليوم الذي يحدث فيه التعلم على التعلم: هل يتعلم الأطفال بشكل أفضل في الصباح أم بعد الظهر؟ وكان عدد الأطفال في التجربة ٣٨ طفلاً. وقد قسم الأطفال في مجموعة باستخدام التعيين العشوائي، واستمعت المجموعة الأولى إلى قصة في التاسعة صباحاً وأجابت بعدها على عدد من الأسئلة حول القصة. واستمعت المجموعة الثانية لنفس القصة وأجابت على نفس الأسئلة في الساعة الثالثة بعد الظهر وكانت وسيلة الاستماع شريطاً مسجلاً عليه القصة والأسئلة. وكان قياس فهم كل طفل للقصة يتم بحساب عدد الأسئلة التي أجاب عليها إجابة صحيحة (من عشرين). وكان متوسط درجة المجموعة الصباحية ١٥,٢، ومتوسط المجموعة المسائية ١٧,٩، بفرق قدره ٢,٧. ولكن ما عظم هذا الفرق؟ إذا كانت عملية القياس قد استخدمت معايير محددة مثل معيار الصف وكان الفرق يعادل نصف سنة دراسية لكان من السهل إدراك معنى هذا الفرق من الناحية التربوية من حيث أثر قراءة قصة على الفهم في فترات مختلفة من اليوم. إلا أن كثيراً من التجارب لا تستخدم مقاييس مألوفة لتسجيل نتائجها. ويضطر الباحث إلى بناء مقياس خاص لاستخدامه أو اللجوء إلى مقياس قريب مما يريد ويعدله، وفي كثير من الأحيان يكون المقياس المستخدم غريباً على الكثيرين ويصعب عليهم تفسيره.

وللتخلص من هذه المشكلة يمكن استخدام كمية التباين في الدرجات للحصول على قرائن بكمية الفرق، وإذا لم يكن هناك أي تداخل بين المجموعتين بحيث كانت درجة كل طفل في المجموعة المسائية أعلى من أي درجة من درجات الأطفال في المجموعة الصباحية، لكان الفرق بين أداء المجموعتين كبيراً. ومن ناحية أخرى إذا كان انتشار الدرجات كبيراً وكان التداخل بين المجموعتين أكبر كثيراً من الفرق بينهما، فقد يبدو الفرق أقل دلالة. ونظراً لأننا نعرف عادة درجة التباين داخل المجموعة التي ندرسها يمكننا استخدام هذا التباين كمعيار نقارن على أساسه الفرق بين المجموعتين. ويمكن تكميم هذه الفكرة أثناء حساب حجم الأثر. ويوضح هذه الفكرة شكل رقم (١-٦) الذي يبين طريقتين محتملتين للتباين بالنسبة للتداخل. فإذا كان الفرق كما هو موضح في الرسم (أ) يكون هذا الفرق كبيراً جداً، أما في الرسم (ب) فمن الصعب ملاحظة الفرق.



شكل (١-٦) – طريقتان لتصوير التداخل المحتمل في التباين بين أداء المجموعتين

تفسير حجم الأثر

ومن أهم خصائص حجم الأثر أنه يمكن تحويله بشكل مباشر إلى عبارات تصف التداخل بين عينتين على هيئة مقارنة بين المئينيات. فحجم الأثر يساوي الدرجة المعيارية (Z-Score) للتوزيع الاعتدالي المعياري (Coe, 2000). مثال ذلك أن حجم أثر قدره ٨, يعني أن درجة الطالب المتوسط في المجموعة التجريبية تزيد على درجة ٧٩٪ من أفراد المجموعة الضابطة.

ويوضح جدول رقم (١-٤) تحويلات من حجم الأثر (٨) إلى مئينيات (م) ورتبة الشخص في مجموعة ضابطة مكونة من ٢٥ فرداً (ر). مثال ذلك أنه بالنسبة لحجم أثر قدره ٦, فإن ٧٣٪ تشير إلى أن الفرد المتوسط في المجموعة التجريبية يحصل على درجة تزيد على ٧٣٪ من المجموعة الضابطة التي كانت متساوية مع المجموعة التجريبية في الاختبار القبلي. وإذا كانت المجموعة مكونة من ٢٥ فرداً فمعنى ذلك أن الفرد المتوسط (وترتيبه ١٣ في المجموعة) يكون مساوياً للفرد السابع في المجموعة الضابطة. ويلاحظ أن حجم أثر قدره ١,٦ يرفع الفرد المتوسط إلى مستوى أعلى فرد في المجموعة الضابطة في مجموعة أكبر. ويبين الجدول أحجام أثر أكبر من هذا بالنسبة لأعلى فرد في مجموعة أكبر. مثال ذلك أن حجم قدره ٣,٠ يرفع الفرد المتوسط في مجموعة قدرها ٧٤٠ إلى مستوى أعلى فرد في المجموعة.

وهناك طريقة أخرى لفهم التداخل وهي تخمين احتمال انتماء الفرد لمجموعة ما على أساس درجته فقط، أو أي درجة تتم مقارنتها. وإذا كان حجم الأثر صفراً (أي أن المجموعتين متساويتان) فإن احتمال تخمين مجال كبير للتداخل ويزيد احتمال التعرف الصحيح على المجموعتين قليلاً إلى ٥٦, . وإذا كان حجم الأثر ١ يصل الاحتمال إلى ٦٩, أي ما يزيد قليلاً على احتمال قدره ثلثين. ويبين جدول رقم (١-٤) هذه الاحتمالات في العمود (ل). ومن الواضح أن التداخل بين المجموعتين التجريبية والضابطة تداخل كبير (ولذلك يظل الاحتمال قريباً من ٥,) مع أن حجم الأثر كبير جداً.

ويعطينا روزنتال وروبين (Rosenthal & Rubin, 1982) تحويلاً أخيراً لأحجام الأثر إلى مقاييس للتداخل، إذ يقترحان أخذ عتبة «نجاح» اعتبارية (مثال ذلك الوسيط الكلي للمجموعتين مجتمعتين) وحساب النسبة المئوية لكل مجموعة فوق هذه القيمة. ويكون الفرق بين النسبتين مؤشراً لحجم الأثر.

ويلاحظ أن القيم الموجودة في جدول رقم (١-٤) تعتمد على مسلم التوزيع الاعتدالي، وتفسير أحجام الأثر باستخدام المئينيات يتأثر كثيراً بانتهاك هذا المسلم.

وهناك طريقة أخرى لتفسير أحجام الأثر وهي مقارنتها بأحجام الأثر للفروق المألوفة. مثال ذلك أن كوهين (Cohen, 1969, p. 23) يصف حجم الأثر الذي قيمته ٢, بأنه صغير ويعطي مثالاً لذلك بالفروق بين أطوال البنات في عمري ١٥ و١٦ سنة، فالفرق بينهما يصعب ملاحظته لأنه فرق «صغير»، ويصف حجم أثر قيمته ٥, بأنه متوسط، وكبير لدرجة أنه يمكن ملاحظته بالعين المجردة، وهو يشبه في هذه الحالة الفروق بين البنات في سن ١٤ سنة والبنات في عمر ١٨ سنة. ويصف كوهين حجم الأثر الذي يبلغ ٨, بأنه كبير وهو واضح تماماً وهو شبيه بالفرق بين طول البنات في عمر ١٣ سنة و١٨ سنة. كما أن كوهين يعطي مثالاً آخر وهو الفرق في نسبة الذكاء بين حملة درجة الدكتوراه والطالب المتوسط المستجد في الجامعة، فهو شبيه بحجم أثر قدره ٨, .

جدول رقم (١-٤) - تفسير حجم الأثر

حجم الأثر	نسبة المجموعة الضابطة التي تقع تحت الفرد المتوسط في المجموعة التجريبية	رتبة الفرد في مجموعة ضابطة من ٢٥ فرداً ويكون مساوياً للفرد المتوسط في المجموعة التجريبية	احتمال تخمين في أي المجموعتين يقع الفرد من معرفة درجته
ح	م	ر	ل
٠,٠	٥٠٪	١٣	,٥٠
٠,١	٥٤٪	١٢	,٥٢
٠,٢	٥٨٪	١١	,٥٤
٠,٣	٦٢٪	١٠	,٥٦
٠,٤	٦٦٪	٩	,٥٨
٠,٥	٦٩٪	٨	,٦٠
٠,٦	٧٣٪	٧	,٦٢
٠,٧	٧٦٪	٦	,٦٤
٠,٨	٧٩٪	٦	,٦٦
٠,٩	٦٢٪	٥	,٦٧
١,٠	٨٤٪	٤	,٦٩
١,٢	٨٨٪	٣	,٧٣
١,٤	٩٢٪	٢	,٧٦
١,٦	٩٥٪	١	,٧٩
١,٨	٩٦٪	١	,٨٢
٢,٠	٩٨٪	١ (أو الأول على ٤٤)	,٨٤
٢,٥	٩٩٪	١ (أو الأول على ١٦٠)	,٨٩
٣,٠	٩٩,٩٪	١ (أو الأول على ٧٤٠)	,٩٣

ومع ذلك فإن كوهين يعترف بالخطر الكامن في استخدام مصطلحات مثل «صغير» و«متوسط» و«كبير» لوصف حجم الأثر. وينتقد جلاس وزميلاه (Glass et al., 1981, p. 104) هذه الطريقة لأنها لا تمكننا من تقدير فاعلية تدخل ما إلا بمقارنتها بفاعلية تدخل معين يعطي نفس الأثر. كما أنهم يذكرون أن الأهمية العملية لأثر ما تتوقف كلية على تكاليفه النسبية وعلى فائدته. فإذا أمكننا مثلاً في التربية أن نبين أنه يمكن رفع مستوى التحصيل الدراسي بنسبة بسيطة وغير مكلفة بحجم أثر لا يزيد على ١, فإن ذلك يكون تحسناً مهماً جداً وبخاصة إذا أمكن تطبيقه على كل الطلاب على السواء، وبخاصة إذا كان هذا الأثر تراكمياً بمرور الزمن.

ويعطي جلاس وزملاؤه (Glass et al., 1981, p. 102) بأن حجم أثر قدره ١ يعادل الفرق في الأداء في اختبارات التحصيل نتيجة الحصول على تعليم يزيد عاماً في المدرسة. ومع ذلك فإن تحليلاً لاختبار إملاء عادي في بريطانيا (Vincent & Crumpler, 1997) يشير إلى أن الزيادة في العمر الإملائي من ١١ إلى ١٢ سنة يعادل حجم أثر قدره ٣,، ولكن هذه النتيجة تختلف باختلاف نوع الاختبار المستخدم. ويبدو أن توزيع درجات الثانوية (GCSE) في المواد الإجبارية (مثل الرياضيات واللغة الإنجليزية) له انحراف معياري قدره ١,٥ إلى ١,٨، ولذلك فإن تحسناً قدره درجة واحدة في امتحان GCSE يمثل زيادة في حجم الأثر من ٠,٥ إلى ٠,٧. ولذلك فإن تحقيق تحسن في الأداء في مجال المدرسة الثانوية يترتب عليه تغير في حجم الأثر حدد بقيمة قدرها ٠,٦، يترتب عليه تحسن في الأداء عملياً قدره درجة واحدة في أداء كل تلميذ في كل مادة دراسية. وبالنسبة لمدرسة يتحسن فيها أداء ٥٠٪ من التلاميذ بمقدار حوالي خمس درجات فإن هذه النسبة قد ترتفع إلى ٧٣٪. وحتى القيمة الصغيرة التي قدرها ٠,٢ لدى كوهين قد تحرز زيادة في حجم الأثر تتراوح بين ٥٠٪ و٦٨٪ – وهذا فرق تعتبره أي مدرسة ولا شك فرقاً كبيراً.

العلاقة بين حجم الأثر ومستوى الدلالة

يعطي حجم الأثر تقديراً كمياً لمدى الفرق بين متوسطي مجموعتين، ولذلك يمكن اعتباره مقياساً صادقاً لدلالة الفروق. مثال ذلك إذا وجدنا أن دراسة دوسون عن تجربة «أثر الفترة من اليوم» يمكن تطبيقها بشكل عام، فقد نسأل هذا السؤال: «ما الفرق الذي يمكن أن يحدث في تعليم الأطفال إذا تعلموا موضوعاً معيناً في المساء بدلاً من الصباح؟» وأفضل إجابة يمكن أن نحصل عليها لهذا السؤال هو استخدام حجم الأثر في تحديد الإجابة.

إلا أن كلمة «دلالة» كثيراً ما تستخدم لتعني «الدلالة الإحصائية»، وهي احتمال أن الفرق بين مجموعتين قد يرجع إلى ظروف المعاينة. فإذا أخذنا عينتين من نفس المجتمع فدائماً ما يكون هناك فرق بينهما. وتحسب الدلالة الإحصائية عادة باستخدام قيمة «ل» وهي احتمال أن نفس الفرق بين عينتين يمكن أن يحدث بالصدفة، رغم أنه قد لا يكون هناك فرق حقيقة بين المجتمعين. وبالنسبة للفرق بين متوسطي مجموعتين تحسب قيمة ل عادة باستخدام اختبار «ت». وقد جرى العرف على استخدام ل < ٠٥, (أي اقل من ٥٪)، فإن الفرق يكون كبيراً إلى الدرجة التي يمكن أن نعتبره «دالاً»، وإلا اعتبرناه «غير دال».

وهناك عدد من المشكلات في استخدام اختبارات الدلالة بهذه الطريقة. والمشكلة الرئيسية هي أن قيمة ل تتوقف على عاملين: حجم الأثر، وحجم العينة. إذ يمكننا الحصول على نتيجة دالة في حالتين: إذا كان حجم الأثر كبيراً جداً (رغم صغر حجم العينة)، أو أن تكون العينة كبيرة جداً حتى ولو كان حجم الأثر صغيراً. ولذلك فمن المهم معرفة الدلالة الإحصائية لنتيجة ما، حيث إن بدونها قد نصل إلى استنتاجات قوية من دراسات يكون حجم العينة فيها صغيراً ولا يبرر هذه الاستنتاجات. إلا أن الدلالة الإحصائية لا تخبرنا عن أهم شيء وهو حجم الأثر. والطريقة التي نتغلب بها على هذا الارتباك هو إعطاء قيمة حجم الأثر مصحوباً بهامش «الخطأ المحتمل» له، أو «حدود الثقة» به.

هامش الخطأ في تقدير حجم الأثر

من الواضح أنه إذا حسبنا حجم الأثر من عينة كبيرة جداً فإن قيمته سوف تكون أكثر دقة مما لو حسبناه من عينة صغيرة. ويمكن تكميم «هامش الخطأ» (Margin of error) هذا باستخدام فكرة حدود الثقة (Confidence interval) التي تعطينا نفس المعلومات الموجودة في اختبار الدلالة: فاستخدام حدود ثقة قدرها ٩٥٪ يعادل استخدام مستوى دلالة ٥٪. ولحساب حدود ثقة قدرها ٩٥٪ فإننا نفترض أن القيمة التي حصلنا عليها (مثال ذلك ٨, هي القيمة الحقيقية ولكي نحسب درجة التغير فيها إذا أخذنا عينات أخرى باستمرار من نفس الحجم (أي عينات مختلفة حجم كل منها ٣٨ طفلاً). ونحصل من كل ١٠٠ عينة من هذه العينات الافتراضية الجديدة، على ٩٥ عينة تعطينا تقديرات لحجم الأثر في «حدود ثقة ٩٥٪». وإذا احتوت حدود الثقة على صفر فإن هذا يعني نفس الشيء كما لو كانت النتيجة غير دالة إحصائياً. ومن ناحية أخرى إذا كان الصفر خارج المدى تكون النتيجة دالة عند مستوى ٥٪. واستخدام حدود الثقة طريقة أفضل لتوصيل هذه المعلومات لأنها تبقي التأكيد على حجم الأثر – وهو الجزء المهم من المعلومات التي نريد إعطاءها – أكثر من قيمة ل.

وقد أعطانا هدجز وأولكين (Hedges & Olkin, 1984) معادلة لحساب حدود الثقة لحجم الأثر. فإذا كان تقدير حجم الأثر للعينة (ح) يكون موزعا توزيعاً اعتدالياً بانحراف معياري قدره:

$$ع[ح] = \sqrt{\frac{ن_ت + ن_ض}{ن_ت \times ن_ض} + \frac{ح^2}{ن_ت + ن_ض}}$$ معادلة ١-٦

حيث:

$ن_ت$ عدد أفراد المجموعة التجريبية

$ن_ض$ عدد أفراد المجموعة الضابطة

وعلى هذا تتراوح حدود الثقة ٩٥٪ لقيمة ح بين

$ح - ١,٩٦ \times ع[ح]$ و $ح + ١,٩٦ \times ع[ح]$ معادلة ١-٧

فإذا كان لدينا في إحدى التجارب نت = نض = ٢٠ وكانت ح = ٨, تكون

ع[ح] = √(١, + ٠٠٨,) = ,٣٣

وعلى هذا الأساس تتراوح حدود الثقة ٩٥٪ بين ١٥, و ١,٤٥. وبمعنى آخر يمكن القول إن نتائج التجربة تؤكد تفوق المجموعة التجريبية على المجموعة الضابطة وأن الفرق بينهما فرق كبير جداً.

دمج المعلومات عن مجموعة من أحجام الأثر:

من أهم مزايا استخدام حجم الأثر فائدته عند القيام بتكرار تجربة من التجارب. ففي هذه الحالة يمكن دمج أحجام الأثر التي نحصل عليها من كل دراسة لنحصل على حجم أثر عام يعتبر أفضل تقدير لحجم الأثر. وتعرف هذه الطريقة في تركيب نتائج التجارب في تقدير واحد لحجم الأثر «بالتحليل البعدي». وقد ظهر هذا المفهوم نتيجة للدراسات التي قام بها إحصائي تربوي هو جين جلاس Gene Glass. ويستخدم الآن على نطاق واسع في التربية والعلوم الأخرى، وبخاصة الطب والعلوم الاجتماعية. ويشتمل كتاب «مناهج البحث في العلوم النفسية والتربوية» على تعريف مختصر بفكرة التحليل البعدي (رجاء أبو علام، ٢٠٠٦).

إلا أن التحليل البعدي يستطيع أن يفعل أكثر من مجرد إعطاء قيمة عامة لحجم الأثر رغم أهمية ذلك. مثال ذلك أننا لو حصلنا من عدد من التجارب تتشابه في التدخل الذي تقوم به على أحجام أثر بعضها كبير وبعضها الآخر صغير، فإن دمج أحجام الأثر هذه يكون ذا فائدة محدودة إذا قلنا إن حجم الأثر العام «متوسط». والأهم من هذا والأكثر فائدة أن نفحص الدراسات الأصلية للحصول على الفروق بين تلك الدراسات التي أعطت حجم أثر كبير والدراسات التي أعطت حجم أثر صغير، ومحاولة فهم العوامل التي قد تكون مسؤولة عن هذه الفروق. ولذلك فإن أفضل تحليل بعدي هو الذي يسعى إلى

معرفة العلاقات بين أحجام الأثر للتدخلات المختلفة والسياق والتصميم اللذين وجد فيهما حجم الأثر (Rubin, 1992).

وأهمية تكرار البحوث للحصول على شواهد أفضل أمر أصبح معروفاً للجميع. ففي تجربة دوسن عن الوقت من النهار وجد أن حجم الأثر كبير جداً يبرر الأهمية الإحصائية والتربوية. ونظراً لأننا نعلم أن الأطفال تم تعيينهم عشوائياً في مجموعتين يمكننا أن نثق في أنه من غير المحتمل أن تكون الاختلافات المبدئية بين المجموعتين أساساً لاختلاف النتائج. كما أن استخدام الاختبار القبلي قبل التجربة يجعل هذا الأمر أقل احتمالاً. إلا أننا لا نستطيع أن نستبعد إمكانية أن تكون الفروق راجعة إلى بعض الخصائص المرتبطة بهؤلاء الأطفال بالذات في هذه التجربة. مثال ذلك إذا لم يحصل جميع الأطفال في المجموعة الصباحية على طعام إفطار في يوم التجربة، فقد يكون هذا سبباً في أدائهم الضعيف. ولا يمكن في هذه الحالة تعميم النتائج على جميع أطفال المدارس الذين يفترض أن يحصل معظمهم على طعام إفطار. كما أن النتائج يمكن أيضاً أن تعتمد على عمر الأطفال. فالأطفال في تجربة تراوحت أعمارهم بين ٧ و٨ سنوات، ومن المحتمل جداً أن تختلف النتائج أو تنعكس إذا استخدمنا أطفالاً أكبر (أو أطفالاً أصغر). وهذا يمثل خطورة تنفيذ سياسة ما على أساس دراسة وحيدة. فالثقة في إمكانية تعميم النتائج تزداد لا تأتي إلا بعد تكرار البحوث تكراراً واسعاً.

ومن أهم نتائج قدرة التحليل البعدي على دمج نتائج الدراسات، أنه حتى الدراسات الصغيرة يمكن أن تضيف للمعرفة. فالتجربة التي قد يقوم بها مدرس واحد قد تتجاوز ٣٠ طفلاً. وإذا لم يكن الأثر عظيماً فإن دراسة بهذا الحجم الصغير للعينة لا تعطي غالباً نتائج دالة إحصائياً، وتكون التجربة حسب المنطق الإحصائي غير جديرة بالمجهود الذي بذلت فيه. ولكن إذا توفر لدينا عدد من مثل هذه الدراسة وقمنا بدمجها باستخدام التحليل البعدي، فإن الاحتمال الأكبر أن نحصل على نتائج على جانب مرتفع من الدلالة الإحصائية.

كما أنها سوف تكتسب القوة الناجمة عن إجراء الدراسات في بيئات مختلفة (مما يزيد الثقة في التعميم) وفي بيئات تنتمي للعالم الواقعي (مما يقوي الثقة في تطبيق النتائج بشكل عام، حيث تبين جدوى مثل هذه السياسة).

ولا بد أن ننوه هنا من خطورة دمج النتائج غير المؤكدة. فإن الحصول على رقمين أو أكثر فإننا نستطيع دائماً حساب متوسطها، إلا أن هذه الأرقام إذا كانت أحكام أثر من تجارب تختلف جذرياً في نتائج المقاييس المستخدمة، فقد لا تكون النتائج ذات معنى. فقد يكون الحصول على قيم مختلفة لأحجام الأثر أمراً مغرياً للغاية بمعاملتها ككل واحد، وهنا نفقد البصيرة في أصولها. ولا شك أن هناك كثير من الدراسات البعدية التي تجعل المرء يشك في تجاور أحجام الأثر التي أفرزتها.

العوامل الأخرى المؤثرة في حجم الأثر:

رغم أن حجم الأثر بسيط ويمكن تفسيره بسهولة كمقياس للفاعلية إلا أنه قد يكون حساساً لعدد من المؤثرات الزائفة مما يستدعي الحرص في استخدامه، ومن هذه العوامل:

١ – محدودية المدى:

لنفرض أننا كررنا تجربة أثر الوقت من اليوم في مدرستين مختلفتين كأن يكون أحدها مدرسة للمتفوقين والثانية مدرسة تجارية. فإذا تم تعيين الأطفال في مجموعات عشوائية صباحية ومسائية، فقد يكون الفرق بينها واحداً في كل حالة، فقد يكون متوسطا المجموعتين في مدرسة المتفوقين مرتفعاً ولكن الفرق بين المجموعتين قد يكون هو نفس الفرق بين المجموعتين في المدرسة التجارية. ولكن من غير المحتمل أن يكون الانحراف المعياري واحداً في الحالتين. فقد يكون انتشار الدرجات في مدرسة المتفوقين محدوداً بعكس الدرجات في المدرسة التجارية. وسوف يكون لهذا الوضع بالطبع تأثيره على حساب حجم الأثر، فإن محدودية انتشار الدرجات الذي يميز مدرسة المتفوقين قد تؤدي إلى الحصول على حجم أثر أكبر كثيراً من الذي نحصل عليه في المدرسة التجارية.

ويجب عند حساب حجم الأثر استخدام الانحراف المعياري للمجتمع بأكمله، حتى تكون المقارنات عادلة. إلا أننا كثيراً ما نصادف حالات لا يمكننا فيها الحصول على قيم غير محدودة، إما عملياً أو من حيث المبدأ. مثال ذلك عند حساب حجم الأثر لتجربة على طلاب الجامعة، أو التلاميذ ذوي صعوبات في القراءة، يجب أن نتذكر أن هذين المجتمعين من المجتمعات المحدودة. وعند حساب حجم الأثر يجب أن نلفت الانتباه إلى هذه الحقيقة، وإذا أمكن تكميم هذه المحدودية فقد يكون من الممكن تعديل البيانات وفقاً لها. ويجب كذلك القيام بأي مقارنة لأحجام الأثر محسوبة على أساس قيم المجتمع كاملاً.

٢ – التوزيعات غير الاعتدالية

يتوقف تفسير حجم الأثر الموضح سابقاً على أن توزيع كل من المجموعة التجريبية والمجموعة الضابطة توزيع اعتدالي. وغني عن البيان أنه إذا لم يكن هذا السلم صحيحاً، فقد يكون من الصعب إعطاء مقارنات دقيقة بين حجم أثر يقوم على المنحنى الاعتدالي وحجم أثر آخر يقوم على توزيعات غير اعتدالية.

٣ – ثبات القياس

والعامل الثالث الذي قد يساعد على إعطاء أحجام أثر زائفة معامل الثبات الذي يعبر كما هو معروف عن دقة القياس واستقراره، والذي كثيراً ما يعرف على أساس الدرجة التي يتفق بها مقياسان يقيسان نفس الشيء، أو الحصول على اتفاق بين مرتي إجراء مقياس معين. ويمكن التعبير عن هذه الفكرة بطريقة أخرى: تتكون درجة المقياس الذي يقيس مخرجاً معيناً من عنصرين، هما الدرجة الحقيقية للصفة التي نقيسها، والخطأ الذي يشوب القياس. والمشكلة هي أن التباين في الدرجات التي نقيسها لعينة معينة (أي انحرافها المعياري) يتوقف على كل من التباين في الدرجات وكمية الخطأ في القياس.

ولذلك عند تفسير حجم الخطأ من المهم أن نعرف معامل ثبات الأداة التي حصلنا منها على البيانات التي كانت أساساً لحساب حجم الأثر. ولهذا السبب يجب أن نعطي دائماً ثبات الأدوات التي نستخدمها كأساس لتفسير البيانات بشكل عام، وعلى الأخص عند استخدام حجم الأثر. ومن الممكن نظرياً تصحيح عدم الثبات، مما يمكننا من الحصول على حجم أثر يقوم على المقياس لو كان ثابتاً. إلا أن هذه ممارسة يجب الابتعاد عنها لأنها تؤدي بنا إلى الحصول على قيم مرتفعة لحجم الأثر.

تلخيص:

ناقشنا في الفصل الأول القضية الجدلية في علم النفس حول الاختبارات الإحصائية، والأحداث التي قادت إليها. وبخاصة فيما يتعلق بالتوجهات التي اتخذتها رابطة علم النفس الأمريكية وضمنتها الطبعة الخامسة من دليل النشر للرابطة. وتبين أن الاستخدام المستمر للدلالة الإحصائية باعتبارها الطريقة الوحيدة لاختبار الفروض واتخاذ القرارات الإحصائية في العلوم السلوكية أمر غير محتمل. كما ناقشنا في هذا الفصل مشكلة الدلالة الإحصائية وأثرها في إعاقة تطور علم النفس والبحوث النفسية، وبيّنا أن المشكلة لا تتعلق بالدلالة الإحصائية في حد ذاتها، لأننا إذا قمنا بشكل روتيني بتحديد فروض معقولة، ودرسنا عينة عشوائية، وراجعنا المسلمات المتعلقة بالتوزيع العيني، وحددنا القوة الإحصائية، وفهمنا المعنى الحقيقي للقيم ل، لا تكون هناك أية مشكلة في استخدام اختبارات الدلالة الإحصائية كوسيلة أساسية في الاستدلال الإحصائي. ولكن الواقع غير ذلك، وهذا ما حاولنا توضيحه، مبينين كيف أن بعض المعتقدات الشائعة حول معنى الدلالة الإحصائية وما يرتبط بها من مفاهيم، هو الذي أدى إلى ما يوجد من تجاوزات وأخطاء في تحليل نتائج البحوث.

ورأينا أن الحل الأفضل حتى الآن هو إضافة حجم الأثر إلى نتائج البحوث التجريبية. وبيّنا مزايا استخدام حجم الأثر كوسيلة لتفسير مستوى

الدلالة الإحصائية. وقد قدمنا عدة تعريفات لحجم الأثر، ومعظم هذه التعريفات مرتبط بطريقة حساب حجم الأثر. وأول معادلة معروفة لحساب حجم الأثر هي معادلة جلاس التي تقوم على قسمة الفرق بين متوسطي المجموعة التجريبية والمجموعة الضابطة على الانحراف المعياري للمجموعة الضابطة على اعتبار أنها لم تتأثر بالمعالجة التجريبية.

وأنهينا هذا الفصل بذكر بعض العوامل المؤثرة في حجم الأثر وهي محدودية المدى، والتوزيع غير الاعتدالي، وثبات المقياس المستخدم في جمع البيانات.

الفصل الثاني
حساب حجم الأثر في الإحصاء المعلمي

يزخر الإحصاء المعلمي بأنواع كثيرة من حجم الأثر ومعظمها يرتبط بعمليات إحصائية معينة. وسوف نتناول فيما يلي بعض طرق حساب حجم الأثر مصنفة حسب العمليات الإحصائية المتربطة بها. وقد سبق لنا الكلام على بعض طرق حساب حجم الأثر وهي أيضاً تنتمي للإحصاء المعلمي أثناء تفسير المقصود بحجم الأثر. وقد ظهر حجم الأثر كما ذكرنا سابقاً في معرض الكلام على التحليل البعدي على يد جلاس، وكان التركيز في البداية على حجم الأثر بالنسبة للمتوسطات في البحوث التجريبية وبخاصة تلك التي تناولتها البحوث البعدية. وأهم أحجام الأثر التي تناولناها حين ذاك هي:

* معادلة كوهين (d).
* معادلة جلاس (Δ).
* طريقة هنتر وشميت في معالجة الخطأ المعياري

وفيما يلي نتناول حجم الأثر وفقاً للاختبارات الإحصائية الأكثر استخداماً في العلوم النفسية والتربوية.

حجم الأثر لمعاملات الارتباط

معامل الارتباط الثنائي الأصيل:

عندما تكون (س) و(ص) متغيرين متصلين فإننا نستخدم معادلة ارتباط بيرسون المألوفة والتي تعطينا تقديراً لمعامل الارتباط (ر). وفي هذه الحالة تكون (ر) تقديراً واضحاً لحجم الأثر ممثلاً في حجم الارتباط (ر)، وتكون علامته مؤشراً باتجاه الارتباط. ولكننا نستخدم في بعض الحالات متغيرات تابعة متصلة ومتغيرات مستقلة قطعية. وفي هذه الحالة لا يمكننا استخدام معادلة بيرسون بل لا بد من استخدام معادلات أخرى تسمح باستخدام متغيرات قطعية

حقيقية (مثل النوع: ذكور وإناث)، ونقصد بكلمة طبيعية أننا ننشئ المتغير بمجرد تصنيف أعضائه وفق معيار معين كالنوع مثلاً، أما المتغيرات القطعية غير الطبيعية فهي المتغيرات التي ننشئها بتحويل مقياس متصل إلى مقياس قطعي باختيار نقطة قطع معينة. وهذه النقطة عادة نقطة اعتبارية قد ينجم عن اختيارها مشكلة ما في قياس المتغير القطعي الناتج. ويؤدي استخدام متغير قطعي كان متغيراً متصلاً إلى انخفاض معامل ارتباطه بالمتغيرات الأخرى نتيجة لانخفاض التباين. وفي هذه الحالة لا يستطيع المتغير القطعي تحقيق أقصى معامل ارتباط كان من الممكن أن يصل إليه لو لم يحول إلى متغير قطعي (Hunter & Schmidt, 2004).

وإجراءات حساب معامل الارتباط بين متغير قطعي ومتغير متصل هي ببساطة ترميز الأفراد في المتغير (أ) أو المتغير (ب) ترميزاً رقمياً، فمثلاً أعضاء المجموعة (أ) يحصلون على الرمز (١) وأعضاء المجموعة (ب) يحصلون على الرمز (٢)، أو العكس. وعلى ذلك فإن ملف البيانات سوف يتكون من عمودين: العمود (س) وتتكون قيمه من قيمتين فقط هما ١ (لأفراد المجموعة أ) و٢ (لأفراد المجموعة ب) أما العمود (ص) فإن قيمه هي الدرجات التي حصل عليها الأفراد في المتغير المتصل (ص). وقد نستخدم رمزين آخرين غير ١ و ٢ للمتغير س (٢ و ٤ مثلاً) إلا أن هذا لن يؤثر في معامل الارتباط الناتج بين المتغيرين. فإذا حصلنا على معامل ارتباط مرتفع موجب فمعنى هذا أن أعضاء المجموعة ذات الرمز الأعلى حصلوا على درجات أعلى في المتغير المتصل، وإذا كان معامل الارتباط سالباً فمعنى هذا أعضاء المتغير المتقطع الذين حصلوا على القيمة الدنيا كانت درجاتهم أعلى في المتغير المتصل.

ويطلق على الارتباط بين متغير قطعي ومتغير متصل معامل الارتباط الثنائي الأصيل (Point-biserial correlation)، ويمكن الحصول على هذا الارتباط بإدخال قيم المتغيرين في الحاسب الآلي وحساب الارتباط بينهما باستخدام أي برنامج يقوم بحساب الارتباط ولا داعي للبحث عن برنامج معين للارتباط الثنائي.

مثال:

لنفرض أننا نريد حساب معامل الارتباط الثنائي الأصيل بين ٥٠ طالباً من الذكور والإناث في نسبة الذكاء، فما علينا إلا أن نعد ملف البيانات على النحو السابق ذكره، وفي هذه الحالة ستكون قيم العمود (س) ١ أو ٢ (للذكور مثلاً و٢ للإناث أو العكس)، وتكون القيم في العمود (ص) نسب ذكائهم التي حصلوا عليها في اختبار الذكاء. فإذا حصلنا مثلاً على معامل ارتباط قدره ٧٠, فمعنى هذا أن الإناث حصلن (في المتوسط) على درجات أعلى في نسب الذكاء من الذكور، وبالعكس إذا كان معامل الارتباط -٧٠, فمعنى هذا أن الذكور حصلوا (في المتوسط) على نسب ذكاء أعلى.

وبرنامج الارتباط الذي يقيس الارتباط (الارتباط الثنائي الأصيل «ر ث أ» في هذه الحالة) يختبر في نفس الوقت الفرض الصفري بأن ر ث أ = صفر، ويعطي مستوى (ل) لهذا الاختبار. وهناك معادلة تبين العلاقة بين معامل الارتباط واختبار (ت)، كما أن مستوى (ل) الذي نحصل عليه من معامل الارتباط عند إجراء اختبار ذي ذيلين لاختبار الفرض الصفري فـ$_0$: ر$_{المجتمع}$ = صفر هو نفسه مستوى الدلالة الذي نحصل عليه من اختبار ت لاختبار ف: م$_1$–م$_2$ = صفر، ومعامل الارتباط الذي حصلنا عليه (٧٠,) يختلف اختلافاً دالاً إحصائياً عن صفر، ولذلك فإننا نتوقع أن تكون قيمة ت بين متوسطي المجموعتين دالة إحصائياً عند نفس المستوى.

وقيم (ر) و (ر ث أ) قيم متحيزة تحيزاً سالباً (أي أنها تنزع إلى أن تعطي تقديراً أقل لقيم الارتباط في المجتمع وإن كان بدرجة قليلة. والتحيز يكون أكثر إذا كان معامل الارتباط في المجتمع أقرب إلى $\pm$ ٥, وبخاصة في حالة العينات الأصغر. وليس لذلك أهمية كبيرة إذا كان حجم العينة أكبر أي يزيد مثلاً على ١٥ (Hedges & Olkin, 1985). وعندما يبلغ حجم العينة ٢٠ يقل التحيز بالنسبة للأخطاء المقربة (Hunter & Schmidt, 2004). ويعطي جريسوم وكيم (Grissom & Kim, 2005, p. 72) المعادلة التالية كتقدير تقريبي للتحيز:

$$ر_{تقريبية} = ر + \frac{ر (١ - ر^{٢})}{٢ (ن - ٣)}$$ معادلة ٢-١

إلا أنه من النادر استخدام هذا التصحيح لأن التحيز عادة ما يكون قليلاً جداً وبخاصة إذا كان حجم العينة كبيراً.

المسلمات عن معامل الارتباط ومعامل الارتباط الثنائي الأصيل:

هناك ثلاث توزيعات للمتغير (ص) يجب الاهتمام بها في حالة معامل الارتباط الثنائي الأصيل وهي: توزيع ص للمجموعة (أ)، وتوزيع (ص) للمجموعة (ب)، والتوزيع الكلي للمتغير (ص) والناتج عن ضم المجموعتين (أ) و(ب). والتوزيعان الأوليان توزيعان شرطيان (Conditional Distributions) للمتغير ص (لأنهما مشروطان بما إذا كنا نتناول قيم المتغير (ص) بالنسبة للمجموعة الأولى أي س = أ أو المجموعة الثانية أي س = ب). ويطلق على التوزيع الكلي للمتغير (ص) التوزيع الهامشي المتغير (ص).

ويتطلب اختبار (ت) العادي مسلم التجانس بين تباين المتغيرين، وتتغلب بعض المعادلات على عدم التجانس باستخدام كل من تباين المتغير الأول والمتغير الثاني بشكل منفصل في مقام معادلة (ت) بدلاً من استخدام التباين المجمع من المتغيرين. (Homoscedasticity) أي أن تباين المتغيرين متساوي في المجتمعين اللذين حصلنا منهما على المجموعتين. أما إذا كان تباين المجتمع الأول لا يساوي تباين المجتمع الثاني فإن المقام في معادلة ت (الخطأ المعياري للفرق بين المتوسطين) لن يكون صحيحاً، مما قد يؤدي إلى قوة إحصائية أقل ودقة أقل في حدود الثقة (Wilcox, 2003). وكذلك إذا كان اختبار ت العادي مستخدماً وكانت قيمة (ل) وقيمة (ل) الفعلية غير المعروفة أقل من ٠١, فإن هذه النتيجة لن تكون في الواقع مؤشراً بارتباط غير صفري، ولكنها قد تكون مؤشراً بعدم التجانس (Heteroscedasticity). وعدم التجانس هو نوع آخر من التبعية بين المتغيرين (س) و (ص)، أي التبعية بين تباين (ص) وقيم (س).

وإذا لم تشر النتائج إلى الدلالة الإحصائية لمعامل الارتباط ولم يتضمن كذلك قيمة (ت) التي تطابق القيمة التي حصلنا عليها (ر ث أ) فإننا نحول قيمة (ر ث أ) إلى (ت) باستخدام المعادلة رقم (٢-٢) التالية:

$$ت = ر\ ث\ أ \left[\frac{ن - ٢}{١ - ر\ ث\ أ^{٢}} \right]$$ معادلة ٢-٢

حيث ن هي المجموع الكلي للمشاركين وفي هذه الحالة نستخدم جدول (ت) عند درجات حرية = ن-٢ في الصفوف. للتأكد من الدلالة الإحصائية لقيمة (ر ث أ). وإذا لم يكن هناك صف خاص بدرجات الحرية المطلوبة فإننا نقدر القيمة المطلوبة.

والاعتدال الثنائي مسلم لمعامل الارتباط. إلا أنه رغم أن الالتواء في الاتجاه المعاكس للمتغيرين (س) و (ص) يخفضان من القيمة القصوى للارتباط، إلا أن عدم الاعتدال في ذاته لا يشكل مشكلة لمعامل الارتباط (Cohen, Cohen et al., 2003). والواقع أنه عند استخدام الارتباط الثنائي الأصيل فإن المتغير (س) القطعي لا يمكن أن يكون توزيعه اعتدالياً. وتوزيع المتغير (س) في هذه الحالة هو مجرد تكدس للقيمتين (١) و (٢). إلا أن وجود قيمة متطرفة حتى ولو كانت قيمة واحدة وكذلك التوزيعات مذنبة الأطراف (أي التوزيعات التي تتكدس فيها الدرجات عند أطرافها) بشكل مختلف عن التوزيع الاعتدالي يمكن أن تؤثر على قيم الارتباط (ر) و(رالمجتمع) وحدود الثقة لهذا الارتباط (Wilcox, 2003). وقد يساعد استخدام عينات كبيرة الحجم في مثل هذا الموقف ولكن ليس في كل الظروف. وحتى التغيرات الطفيفة في شكل التوزيع العام للمتغير (ص) في المجتمع يمكن أن يغير من قيمة الارتباط تغييراً كبيراً.

ويجب في هذه الحالة التمييز بين الاختلاف في شكل التوزيعات الشرطية في التكوين الذي نقيسه (مثال ذلك الاستعداد الحقيقي أو أحد عوامل الشخصية) في المجتمعين والاختلاف في التوزيعين الشرطيين لمقياس هذا التكوين (أي درجات اختبار الاستعداد أو اختبار عامل الشخصية)، في العينتين (Cohen, Cohen et al., 2003). وفي حالة وجود اختلاف في شكل التوزيع للتكوين

في المجتمعين فإن الانخفاض الناتج في الحد الأعلى لأقل من (١) لن يشكل مشكلة لمعامل الارتباط في المجتمع، لأنها ظاهرة طبيعية. إلا أن هناك مشكلة بالنسبة لمعامل الارتباط الثنائي الأصيل حيث يعطي قيمة أقل للارتباط إذا اختلف شكل توزيع العينتين في حين أن شكل توزيع المجتمعين لا يختلف أو إذا كان توزيع العينتين مختلفاً أكثر في شكله من توزيع كل من المجتمعين.

ويجب أن تشتمل تقارير البحوث التي تتناول معامل الارتباط أو معامل الارتباط الثنائي الأصيل على شكل التبعثر للمتغيرين، وعلى عبارات تحذيرية حول المؤثرات المحتملة للدرجات المتطرفة، أو تكدس الدرجات عند ذيل التوزيعات، أو عدم تجانس التباين. وفي حالة معامل الارتباط الثنائي الأصيل فإن شكل التبعثر قد يبين اختلاف التباين بالنسبة للمتغيرين الشرطيين والتواء في التوزيع الهامشي للمتغير (ص) أو تجانس تباين المتغيرين وعدم وجود التواء في التوزيع، لأن عدم تجانس التباين والتواء التوزيع كثيراً ما يكونا معاً.

اختلاف حجم العينتين:

إذا كانت $ن_1 \neq ن_2$* في بحث تجريبي فقد تقل قيمة معامل الارتباط الثنائي الأصيل مما يترتب عليه خفض في تقدير قيمة الارتباط في المجتمع. وتزيد درجة هذا الانخفاض مع زيادة عدم التناسب بين حجم العينتين وارتفاع قيمة معامل الارتباط في المجتمع. ويلاحظ أنه عندما يكون حجم العينتين غير متساو فهناك فرصة كبيرة في التواء المتغيرين (س) و(ص) في الاتجاه المعاكس لأن العينات غير المتساوية في حالة معامل الارتباط الثنائي الأصيل يترتب عليها التواء في المتغير (س). وسبق أن ذكرنا أن الالتواء في س و ص في الاتجاه المعاكس يخفض من القيمة المطلقة للارتباط. ويمكن حساب معامل تصحيح لهذا الانخفاض في معامل الارتباط الثنائي الأصيل بالمعادلة ٢-٣:

$$ر_{مص} = \left\{ \frac{ن-٢}{١-ر_{ثأ}^{٢}} \right\}$$ معادلة ٢-٣

* استخدم الكاتب الرمز ن ليعبر عن حجم المجتمع، والرمز $ن$ للتعبير عن حجم العينة.

حيث رمص = معامل الارتباط المصحح، و أ = [٢٥, \ل (ل-١)$^{١/٢}$]، و (ل) و (ل-١) نسبتي الحجم الكلي للعينة في كل مجموعة (Hunter & Schmidt, 2004). مثال ذلك إذا كانت نمجتمع = ١٠٠ و ن١ = ٦٠ و ن٢ = ٤٠، فإن ل = ٦٠\١٠٠ = ٦, و ل-١ = ٤٠\١٠٠ = ٤, . وبالطبع ليس من المهم إذا كانت ن١ أو ن٢ هي المرتبطة بـ ل أو ل-١، لأن ل (ل-١) تساوي (ل-١) ل. ويلاحظ أنه إذا كانت العينات مختلفة في بحوث تدريبية فقد يحصل الباحثون الذين يدرسون نفس المتغير (س) ونفس المتغير (ص) على قيم مختلفة لمعامل الارتباط الثنائي الأصيل غير المصحح ويرجع ذلك إلى اختلاف ن١ \ ن٢ من دراسة لأخرى. ولذلك لا يجب تحليل معامل الارتباط الثنائي أو إدخاله في دراسة للتحليل البعدي في مثل هذه الظروف إلا بعد تصحيحه باستخدام المعادلة (٢-٣).

انخفاض الثبات:

انخفاض الثبات عامل آخر قد يؤدي إلى انخفاض معامل الارتباط الثنائي الأصيل مما يضعف تقديرات حجم أثر الفروق المقننة. ويقصد هنا «بانخفاض الثبات» المدى الذي تعكس به الدرجة خطأ القياس وهو شيء آخر غير القيمة الحقيقية التي نهدف إلى قياسها. ويؤدي خطأ القياس من موقف لآخر إلى عدم تجانس القياس، وعدم قابليته للتكرار، بمعنى أن الاختبار غير الثابت لا يمكن الاعتماد عليه في الحصول على درجات متجانسة من الاختبار أو المقياس رغم عدم اختلاف الصفة التي نريد قياسها.

ونظراً لأن معامل الارتباط الثنائي الأصيل يقيس التباين المشترك بين (س) و(ص) أي الدرجة التي يرجع فيها التباين الحقيقي في (ص) إلى التباين الحقيقي في (س). فإن عدم ثبات الاختبار يؤدي إلى خفض معامل الارتباط. وهذا يرجع بالطبع إلى أخطاء القياس. وهناك معادلة لتصحيح معامل الارتباط وغيره من القيم التي تقدر حجم الأثر بسبب أخطاء القياس في المتغير التابع (Schmidt & Hunter, 2004). ويترتب على هذه المعادلة حجم أثر معدل من المتوقع أن يحدث إذا أمكن قياس المتغير ص بثبات كامل.

$$ح_{معدلة} = \frac{ح}{ر_{ص\ ص}^{\frac{1}{2}}} \qquad \text{معادلة ٢-٤}$$

حيث:

$ح_{معدلة}$ = حجم الأثر المعدل.

ح = حجم الأثر.

$ر_{ص\ ص}$ = معامل ثبات الاختبار ص.

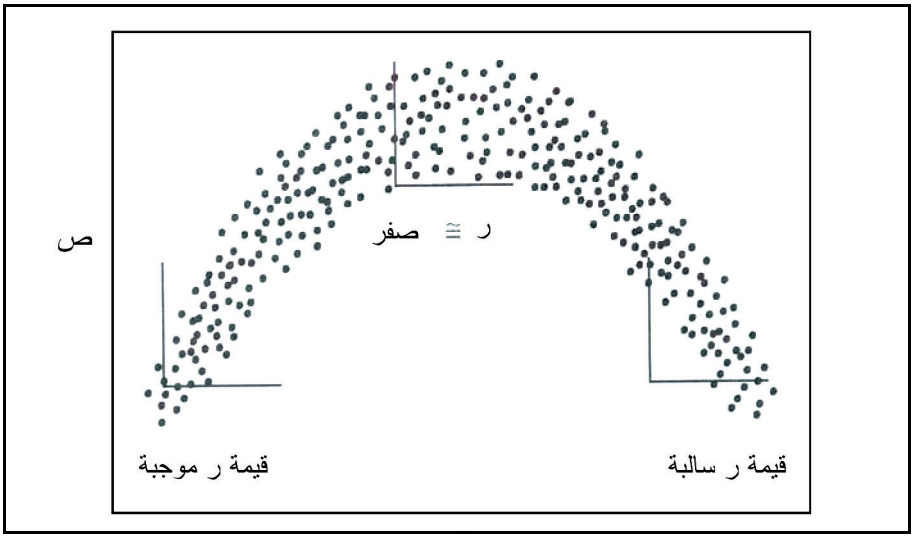
وفي حالة الدراسات غير التجريبية يصحح ثبات المتغير (س) بوضع $ر_{س\ س}$ مكان $ر_{ص\ ص}$ في المعادلة. ويمكن كذلك وضع $(ر_{ص\ ص}\ ر_{س\ س})^{\frac{1}{2}}$ لتصحيح ثبات الاختبارين س و ص دفعة واحدة.

على أنه من النادر اليوم استخدام معادلة تصحيح معامل الثبات ربما لانخفاض الاهتمام بالقياس النفسي بشكل عام وربما لاعتقاد الباحثين بأنه من الخير استخدام معامل الثبات كما حصلوا عليه لأن هذا أمر واقعي ولا داعي للدخول في عمليات إحصائية نظرية.

محدودية المدى:

أحد العوامل الأخرى التي قد تؤدي إلى انخفاض معامل الارتباط الثنائي الأصيل هي المدى المحدود، ويعني هذا عادة استخدام عينات يقل فيها مدى التغير في المتغير التابع عن مدى تغيره في المجتمع الذي نعمم عليه النتائج. ومن أمثلة المدى المحدود إذا أجرينا بحثاً يتلقى فيه المرضى علاجاً لمدة أقصاها ٢٦ أسبوعاً باستخدام علاج معين في عيادة أحد الأطباء النفسيين. ويقارن باحث آخر يدرس أثر فترة العلاج مجموعة تعرضت للعلاج لمدة ١٦ أسبوعاً، بمجموعة ضابطة لم تتلق أي علاج. وهذا نوع من المدى المحدود المباشر لأن الباحث يعرف مسبقاً أن مدى المتغير المستقل محدود. وفي حالات أخرى يكون المدى محدوداً لأن المشاركين الموجودين وجدوا بالصدفة، وهذا النوع من المدى المحدود غير مباشر. وقد ذكر هنتر وشميدت (Hunter & Schmidt, 2004) بعض الطرق التي يمكن بها تصحيح محدودة المدى المباشر وغير المباشر.

وأحياناً ما تؤدي محدودية المدى إلى زيادة معامل الارتباط، رغم أن هذه حالة غير طبيعية. فلو كان عندنا علاقة منحنية فقد تؤدي محدودية المدى فيه إلى زيادة معامل الارتباط وتغير في إشارته عندما ينخفض المدى نتيجة لإزالة الحالات المتطرفة. لنفرض أن لدينا مجتمعاً العلاقة بين متغيرين من متغيراته علاقة منحنية وحصلنا على العينة (س) منه ولها مدى محدود، ففي هذه الحالة يتوقف حجم الارتباط وإشارته في العينة على القيم المنخفضة أو المتوسطة أو المرتفعة للمتغير (س). ويوضح شكل رقم (٢-١) هذه الحالة، ولنتذكر دائماً أن (ر) تعكس العنصر الخطي في علاقة بين متغيرين.



شكل (٢-١): إذا كانت العلاقة بين (س) وص) علاقة منحنية في المجتمع، فإن اقتصار مدى (س) في العينة على قيم منخفضة فقط، أو متوسطة فقط، يمكن أن يؤثر على حجم وإشارة (ر) في العينة.

وقد يرجع المدى المحدود إلى أخطاء في التصميم التجريبي. فمن غير المعقول أن يجري باحث تجربة عن طريقة جديدة في تدريس مادة معينة على مجموعة تجريبية مع استخدام مجموعة ضابطة لا تدريس هذه المادة إطلاقاً. وفي هذه الحالة نحصل على مدى محدود نتيجة لانخفاض مدى أداء المجموعة الضابطة التي لا تدري شيئاً عن المقرر المستخدم في التجربة.

قيم حجم الأثر الصغيرة والمتوسطة والكبيرة:

لا يوجد حدود نظرية لحجم الأثر للفروق المقننة نظراً لأنه عبارة عن درجة معيارية (ز)، كما أن قيمة الارتباط في المجتمع قد تتراوح نظرياً أيضاً بين -١ و+١. وقد يرغب البعض في الحصول على إحساس أفضل بمستوى حجم الأثر حتى يمكنهم أن يضعوا القيم الجديدة التي تقابلهم في مكانها. ومن النادر أن نحصل في البحوث النفسية أو التربوية أو السلوكية على فروق مقننة تزيد على ٢,٠٠ (مع تجاهل العلامة)، كما أن معاملات الارتباط الثنائي الأصيل نادراً ما تزيد على ٠,٧٠، والواقع أن القيم التي نحصل عليها في كلتا الحالتين تقل كثيراً عن القيمتين المذكورتين.

وتصنيف قيم حجم الأثر في صغيرة ومتوسطة وكبيرة تصنيف اعتباري إلى حد ما. ومثل هذا التصنيف كما قال كوهين (Cohen, 1988) يستخدم مصطلحات نسبية تتعلق بالعوامل المرتبطة بمجال البحوث وإلى الدرجة التي يتم بها ضبط المتغيرات الخارجية، وثبات درجات مقياس المتغير التابع. مثال ذلك أن حجم الأثر بقيمة معينة قد يعتبر كبيراً نسبياً إذا حدث في مجال بحثي في علم النفس الاجتماعي مثلاً، في حين أن نفس القيمة قد تعتبر صغيرة نسبياً إذا حدثت في مجال بحثي أكثر انضباطاً مثل علم النفس العصبي. وحتى نفس القيمة قد ينظر إليها باحثان في نفس المجال البحثي بطريقتين مختلفتين.

ويعترف كوهين كما ذكرنا سابقاً بأن فئات التقدير الثلاثة التي وضعها معايير تقريبية. وهو يحدد هذه القيم على النحو التالي:

المستوى الصغير

$\Delta \geqslant ٠,٢$

ر ث أ $\geqslant$ ٠,١٠

ويعطي أمثلة لذلك التفوق البسيط في نسب ذكاء غير التوائم مقارنة بالتوائم، والمتوسط الأعلى قليلاً في بنات السادسة عشرة مقارنة ببنات الخامسة عشرة. وبعض الاختلافات بين الرجال والنساء في مقياس وكسلر للراشدين.

وقد وجد ليبسي وويلسون (Lipsey & Wilson, 1993) أن توزيع الربع الأدنى (٢٥٪) لتقديرات الفروق المقننة لحجم الأثر للمقاييس النفسية والتربوية والسلوكية كان حوالي $\Delta \geqslant$ ٠,٣٠ وهذه القيمة تعادل معامل ارتباط ثنائي أصيل قدره ≥ ١٥, مما يؤيد تقريبا معايير كوهين.

وحدد كوهين المستوى المتوسط بأنه

Δ = ٠,٥

ر ث أ = ٠,٢٤٣

وقد وجد ليبسي وويلسون ما يؤيد ذلك إذ حصلا على وسيط قدره ٠,٥ ، وهو ما ينطبق أيضاً على ما وجد في علم النفس الإرشادي وفي علم النفس الاجتماعي. وقد أعطى كوهين أمثلة لهذا المستوى من حجم الأثر في الفرق بين متوسط طول البنات في سن الثامنة عشرة مقارنة بالبنات في سن الرابعة عشرة، والزيادة في متوسط نسبة ذكاء موظفي الأعمال الكتابية مقارنة بمتوسط نسب ذكاء العمال شبه المهرة. ويذكر جريسوم (Grissom, 1996) أنه حصل على وسيط قدره ٤٤, لقيم Δ عندما قارن مجموعتين تجريبية وضابطة في البحوث في مجال علم النفس العلاجي.

أما المستوى المرتفع في رأي كوهين فهو:

$\Delta \leqslant$ ٠,٨

ر ث أ ≤ ٠,٣٧١

ويعطي كوهين أمثلة لهذا المستوى في الفرق بين متوسط طول بنات الثامنة عشرة مقارنة بطول بنات الثالثة عشرة. والفرق في متوسط نسب ذكاء حملة دكتوراه الفلسفة مقارنة بطلاب السنة الأولى في الجامعة.

ولا ينطبق تصنيف حجم الأثر في صغير ومتوسط وكبير بالضرورة على مستوى الدلالة العملية لأثر ما، إذ يتوقف مستوى الدلالة العملية على نوع البحث وعلى خبرة الباحث وحكمه على النتائج. وأحياناً ما تقاس الدلالة العملية للأثر قياساً ملموساً، ولا يرتبط حكم الباحث على الدلالة العملية

بالضرورة بمستوى حجم الأثر، فقد يكون حجم الأثر صغيراً ومع ذلك يرى الباحث أن النتائج لها دلالة عملية. وهذا يتوقف على عوامل أخرى تدخل في تقدير الباحث. وكذلك لا يجب أن نتسرع في الحكم على معاملات الارتباط التي تختلف عن الصفر عندما لا نجد تفسيراً منطقياً للعلاقة المفترضة وبخاصة في حالة عدم تكرار الدراسة. فمثل هذه النتائج عندما تكون العينة صغيرة الحجم قد تكون راجعة لعوامل الصدفة البحتة.

عرض حجم الأثر ذي الحدين:

قدم روزنتال وروبين (Rosenthal & Rubin, 1982) جدولاً للمعاونة في تفسير أي نوع من معاملات الارتباط بما فيها معامل الارتباط الثنائي الأصيل. وقد أطلقا على هذا الجدول «إظهار حجم الأثر ذي الحدين binomial effect size display» وأطلق على هذا الجدول اختصاراً BESD. وكان الغرض الأساسي من هذا الجدول بيان الأهمية العملية الممكنة لقيمة صغيرة لمعامل الارتباط. وليس الغرض من هذا الجدول تقدير حجم الأثر ولكن الغرض منه شرح فرضي لما يمكن أن يستدل عليه من حجم الأثر من قيمة أي معامل ارتباط. وقد أصبح جدول BESD أداة شائعة بين الباحثين، رغم أن له بعض نواحي القصور كما سنرى.

وقد نشأ عرض حجم الأثر ذي الحدين (BESD) من حقيقة أن معامل الارتباط (ر) يمكن تطبيقه على البيانات التي يكون فيها (س) و(ص) متغيرين قطعيين. وفي هذه الحالة فإن اسم (ر) للعينة هو معامل فاي ϕ. مثال ذلك أن المتغير (س) قد يحتوي على مستويين هما: المعالجة (أ) في مقابل المعالجة (ب)، ويتكون (ص) أيضاً من مستويين هما الفئة التي تحسن أفرادها بعد العلاج والفئة التي لم يتحسن أفرادها بعد العلاج. وفي هذه الحالة نعطي المتغير (س) الرمز ١ للمعالجة (أ)، والرمز ٢ للمعالجة (ب)، كما يحدث في حالة حساب معامل الارتباط الثنائي الأصيل. إلا أننا هنا في حالة حساب معامل فاي نُرَمِّز أيضاً المتغير ص بنفس الطريقة بحيث يحصل على ١ الأفراد الذين تحسنوا، وعلى ٢ الأفراد الذين لم يتحسنوا. ومعامل فاي هو ببساطة الارتباط بين المتغيرين (س) و (ص) اللذين يتكونان من مجموعتين من الأرقام هي ١ و٢.

وبمعنى آخر فإن العمود (س) يتكون من الرقمين ١ و ٢، وكذلك العمود (ص) يتكون من رقمين هما ١ و ٢. وعندما نستخدم برنامج الارتباط مع هذين المتغيرين فإنه يحسب معامل الارتباط الذي هو في الواقع معامل ارتباط فاي. وسوف نناقش في الفصل الثالث معامل ارتباط فاي كتقدير لحجم الأثر مع طريقة أخرى لحسابه عند الكلام على حساب حجم الأثر لمربع كاي.

وإذا افترضنا أن نأ = نب وبمعاملة قيمة (ر) باعتبارها قيمة فاي، يمكن تكوين جدول فرضي (BESD) يشرح طريقة أخرى لتفسير قيمة (ر). لنفرض أننا حصلنا على معامل ارتباط قدره ٠,٢٠ وهي قيمة متواضعة. وبالرغم من أن (ر) حسبت على أساس المتغير المتصل (ص)، للحصول على منظور مختلف لهذه النتيجة فإن جدول (BESD) يتظاهر للحظة أن المتغيرين (س) و(ص) متغيران قطعيان، وأن معامل الارتباط الذي قيمته ٠,٢٠ حصلنا عليه من معامل فاي. ويصور جدول (٢-١) كيف تكون النتائج لو أن فاي = ٠,٢٠ وأن نأ = نب = ١٠٠. إن معامل ارتباط قدره ٠,٢٠ قد يبدو للبعض أنه لا يمثل حجم أثر ذا أهمية كبيرة، إلا أننا نلاحظ من جدول ٢-١ أن ر = ٠,٢٠ أصبحت فاي = ٠,٢٠ (أساس جدول ٢-١)، ومثل هذه النتائج تشير إلى أن نسبة أكبر (بلغت ٢٠٪) من المشاركين تحسنوا تحت العلاج (أ) من العلاج (ب) (أي أن ٦٠٪ – ٤٠٪ = ٢٠٪).

ويلاحظ من جدول ٢-١ أن ٦٠ من ١٠٠ (٦٠٪) من المشاركين في العلاج (أ) قد صنفوا كذلك لتحسن علاجهم مما كان عليه قبل بدء العلاج، وأن ٤٠ مما مجموعه ١٠٠ مشارك صنفوا بأنهم تحسنوا باستخدام العلاج (ب) مما كان عليه قبل العلاج. ويطلق على هاتين النسبتين «نسبتا نجاح العلاج». وتظهر النتائج الآن على اعتبار أنها نسبة النجاح الناتجة لجدول (BESD) ذات تأثير كبير. وإذا كان عدد الأفراد الذين يتم علاجهم يبلغ آلاف أو ملايين المرضى في موقف ممارسة إكلينيكية حقيقية وأعطوا العلاج (أ) بدلاً من العلاج القديم (ب) فإنهم بسبب النتائج الواردة في جدول ٢-١ (ولنفترض أن معامل ارتباط فاي الذي قيمته ٠,٢٠ هو معامل ارتباط فاي في المجتمع) نكون قد قمنا بتحسين ٢٠٪ إضافية من الآلاف أو الملايين العديدة زيادة عن الذين تم علاجهم بالعلاج (ب).

جدول (٢-١)

عرض حجم الأثر ذي الحدين (BESD)

نوع المعالجة س	تحسن المشارك (ص=١)	عدم تحسن المشارك (ص=٢)	المجموع
المعالجة أ (س=١)	٦٠	٤٠	١٠٠
المعالجة ب (س=٢)	٤٠	٦٠	١٠٠

وتزداد الأهمية الاجتماعية والطبية لهذه النتائج إذا زادت خطورة المرض الذي يتم علاجه (وذلك مع افتراض أن تكاليف العلاج ليست باهظة أو أن هناك خطورة في الأمر). والطريقة المبينة في جدول ٢-١ هي طريقة مفترضة لتفسير معامل الارتباط أو معامل الارتباط الثنائي الأصيل وذلك عن طريق الافتراض أن كلاً من (س) و(ص) متغيران قطعيان، وكان معامل الارتباط الناتج معامل فاي، مما يؤدي إلى الحصول على جدول ٢ × ٢ متناسق المجموع، ويدلنا على الزيادة في نسبة النجاح التي نحصل عليها، لو استخدمنا العلاج (أ) بدلاً من العلاج (ب). ويلاحظ أن البيانات التي نحصل عليها في معظم الحالات بيانات تكونت في جدول ٢ × ٢ ولكنها ليست دائماً من نوع البيانات التي تحقق المطلوب للحصول على جدول (BESD).

وللحصول على نسبة النجاح (تحسن العلاج) لأي معامل ارتباط (ر) يرمز إلى المتغير بالرمز س = ١، نستخدم ١٠٠ [٠,٥٠ + (ر\٢)]٪. ونظراً لأن مجموع النسبتين في جدول BESD يجب أن يكون ١٠٠٪ فإن نسبة الفشل في الصف س=١ يجب أن تكون بالطبع ١٠٠٪ – نسبة النجاح في الصف. ونسبة النجاح بالنسبة للصف س-٢ يمكن الحصول عليها باستخدام ١٠٠ [٠,٥٠+(ر\٢)]٪=٦٠٪، ونسبة الفشل المقابلة لها = ١٠٠-٦٠=٤٠. ويترتب على زيادة قيمة ر زيادة الفرق بين نسبتي النجاح للعلاجين. وبالتحديد فإن الفرق بين النسبتين يمكن الحصول عليه من (١٠٠ر)٪. ولذلك فإنه حتى قبل بناء جدول BESD فإننا نعلم أنه عندما تكون ر=٢٠, فإن الفرق بين نسبتي النجاح سوف تكون [١٠٠(٢٠,)]٪=٢٠٪، وذلك إذا وضعت البيانات الأصلية في جدول BESD بشكل مناسب. ويلاحظ أنه

يمكن أيضاً بناء جدول BESD وتقدير الفرق في نسبة النجاح لقيمة الفرض البديل لمعامل الارتباط وذلك بدءاً من استخدام المعادلة (٢-٥) التالية والخاصة بالفرض البديل لمعامل الارتباط:

$$\text{ف}_{\text{ب}} = \frac{\text{ر}^{٢}}{\left(١+٣\text{ر}^{٢}\right)^{\frac{١}{٢}}} \qquad \text{معادلة ٢-٥}$$

حيث:

ف$_{ب}$ = الفرض البديل لمعامل الارتباط.

وفي مثالنا الحالي تبلغ قيمة معامل الارتباط ٠,٢٠ ولذلك فإن:

ر$_{ب}$ = ٢ (٠,٢٠) \ [١+٣ (٠,٤٠)]$^{\frac{١}{٢}}$ = ٠,٨٨

عيوب BESD

هناك بعض نواحي القصور في BESD وما ينتج عنها من تقديرات عن الفرق بين نسبة نجاح علاجين. فأولاً إن الفرق في نسب النجاح من BESD لا يعادل فاي إلا إذا كانت نسبة الفشل العامة أي ٥٠٪ وإذا كانت المجموعتان متساويتين في العدد. ونتيجة لذلك نحصل على جدول يقال إن له هامشين متساويين. وإذا نظرنا إلى جدول ٢-١ نجد أن هذه الحالة تنطبق عليه لأن العينتين لهما نفس الحجم. والنسبة (الهامشية) الكلية للنجاح تساوي (٦٠+٤٠) \٢٠٠=٥٠٪، والنسبة (الهامشية) الكلية للفشل (٤٠+٦٠) \٢٠٠=٥٠٪. إلا أنه يلاحظ أن النقد الأول ليس نقصاً أو قصوراً في BESD، ولكنه مجرد جزء من تعريفه.

وعندما تتحقق معايير BESD فإن الفرق الناتج في نسبتي النجاح تتفق مع المجتمع الفرضي الذي تمثله بيانات BESD. ولكن هل النتائج تتفق مع المجتمع الذي حصلنا منه على البيانات الأصلية الفعلية التي وضعت في الجدول، أو متفقة مع أي مجتمع فعلي؟ إن المجتمع الذي يتفق مع البيانات التي وضعت في جدول BESD مجتمع تلقي فيه النصف العلاج (أ)، والنصف الآخر العلاج (ب)، ونجح العلاج مع نصف المجتمع ولم ينجح مع النصف الآخر. وهذا القصور أيضاً هو في نظر البعض مظهر كامن من مظاهر تعريف BESD.

إن الحالات التي تتوفر عنها البيانات هي الحالات التي تتفق مع هذا التعريف لأن هذا الأمر موجه إلى الذين ينتجون هذه البيانات (أي الباحثين). ومن المعقول في مثل هذه الحالات للجداول الأصلية الثنائية (٢ × ٢) مقارنة نسب النجاح للعلاجين القائمين على بيانات فعلية بدلاً من البيانات الفرضية لجدول BESD. ففي مثل هذه الحالات قد يقبل الفرد المخاطرة النسبية أو أحجام الأثر الأخرى للبيانات في جدول ثنائي والتي سوف نناقشها في الفصل التالي. ولنفرض أن نسبة النجاح ونسبة الفشل في المجتمع الذي سوف نعمم عليه البيانات لا تساوي ٥٠٪، ففي هذه الحالة فإن الفرق القائم على جدول BESD بين نسبتي النجاح والفشل في العينة سوف يتحيز نحو زيادة تقدير الفرق بين نسبتي النجاح للمعالجتين في ذلك المجتمع (Grissom and Kim, 2005).

وقد تنشأ مشكلات إضافية عندما تكون المقاييس المستخدمة لقياس المتغير التابع مقاييس متصلة وليست قطعية. ففي هذه الحالة يقوم الباحثون عادة بتقسيم بيانات العينتين عند الوسيط العام لكل عينة لتكوين مجموعتي نجاح وفشل متساويتين، وبذلك تتحقق معايير جدول BESD. إلا أن تعريف النجاح أو الفشل بالنسبة للدرجات الواقعة فوق أو تحت الوسيط قد لا يكون أمراً واقعياً. مثال ذلك أن كل فرد مكتئب حصل على درجة تقل عن الوسيط في مقياس للاكتئاب وعولج قد يحصل على الدرجة التي تضعه في فئة الناجحين. كما أنه ليس أمراً جيداً تقسيم متغير في مجموعتين لأن ذلك يقلل القوة الإحصائية. ولقد تناول هنتر وشميت تصحيح الانخفاض الذي يحدث في معامل الارتباط نتيجة لتقسيم متغير متصل في فئتين (Hunter & Schmidt, 2004).

معامل التحديد

يطلق على مربع معاملات الارتباط المختلفة في العينة «معامل التحديد» Coefficient of Determination للعينة، وقد استخدم مربع معامل الارتباط على نطاق واسع لتقدير معامل التحديد للمجتمع. وتستخدم بعض المصطلحات للتعبير عن معامل التحديد وإن كان بعضها تعبيراً غير سليم. والتفسير العادي لمربع معامل الارتباط (معامل التحديد) في المجتمع هو نسبة التباين في المتغير

التابع التي يمكن أن نعزوها أو ننسبها للمتغير المستقل. ويمكن التعبير عن ذلك أيضاً بالتباين المشترك بين المتغيرين التابع والمستقل، أو التباين الذي يمكن التنبؤ به من المتغير المستقل إلى المتغير التابع.

ومن الممكن البرهنة تحت شروط ومسلمات معينة أن معامل الارتباط في المجتمع هو نسبة:

أ - جزء تباين الدرجات في المتغير التابع والذي يرتبط بالتباين في المتغير المستقل (التباين المفسر).

ب - التباين الكلي للدرجات (المرتبط وغير المرتبط بالمتغير المستقل).

وبالنسبة للمثال الأول من هذين المثالين المتطرفين إذا كانت ر المجتمع=صفر، ر^{٢}المجتمع = صفر فإنه لا يمكن تفسير تباين أي من درجات بتباين المتغير المستقل. ومن ناحية أخرى إذا كانت ر المجتمع=١، ر^{٢} المجتمع = ١ فإن كل تباين في الدرجات يمكن أن ننسبه للتباين في المتغير المستقل. وبمعنى آخر عندما يكون معامل التحديد صفراً فإننا بمعرفة قيم المتغير المستقل فإننا نعرف صفر٪ من القيم التي يمكن التنبؤ بها بالمتغير التابع، ولكن إذا كان معامل التحديد ١,٠٠ فإننا نعرف نسبة قدرها ١٠٠٪ من القيم التي يمكن بها التنبؤ بالمتغير التابع. ففي هذه الحالة الأخيرة تقع جميع نقاط شكل التبعثر التي تربط كلاً من (س) و(ص) على خط مستقيم (خط الانحدار أو خط التنبؤ في هذه الحالة). ولا يوجد أي تباين في درجات ص عند أي نقطة من نقاط س، مما يجعل التنبؤ بدرجات ص من المتغير من درجات س تنبؤاً تاماً.

ويلاحظ أن مصطلحي تحديد وتفسير قد يكونا مصطلحين غير دقيقين إذا كنا نتكلم على بحوث غير تجريبية لأن ذلك قد يعني علاقة سببية، وهذه العلاقة لا يمكن الجزم بها خارج إطار البحوث التجريبية. فمعامل الارتباط في البحوث غير التجريبية يعكس التباين المشترك بين المتغيرين (س) و(ص)، وليس

علاقة سببية ففي هذه الحالة إذا كان معامل التحديد ٤٩, فإن هذا يعني أن ٤٩٪ من تباين الدرجات (وليس حجمها) يمكن تفسيره بالتباين في المتغير (س). وتفسير التباين المشترك في الدرجات لا يعني تحديد حجم الدرجات.

ولا يمكن الكلام على التباين في المتغير المستقل (المعالجة) باعتباره محدداً أو سبباً في الدرجات إلا في البحوث التي يتم فيها تعيين المجموعات عشوائياً في مجموعة تجريبية ومجموعة ضابطة. ويتم فيها كذلك ضبط المتغيرات الخارجية والدخيلة. وعلى هذا فمن الأفضل عدم استخدام مصطلح تحديد في البحوث غير التجريبية ونستخدم بدلاً من ذلك عند الكلام على $ر^{٢}$ نسبة التباين في الدرجات المرتبطة بالتباين في المتغير المستقل. ومع ذلك فهناك رأي يقول إن تربيع معامل الارتباط للحصول على معامل تحديد ليس مناسباً أيضاً في البحوث غير التجريبية، وأن هذه القيمة يستخدمها بعض الباحثين كتقدير لحجم الأثر في البحوث التجريبية.

ويذكر جريسوم وكيم (Grissom & Kim, 2005) أربعة أسباب للتخلي عن استخدام $ر^{٢}$ هي:

١ – تربيع القيم الصغيرة أو المتوسطة لمعامل الارتباط (أي القيم الأقرب إلى الصفر منها إلى الواحد) والتي نجدها عادة في البحوث النفسية والتربوية والسلوكية بشكل عام ينتج عنه الحصول على قيم أقل لمربع الارتباط. فإذا حصلنا على ر=٢٤, فإن مربع هذه القيمة يبلغ ٠٦, ومثل هذه القيم الصغيرة تقود إلى تقدير حجم الأثر تقديراً أقل من قيمته العملية.

٢ – السبب الثاني الذي يجعل بعض الأوساط الإحصائية تستغني عن $ر^{٢}$ كمؤشر بحجم الأثر أن مربع الارتباط ليس له اتجاه على العكس من (ر) أو (ر ث أ) فلا يمكن أن يكون سالباً. مثال ذلك أنه إذا كنا نقوم بدراسة حول العلاقة بين الذكور والإناث، وأعطينا الذكور (س=١)، فعندما تكون (ر ث أ) سالبة فمعنى ذلك أن متوسط الذكور كان أعلى.

أما بالنسبة لمربع الارتباط فلا يمكن الحكم على أي النوعين حصل على متوسط أعلى. لأن مربع معامل الارتباط لا بد أن يكون موجباً سواء كان الارتباط سالباً أو موجباً. ولذلك لا يمكن في بحوث التحليل البعدي تحديد متوسط معاملات الارتباط من مجموعة من الدراسات كان معامل الارتباط في بعضها موجباً وفي بعضها الآخر سالباً. ولذلك لا بد أن يقوم الباحثون الأصليون عندما يذكرون مربع الارتباط أن يذكروا معه معامل الارتباط نفسه، وفي هذه الحالة يستطيع القائم بالتحليل البعدي تحديد اتجاه المتوسط بناء على قيمة معامل الارتباط.

٣ – السبب الثالث لعدم تفضيل بعض الباحثين لمربع معامل الارتباط عند دراسة حجم الأثر توفر مقاييس أخرى لحجم الأثر غير مربع معامل الارتباط لم تكن معروفة من قبل أو أن شيوع استخدام مربع معامل الارتباط طغى عليها. وأولئك الذين ينادون باستخدام مقاييس أكثر قوة من معامل ارتباط بيرسون، يرون أن حجم معامل التحديد يمكن أن يتأثر بالظروف التي كانت سائدة والتي يمكن أن تؤثر على معامل الارتباط، مثل العلاقة المنحنية التي ليست لها صلة بمعامل ارتباط بيرسون.

٤ – فيما يتعلق بالقيم الصغيرة لمربع معاملات الارتباط في العلوم الإنسانية فنحن نعلم أن العامل المحدد لسلوك الإنسان عامل متعدد، لأن هناك كثيراً من الفروق الوراثية والبيئية بين الناس، ومثل هذه الفروق لا بد أن تؤثر على تباين المتغيرات التابعة المستخدمة في البحوث السلوكية مما يجعل من الصعب على الباحث أن ينسب ما يراه من تغيرات في المتغير التابع إلى متغير مستقل وحيد.

حجم الأثر لاختبار ت

نتناول في هذا الجزء حجم الأثر عند دراسة الفروق بين متوسطين، حيث نختبر الفروض المتعلقة بثلاثة أنواع من الفروق نستخدم فيها اختبار (ت) للفروق بين متوسطين، وهي:

* اختبار (ت) لعينة واحدة.
* اختبار (ت) لعينتين مستقلتين.
* اختبار (ت) لعينتين مرتبطتين أو متكافئتين.

اختبار (ت) لعينة واحدة

تهدف هذه العملية الإحصائية إلى اختبار الفرض باختلاف متوسط التوزيع عن قيمة ثابتة يطلق عليها القيمة الاختبارية Test Value (رجاء أبو علام، ٢٠٠٣). ويجب أن يكون لكل حالة درجة على متغير واحد هو المتغير الذي نريد اختباره.

والقرار الرئيسي في هذا النوع من الاختبار هو اختيار القيمة الاختبارية. وتمثل هذه القيمة عادة درجة محايدة. والأفراد الذين يحصلون على درجات أعلى من نقطة الحياد يعطون مسمى معيناً، في حين أن الأفراد الذين يحصلون على درجات تقل عن هذه النقطة يعطون مسمى مختلفاً. أما أولئك الذين يحصلون على نقطة الحياد بالضبط فلا يعطون أي مسمى.

ونظراً لأن القيمة الاختبارية هي أهم خاصية في اختبار (ت) لعينة واحدة، أصبحت القيمة الاختبارية تميز هذا الاختبار عما عداه من التطبيقات في طريقة اختيار هذه القيمة. ويمكن للقيمة الاختبارية أن تكون واحدة مما يلي:

* النقطة الوسيطة في المتغير الذي نختبره.
* القيمة المتوسطة في المتغير المراد قياسه بناء على نتائج البحوث السابقة.
* تغير مستوى الأداء في اختبار ما.

أسس اختبار (ت) لعينة واحدة

هناك عدة مسلمات ترتبط باختبار (ت) لعينة واحدة وحجم الأثر المقترن به:

مسلمات اختبار (ت) لعينة واحدة

يرتبط مسلمان باختبار (ت) لعينة واحدة، هما:

المسلم رقم ١: أن يكون المتغير الذي نختبره موزعاً توزيعاً اعتدالياً في المجتمع الذي ندرسه.

وعندما يكون لدينا عينة متوسطة الحجم فقد ينتهك هذا المسلم، وقد ينتهك أيضاً مع العينات الأكبر، إلا أن الاختبار يظل يعطي نتائج دقيقة نسبياً. ويحتاج الأمر إلى الحصول على عينة أكبر للحصول على نتائج صادقة نسبياً إذا كان توزيع المجتمع بعيداً عن أن يكون اعتدالياً. إلا أنه في معظم الحالات يكفي الحصول على عينة مكونة من ٣٠ فرداً للحصول على قيم دقيقة لـ (ل).

المسلم رقم ٢: أن العينة المسحوبة من المجتمع سحبت بشكل عشوائي وأن درجات المتغير موضوع الاختبار الإحصائي مستقلة عن بعضها البعض.

يعطي اختبار (ت) لعينة واحدة نتائج غير دقيقة إذا انتهك المسلم المتعلق باستقلالية الدرجات عن بعضها البعض.

حساب حجم الأثر

يمكن حساب حجم الأثر (Δ) كما يظهر في المعادلة التالية:

$$\Delta = \frac{\text{متوسط الفروق}}{\text{الانحراف المعياري}}$$ معادلة ٢-٥

وإذا لم يكن لدينا متوسط الفروق والانحراق المعياري يمكننا حساب حجم الأثر من قيمة (ت) حيث يساوي حجم الأثر في هذه الحالة:

Δ = ت ÷ $\sqrt{\text{ن}}$ حيث ن هي العدد الكلي للعينة.

وتساعدنا Δ على تقويم الدرجة التي يختلف بها متوسط الدرجات في المتغير الذي نختبره عن القيمة التي نختبرها وذلك في وحدات من الانحراف المعياري. وإذا كانت قيمة Δ تساوي صفراً فإن ذلك يعني تساوي متوسط الدرجات والقيمة التي نختبرها. وكلما زاد الفرق عن صفر يزداد حجم الأثر. ومن الممكن أن تتراوح قيمة حجم الأثر بين $\pm\infty$ ($\pm$ ما لا نهاية). وبغض النظر عن علامة – أو + فإن القيمة ٠,٢ تمثل حجم أثر صغير ويعتبر حجم الأثر ٠,٥ قيمة متوسطة، أما ٠,٨ فهي تمثل حجم أثر كبير وذلك حسب التفسير الذي وضعه كوهين (Cohen, 1989).

اختبار (ت) لعينتين مستقلتين

يدرس هذا الاختبار الفروق بين متوسطي مجموعتين مستقلتين. وفي هذا الاختبار يجب أن يكون لكل فرد درجتان في متغيرين: أحدهما متغير تصنيفي أو قطعي هو المتغير المستقل، والثاني المتغير الذي نريد اختباره وهو عادة المتغير التابع. ويقسم المتغير التصنيفي الحالات إلى مجموعتين أو فئتين جامعتين مانعتين مثل النوع (ذكور وإناث). في حين أن المتغير الآخر يصف الفرد بالنسبة لمتغير كمي مثل الذكاء أو الفهم اللفظي. ويعطينا اختبار (ت) بيانات عما إذا كان متوسط درجات المتغير المستقل يختلف اختلافاً دالاً إحصائياً عن متوسط المجموعة الضابطة أو المقارنة.

أسس اختبار (ت) لعينتين مستقلتين

هناك عدة مسلمات ترتبط باختبار (ت) لعينتين مستقلتين وحجم الأثر المقترن به.

مسلمات اختبار (ت) لعينتين مستقلتين

هناك ثلاث مسلمات مرتبطة باختبار (ت) لعينتين مستقلتين، هي:

المسلم رقم ١: أن يكون المتغير الذي نختبره موزعاً توزيعاً اعتدالياً في المجتمع الذي ندرسه.

وعندما يكون لدينا عينة متوسطة الحجم قد ينتهك هذا المسلم، وقد ينتهك أيضاً مع العينات الأكبر، إلا أن الاختبار يظل يعطي نتائج دقيقة نسبياً. وإذا بلغ حجم العينة في كل مجموعة ١٥ فرداً فإن هذا يكون كافياً للحصول على قيم دقيقة لـ (ل). وإذا كان توزيع المجتمع بعيداً جداً عن الاعتدال، فقد نحتاج إلى عينات كبيرة الحجم.

المسلم رقم ٢: أن تباين المتغيرين الموزعين توزيعاً اعتدالياً اللذين نختبرهما، متساو في المجتمعين اللذين سحبت منهما العينتين.

وإذا انتهك هذا المسلم وكان حجم العينتين مختلفاً فإن اختبار (ت) لعينتين مستقلتين لن يكون صادقاً ولا يمكن الثقة في قيمة (ل). وإذا كنا نستخدم برنامجاً إحصائياً مثل برنامج SPSS فإنه يقوم بحساب قيمة تقريبية لـ (ت) لا تسلم بتساوي التباين في المجتمعين، وذلك بالإضافة إلى اختبار (ت) التقليدي الذي يقوم على تساوي تباييني المجتمعين.

المسلم رقم ٣: أن أفراد العينة المسحوبة من المجتمع موزعون توزيعاً عشوائياً، وإن درجات المتغير موضوع الاختبار الإحصائي مستقلة عن بعضها البعض.

وإذا انتهك مسلم استقلال الدرجات يجب عدم الثقة في قيمة (ل) التي نحصل عليها.

حساب حجم الأثر

يمكن حساب حجم الأثر (Δ) كما سبق أن ذكرنا بالمعادلة (٢-٥):

$$\Delta = \frac{\text{متوسط الفروق}}{\text{الانحراف المعياري للمتغيرين معاً}}$$

وتساعدنا Δ على تقويم الدرجة التي يختلف بها متوسط الدرجات في المتغير الذي نختبره عن القيمة التي نختبرها وذلك في وحدات من الانحراف

المعياري. وإذا كانت قيمة Δ تساوي صفراً فإن ذلك يعني تساوي متوسط الدرجات والقيمة التي نختبرها. وكلما تباعد الفرق عن صفر يزداد حجم الأثر. ومن الممكن أن تتراوح قيمة حجم الأثر بين $\pm\infty$ ($\pm$ ما لا نهاية). وبغض النظر عن علامة – أو + فإن القيمة ٠,٢ تمثل حجم أثر صغير ويعتبر حجم الأثر ٠,٥ قيمة متوسطة، أما ٠,٨ فهي تمثل حجم أثر كبير.

ويمكن استخدام مربع إيتا (η^2) كبديل لحجم الأثر Δ. وتتراوح قيمة η^2 بين صفر و واحد. ويمكن تفسيرها باعتبارها نسبة تباين متغير الاختبار التي هي دالة متغير المجموعة. وإذا كانت هذه القيمة صفراً فمعنى هذا أن متوسط الفروق يبلغ صفراً. في حين إذا كانت هذه القيمة ١ فمعنى هذا أن متوسطي المجموعتين مختلفان. ولكن درجات الاختبار لا تختلف داخل كل مجموعة (أي أن هناك تطابقاً تاماً بين مجموعتي الدرجات). ويمكن حساب مربع إيتا كما يلي:

$$\eta^2 = \frac{ت^2}{ت^2 + (ن_1 + ن_2 - 2)}$$ معادلة ٢-٦

وإذا بلغت قيمة η^2 ٠,٠١ فإنها تعتبر قيمة صغيرة، وإذا كانت هذه القيمة ٠,٠٦ كانت قيمة متوسطة، أما إذا كانت ٠,١٤ فإنها تعتبر قيمة كبيرة.

ويذكر صلاح مراد (٢٠٠٠، ص ٢٤٥-٢٤٦) أنه في حالة الحصول على قيمة ت دالة إحصائياً فإننا نستخدم المعادلة التالية:

$$ح = ت \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}$$ معادلة ٢-٧

حيث:

ح = حجم الأثر

ت = قيمة ت التي حصلنا عليها

ن١ ، ن٢ حجما العينتين

اختبار (ت) لعينتين متطابقتين (عينتين متكافئتين)

في هذا الاختبار يجب أن يكون لكل حالة درجتان في متغيرين. ويدرس هذا الاختبار ما إذا كان متوسط الفروق بين المتغيرين يختلف اختلافاً دالاً إحصائياً عن الصفر. ويمكن استخدام هذا الاختبار في الدراسات التي يجري فيها إعادة القياس، أو الدراسات التي تستخدم تصميم العينات المتطابقة.

وفي دراسات إعادة القياس نحصل من كل فرد على درجتين لمقياس واحد في مناسبتين مختلفتين أو تحت شرطين مختلفين. ويجب أن يكون في ملف البيانات الذي نستخدمه لإجراء اختبار (ت) لعينتين متطابقتين درجتان لكل فرد في متغيرين مختلفين. ويمثل المتغير الأول درجة الاختبار الأول ويمثل المتغير الثاني درجة الاختبار الثاني. والفرض الذي يهمنا اختباره هنا هو إذا ما كانت هناك فروق دالة إحصائياً بين متوسطي الإجراءين.

وبالنسبة لتصميم العينات المتطابقة يجري مطابقة الأفراد في أزواج، ويتم قياس الصفة التي نريدها مرة واحدة بالنسبة للفرد في كل زوج. ويعتبر كل فردين في زوج واحد حالة واحدة ولهذه الحالة درجتان في متغيرين، وهاتان الدرجتان هما الدرجة التي نحصل عليها من مشارك ما تحت شرط معين، والدرجة التي نحصل عليها من المشارك الآخر تحت الشرط الآخر. والفرض الأساسي هنا ايضاً هو اختبار ما إذا كان المتوسطان اللذان حصلنا عليهما في الإجراءين مختلفين اختلافاً دالاً إحصائياً.

ويستخدم اختبار (ت) لعينتين متطابقتين في الدراسات التالية:

* تصميمات إعادة القياس مع وجود فترة زمنية بين الإجراءين.
* تصميمات إعادة القياس دون وجود فاصل زمني.
* تصميمات الأفراد المتطابقين مع وجود فترة زمنية بين الإجراءين.
* تصميمات الأفراد المتطابقين دون وجود فاصل زمني.

أسس اختبار (ت) لعينتين متطابقتين

هناك عدة مسلمات ترتبط باختبار (ت) لعينتين متطابقتين وحجم الأثر المقترن به.

مسلمات اختبار (ت) لعينتين متطابقتين

هناك مسلمان مرتبطان باختبار (ت) لعينتين متطابقتين، هما:

المسلم رقم ١: أن يكون المتغير الناتج عن حساب الفروق بين المتغيرين موزعاً توزيعاً اعتدالياً في المجتمع الذي ندرسه.

وقد ينتهك هذا المسلم رغم وجود عينة كبيرة الحجم، إلا أن الاختبار يظل يعطي نتائج دقيقة نسبياً. ويحتاج الأمر إلى الحصول على عينة أكبر للحصول على نتائج صادقة نسبياً إذا كان توزيع المتغير في المجتمع بعيداً عن أن يكون اعتدالياً. إلاأنه في معظم الحالات يكفي الحصول على عينة مكونة من ٣٠ فرداً للحصول على قيم دقيقة لـ (ل).

المسلم رقم ٢: أن العينة المسحوبة من المجتمع سحبت بشكل عشوائي وإن درجات المتغيرين موضوع الاختبار الإحصائي مستقلة عن بعضها البعض.

يعطي اختبار (ت) لعينتين متطابقتين ننتائج غير دقيقة إذا انتهك المسلم المتعلق باستقلالية الدرجات.

حساب حجم الأثر

يمكن حساب حجم الأثر (Δ) كما يظهر في المعادلة التالية:

$$\Delta = \frac{\text{متوسط الفروق}}{\text{الانحراف المعياري}}$$ معادلة ٢-٨

وإذا لم يكن لدينا متوسط الفروق والانحراف المعياري ضمن النتائج يمكننا حساب حجم الأثر من قيمة (ت) حيث يكون حجم الأثر في هذه الحالة:

$$\Delta = \text{ت} \div \sqrt{\text{ن}}$$ معادلة ٢-٩

حيث ن هي العدد الكلي للعينة.

وتساعدنا Δ على تقويم الدرجة التي يختلف بها متوسط الدرجات في المتغير الذي نختبره عن القيمة التي نختبرها وذلك في وحدات من الانحراف المعياري. وإذا كانت قيمة Δ تساوي صفراً فإن ذلك يعني تساوي متوسط الدرجات. وكلما تباعد الفرق عن صفر يزداد حجم الأثر. ومن الممكن أن تتراوح قيمة حجم الأثر بين $\pm\infty$ ($\pm$ ما لا نهاية). وبغض النظر عن علامة – أو + فإن القيمة ٠,٢ تمثل حجم أثر صغير ويعتبر حجم الأثر ٠,٥ قيمة متوسطة، أما ٠,٨ فهي تمثل حجم أثر كبير.

ويمكن استخدام مربع إيتا (η^2) كبديل لحجم الأثر Δ. وتتراوح قيمة η^2 بين صفر وواحد. ويمكن حسابها كما يلي:

$$\eta^2 = \frac{ت^2}{ت^2 + ن - ١}$$ معادلة ٢-١٠

وإذا كانت هذه القيمة صفراً فمعنى هذا أن متوسط الفروق يبلغ صفراً. وإذا كانت هذه القيمة ١ فهذا يعني أن جميع فروق الدرجات غير صفرية، أي إن هناك تطابقاً بين مجموعتي الدرجات. وإيتا عبارة عن معامل ارتباط ثنائي أصيل بين المتغير المستقل بفئتيه والمتغير التابع المتصل. ويشير مربع إيتا إلى حجم الأثر ممثلاً في التباين المشترك بين المتغيرين.

ويعطينا صلاح مراد (٢٠٠٠) معادلة بديلة لحجم الأثر باستخدام معامل الارتباط بين المتغيرين:

$$ح = ت \sqrt{\frac{٢(١-ر)}{ن}}$$ معادلة ٢-١١

حيث:

ر = معامل الارتباط بين درجات القياسين.

ويذكر صلاح مراد (٢٠٠٠، ص٢٤٨) أن كيس (Kiess, 1989) أعطى التفسير المقارن التالي لحجم الأثر بطريقة كوهين وبطريقة مربع إيتا:

١ – يقابل حجم الأثر ٢, عند كوهين مربع إيتا قدره ٠١, وهي قيمة صغيرة جداً (١٪ من التباين).

٢ – إذا كان مربع إيتا ٠٦, فإنه يقابل قيمة حجم الأثر عند كوهين ٥٠٥, وهي قيمة تدل على حجم أثر متوسط.

٣ – إذا بلغ مربع إيتا ١٥, فإن قيمة حجم الأثر عند كوهين تبلغ ٨٤, وهي قيمة مرتفعة.

حجم الأثر في تحليل التباين

تناولنا فيما سبق حجم الأثر عند دراسة الفروق بين متوسطين باستخدام اختبار (ت)، وننتقل الآن إلى دراسة الفروق بين أكثر من متوسطين، أي دراسة الفروق بين متوسطات ثلاث مجموعات أو أكثر، وفي هذه الحالة نستخدم تحليل التباين لدراسة الفروق بين المتوسطات على النحو التالي:

* متغير مستقل واحد (بين المجموعات):
  - تحليل التباين الأحادي باستخدام قائمة المتوسطات.
  - تحليل التباين الأحادي باستخدام أمر Oneway.
  - تحليل التباين الثنائي باستخدام النموذج الخطي العام (GLM).
* متغير مستقل واحد (داخل المجموعات).
* تحليل التباين الثنائي باستخدام متغيرين مستقلين (أو أكثر)، وكلها بين المجموعات.
* متغيرين مستقلين (أو أكثر) باستخدام بين المجموعات وداخل المجموعات (التصميم المختلط).
* متغيرين مستقلين أو أكثر كلها داخل المجموعات (إعادة إجراء المقاييس).
* تحليل متغيرين تابعين أو أكثر (تحليل التباين المتعدد MANOVA).

المبادئ الأساسية لتحليل التباين:

عندما يكون لديك ثلاث مجموعات أو أكثر من البيانات المعلمية، فقد ترغب في وضع فرض حول اختلاف متوسطات المجموعات. وفي هذه الحالة لا تستطيع استخدام اختبار (ت) الذي استخدمناه لمقارنة متوسطي مجموعتين، ولذلك فإن تحليل التباين هو الأسلوب الإحصائي الذي يستخدم في هذه الحالة.

وكما هو واضح من التسمية فإن تحليل التباين يعني فحص التباين داخل مجموعة كاملة من الدرجات. تخيل أن لدينا بيانات من ثلاث مجموعات

مستقلة من الأفراد، ونريد أن نعرف إذا ما كان هناك فرق بين المجموعات الثلاث. فإذا لم يكن هناك فرق بين المجموعات (الفرض الصفري صحيح)، فإن هذا يعني أن هذه البيانات أتت من نفس المجتمع، وتكون متوسطات المجموعات الثلاث متساوية، وكذلك تبايناتها. وفي هذه الحالة يكون تباين كل مجموعة تقديراً لتباين المجتمع (أي أن اختلاف التباين يرجع إلى التذبذبات العشوائية بين المستجيبين، وهو ما يعرف بتباين الخطأ لأنه يرجع إلى التغيرات العشوائية في البيانات). وأفضل تقدير لتباين المجتمع في هذه الحالة هو متوسط تباينات المجموعات الثلاث. ولذلك فإننا بالنظر إلى متوسط تباين المجموعات الثلاث نستطيع تقدير تباين الخطأ.

وإذا كان الفرض الصفري صحيحاً تكون متوسطات المجموعات الثلاث متساوية، ويكون تباين المتوسطات (أي مدى اختلاف المتوسطات عن بعضها البعض) صغيراً جداً. (ونتوقع في هذه الحالة أن يكون مساوياً لتباين المجتمع). ويعرف تباين متوسطات المجموعات الثلاث بتباين المعالجة. ولذلك إذا كان الفرض الصفري صحيحاً، ولم تختلف المجموعات الثلاث عن بعضها البعض فإن التباين بين المتوسطات (تباين المعالجة) يكون مساوياً لتباين الخطأ. فإذا قسمنا تباين المعالجة على تباين الخطأ تكون الإجابة =١٠٠,. وتعرف نتيجة قسمة تباين المعالجة على تباين الخطأ بالنسبة الفائية «F».

وإذا كان الفرض الصفري غير صحيح، فإن معنى ذلك وجود فروق بين المتوسطات الثلاثة، ويكون تباين المتوسطات في هذه الحالة أكبر من تباين الخطأ. فإذا قسمنا تباين المتوسطات على تباين الخطأ، نحصل على قيمة «ف» تزيد على ١,٠٠.

وفي تحليل التباين نقارن تباين المعالجة بتباين الخطأ لاختبار الفرض بوجود فروق دالة إحصائياً بين المتوسطات. وهناك أنواع مختلفة من تحليل التباين، ولذلك يجب اختيار نوع التحليل الذي يناسب البيانات التي لديك. فعندما ترغب في مقارنة ثلاثة مستويات أو أكثر بين المجموعات (متغير مستقل

واحد ذو ثلاثة مستويات أو أكثر) فإننا نستخدم تحليل التباين الأحادي. إلا أن تحليل التباين يمكن أن يمتد لمواقف يوجد بها متغيران مستقلان أو أكثر. فقد نقيس مثلاً أداء مجموعة من الأطفال يقل عمرها عن عشر سنوات، ومجموعة أخرى يزيد عمرها على عشر سنوات، في وقتين مختلفين (الثامنة صباحاً، والثانية مساء). واستخدمنا أفراداً مختلفين في كل مرة، فيصبح لدينا أربع مجموعات مستقلة. وقد نريد معرفة هل يختلف الأداء تبعاً لسن الفرد، وتبعاً للوقت وإذا ما كان هناك تفاعل بين هذه المتغيرات. والمقصود بالتفاعل أن أثر المتغير متأثر بالمتغير الآخر، مثال ذلك أننا قد نجد أن الفرق بين الأداء في الساعة الثامنة صباحاً والساعة الثانية مساء أقل بالنسبة للأطفال الصغار منه بالنسبة للأطفال الكبار. وفي هذه الحالة سوف يُظهر تحليل التباين أن هناك تفاعلاً دالاً إحصائياً بين المتغيرين.

وفي المثال السابق فإن كلا المتغيرين يهتم بالفروق بين المجموعات، لأن أفراد كل مجموعة يختلفون عن أفراد المجموعات الثلاث الأخرى. ولكن تحليل التباين يمكن أن يتم أيضاً داخل المجموعات (تكرار القياس على نفس المجموعة)، وهنا نستخدم نفس الأفراد في ظروف مختلفة. مثال ذلك، يكون لدينا دراسة لإعادة القياس إذا قمنا بدراسة أداء نفس الأفراد في كل عمر في الساعة الثامنة صباحاً والساعة الثانية مساء.

العوامل الثابتة والعوامل العشوائية

عندما نقوم بإجراء تحليل التباين، يجب تحديد هوية المتغيرات المستقلة (العوامل): هل هي متغيرات ثابتة أم متغيرات عشوائية. فإذا كانت مجموعات المتغير المستقل مجموعات تكونت بالتعيين العشوائي يكون لدينا عوامل عشوائية. وتكون مستويات المتغير المستقل في هذه الحالة مستويات عشوائية تكونت بالتعيين العشوائي، ولذلك يكون المتغير متغيراً عشوائياً ويطلق عليه في تحليل التباين عاملاً عشوائياً. ولكن إذا كانت مستويات المتغير المستقل مستويات ثابتة لم تتكون بالتعيين العشوائي، فإنها تكون من العوامل الثابتة، فإذا

استخدمنا متغير النوع في تحليل التباين مثلاً، فإن مستويات هذا المتغير (ذكوراً أو إناثاً) هي مستويات محددة ولا يمكن أن تكون عشوائية، ولذلك فإن متغير النوع يعتبر من العوامل الثابتة في تحليل التباين.

ويلاحظ أن الأرقام في نتائج تحليل التباين قد تختلف بعض الشيء في العوامل العشوائية عنها في العوامل الثابتة، حيث إن حساب الخطأ في العوامل العشوائية يختلف عن حساب الخطأ في العوامل الثابتة.

الاختبارات المسبقة والاختبارات التتبعية

قد تشير نتائج تحليل التباين إلى أن هناك فروقاً دالة بين متوسطات المجموعات ولكنها لا تبين لك أين يوجد هذا الفرق. هل المجموعة الأولى تختلف عن المجموعتين الثانية والثالثة؟ هل تختلف المجموعة الثانية عن كل من المجموعتين الأولى والثالثة؟ ويمكن الإجابة عن مثل هذه الأسئلة بمقارنة متوسطات المجموعات الفرعية، وهناك طريقتان للمقارنة: ويشار للطريقة الأولى بأنها طريقة المقارنات المسبقة (أو القبلية) a priori comparisons، وفي هذه الطريقة مثلاً التنبؤ بأن المجموعة الثالثة تختلف اختلافاً دالاً إحصائياً عن المجموعة الأولى، ولكنها لا تختلف عن المجموعة الثانية. أما الطريقة الثانية فيشار إليها بالطريقة التتبعية post hoc tests، وفيها تعقد المقارنات بعد تحليل التباينات.

أما الاختبارات التتبعية Post hoc tests مثل اختبار دنكان Duncan's test، واختبار توكي Tukey's test واختبار شافيه Scheffe's test (وغيرها) فيمكن استخدامها عندما تشير نتائج تحليل التباين إلى وجود فروق بن متوسطات المتغير المستقل.

تحليل التباين الأحادي (متغير واحد مستقل بين المجموعات)

توجد درجة لكل فرد أو حالة في تحليل التباين الأحادي في متغيرين: العامل (أي المتغير المستقل) والمتغير التابع (أي المحك). ويقسم العامل أفراد العينة إلى مجموعتين أو أكثر، بينما يميز المتغير التابع بين الأفراد في صفة من

الصفات الكمية. وتختبر النسبة الفائية في تحليل التباين الأحادي ما إذا كانت متوسطات المجموعات مختلفة عن بعضها البعض اختلافاً دالاً إحصائياً. ويجب أن يتوفر في ملف البيانات في برنامج SPSS لإجراء تحليل التباين درجة لكل حالة في العامل (تواجد في إحدى المجموعات) ودرجة في المتغير الكمي التابع.

أسس تحليل التباين الأحادي:

نجري تحليل التباين لاختبار الفروق بين متوسطات المجموعات المختلفة في المتغير التابع. وإذا كان تحليل التباين العام دالاً إحصائياً وكان للعامل أكثر من مجموعتين أو مستويين يجب إجراء اختبار تتبعي كما سبق ذكره. ويقوم الاختبار التتبعي بعقد مقارنات زوجية بين المتوسطات. مثال ذلك إذا كان للعامل ثلاثة مستويات فمن الممكن عقد ثلاثة أزواج من المقارنات: مقارنة بين متوسطي المجموعتين ١ و٢، ومقارنة بين متوسطي المجموعتين ٢ و ٣، ومقارنة بين متوسطي المجموعتين ١ و ٣. ويطلق SPSS على هذه الاختبارات التتبعية (Post hoc multiple comparisons).

مسلمات تحليل التباين:

مسلم رقم ١: المتغير التابع موزع توزيعاً اعتدالياً في المجتمعات الأصلية وفي أي مستوى من مستويات المتغير المستقل (العامل).

ويتطلب هذا المسلم أن يكون توزيع المتغير التابع توزيعاً اعتدالياً في جميع مستويات المتغير المستقل. وفي العينات المتوسطة والكبيرة يمكن انتهاك هذا المسلم مع الحصول على نتائج دقيقة نسبياً. ويمكن اعتبار حجم العينة التي يبلغ عدد أفرادها ١٥ فرداً في كل مجموعة عينة كبيرة بشكل كاف للحصول على قيم (ل) تتصف بالدقة. ويمكن أن يحتاج الأمر إلى عينات أكبر للحصول على نتائج دقيقة إذا ابتعد توزيع المجتمع ابتعاداً كبيراً عن التوزيع الاعتدالي.

المسلم رقم ٢: تباينات المتغير التابع للتوزيعات المذكورة في المسلم رقم ١ متساوية.

إذا انتهك هذا المسلم مع اختلاف أحجام العينات تصبح نتائج تحليل التباين الأحادي موضع شك. وحتى إذا تساوت أحجام العينات يجب الشك في نتائج الاختبارات التتبعية (post hoc) إذا اختلفت تباينات المجتمع.

المسلم رقم ٣: أفراد العينات مسحوبة بشكل عشوائي من المجتمع كما أن درجات المتغير التابع مستقلة عن بعضها البعض.

يعطي تحليل التباين الأحادي نتائج غير صحيحة إذا انتهك هذا المسلم.

حجم الأثر لتصميمات تحليل التباين الأحادي:

ذكرنا من قبل أنه في حالة مجموعتين فإننا يمكن دراسة معامل التحديد باستخدام مربع معامل الارتباط الثنائي $ر^{٢}$ث أ الذي يعطينا نسبة التباين في المتغير التابع الذي يرتبط بالتباين في المتغير المستقل. وتستخدم في تحليل التباين تقديرات لحجم الأثر شبيهة بذلك يزيد عدد مجموعات المتغير المستقل عن مجموعتين. وتقديرات حكم الأثر تعكس قوة الارتباط على مقياس يتراوح من صفر (لا ارتباط) إلى ١ (ارتباط تام).

أما بالنسبة لتحليل التباين الأحادي فهناك عدة طرق لحساب حجم الأثر:

مقياس الفرق المقنن لحجم الأثر العام:

تعطينا المعادلة التالية حجم الأثر العام:

$$ح_{عام} = \frac{\mu_{قصوى} - \mu_{دنيا}}{\sigma}$$ معادلة ٢-١٢

حيث:

$ح_{عام}$ حجم الأثر العام في المجتمع.

$\mu_{قصوى}$ أعلى متوسط في المجتمعات.

$\mu_{دنيا}$ أدنى متوسط في المجتمعات.

σ الانحراف المعياري للمجتمعات ويمكن تقديره باستخدام متوسط المربعات (م مر$_{د}^{1/2}$) والتي يمكن الحصول عليها من نتائج برنامج تحليل التباين باستخدام معادلة معدلة لتجميع التباينات.

$$\text{م مر}_{د} = \frac{(\nu_{1} - 1)ع_{1}^{2} + \cdots + (\nu_{ك} - 1)ع_{ك}^{2}}{ن - ك}$$ معادلة ٢-١٣

حيث:

ن حجم المجتمع الكلي.

ك عدد المجموعات.

ν حجم العينة.

ع الانحراف المعياري للعينة.

وتقدير حجم الأثر الذي نحصل عليه من المعادلة ١٧ هو:

$$ح_{\text{عام}} = \frac{\text{م ص}_{\text{قصوى}} - \text{م ص}_{\text{دنيا}}}{ع}$$ معادلة ٢-١٤

حيث م ص متوسط ص

وتختلف ح$_{\text{عام}}$ عن ح في أن الأولى تعبر عن التباين الكلي المشترك بين المتغيرات، في حين أن ح الثانية تسلم بتساوي التباين في العينات المختلفة.

حجم الأثر العام المقنن باستخدام جميع المتوسطات:

تتجاهل المعادلتان ٢-١٤ و٢-١٦ جميع المتوسطات باستثناء أعلى متوسط وأقل متوسط. وهناك مقياس لحجم الأثر العام في تحليل التباين الأحادي يستخدم جميع المتوسطات. ويسلم حجم الأثر هذا بتجانس التباين ويطلق عليه كوهين f وتحسب باستخدام المعادلة التالية:

$$f = \frac{\mu\sigma}{\sigma'}$$ معادلة ٢-١٥

حيث:

$\sigma\mu$- الانحراف المعياري لكل المتوسطات في المجتمعات المماثلة بكل العينات (على أساس انحراف كل متوسط عن متوسط جميع المتوسطات كما هو الحال في المعادلة رقم ٢٢) و σ هي الانحراف المعياري العام في المجتمعات.

ويمكن تقدير f بالمعادلة:

$$\hat{f} = \frac{\text{ع ص}}{\text{م مر}_{د}} \qquad \text{معادلة ٢-١٦}$$

حيث:

ع م ص الانحراف المعياري لمجموعة كل متوسطات ص ابتداء من م ص١ إلى م صك، ولذلك فإنه بالنسبة للعينات متساوية الحجم:

$$\text{ع}_{ص} = \left[\frac{\text{مجـ (م ص}_{١} - \text{م ص}_{كل})^{٢}}{ك - ١} \right]^{\frac{١}{٢}} \qquad \text{معادلة ٢-١٧}$$

قوة الارتباط

ذكرنا من قبل أن معامل التحديد استخدم بشكل تقليدي لتقدير نسبة التباين الكلي في المتغير التابع والذي يرتبط بتباين المتغير المستقل. كما استخدمت بعض أحجام الأثر بشكل تقليدي أيضاً مع تحليل التباين الأحادي حيث ك > ٢. ويقصد من هذه المقدرات تقدير قوة الارتباط على مقياس يتراوح صفر (لا ارتباط) إلى ١ (ارتباط تام).

مربع إيتا (η'^2):

استخدمنا من قبل مربع إيتا عند الكلام على اختبار ت ورأينا أنه مقياس لنسبة التباين في المجتمع الذي يمكن تفسيره على أساس المعالجة التجريبية. وفي حالة تحليل التباين يمكن استخدام المعادلة التالية:

$$\eta'^2 = \frac{\text{مجموع مربعات المتغير المستقل}}{\text{المجموع الكلي للمربعات}} \qquad \text{معادلة ٢-١٨}$$

والاسم الأصلي لإيتا (η') ذاتها كان نسبة الارتباط ولكن أطلق نفس الاسم بعد ذلك على مربع إيتا (Grissom & Kim, 2005). وعندما يكون لدينا مجموعتان فإن لإيتا نفس القيمة المطلقة لمعامل الارتباط في المجتمع.

ومن عيوب مربع إيتا أنها كمقياس لقوة الارتباط أنها متحيزة تحيزاً موجباً، أي أنها تميل إلى زيادة تقدير η'^2 لأن بسط المعادلة (مجموع مربعات المتغير المستقل) يتضخم بسبب الخطأ في التباين. والتحيز أقل في حالة العينات الكبيرة وبالنسبة لقيم مربع إيتا الكبيرة.

مربع أبسيلون (ε^2) ومربع أوميجا (ω^2):

وهاتان القيمتان على العكس من قيمة η'^2 تقديران غير متحيزان. ومعادلة أبسيلون هي على النحو التالي (Grissom & Kim, 2005):

$$\varepsilon^2 = \frac{\text{مجـ مر}_{م} - (\text{ك} - ١)\ \text{م مر}_{د}}{\text{مجـ مر}_{كل}}$$ معادلة ٢-١٩

حيث:

مجـ مر$_{م}$ مجموع مربعات المتغير المستقل.

ك عدد المجموعات.

م مر$_{د}$ متوسط المربعات داخل المجموعات.

مجـ مر$_{كل}$ مجموع المربعات الكلي.

$$\omega^2 = \frac{\text{مجـ مر}_{م} - (\text{ك} - ١)\ \text{م مر}_{د}}{\text{مجـ مر}_{كل}}$$ معادلة ٢-٢٠

ونحن نسلم هنا بتساوي حجم العينات وتجانس التباين، وتشتمل بعض نتائج البرامج الإحصائية مثل STATISTICA على قيمتي ε^2 أو ω^2 أو كليهما. إلا أن حسابهما يدوياً سهل للغاية لتوفر قيم مجـ مر$_{م}$ و م مر$_{د}$ مع نتائج تحليل التباين.

وبمقارنة البسط في المعادلتين (٢-١٩) و(٢-٢٠) مع بسط المعادلة (٢-١٨) (η'^2) نجد أن هاتين المعادلتين تحاولان التعويض عن زيادة التقدير في

(η'^2) باستخدام (ك-١)م مر. وتحاول المعادلة (٢-٢٠) أكثر من ذلك بإضافة م مر. ولذلك نجد أن مربع أوميجا (ω^2) أوسع استخداماً من مربع أبسيلون (ε^2) في تقدير حجم أثر المعالجات التجريبية.

ويمكن أن يكون الحصول على قيمة (ف) دالة بشكل عام دليلاً على أن قيمة مربع أوميجا أكثر من الصفر بشكل دال إحصائياً. إلا أن استخدام حدود الثقة مهم على وجه الخصوص في مثل هذه الظروف بسبب تباين المعاينة المرتفع الذي يرجع إلى هذه التقديرات. وقد وجد كارول ونوردولم (Carroll & Nordholm, 1975) تبايناً أعظم في المعاينة حتى عندما تكون ن = ٩٠ و ك = ٣. وبالطبع فإن ارتفاع تباين المعاينة ينتج عنه تقديرات أعلى أو أقل بكثير من حجم الأثر الذي نحصل عليه.

ويلاحظ أن قيمة (η'^2) كمقياس لنسبة (التباين الكلي للمتغير التابع والتي ترتبط بتباين المتغير المستقل) لا يمكن أن تكون أقل من صفر ولكن دراسة المعادلتين (٢-١٩) و(٢-٢٠) يبين أن قيمتي مربع أوميجا (ω^2) ومربع أبسيلون (ε^2) يمكن أن تكونا أقل من الصفر، وفي هذه الحالة يجب أن تكون القيمة التي تذكر في التقرير صفراً. إلا أن بعض الباحثين في التحليل البعدي يرون أن إحلال القيمة صفر محل القيم السالبة قد يترتب عليه زيادة التحيز الموجب في التقدير الذي يقوم أساساً على متوسط التقديرات في التحليل البعدي. ولذلك يقول فيدلر وتومسون (Fidler & Thompson, 2001) أنه يجب ذكر أي قيمة سالبة في التقرير بدلاً من تحويلها إلى صفر حتى يمكن بالطبع تحديد الاتساع الفعلي لحدود الثقة. وعند ذكر أي قيمة سالبة يكون لدى القارئ الحرية في تفسيرها صفراً إذا اختار ذلك.

حجم الأثر في تحليل التباين العاملي

يجب أن يكون لكل فرد في العينة في تحليل التباين العاملي درجات في متغيرات العوامل والمتغير التابع. ويقسم كل عامل الحالات إلى مستويين أو أكثر، أما المتغير التابع فإنه يصف الحالات على طول متغير كمي متصل. ويطبق اختبار (ف) على المؤثرات الرئيسية للعوامل وعلى التفاعل بينها. ويمكن القيام باختبارات تتبعية لاختبار الفروض الخاصة إذا كانت المؤثرات الرئيسية أو التفاعل أو كليهما دالاً إحصائياً.

أسس تحليل التباين العاملي:

الاختبارات الأولى التي تجرى في تحليل التباين العاملي هي الاختبارات العامة أو الشاملة للمؤثرات الرئيسية والتفاعل. وتختبر هذه الاختبارات العامة الفروض المتعلقة بما يلي:

* **المؤثر الرئيسي لكل عامل:** اختلاف متوسطات المتغير التابع في المجتمع بالنسبة لمستويات كل عامل مع تثبيت مستويات العوامل الأخرى.
* **مؤثرات التفاعل:** اختلاف متوسطات المتغير التابع في المجتمع بين مستويات كل عامل كدالة لمستويات العوامل الأخرى.

وإذا كان واحد أو أكثر من المؤثرات العامة دالاً إحصائياً تجرى اختبارات تتبعية متنوعة. ويتوقف اختيار الاختبار التتبعي على أين توجد الفروق الدالة. فإذا كانت الدلالة في تفاعل المؤثرات يمكن إجراء الاختبارات التتبعية لتقويم المؤثرات الرئيسية البسيطة، أو المقارنات التفاعلية أو كليهما. واختيار الاختبار المطلوب يتوقف على أيها أنسب للفروض التي يجري اختبارها.

ويمكن تقويم عدد من المؤثرات الرئيسية البسيطة ويتوقف ذلك على فروض البحث: الفروق في متوسطات المجتمع بين مستويات العامل الأول وذلك بالنسبة لكل مستوى من مستويات العامل الثاني. كما يمكن تقويم الفروق في المتوسطات بين مستويات العامل الثاني لكل مستوى من مستويات العامل الأول.

وإذا كان أي من هذه المؤثرات البسيطة دالاً إحصائياً وكان هناك أكثر من متوسطين، فإن الأمر يحتاج إلى اختبارات إضافية تشتمل على مقارنات زوجية لتقويم الفروق بين أزواج المتوسطات.

ويمكن إجراء مقارنات بين المتغيرات المتفاعلة بدلاً من المؤثرات البسيطة أو بالإضافة إليها. وتتكون أبسط المقارنات بين التفاعلات على أربعة متوسطات ويشار إليها بالمقارنات الرباعية. وتختبر المقارنات الرباعية إذا ما كانت الفروق في متوسطات المجتمع بين مستويين من مستويات العامل الأول تختلف بالنسبة لمستويين من مستويات العامل الثاني.

وإذا كان التفاعل غير دال فإن الانتباه يتحول إلى المؤثرات الرئيسية. فإذا كانت المؤثرات الرئيسية لعامل له أكثر من مستويين دالة يجب القيام باختبارات تتبعية. والغرض من هذه الاختبارات التتبعية معرفة ما إذا كانت هناك فروق بين متوسطات مستويات هذا العامل بعد تثبيت متوسطات العوامل الأخرى. وغالباً ما تتضمن هذه الاختبارات التتبعية مقارنة متوسطات أزواج مستويات العامل الذي ظهرت فروق دالة في مؤثراته الرئيسية.

مسلمات تحليل التباين العاملي:

المسلم الأول: أن يكون المتغير التابع موزعاً توزيعاً اعتدالياً بالنسبة لكل المجتمعات المتضمنة في التحليل.

وتعرف هذه المجتمعات بعدد خلايا مستويات العوامل. مثال ذلك إذا كان تحليل التباين العاملي يتكون من ٢ × ٣ من المستويات، يكون لدينا ست خلايا، وعلى ذلك فإن المسلم يتطلب أن يكون توزيع المتغير التابع اعتدالياً في المجتمع في جميع الخلايا الست. وإذا لم تكن التوزيعات اعتدالية يمكن لتحليل التباين العاملي إعطاء نتائج دقيقة بشكل معقول إذا كانت أحجام العينة متوسطة أو كبيرة. وفي معظم الحالات يمكن اعتبار العينات التي تتكون من ١٥ حالة في كل

مجموعة كبيرة بشكل كاف لإجراء التحليل والحصول على قيم دالة إحصائياً. وقد يتطلب الأمر عينات أكبر للحصول على نتائج صادقة إذا ابتعدت توزيعات المجتمع كثيراً عن التوزيع الاعتدالي.

المسلم رقم ٢: تتساوى تباينات المتغير التابع في جميع الخلايا.

لا يجب الثقة في نتائج تحليل التباين الثنائي إذا انتهك هذا المسلم وكانت أعداد العينات غير متساوية في الخلايا، وبالإضافة إلى ذلك يجب عدم الثقة في الاختبارات التتبعية التي تتطلب تساوي التباين إذا اختلفت تباينات المجتمع.

المسلم رقم ٣: الحالات عبارة عن عينات عشوائية من المجتمعات التي سحبت منها كما أن درجات المتغير التابع مستقلة عن بعضها البعض.

يعطي تحليل التباين العاملي نتائج غير دقيقة إذا انتهك مسلم الاستقلالية.

حجم الأثر:

يمكن استخدام مربع إيتا كمؤشر لحجم الأثر في تحليل التباين العاملي، وهذه القيمة عبارة عن مربع ارتباط جزئي (Partial η^2). ويتراوح هذا المعامل بين صفر وواحد. ويفسر هذا الارتباط الجزئي على أنه نسبة تباين المتغير التابع التي ترتبط بمصادر التباين للمؤثرات الرئيسية أو مؤثرات التفاعل، مع استبعاد مصادر التباين الرئيسية أو مصادر التفاعل الأخرى. ومن غير الواضح ما هي قيمة الارتباط الجزئي التي تعتبر كبيرة أو متوسطة أو صغيرة، ومن المحتمل جداً أن نقاط القطع التقليدية وهي ٠١, و٠٦, و١٤, تعتبر كبيرة جداً بالنسبة للارتباط الجزئي Partial η^2.

وسوف نتناول أولاً أحجام الأثر التي تعالج التباين بين المجموعات، ثم نعقبها بالتباين داخل المجموعات. ومعظم ما نورده هنا متأثر بالأعمال التي قدمها جريسوم وكيم (Grissom & Kim, 2005).

ويطلق على العامل الذي يقدر له حجم الأثر **العامل المستهدف** (Targeted Factor). ويطلق على أي عامل آخر في التصميم **عامل سطحي** (Peripheral Factor). وإذا قام الباحث فيما بعد أثناء تحليل نفس البيانات بتقدير حجم الأثر بالنسبة لأي عامل كان معتبراً عاملاً سطحياً فإننا نغير أسماء الأدوار المنسوبة لكل عامل، ويصبح العامل السطحي عاملاً مستهدفاً، والعالم المستهدف عاملاً سطحياً.

وتتوقف الإجراءات المناسبة لتقدير حجم الأثر في التصميم العاملي على ما إذا كان العاملان المستهدفان والسطحي **عوامل داخلية** (Intrinsic Factors) أم **عوامل خارجية** (Extraneous Factors). والعوامل الخارجية عوامل لا تتغير في المجتمع عادة بالنسبة للنتائج التي تعمم على المجتمع. فالعوامل الخارجية عادة من عوامل المعالجة المفروضة على المشاركين. أما العوامل الداخلية فهي عوامل تتغير عادة في المجتمع الذي تعمم عليه النتائج، وهي متغيرات مثل النوع أو المهنة أو مستوى التعليم. وغيرها من التصنيفات التي تستخدم عادة في المعالجات التجريبية، والعوامل الداخلية **متغيرات تصنيفية قطعية** (Classificatory Categorical Variables).

ومن العوامل المهمة للغاية عند اختيار طريقة لتقدير الفروق المقننة أو نسبة التباين المفسر لحجم الأثر بالنسبة للعامل المستهدف هو ما إذا كان العامل السطحي عاملاً داخلياً أو عاملاً خارجياً. وعندما يكون العامل السطحي عاملاً خارجياً فإننا نختار طريقة ينسب فيها التباين في البيانات لهذا العامل نقوم بتثبيته (فلا يقوم بأية إضافة لتباين حجم الأثر أو التباين الكلي الذي تقوم بتفسيره). وعندما يكون العامل السطحي عاملاً داخلياً فإننا نستخدم طريقة ينسب فيها التباين للعامل السطحي ويسمح له في هذه الحالة بإضافة التباين لحجم الأثر أو التباين الكلي بنسبة محددة يمكن تفسيرها في العامل المستهدف.

ومن النواحي العملية عندما يقرر الباحث الدور الذي سوف يلعبه عامل سطحي معين يجب أن يأخذ في اعتباره طبيعة المجتمع الذي سوف نقوم بتقدير حجم الأثر بالنسبة له. فإذا لم يكن العامل السطحي من العوامل المتغيرة في المجتمع (قد

يستخدم كمتغير في التجربة ولكنه ليس كذلك في المجتمع)، فإننا نختار طريقة تتجاهل التباين في البيانات المنسوبة للعامل السطحي. وإذا كان المتغير السطحي متغيراً في المجتمع (وهذا ما يحدث غالباً بالنسبة للمتغيرات التصنيفية كالنوع) فإننا نختار طريقة يسمح فيها للتباين في المتغير السطحي بالإضافة إلى التباين الكلي.

قوة الارتباط: نسبة التباين المفسر

تقدير مربع إيتا (نسبة التباين في المجتمع) أكثر تعقيداً بالنسبة للتصميمات العاملية منها في حالة تحليل التباين الأحادي، وحتى في الحالات البسيطة التي نسلم فيها بتساوي أحجام جميع العينات. وبشكل عام فإنه بالنسبة لأثر بعض العوامل أو أثر التفاعل إذا كانت العوامل السطحية داخلية يمكن أن نحسب نسبة التباين في المجتمع بالمعادلة التالية:

$$\omega^2 = \frac{\text{مجـ مر}_{\text{أثر}} - (\text{دح}_{\text{أثر}} \times \text{م مر}_{\text{د}})}{\text{مجـ مر}_{\text{كلي}} + \text{م مر}_{\text{د}}}$$

معادلة ٢-٢١

حيث:

مجـ مر$_{\text{أثر}}$ مجموع مربعات الأثر

دح$_{\text{أثر}}$ درجات الحرية للأثر

مجـ مر$_{\text{كلي}}$ مجموع المربعات الكلي

م مر$_{\text{د}}$ متوسط المربعات داخل المجموعات

ونحصل على كل هذه القيم من نتائج برنامج تحليل التباين في SPSS وغيره من البرامج الإحصائية. وبالنسبة للمؤثر الرئيسي للعامل (أ)، فإن:

$$\text{مجـ مر}_{\text{أثر}} = \text{مجـ مر}_{\text{أ}} \text{ ، دح}_{\text{أثر}} = \text{أ} - ١$$

حيث يكون (أ) عدد مستويات العامل (أ). وبالنسبة للمؤثر الرئيسي للعامل (ب) نستبدل (ب) بـ (أ) في العبارة السابقة، وهكذا بالنسبة لأي عامل آخر في التصميم العاملي. وبالنسبة لأثر التفاعل في التصميم الثنائي فإن:

$$\text{مجـ مر}_{\text{أثر}} = \text{مجـ مر}_{\text{أب}} \text{ ، و دح}_{\text{أثر}} = (\text{أ}-١)(\text{ب}-١)$$

وتمدنا المعادلة (٢-٢١) بتقدير لنسبة التباين الكلي لمقياس المتغير التابع الذي يفسره متغير مستقل (أو بالتفاعل أو بأي مثال آخر في حالة التصميم العاملي). على أنه في هذا النوع من التصميم نجد مصادر كثيرة للتباين أكثر من تحليل التباين الأحادي بسبب الإضافات التي يضيفها عامل أو أكثر أو التفاعلات إلى التباين الكلي. مثال ذلك أن أثر العامل أ قد يعطي قيمة مختلفة لـ ω'^2 إذا كانت تستخدم في تحليل التباين الثنائي عما نحصل عليه من استخدامها في تحليل التباين العاملي حيث نجد كثيراً من مصادر التباين التي تضيف للتباين الكلي. وكما ذكرنا من قبل يجب تفسير حجم الأثر في سياق التصميم الذي يعطينا تقديراً له. وسوف نناقش في الجزء التالي طريقة أخرى يقصد منها الحصول على تقديرات لنسبة التباين في المجتمع من التصميمات العاملية شبيهاً بتلك التي نحصل عليها من تحليل التباين الأحادي.

مربع أوميجا الجزئي:

هناك تقدير آخر لنسبة التباين في المجتمع من تحليل التباين العاملي والذي يعدل من المعادلة رقم (٢-٢١) وذلك في محاولة لاستبعاد الإضافة التي قد يقوم بها العامل السطحي الخارجي للتباين الكلي. والمقياس الناتج يطلق عليه نسبة التباين الجزئي في المجتمع. وبينما نجد أن نسبة التباين تقيس قوة الأثر نسبة إلى التباين الكلي للخطأ ولكل المؤثرات، فإن نسبة التباين الجزئي يقيس قوة الأثر نسبة للتباين الذي لا يرجع إلى المؤثرات الأخرى، والمؤثرات المستبعدة هي تلك التي لم يكن لتوجد إذا كانت مستويات العامل المستهدف قد بحثت في تصميم تباين أحادي. ولذلك فإن مربع معامل إيتا الجزئي ومربع معامل أوميجا الجزئي أصبحت تستخدم بشكل تقليدي. ومعامل مربع أوميجا الجزئي يمكن حسابه لأي مؤثر بالمعادلة التالية:

$$\omega'^2_{\text{الجزئي}} = \frac{\text{مجـ مر}_{\text{أثر}} - (\text{دح}_{\text{أثر}} \times \text{م مر}_{\text{د}})}{\text{مجـ مر}_{\text{أثر}} (\text{ن} - \text{دح}_{\text{أثر}}) \text{ م مر}_{\text{د}}} \qquad \text{معادلة ٢-٢٢}$$

حيث:

ن = حجم العينة الكلي

ونجد هنا أيضاً أن العمليات الحسابية لهذه المعادلة متوفرة في تحليل التباين الأحادي.

ويجب أن يكون تقرير البحث واضحاً تماماً في بيان ما إذا كان تقدير نسبة التباين في المجتمع محسوبة باستخدام ω'^2 أو باستخدام ω'^2الجزئي. ولسوء الحظ فإنه نظراً لأن تقدير نسبة التباين الكلي في المجتمع أو نسبة التباين الجزئي في المجتمع قد يكونا مختلفين تماماً وبخاصة كلما زاد تعقيد التصميم، فإن بعض كتب الإحصاء وبعض البرامج الإحصائية ليست واضحة أو قد تعطي قيماً خاطئة لهذين التقديرين (Levine & Hullett, 2002). وأحد النتائج الخطيرة لهذا الخلط قد يؤثر سلباً على التحليل البعدي، إذ إننا في التحليل البعدي قد ندمج دون دراية مجموعتين من التقديرات المختلفة لنسبة التباين في المجتمع.

مقارنة قيم مربع أوميجا

يجب أن يكون الباحث الذي يريد تفسير القيم النسبية لتقديرين أو أكثر من تقديري نسبة التباين في المجتمع (نسبة التباين الكلي أو الجزئي) لمختلف المؤثرات في التصميم العاملي حذراً جداً عندما يعرض البيانات. إذ ليس من الضروري أن تكون قيمة ω'^2أثر التي تكون قيمة (ف) المقابلة لها دالة إحصائياً أن تكون نسبة التباين فيها أكبر من قيمة (ف) غير الدالة إحصائياً. فقد يكون لكل من اختباري (ف) قوة إحصائية مختلفة. وبالمثل فإن قيمة أكبر دالة إحصائياً لـ (ف) ليس من الضروري أن يقابلها نسبة أكبر من التباين في المجتمع من القيمة المقابلة لقيمة (ف) الأصغر الدالة إحصائياً. ويرجع ذلك إلى أن تقديرات نسبة التباين في المجتمع لها تباين كبير يرجع إلى المعاينة، ولذلك فإن القيم المذكورة قد تكون غير دقيقة.

وفي حالة المستويات المتعددة للمتغيرات المستقلة المتصلة (مثل مستويات زمن المعالجة) فإن أي فرق ظاهر في تقدير نسبة التباين في المجتمع (أو قيم d أو g) لعاملين قد يعكس فقط قوة معالجة العاملين. وترتبط هذه المشكلة

بمشكلة المدى المحدود التي سبقت مناقشتها. كما أنه قد يكون من غير المناسب ومن غير المفيد مقارنة معالجتين مختلفتي القوة فإن هذا يكون كمن يقارن التفاح بالبرتقال. لنفرض مثلاً أن العامل (أ) يمثل فترة العلاج النفسي، والعامل (ب) عقار ضد الاكتئاب. ما عدد أسابيع العلاج التي يمكن أن تمثل معالجة تبلغ قوتها نفس قوة بضع ميلليجرامات من العقار؟ ومن ناحية أخرى إذا كانت مستويات العاملين اللذين نقارنهما مستويات مقننة لكل عامل في الممارسات الإكلينيكية فقد يكون من الممكن مقارنة تقديري نسبة التباين في المجتمع أو حجم الأثر Δ. كمكا أنه نظراً لأن نسبة التباين في المجتمع لها تباين أكبر، فحتى عندما يكون من الممكن مقارنة قوتين للمعالجة فمن الصعب تعميم الفرق بين قيمتين لنسبة التباين في المجتمع لمجرد أننا نقارن التقديرين.

ويلاحظ أن الاختلاف الكبير في تباين المعاينة فيما يتعلق بنسبة التباين في المجتمع قد تتطلب استخدام حدود الثقة لهذا التباين. وفي هذا المجال يذكر أولجنيك وآلجينا (Olejnik & Algina, 2000) أنه لا يجب مقارنة التقديرات الجزئية لنسبة التباين في المجتمع بالنسبة لعاملين في نفس الدراسة لأنه كما يتبين من المعادلة رقم (٢-٢٢) أن مقام تقدير نسبة التباين الجزئي في المجتمع قد يكون له قيمة مختلفة بالنسبة لكل عامل (مصادر مختلفة للتباين). ولنفس السبب لا يجب مقارنة تقديرات نسبة التباين في المجتمع بالنسبة لأثر عامل معين في دراستين لا تستخدمان نفس العوامل السطحية ونفس مستويات هذه العوامل.

نسبة تقديرات حجم الأثر

يجب أن يكون الباحث حريصاً كذلك عندما يحدد الأهمية النسبية لعاملين عن طريق معرفة نسبة تقديرهما لحجم الأثر. فإن نسبة قيم ω^2 لعاملين يمكن أن تكون مختلفة جداً عن نسبة تقديرين للفرق المقنن لحجم الأثر لهذين العاملين. ولذلك فإن هذين النوعين من التقديرات يمكن أن يعطيا منظوراً مختلفاً لأحجام الأثر النسبية لهذين العاملين. فقد تصل نسبة تقديري ω^2 إلى حوالي ١:٤، وهذه القيمة يمكن أن تجعل البعض يستنتجون أن أهمية

أحد العاملين تبلغ أربعة أمثال أهمية العامل الآخر. ومثل هذا التفسير لا يأخذ في اعتباره قوة المعالجة النسبية للعاملين واحتمال التباين الكبير في المعاينة بالنسبة للتقديرين، كما ذكرنا من قبل.

تلخيص:

ناقشنا في هذا الفصل حجم الأثر في الإحصاء المعلمي حيث تناولنا أحجام الأثر المتعلقة بالفروق بين المتوسطات. وقد بدأنا بمناقشة أحجام الأثر بالنسبة لمعاملات الارتباط. ثم ناقشنا أحجام الأثر المتعلقة بالفروق بين متوسطين (اختبار (ت) بأنواعه الثلاثة). كما ناقشنا أحجام الأثر المتعلقة بالفروق بين أكثر من متوسطين حيث ناقشنا أحجام الأثر لاختبارات تحليل التباين البسيط والعاملي.

وتبين لنا من تلك المناقشات أن هناك طرقاً متعددة لتقدير حجم الأثر. بل إن هناك بالقطع طرقاً أخرى لم يتسن لها مناقشتها وبخاصة تلك التي ترتبط بالعمليات الإحصائية التي لم نناقشها هنا.

ونظراً لأن استخدام طرق مختلفة قد يترتب عليه الحصول على تقديرات متعارضة لحجم الأثر في التراث البحثي، ولذلك لا بد أن يذكر الباحث في تقريره أي طرق استخدمها لتقدير حجم الأثر. لأنه إذا فعل ذلك سوف يساعد القراء على عدم المزج بين طرق متعارضة لتقدير حجم الأثر. وهذا له أثر دون شك على الباحثين الذين يستخدمون التحليل البعدي كأسلوب لدراسة المتغيرات. ويجب كذلك على الباحثين الذين يستخدمون التحليل البعدي كأسلوب لدراسة المتغيرات. ويجب كذلك على الباحثين أن يفكروا ملياً في الغرض من بحثهم، وطبيعة المجتمعات التي يدرسونها. وذلك قبل التفكير في الطريقة التي سوف يستخدمونها لتقدير حجم الأثر، خاصة وأن التراث البحثي مليء بالبحوث التي تتعارض مع بعضها البعض، فقد نجد عدم اتفاق على كيفية حساب Δ أو غيرها من تقديرات حجم الأثر.

ورأينا كذلك كيف أن أحجام الأثر يمكن أن تختلف باختلاف التباين في أفراد العينة في الدراسات المختلفة، لأن نوع المجتمع الذي ندرسه أو طبيعة العينة التي نختارها وما يرتبط بها من تباين قد يكون له أثر كبير على حجم الأثر. ولذلك يجب أن يكون الباحث حريصاً عند مقارنة أحجام الأثر التي يحصل عليها من دراسات مختلفة، لأن اختلاف التباين المحتمل بين أفرادها يمكن أن يكون السبب فيما قد يلاحظ من اختلاف بين أحجام الأثر في الدراسات المختلفة.

الفصل الثالث
حساب حجم الأثر في الإحصاء اللامعلمي

تستخدم المتغيرات القطعية من مستوى الرتبة أو المستوى الإسمي كثيراً في البحوث النفسية والتربوية. ومثل هذه المتغيرات لا يصلح معها استخدام الإحصاء المعلمي الذي يقوم على المتغيرات المتصلة من مستوى المسافة أو مستوى النسبة (رجاء أبو علام، ٢٠٠٦). ومستويات المتغيرات القطعية مستويات جامعة مانعة، فكل حالة تصنف في مستوى واحد فقط. وسوف نتناول في هذا الفصل مؤشرات حجم الأثر الأكثر استخداماً في العلوم النفسية والتربوية وكذلك العلوم الطبية وعلم الوراثة وعلم الأدوية. كما أنها تستخدم كثيراً في التحليل البعدي. ويلاحظ أن بعض هذه المؤشرات تستخدم في بعض الأساليب الإحصائية مثل اللوغاريتم الخطي والانحدار اللوغاريتمي.

وأبسط المتغيرات القطعية متغيرات ثنائية لها مستويان فقط، مثل (ناجح – راسب). وعند مقارنة مجموعتين مقارنة قطعية فإن البيانات تكون تكرارات مصنفة في جدول احتمالي ثنائي التغير، ويعرف أيضاً بالجدول الرباعي. وقد يكون للمتغيرات القطعية أكثر من مستويين مثل موافق، غير متأكد، غير موافق. ويزيد حجم الجدول الاحتمالي عن ٢ × ٢ إذا قارنا بين مجموعتين على أكثر من فئتين. وهناك بعض مؤشرات حجم الأثر التي يمكن امتدادها لتشمل الجداول الأكبر، إلا أن هذه محدودة العدد.

وقد تكون مستويات المتغيرات القطعية مستويات مرتبة أو مستويات غير مرتبة. والمستويات غير المرتبة ليس لها أي ترتيب، ومن أمثلة المستويات غير المرتبة الحالة الاجتماعية: متزوج/ غير متزوج. أو النوع: ذكور/ إناث. والمستويات القطعية المرتبة تعطي معنى معيناً مرتبطاً برتبة ما، مثال ذلك مقياس الاتجاهات من نوع ليكرت: موافق تماماً، موافق، غير متأكد، غير موافق، غير موافق بالمرة.

ومعظم الإحصاء المستخدم يمكن تصنيفه بأنه إحصاء معلمي نستخدم فيه عمليات مثل اختبار (ت) وتحليل التباين وتحليل الارتباط، وغير ذلك من الأساليب الإحصائية. ففي مواقف دراسة تقديرات حجم الأثر التي سبق مناقشتها فيما يتعلق بالارتباط واختبار ت وتحليل التباين بأنواعه المختلفة كان من المسلم به ضمناً أن قياس المتغير التابع هو من مستوى المسافة. وبالإضافة إلى ذلك تطلبت الاختبارات الإحصائية (مثل (ت) وتحليل التباين) أن المجتمع الأصلي الذي سحبت منه العينات ذا توزيع اعتدالي وتباين متجانس. وفي مواقف اختبار الفرض الصفري التي نستخدم فيها الارتباط نسلم ضمناً أن المتغيرين (س) و(ص) من مستوى المسافة، كما أن الاختبارات الإحصائية تطلبت التسليم بتجانس التباين واعتدال توزيع المتغيرات في المجتمع الأصلي الذي سحبت منه العينات. أي أن استخدام الإحصاء المعلمي لأغراض اختبار الفرض الصفري وتحديد مستوى الدلالة يحتاج إلى توافر عدد من المسلمات. وعندما ينتهك واحد أو أكثر من هذه المسلمات التي تتعلق بمعالم وتوزيع المجتمع الأصلي لا بد من استخدام طرق بديلة من الإحصاء الاستدلالي (رجاء أبو علام، ٢٠٠٣).

وحيث إن المجتمعات لا تحقق دائماً المسلمات التي تعتبر أساساً للاختبارات المعلمية، فإننا كثيراً ما نحتاج وسائل أخرى للقيام بالعمليات الإحصائية لا تتطلب مسلمات جامدة. والإحصاء اللامعلمي يحقق لنا هذا الأمر لأنه يصلح في كثير من الظروف التي لا تتطلب إلا مسلمات عامة للغاية. هذا بالإضافة إلى أن الإحصاء اللامعلمي يوفر حاجات أخرى للباحث.

وهناك نوعان من العمليات الإحصائية تعامل معاملة الإحصاء اللامعلمي وهما:

١ – الإجراءات اللامعلمية الحقيقية.

٢ – الإجراءات ذات التوزيع الحر.

والإحصاء اللامعلمي الحقيقي لا يهتم بمعالم المجتمع. ففي اختبارات حسن التطابق نهتم ببعض الخصائص التي تختلف عن قيمة معلم من معالم

المجتمع. وكما يشير هذا المصطلح فإن صدق عمليات «التوزيع الحر» لا يعتمد على شكل توزيع المجتمع الذي سحبت منه العينة. إلا أنه من المعتاد لدى الباحثين والإحصائيين اعتبار هاتين العمليتين من عمليات الإحصاء اللامعلمي.

وقد نما الاهتمام بالإحصاء اللامعلمي من النواحي النظرية والتطبيقية منذ نشأته في منتصف القرن العشرين، حتى إنه يعتبر اليوم من أهم فروع الإحصاء. وتستخدم الأساليب الإحصائية التي تنتمي لهذا الفرع في معظم العلوم الطبيعية والسلوكية والاجتماعية إن لم يكن فيها جميعاً (رجاء ابو علام، ٢٠٠٣).

مزايا استخدام الإحصاء اللامعلمي:

١ – ينخفض احتمال إساءة استخدام الإحصاء اللامعلمي لأن المسلمات التي يقوم عليها محدودة للغاية.

٢ – يمكن إجراء العمليات الحسابية التي تستخدم في الإحصاء اللامعلمي بسهولة وبسرعة وبخاصة إذا أجريت هذه العمليات يدوياً، مما يوفر الوقت. ويمكن أن تظهر أهمية هذه الميزة إذا كانت العمليات الإحصائية مطلوبة بسرعة ولا يتوفر حاسب آلي للقيام بها.

٢ – يسهل الإحاطة بمفاهيم الإحصاء اللامعلمي والقيام بالعمليات الإحصائية التي تتطلبها إجراءاته.

٤ – يمكن استخدام الإحصاء اللامعلمي عندما يكون مستوى القياس ضعيفاً، مثال ذلك عندما تكون البيانات التي لدينا كلها من المستوى الاسمي أو مستوى الرتبة.

عيوب استخدام الإحصاء اللامعلمي:

١ – يفرط بعض الباحثين في استخدام الإحصاء اللامعلمي نظراً لسهولة وسرعة العمليات الحسابية التي يحتاجها، مما يجعلهم يفضلون الإحصاء اللامعلمي رغم أن الإحصاء المعلمي قد يكون هو الأنسب لبياناتهم. وكثيراً ما تؤدي مثل هذه الممارسة إلى فقد كثير من المعلومات.

٢ – رغم أن الإحصاء اللامعلمي لا يتطلب سوى عمليات حسابية بسيطة، إلا أن هذه العملية كثيراً ما تكون متعبة وشاقة، وبخاصة عندما يكون حجم العينة كبيراً.

٣ – قوة الاختبار الإحصائي في الإحصاء اللامعلمي أقل من قوة الاختبار الإحصائي في الإحصاء المعلمي المناظر إذا كان من الممكن استخدامه.

٤ – الفروض التي تختبر باستخدام الإحصاء اللامعلمي ليست مماثلة بالضبط للفروض التي يتم اختبارها باستخدام الإحصاء المعلمي. مثال ذلك أنه بالرغم من أن اختبار (ت) المعلمي لعينتين مستقلتين يقيس على وجه الخصوص الفرض بتساوي متوسطين من متوسطات المجتمع، فإن الإحصاء اللامعلمي المناظر (والذي سوف ندرسه في هذا الفصل) يختبر الفرض بتساوي توزيع مجتمعين، وتأتي المعلومات بتساوي الوسيطين كنتيجة ضمنية.

متى نستخدم الإحصاء اللامعلمي؟

١ – ألا يتضمن اختبار الفروض معلماً من معالم المجتمع.

٢ – أن تكون المقاييس المستخدمة في جمع البيانات من المستوى الاسمي أو مستوى الرتبة مما يجعل من غير الممكن استخدام أساليب إحصائية مناسبة لمستوى أعلى من القياس.

٣ – عندما لا تستوفي البيانات التي لدينا المسلمات الضرورية للإحصاء المعلمي، ففي كثير من الأحيان يتطلب تصميم البحث استخدام أسلوب إحصائي معين، إلا أن فحص البيانات بعد جمعها يكشف أن المسلمات المطلوبة قد انتهكت بشدة. وفي هذه الحالة يكون استخدام الأساليب اللامعلمية هو البديل الوحيد.

وتستخدم الوسائل اللامعلمية لاختبار الفروض أساليب إحصائية مختلفة عن تلك المستخدمة في الإحصاء المعلمي. وقد تتضمن هذه الأساليب تحليلاً إحصائياً يتضمن:

* القيم المرتبة.
* كماً من مجموع القيم في توزيع ما، أكبر (أو أقل) من مجموع القيم في توزيع آخر.
* عمل مقارنات وزنية.
* اختباراً لتحديد إذا ما كان توزيع مجموعة من القيم ينحرف عن التوزيع العشوائي أو التوزيع.
* ذا الحدين.
* اختبار انحراف توزيع أحادي عن التوزيع الاعتدالي.
* المقارنات بين التوزيعات التكرارية.
* مقارنة المجموعات بحساب تكرارات القيم أعلى وأدنى الوسيط العام لتوزيع ما.

وبالإضافة إلى ذلك قد يستخدم الإحصاء اللامعلمي عمليات إحصائية تتعلق بعينة واحدة أو يقوم بعقد مقارنات بين عينتين أو أكثر. وبالرغم مما قد يبدو من تعقيد في تعدد الأساليب الإحصائية اللامعلمية إلا أن معظم الاختبارات اللامعلمية مفهومة تماماً وسهل القيام بها.

واستخدام حجم الأثر في الإحصاء اللامعلمي أمر تال على استخدامه في الإحصاء المعلمي. وهذا أمر طبيعي فقد نشأ التفكير في حجم الأثر عند نشأة التحليل البعدي على يد جلاس. وكان الغرض منه معرفة مدى الفرق بين متوسطي المجموعة التجريبية والمجموعة الضابطة. أي إن نشأة حجم الأثر جاءت مرتبطة بالمتوسط وظلت لذلك فترة من الزمن حتى بدأ التفكير في استخدام حجم الأثر مع دراسة المتوسطات في الاختبارات الإحصائية المعلمية. كما رأينا في الفصلين السابقين.

وسوف نتناول فيما يلي حجم الأثر مع بعض الاختبارات الإحصائية اللامعلمية بادئين بمربع كاي، ثم نليها ببعض الاختبارات اللامعلمية الأخرى. ونظراً لأننا في معظم الاختبارات اللامعلمية لا نستخدم مقاييس النزعة المركزية، بل نستخدم احتمال حدث معين نسبة إلى مجموع التكرارات الممكنة

لهذا الحدث، نجد أن معظم أحجام الأثر المستخدمة في الإحصاء اللامعلمي تستخدم احتمال حدث معين بصفة أساسية نسبة إلى ما يمكن أن نتوقعه من أحداث في المجتمع، أو ما يمكن أن نطلق عليه: **نسبة الأرجحية.**

حساب حجم الأثر لمربع كاي

مربع كاي أكثر الاختبارات استخداماً في العلوم الاجتماعية والعلوم التربوية. ولعل ذلك يرجع إلى سهولة استخدامه في اختبار الفروض كما أن البيانات التي تجري عليها الاختبارات الإحصائية هي عادة بيانات من مستوى الرتبة والمستوى الاسمي. واختبار مربع كاي هو أحد الاختبارات اللامعلمية التي لا تتطلب مسلمات معينة ودقيقة حول شكل توزيع المتغيرات. ونظراً لأن اختبار مربع كاي مناسب تماماً للمستوى الاسمي للقياس فإننا لا نحتاج لعمل مسلمات تتعلق بمستوى القياس. ومن أهم استخدامات مربع كاي قياس التجانس بين متغيرين، واختبار حسن تطابق التوزيعات، إلى غير ذلك من الاستخدامات. ومن أهم استخدامات اختبار مربع كاي اختبار الفرض باستقلال توزيع متغيرين من المستوى الاسمي أو مستوى الرتبة عن بعضهما البعض.

ويرجع الانتشار الواسع لمربع كاي إلى تنوع استخداماته في العديد من مواقف البحث، ربما أكثر من أي اختبار آخر للدلالة الإحصائية. ويمكن استخدام مربع كاي في الاختبارات التي تتعلق بعينتين، إلا أنه يمكن استخدامه أيضاً في المواقف التي تتكون من أكثر من عينتين، أو تتكون من أكثر من فئتين.

أسس استخدام اختبار مربع كاي:

يمكن استخدام مربع كاي في العديد من المواقف، من أهمها اختبار الاستقلالية لمتغيرين نظما في جدول ثنائي البعد.

ويقصد بالاستقلالية في معرض استخدام مربع كاي أن تصنيف الحالة في خلية أو فئة ما من فئات متغير لا تأثير لها على احتمال وقوع هذه الحالة في خلية من خلايا المتغير الآخر في نفس الجدول.

ويلاحظ أن الإجراءات الإحصائية لحجم الأثر المقدمة هنا تستخدم جداول ثنائية (٢ × ٢). وهذه لا تنطبق إلا على المتغيرات القطعية الحقيقية أو الأصلية، وليس على المتغيرات المحولة إلى متغيرات قطعية ثنائية. وهناك مشكلات ترتبط بالعمليات الإحصائية المختلفة عند استخدام متغيرات قطعية ثنائية لم تكن كذلك في الأصل، فالنوع متغير قطعي حقيقي لأنه يتكون بشكل طبيعي من فئتين (ذكور وإناث). أما إذا حولنا درجات متغير متصل (مثل التحصيل في اللغة العربية) عند الوسيط مثلاً لنحصل على فئتين مثل مرتفع ومنخفض فهذا المتغير الناتج عن عملية التحويل إلى متغير قطعي ليس متغيراً قطعياً حقيقياً ولا يصح أن نستخدمه مع العمليات الإحصائية التي سوف نذكرها هنا. لنفرض أن نرغب في مقارنة نوعين من العلاج النفسي من حيث أثرهما على القلق. ولنفرض كذلك أن الباحث كون فئتين من القلق تصنف المرضى في قلق مرتفع وقلق منخفض باستخدام الدرجات التي تزيد أو تقل عن الوسيط (أو أي نقط قطع أخرى على مقياس القلق المتصل). ومثل هذا التقسيم الاعتباري قد يترتب عليه إبطال صدق الإجراءات الإحصائية لأن نتائج التحليل لن تتوقف فقط على فاعلية العلاج في المجموعتين كما هو مفروض، ولكن كذلك على نقطة القطع الاعتبارية التي قرر الباحث اختيارها وصنف جميع الحالات بالنسبة لها في مجموعتين مجموعة أعلى النقطة (اعتبرها مجموعة القلق المرتفع)، ومجموعة أخرى أدناها (اعتبرها مجموعة القلق المنخفض). وإذا كان قد استخدم نقطة قطع أخرى، مثل أدنى ٢٥٪ من الدرجات على مقياس القلق المتصل (القلق المنخفض)، وأعلى ٢٥٪ من الدرجات (القلق المرتفع)، فإن نتائج الاختبارات الإحصائية، وتقدير حجم الأثر سوف تختلف عن النتائج القائمة على تقسيم درجات المقياس المتصل حسب الوسيط تقسيماً اعتبارياً أيضاً. وهناك وسائل أخرى لتصنيف العينة حسب درجات المتغير التابع ويمكن الرجوع إليها في بعض المصادر مثل (Sanchez-Meca, Marin-Martinez, & Chacon-Moscoso, 2003). حيث نجد في هذا المصدر أساليب تؤدي إلى اختيار نقطة القطع بحيث لا تؤثر على تقديرات حجم الأثر. وهذه الوسائل لا تعتمد على تقسيم العينة تقسيماً اعتبارياً في مجموعتين.

وأفضل اختبارات الدلالة الإحصائية لمقاييس الارتباط بين الصفوف والأعمدة في جدول مثل جدول ٣-١، اختبار مربع كاي للارتباط. ودرجات الحرية لمثل هذا الجدول د (ص-١) (ع-١)* وهي في حالة الجدول (٣-١) = (٢-١) (٢-١) = ١. ولكن عندما يكون الاهتمام موجهاً نحو اختبار مربع كاي الذي يعالج قضية وجود ارتباط من عدمه، يكون التركيز على تقدير قوة الارتباط بمقياس مناسب حجم الأثر.

جدول (٣-١) - تكرار نتائج العلاج

العلاج	الأعراض		المجموع
	باقية	اختفت	
علاج نفسي	ك١١ = ١٤	ك١١ = ٢٢	٣٦
علاج بالعقاقير	ك١١ = ٢٢	ك١١ = ١٠	٣٢
المجموع	٣٦	٣٢	٦٨

وكما رأينا من قبل في حالة اختبار (ت) فإن حجم مربع كاي لا يشير بالضرورة إلى قوة الارتباط بين الصفوف والأعمدة. فإن القيمة الرقمية لمربع كاي لا تتوقف فقط على قوة الارتباط ولكن على حجم العينة كذلك، ولذلك فإنه إذا ظل نمط الجدول على ما هو عليه ولكن حدثت زيادة في حجم العينة، فإن مربع كاي سوف تزيد قيمتها.

ونحن نحتاج هنا إلى مقياس لقوة الارتباط بين الصفوف والأعمدة أقل تأثراً بحجم العينة، أو لا يتأثر بها. والمقياس المعروف لحجم الأثر في حالة الجدول الثنائي هو معامل الارتباط في المجتمع. ومعامل الارتباط في المجتمع والذي يمكن استخدامه مع الجداول الثنائية التوزيع هو معامل ارتباط (فاي Phi) وسوف نستخدم معامل ارتباط فاي للعينة لتمثيل معامل ارتباط فاي في

* (الصفوف - ١) (الأعمدة - ١).

المجتمع والرمز المستخدم عادة مع هذا الارتباط هو Φ. ويقدر عادة بمعامل ارتباط فاي للعينة ورمزه ϕ. ورغم أنه من الأسهل اعتبار معامل ارتباط فاي حالة خاصة من معامل ارتباط بيرسون عندما يكون المتغيران (س و ص) قطعيين يجب أن نلاحظ أولاً أن مربع كاي يمكن اعتباره في الواقع مجموع مربعات الآثار، أي أن:

$$\chi^2 = \text{مجـ} \frac{(\text{ك}_{\text{ظ}} - \text{ك}_{\text{ق}})^2}{\text{ك}_{\text{ق}}} \qquad \text{معادلة ٣-١}$$

حيث:

ك$_{ظ}$ = مجموع التكرارات الملاحظة في الخلية

ك$_{ق}$ = مجموع التكرارات المتوقعة في الخلية

مجـ = مجموع تكرارات الخلايا

ولذلك يمكن اعتبار مربع فاي نوعاً من متوسط الأثر، أي الجذر التربيعي لمتوسط مربعات الآثار (Grissom & Kim, 2005). وليس مستغرباً أن تكون فاي متوسطاً لأن ر$_{مجتمع}$ هي نفسها عبارة عن متوسط حاصل ضرب درجات ز:

$$\text{ر}_{\text{مجتمع}} = \frac{\text{مجـ ز}_{\text{س}}\ \text{ز}_{\text{ص}}}{\text{ن}} \qquad \text{معادلة ٣-٢}$$

ويمكن حساب معامل ارتباط فاي بحساب الجذر التربيعي لمربع كاي مقسوماً على ن. ومعامل ارتباط فاي هو تقدير حجم الأثر لاختبار مربع كاي.

ويلاحظ في هذه المعادلة أن معامل فاي باعتباره تقديراً لحجم الأثر يتخلص من تأثير حجم العينة وذلك بقسمة مربع كاي على ن (حجم العينة الكلي).

$$\Phi = \sqrt{\frac{\chi^2}{\text{ن}}} \qquad \text{معادلة ٣-٣}$$

وباستخدام بيانات جدول ٣-١ نجد أن قيمة مربع كاي تبلغ ٦,٠٦ وهذه القيمة دالة عند مستوى ل = ٠١٣,، وعلى هذا تبلغ قيمة فاي:

$$\Phi = \sqrt{\frac{٦,٠٦}{٦٨}} = ,٣٠$$

وهذه القيمة تعتبر دالة إحصائياً بمستوى دلالة قدره ٠١٣, .

وتتراوح قيمة كل من رمجتمع و Φمجتمع نظرياً بين -١ و+١، إلا أنه باستخدام المعادلة رقم ٣-٣ من غير الممكن الحصول على قيم فاي سالبة نظراً لأنها حسبت باستخدام الجذر التربيعي. ويمكن الحصول على هذا المدى باستخدام المعادلة التي ذكرناها في الفصل السابق والتي لا تقوم على الجذر التربيعي لمربع كاي. وعلى أي حال فإن الإشارة المستخدمة مع معامل فاي ليس لها أهمية لأنها نتيجة لطريقة ترتيب بيانات الجدول الذي نستخدمه في حساب معامل فاي. فإذا كانت بيانات جدول ٣-١ قد نظمت بحيث يكون العلاج بالعقاقير ونتائجه في الصف الأول والعلاج النفسي ونتائجه في الصف الثاني فإن إشارة معامل الارتباط تتغير، أما حجمه فيظل كما هو. ولتفسير قيمة معامل فاي التي تساوي ٣٠, (سالبة أو موجبة) فإننا نلاحظ أن اختفاء الأعراض هي أفضل الفئتين. ونلاحظ أيضاً أن ٣٦\٢٢ = ٦١, من مجموع المرضى المعالجين نفسياً حققوا هذه النتيجة الجيدة، في حين حقق هذا التحسن ٣٦\٢٢ = ٣١, من مجموع الذين عولجوا بالعقاقير. وبذلك يكون لدينا التفسير المناسب لقيمة فاي التي حصلنا عليها. ونظراً لأن مربع كاي، وبالتبعية قيمة فاي دالان إحصائياً ونسبة أكبر في مجموع العلاج النفسي من مجموعة العلاج بالعقاقير هي التي تحسنت، فإننا نستنتج أن العلاج النفسي أفضل بشكل دال إحصائياً من العلاج بالعقاقير في المثال الوارد في جدول ٣-١.

ونظراً لأن لدينا الآن التفسير السليم للنتائج فإن مسألة إشارة معامل ارتباط فاي غير مهمة. إلا أنه باستخدام المنطق الوارد في الفصل الثاني بالنسبة لإشارة معامل الارتباط الثنائي الأصيل يمكن للقارئ أن يتبين أن إشارة معامل فاي سالبة بالنسبة لبيانات جدول ٣-١ باستخدام الترميز العادي لكل من (س) و(ص). وإذا كنا قد حددنا رمز الصف الأول بأن س = ١، والصف الثاني بأن س = ٢، والعمود الأول ص = ١ والعمود الثاني ص = ٢ تكون فاي

سالبة لوجود نزعة لدى أولئك الموجودين في الفئة الأدنى للمتغير س (أي الصف الأول) أن توجد في الفئة الأعلى للمتغير ص (أي العمود الثاني) ولدى أولئك الموجودين في الفئة الأعلى للمتغير س (أي الصف الثاني) أن توجد في الفئة الأدنى للمتغير ص (العمود الأول). وهذا النمط من النتائج يؤدي إلى الحصول على علاقة سالبة بين المتغيرين.

وللأسف فإن قيمة معامل فاي لا تتأثر فقط بقوة الارتباط بين متغيرات الصف ومتغيرات الأعمدة، كما يجب أن تكون عليه الحال، ولكنها تتأثر أيضاً بالمجموع الهامشي، مما قد يكون له تأثير سيئ على فاي كمقياس لحجم الأثر. فإن استخدام فاي مفضل فقط في البحوث التي تستخدم عينات طبيعية يختار فيها الباحث حجم العينة الكلي فقط، وليس الصفوف أو الأعمدة، بحيث يكون أي تغير بين مجموع العمودين، أو بين مجموع الصفين تغيراً طبيعياً وليس نتيجة للاختيار الاعتباري للباحث لحجم العينتين. والحصول على قيمة فاي من دراسة أخرى لنفس المتغيرين القطعيين ولكن باستخدام طريقة معاينة تختلف عن المعاينة الطبيعية لن تكون نتائجه مماثلة لمعامل الفاي الذي نحصل عليه من المعاينة الطبيعية. ولذلك يجب على الباحث في البحوث البعدية ألا يقوم فقط بالحصول على متوسط معاملات فاي من دراسات استخدمت وسائل معاينة مختلفة. كما أن فاي لا يمكن أن تصل إلى القيم القصوى لها (+١ ، -١) أي الارتباط التام عندما يكون المتغيران متغيرين قطعيين حقيقيين، وعندما تكون نسبة المجموع الهامشي للصفين مساوية لنسبة المجموع الهامشي للعمودين (Grissom & Kim, 2005).

الفرق بين نسبتين

أحد الأغراض المهمة لحجم الأثر أن يعطي على قدر الإمكان نتائج البحث بطريقة مفهومة للغاية للأفراد ذوي المعلومات المحدودة في الإحصاء مثل العملاء والمرضى ورعاة المرضى وبعض العاملين في التربية وغيرها من المجالات. ولهذا الغرض ربما كان أبسط تقدير للارتباط بين المتغيرات في

جدول ثنائي (٢ × ٢) هو الفرق بين نسبتين تقدران الفرق في الاحتمالات لنتيجة ما بين متغيرين مستقلين. وعلى العكس من فايالمجتمع التي تتطلب معاينة طبيعية فإن هذا المقياس لحجم الأثر يتطلب إما معاينة عشوائية لإحدى مجموعتي المعالجة (في دراسة تجريبية)، أو معاينة غرضية. وفي المعاينة الغرضية لمجموعتين يختار الباحث عينة محددة من المشاركين، بحيث تتكون ن١ من خصائص معينة، ون٢ من خصائص بديلة (مثال ذلك الذكور والإناث أو معالجة سابقة تتكون من مجموعتين: مجموعة العقار أ، ومجموعة العقار ب).

وسوف نستخدم هنا أيضاً بيانات جدول ٣-١، ولكن سوف نفترض أن المشاركين عينوا عشوائياً في كل مجموعة من مجموعات المعالجة. ويلاحظ أن حجم العينتين يختلف في جدول ٣-١ (٣٦ و ٣٢). وبالرغم من أننا عادة نتوقع وجود عينتين متساويتين عندما يكون التعيين عشوائياً، إلا أن المعاينة العشوائية لا تتطلب تساوي حجم العينات. والواقع أن كل ما تحتاجه المعاينة العشوائية هو تعيين الأفراد عشوائياً في مجموعات عشوائية يمثل كل منها ظرفاً أو حالة مختلفة وليس شرطاً أن تكون المجموعات متساوية العدد. ولكن إذا كان اختلاف حجم العينات يرجع إلى إنهاك المشاركين وعدم قدرتهم على المواصلة، قد يترتب على ذلك مشكلات في الدلالة الإحصائية وتقدير حجم الأثر إلا إذا كان هذا الإنهاك موزعاً عشوائياً.

والخطو الأولى هي اختبار أحد فئتي النتيجتين لاستخدامها فيما نسميه الفئة المستهدفة (Target category) أو النتيجة المستهدفة (Target outcome) (Grissom & Kim, 2005, p. 177). وفي جدول ٣-١ يمكننا اختيار «الأعراض اختفت» باعتبارها النتيجة المستهدفة. وسوف نرى فيما بعد أنه من غير المهم أي فئتي النتائج نختار لهذا الغرض. والخطوة التالية هي حساب نسبة مجموع المشاركين في العينة الأولى (العلاج الأول) الذين لديهم هذه النتيجة المستهدفة، ونسبة مجموع المشاركين في العينة الثانية (العلاج الثاني) الذين لديهم هذه النتيجة المستهدفة كذلك. وفي مثالنا نجد أن ٦١, من مرضى العلاج النفسي

و٣١, من مرضى العلاج بالعقاقير تخلصوا من أعراضهم. وبحساب الفرق بين هاتين النسبتين وهي ٦١, – ٣١, = ٣٠,. وتعني هذه النتيجة أن احتمال تخلص أحد أفراد المجتمع من الذين يتلقون العلاج النفسي يبلغ ٦١,، وان احتمال تخلص أحد أفراد المجتمع ممن يتلقون العلاج بالعقاقير هي ٣١,. وهناك تفسير أبسط وهو أن النتائج تشير إلى أنه من بين كل مائة عضو في المجتمع من الذين يتعرضون للعلاج النفسي بالنسبة للأعراض التي لدينا هي ٣٠, (٦١, – ٣١, = ٣٠,) أي نسبة المرضى الذين يشفون من أعراضهم إذا لم يتلقوا العلاج بالعقاقير تزيد بنسبة ٣٠٪ عن أولئك الذين تلقوا علاجاً بالعقاقير.

ونستمر في استخدام العمود الثاني من جدول ٣-١ (الأعراض اختفت) كفئة مستهدفة. وسوف نطلق على نسبة مجموع المشاركين في الصف الأول الذين يوجدون في العمود الثاني (ل١) وعلى نسبة مجموع المشاركين في الصف الثاني الذين يوجدون في العمود الثاني (ل٢)، ولذلك فإن النسبتين اللتين حصلنا عليهما من قبل هما ل١ = ٦١, و ل٢ = ٣١,.

ويلاحظ أن الفرق المطلق بين نسبتين هو نفسه قيمة فاي بالنسبة لجدول ٢ × ٢، إذ إن كلاً منهما = ٣٠, بالنسبة للبيانات الموجودة بجدول ٣-١. وقد لاحظنا من الكلام على BESD في الفصل السابق بالنسبة لجدول مثل جدول ٣-١ أن (ل١) و(ل٢) يمكن أن نطلق عليهما نسبتي النجاح، وأن الفرق بينهما مساو فاي عندما يكون المجموعين الهامشيين متساويين. (إلا أنه من غير المحتمل تساوي المجموع الهامشي في التعيين العشوائي أو المعاينة الطبيعية). وقد لاحظنا أيضاً كما هي الحالة غالباً وكما سبق شرحه عند مناقشة BESD أن مقاييس مختلفة لحجم الأثر يمكن أن تعطي آراء مختلفة بالنسبة للبيانات. مثال ذلك أن فاي = ٣٠, قد لا تكون ذات قيمة بالنسبة للبعض، وأن معامل التحديد ر^{٢} = فاي^{٢} = ٣٠,^{٢} = ٠٩, قد تكون قيمة منخفضة بالنسبة لهم. إلا أن نسبة نجاح قدرها ٦١, بالنسبة لنوع واحد من العلاج وهي تقريباً ضعف نسبة النجاح في النوع الثاني من العلاج تبدو مهمة جداً. والفرق بين نسبتي النجاح يطلق عليه عادة **مجازرة الفرق** (Risk difference).

ولأن طريقة المعاينة في مثالنا السابق تضمنت المعاينة العشوائية للمعالجات بدلاً من المعاينة الطبيعية، فإن هناك وسائل أفضل لتقدير حجم الأثر من الوسيلة المعتمدة على مربع كاي وفاي. والطريقة المقترحة في مثل هذه الأحوال هو التركيز على الفرق بين النسبتين (Grissom & Kim, 2005). وفي هذه الحالة نجد أن ل (تقدير للاحتمال ل في المجتمع. ولذلك فإن أفضل طريقة هي استخدام الطريقة التقليدية لاختبار فـ$_0$: ل$_1$ = ل$_2$ في مقابل فـ$_ب$: ل$_1$ ≠ ل$_2$. ذات ذيلين حيث: (ل$_1$) و(ل$_2$) بالنسبة للعينة تقدير للاحتمال المناظر في المجتمع. وبشكل عام فإن (ل$_ع$) هي احتمال أن فرداً ما في المجتمع والذي عين في فئة المعالجة بالصف الأول سوف يحصل على النتيجة المستهدفة وأن (ل$_غ$) هي احتمال أن فرداً ما في المجتمع والذي عين في فئة المعالجة بالصف الثاني سوف يحصل على النتيجة المستهدفة.

وقد يختار باحث الفئة التي يمثلها العمود الأول فئة مستهدفة بدلاً من استخدام الفئة التي يمثلها العمود الثاني. وليس لهذا الاختيار أي تبعة إحصائية لأن الباحث سوف يحصل على نفس مستوى الدلالة عندما يكون الفرق بين النسبتين محسوباً على أساس العمود الأول أو على أساس العمود الثاني. وبالطبع فإن إيجاد معدل النجاح (نسبة النجاح) مثلاً للعلاج (ع) يكون ذا مستوى دلالة إحصائية أعلى من معدل النجاح للعلاج (غ) معادل لمعرفة أن معدل الفشل للعلاج (ع) يكون ذا مستوى دلالة إحصائية أقل من معدل الفشل للعلاج (غ). وفي جدول ٣-١ يتمثل الفشل في العمود الأول.

وهناك طرق أخرى لاختبار فـ$_0$: ل$_1$ = ل$_2$ تناولتها بعض المصادر مثل (Argesti, 2000) إلا أن الطريقة التي يعطيها (Grissom & Kim, 2005) أبسط وأسهل رغم أنها أقل دقة من بعض الطرق الأخرى، وهذه الطريقة مثال على ما يمكن أن نطلق عليه طريقة العينة الكبيرة (Large sample) أو التقريبية (Approximate) أو المقاربة (Asymptoic) لأن دقتها تزداد مع زيادة حجم العينتين (ن$_1$) و(ن$_2$) (أي مجموعي الصفين في جدول ٣-١). وسوف نذكر معياراً لحجم العينة الكبير في نهاية هذا الجزء. وبعد شرح مفهوم آخر نعطي مثالاً تفصيلياً لهذه الطريقة.

ومتوسط النسبة، $\bar{ل}$ هي نسبة جميع المشاركين (للعينتين) الموجودة بالفئة المستهدفة، وفي جدول ٣-١ حيث يمثل العمود الثاني الفئة المستهدفة،

$$\bar{ل} = \frac{ك_{٢١} + ك_{٢٢}}{ن}$$ معادلة ٣-٤

حيث ن العدد الكلي للعينة ($ن_١ + ن_٢$). وبالنسبة لجدول ٣-١

$\bar{ل} = (٢٢ + ١٠) \setminus ٦٨ = ٠,٤٧$

وهذه القيمة نحتاجها لاختبار الفرض الصفري الحالي. ويمكن أن نطلق على متوسط النسبة، النسبة المجمعة $ل_١$ أي النسبة العامة للمجتمع للذين يوجدون في الفئة المستهدفة. ونظراً لأننا نعتبر الفرض الصفري صحيحاً بصفة مبدئية فإننا نسلم بأن $ل_١ = ل_٢ = ل$ ولذلك فإن أفضل تقدير لقيمة (ل) يمكن الحصول عليها من تجميع (المتوسط العام) $ل_١$ و $ل_٢$ كما هو مبين في المعادلة رقم ٣-٤.

ومعروف أننا لتحويل إحصاءة معينة إلى درجة معيارية (ز) فإننا نقسم الفرق بين تلك القيمة ومتوسط المتغير الذي تنتمي إليه على الانحراف المعياري لهذا المتغير. والإحصاءة التي نركز عليها حالياً هي $ل_١ - ل_٢$ ومتوسط هذه الإحصاءة بعد تكرار معاينتها ومع التسليم كذلك بأن الفرض الصفري صحيح أي صفر. فإن الانحراف المعياري للتوزيع العيني للقيمتين $ل_١-ل_٢$ مبين في مقام المعادلة ٣-٥.

$$ز_{ل_١ - ل_٢} = \frac{ل_١ - ل_٢ - ٠}{\frac{١}{٢}\left(\frac{\bar{ل}_١(١ - \bar{ل}_١)}{ن_١} + \frac{\bar{ل}_٢(١ - \bar{ل}_٢)}{ن_٢}\right)}$$ معادلة ٣-٥

وقد احتفظنا بالقيمة (٠) في المعادلة ٣-٥ لكي يكون واضحاً أن المعادلة تمثل نوعاً من الدرجات المعيارية (ز) ولكننا سوف نناقش بعد قليل سبباً لإحلال قيمة تصحيحية محل القيمة (٠). وتقترب كثيراً قيمة ($زل_١-ل_٢$) من المنحنى الاعتدالي كلما زاد حجم العينة. وبتطبيق القيم الموجودة بجدول ٣-١ على المعادلة ٣-٥ فإننا نحصل على ما يلي:

$$\text{ز}_{\text{ل}_1 - \text{ل}_2} = \frac{\text{٠} - \text{,٣١} - \text{,٦١}}{\frac{1}{2}\left(\frac{(\text{,٤٧} - 1)\text{,٤٧}}{31} + \frac{(\text{,٤٧} - 1)\text{,٤٧}}{36}\right)} = \text{٢,٤٧}$$

وبالرجوع إلى قيمة ز = ٢,٤٧ في جداول المنحنى الاعتدالي فإننا نجد أن (ز) هذه وبالتالي ل١-ل٢ دالة إحصائياً عند مستوى أقل ٠,,٠١٣٦. وهناك تعديل للمعادلة ٣-٥ يوضع فيها في البسط بدلاً من الصفر القيمة التالية: ٠,٥(١ \ ن١ + ١ \ ن٢) للحصول على قيمة تقريبية أكثر للمنحنى الاعتدالي. وإذا وضعنا هذا التعديل في المعادلة بدلاً من الصفر فإننا نحصل على ز=٢,٢٣ وهي قيمة دالة إحصائياً عند مستوى يقل عن ٠,,٠٢٥٨. وينصح جريسوم وكيم (Grissom & Kim, 2005, p. 180) باستخدام هذا التعديل لاختبار (ز). وكقاعدة عامة فإن اختبار (ز) الذي أعطيناه عندما تكون كل القيم التي لدينا ⩽ ٥: ن١ل، ن١(١-ل)، ن٢ل، ن٢(١-ل) وبالنسبة لبيانات جدول ٣-١ تبلغ قيمة ن١ل=٣٦(٠,٤٧) = ١٦,٩٢، ن١(١-ل) = ٣٦(١-٠,٤٧) = ١٩,٠٨، ن٢ل = ٣٢(٠,٤٧) = ١٥,٠٤، ن٢(١-ل) = ٣٢(١-٠,٤٧) = ١٦,٩٦، وجميع القيم تزيد كثيراً على معيار الحد الأدنى الذي قدره ٥.

حدود الثقة التقريبية للقيمة ل١-ل٢

سوف نستعرض هنا أبسط الطرق لبناء حدود الثقة للفرق بين نسبتين (احتمالين) في مجتمعين مستقلين. وبشكل عام فإن أبسط حدود ثقة للفرق بين نسبتين يمكن استخدامها هي حدود الثقة (١-α) للقيمة ل١-ل٢ وهذه يمكن تقريبها على النحو التالي:

ح ث ل١-ل٢: (ل١ - ل٢) ± هـ خ معادلة ٣-٦

حيث هـ خ هامش الخطأ الناتج عن استخدام ل١-ل٢ لتقدير ل١-ل٢

هـ خ = ز ع ل١-ل٢ معادلة ٣-٧

حيث:

زْ قيمة (ز) الموجبة التي لها α\٢ من مساحة المنحنى الاعتدالي بعدها.

ع ل١-ل٢ الانحراف المعياري التقريبي للتوزيع العيني للفرق بين ل١ ول٢

وإذا كنا نرغب كما هو الحال عادة في تحديد ٩٥, ح ث

(أي α\٢ = ٠٥,\٢ = ٠٢٥,) فإننا ننظر إلى قيمة ز في جدول للمنحنى الاعتدالي حيث زْ = +١,٩٦.

نظراً لأننا وجدنا دليلاً على أن ل١ ≠ ل٢ بالنسبة لحدود الثقة فإننا لا نستخدم نفس المعادلة لحساب ع ل١-ل٢ الذي استخدم في مقام المعادلة رقم ٣-٥ عند اختبار أن فـ٠: ل١ ≠ ل٢. ولحساب حدود الثقة فإننا لا نجمع ل١ ول٢ لتقدير القيمة المفترضة مسبقاً بأن ل١ = ل٢ = ل التي سلمنا بها قبل رفض الفرض الصفري. وبدلاً من ذلك فإننا نقدر قيمة كل من ل١ ول٢ بشكل منفصل باستخدام ل١ ول٢ في المعادلة لحساب ع ل١-ل٢

$$ع_{ل_١ - ل_٢} = \frac{١}{٢}\left(\frac{ل_١(١ - ل_١)}{ن_١} + \frac{ل_٢(١ - ل_٢)}{ن_٢}\right) \quad \text{معادلة ٣-٨}$$

ونجمع ل١ و ل٢ عند اختبار الدلالة الإحصائية لأننا في تلك الحالة نفترض صحة (فـ٠)، ولكن لا يوجد مثل هذا الافتراض عند حساب حدود الثقة.

وبالنسبة للبيانات في جدول ٣-١ فإن ل١ – ل٢ = ٦١, – ٣١, = ٣٠,، زْ = +١,٩٦ لأننا نريد حساب حدود الثقة ٩٥,، ١-ل١ = ١ – ٦١, = ٣٩,، ن١ = ٣٦، ١ – ل٢ = ١ – ٣١, = ٦٩,، ن٢ = ٣٢، ولذلك فإننا بتطبيق المعادلة رقم ٣-٧ فإن هـ خ التي نطرحها من ل١ – ل٢ ونضيفها إليها =

١,٩٦ [(٦١,(١-٦١,)\٣٦)+(٣١,(١=٣١,)\٣٢)]^½ = ٢٣, وبذلك تكون حدود الثقة ٣٠, ± ٢٣,. ولذلك يمكن القول إن حدود ثقة ٩٥٪ تقريباً تبلغ من ٣٠,- ٢٣, =٠٧, إلى ٣٠, +٢٣, =٥٣, وفي هذه الحدود يقع الفرق بين ل١ و ل٢. ولسوء الحظ فإن هذه الحدود واسعة جداً، ولكن هذا هو الوضع الغالب. ومع

ذلك فإن حدود الثقة لا تحتوي على القيمة صفر، وهذه النتيجة متسقة مع النتيجة التي حصلنا عليها من اختبار فـ$_0$: ل$_1$ = ل$_2$. ومع ذلك فإنه يلاحظ أنه في بعض الأحيان لا يعطينا اختبار الدلالة عند مستوى معين من ألفا (α) و ($1-\alpha$) ح ث نتائج متساوية للقيمة ل$_1$–ل$_2$.

المجازفة النسبية والعدد الذي نحتاجه للمعالجة

لنفرض أن بيانات جدول ٣–١ جاءت من بحث عين فيه المشاركون تعييناً عشوائياً للعلاج ١ والعلاج ٢، وهذا الافتراض هو في الواقع صحيح بالنسبة لهذه البيانات. في هذه الحالة فإن حساب حجم الأثر يطلق عليه المجازفة النسبية (Relative risk)، وهذا ما سوف نناقشه فيما يلي.

إن الفرق بين ل$_1$ ول$_2$ قد يكون له أهمية عملية عندما يكون تقدير قيمتي (ل) قريباً جداً من صفر أو ١ مما لو كان أقرب إلى ٠,٥. لنفرض أن ل$_1$=٠١٠,٠ و ل$_2$ = ٠٠١,٠ أو أن ل$_1$ = ٥٠٠,٠ و ل$_2$ = ٤٩١,٠، في كلتا الحالتين تبلغ ل$_1$–ل$_2$ = ٠٠٩,٠ ولكن في الحالة الأولى تبلغ قيمة ل$_1$ عشرة أمثال ل$_2$(ل$_1$\ل$_2$ = ٥٠٠,٠\٤٩١,٠ = ١,٠١٨). ولذلك فإن نسبة كل من الاحتمالين قد تكون ذا فائدة. وبالنسبة للجداول ٢×٢ فإن نسبة الاحتمالين يرمز لها بالقيمة ن م أي نسبة المجازفة (rate ratio أو risk ratio) وتقدر ن م بالقيمة نمـ التي تحتسب من بيانات العينة:

$$\text{نمـ} = \frac{\text{ل}_1}{\text{ل}_2} \qquad \text{معادلة ٣–٩}$$

وتمثل ل$_1$ و ل$_2$ نسبة المشاركين في العينتين ١ و ٢ على التوالي والذين يقعون في الفئة المستهدفة والتي يمثلهما العمود ١ أو العمود ٢ في جدول مثل جدول ٣–١. وبالنسبة لهذا الجدول الأخير إذا كان العمود ١ يمثل الفئة المستهدفة فإن نمـ$_1$ = (١٤\٣٦) ÷ (٢٢\٣٢) = ٥٧,٠ وإذا كان العمود ٢ يمثل الفئة المستهدفة فإن نمـ$_2$ = (٢٢\٣٦) ÷ (١٠\٣٢) = ١,٩٦. وبالنسبة للحالة

الأخيرة هناك احتمال قدره حوالي ٢ إلى ١ احتمالاً أكبر لنجاح العلاج النفسي من العلاج بالعقاقير وذلك بالنسبة للمشكلة الإكلينيكية الحالية. (وكما سبق أن ذكرنا فإن معنى الفرق بين ل١ و ل٢ يختلف باختلاف ل١ و ل٢ ولذلك تكون ن م أكثر فائدة كمقياس لحجم الأثر في بحوث التحليل البعدي من (ل١-ل٢) (Grissom & Kim, 2005, p. 184).

وتنتمي التسمية م ن (المجازفة النسبية Relative risk) إلى مجال البحوث الطبية حيث تتكون الفئة المستهدفة من أناس يتم تصنيفهم في فئة حاملة للمرض، وفئة غير حاملة للمرض. ولإحدى العينتين عامل مجازفة مفترض بالنسبة للمرض (مثال ذلك المدخنين)، أما العينة الثانية فليس لديها هذا العامل المفترض. ولكن لأنه يبدو غريباً أن تعطي التسمية مجازفة نسبية عند تطبيق النسبة على عمود مثل عمود ٢ في جدول ٣-١ والذي يمثل نجاح العلاج، وفي مثل هذه الحالة يمكن الإشارة إلى ن م (أو نمـ بأنها تعني نسبة النجاح)، أو نسبة احتمالين مستقلين، أو نسبة نسبتين مستقلتين.

وأحد نواحي القصور في (م ن) أننا نحصل على قيم مختلفة طبقاً لاختيار الباحث لوضع المجموعتين في البسط أو المقام مما يؤدي إلى نتائج مختلفة. وتنبع المشكلة من أنه نظراً لأن (ن م) عبارة عن نسبة فإنها أي (نمـ) تتراوح بين صفر و ١ إذا وضعت المجموعة ذات النسبة الأقل في البسط، ولكن قد تتراوح بين ص و ∞، إذا وضعت هذه المجموعة في المقام. ويمكن حل هذه المشكلة بشكل جزئي إذا أعطى الباحث لوغاريتم (نمـ) كتقدير للوغاريتم (ن م). وعندما تكون النسبة الأصغر في البسط فإن (لو نمـ) يمكن أن تتراوح بين صفر و -∞ أما إذا وضعت النسبة الأكبر في البسط فإن (لو نمـ) قد تتراوح بين صفر و +∞. ويجب باستمرار إعطاء النسب الخام مهما كانت الطريقة التي حسبت بها (نمـ). وتختلف أيضاً قيمة المجازفة بناء على الفئة التي تبنى عليها، ولذلك يجب إعطاء النتائج بالطريقتين حتى لا نضلل القارئ، لأنه سوف يعلم كيف حسبت النتائج بالضبط، وكيف أن (نمـ) تختلف لو حسبت بالطريقة الأخرى.

نسبة الأرجحية

وحجم الأثر الأخير الذي نتناوله بالنسبة للجداول الثنائية ٢×٢ هي نسبة الأرجحية (ن أ) (Odds Ratio, OR) وهي مقياس لعدد المرات الأكثر أرجحية أن فرداً من أفراد مجتمع معين سوف يقع في فئة محددة أكثر من فرد آخر ينتمي لمجتمع آخر. وينطبق حجم الأثر هذا على البحوث التي تستخدم التعيين العشوائي (Random assigmment)، والبحوث الطبيعية (Naturalistic research)، والبحوث التوقعية (Prospective ersearch)، والبحوث التراجعية (Retrospective research) (Fleiss et al., 2003). وعلى العكس من معامل فاي فإن المدى المحتمل لقيم نسبة الأرجحية ليس محدوداً بالتوزيع الهامشي للجداول الاحتمالية وسوف نشرح حساب حجم الأثر كنسبة أرجحية وذلك بالنسبة لبحث طبيعي (جدول ٣-٢) وفي هذا المثال نجد أن نسبة الأرجحية للعينة توفر لنا تقديراً لنسبة:

أ – أرجحية أن المشاركين من نوع معين (النساء مثلاً) سوف يتمسكون بوجودهم في فئة معينة (التصويت لحزب الوفد بدلاً من الحزب الوطني).

ب – نفس الأرجحية للنوع الآخر (الرجال مثلاً).

ويمكن حساب نسبة الأرجحية لأي زوج من الفئات لمتغير ما (النوع مثلاً) يرتبط بزوج آخر من الفئات لمتغير آخر (الانتماء الحزبي مثلاً).

ويمكن تعريف نسبة الأرجحية أن أمجتمع إذا نظرنا إلى حالة عادية يسبق فيها التقسيم بالنسبة لأحد المتغيرين التقسيم بالنسبة للمتغير الآخر. مثال ذلك أن التقسيم حسب النوع (ذكور أو إناث) يسبق التقسيم حسب الاستجابة (موافق أو غير موافق) كما هو الحال في جدول ٣-٢. فإذا أسمينا أحد فئتي النتيجة المستهدفة م (موافق) فإن الفئة الناتجة الأخرى تسمى غ م (غير موافق). ثم نسمي فئة لها أسبقية التقسيم (رجال مثلاً) (أت). وحيث ترمز (ل) إلى الاحتمال فإن مقياسا بأرجحية أن (م) سوف تحدث شريطة حدوث أت=ل(م|أت)\ل (غ م | أت). وبالمثل فإن الأرجحية بأن (م) سوف تحدث شريطة عدم حدوث الفئة أت (النساء مثلاً) نحصل عليها من الأرجحيةغ أ ت = ل (م| غ أت)\ ل (غ م| غ أت). ونسبة هاتين الأرجحيتين في المجتمع هي نسبة الأرجحية: ن أ = أرجحيةأت | أرجحية غ أ ت.

جدول (٣-٢)

الفرق بين النوعين في الاتجاه نحو عبارة جدلية

النوع	موافق	غير موافق
رجال	ك١١ = ١٠	ك٢١ = ١٣
نساء	ك١٢ = ١	ك٢٢ = ٢٣

تتكون خلايا جدول ٣-١ من تكرارات المشاركين، في الصف الأول والعمود الأول، والصف الثاني والعمود الثاني، والصف الثاني والعمود الأول، والصف الثاني والعمود الثاني على التوالي. وسوف نستخدم الفئة الممثلة بالعمود الأول كفئة مستهدفة. وأرجحية العينة أن مشاركاً من الصف الأول سوف يكون في العمود الأول بدلاً من العمود الثاني هي ك١١\ك٢١ وهذه تساوي تقريباً ١٠\١٣ = ,٧٧ ، في حالة جدول ٣-٢. وإذا كانت الدراسة تقارن نوعين مختلفين من المشاركين الممثلين بالصفين يمكننا تقويم هذه الأرجحية بالنسبة لأرجحية مماثلة للذين يوجدون بالصف الثاني. وأرجحية أن مشاركاً من الصف الثاني سوف يكون في العمود الأول بدلاً من العمود الثاني هي ك١٢\ن٢٢ وهذه تساوي تقريباً ١\٢٣ = ,٠٤ في جدول ٣-٢. ونسبة أرجحية العينتين والتي أطلقنا عليها (ن أ) يمكن الحصول عليها من حساب (ك١١\ك١٢)\(ك١٢\ك٢٢) وهي تعادل (أ\ب)\(ج\د) = (أ د)\(ب ج) وهذه تساوي:

$$\text{ن أ} = \frac{\text{ك}_{11}\ \text{ك}_{22}}{\text{ك}_{21}\ \text{ك}_{12}}$$ معادلة ٣-١٠

ويلاحظ أنه في المعادلة ٣-١٠ نقوم بضرب تكرار كل خلية في تكرار الخلية الواقعة على الخط القطري لها في جدول مثل جدول ٣-٢. ولذلك يطلق أيضاً على نسبة الأرجحية نسبة حاصل ضرب القطرين. ويلاحظ أن الأرجحية ليست مرادفة للاحتمال. وقد لاحظنا ذلك من أن أرجحية وجود رجل في فئة الموافقين = ١٠\٢٣ = ,٧٧ . إلا أن الاحتمال أن رجلاً سوف يوجد في فئة الموافقين بالنسبة ١٠\٢٣ = ,٤٣ ، حيث إن ٢٣ هو العدد الكلي للرجال في العينة.

ويعطينا جدول ٣-٢ بيانات حقيقية، ولكن المثال المذكور مثال افتراضي لأن العنوان تغير وكذلك عنوان الصفوف والأعمدة لتحقيق الغرض من هذا الجزء. ويظهر جانب مهم جداً من إذا ربطنا أرجحية أن رجلاً يوافق بدلاً من غير موافق لعبارة اختبار جدلية قدمت لجميع المشاركين بواسطة الباحث. وبتطبيق المعادلة ٣-١٠ فإننا نجد أن: ن أ = ١٠(٢٣)\ ١٣(١) = ١٧,٦٩ ومعنى هذا أن أرجحية موافقة رجل مع العبارة الجدلية تقدر بحوالي ١٨ مرة من أرجحية موافقة امرأة معها. ومع ذلك إذا نزعنا هذه العبارة من سياقها فقد تكون مضللة لأننا إذا نظرنا لجدول ٣-٢ الذي يجب على الباحث وضعه في تقرير البحث، فإننا نلاحظ أيضاً أنه يوجد في الواقع أغلبية من الرجال (١٣ من ٢٣)، وكذلك أغلبية أكثر من النساء (٢٣ من ٢٤) غير موافقين على العبارة.

ويتراوح كل من (ن أ) العينة والمعلم ن أمجتمع بين صفر وما لا نهاية، وتصل القيم إلى أحد هاتين القيمتين المتطرفتين عندما يكون تكرار أحد الخلايا صفراً. وعندما لا يكون هناك ارتباط بين متغيرات الصف والأعمدة تكون قيمة ن أمجتمع = ١. ووجود خلية تكرارها صفر في المجتمع (ويطلق عليه صفر بنائي Structural zero) غير محتمل لأنه في معظم البحوث في العلوم السلوكية والاجتماعية من غير المحتمل أن يستخدم الباحث متغيراً لا يوجد في أحد فئاته أي عضو في المجتمع. ومع ذلك فإننا نلاحظ من البيانات الحقيقية الموجودة بجدول ٣-٢ أننا اقتربنا جداً من الصفر في أحد خلايا العينة (الخلية ١٢) التي يبلغ تكرارها ١. وفي البحوث التي يكون من غير المحتمل أن تكون "ن أمجتمع» صفراً أو ما لا نهاية من غير المرغوب فيه أن يحتوي أحد خلايا الجدول على صفر أو ما لا نهاية. وعندما يكون هناك خلية خالية من بيانات العينة ولا تعكس وجود خلية تكرارها صفر في المجتمع، فلا بد من حل هذه المشكلة التي تتطلب وجود خلية تكرارها صفر. وأحد الحلول الممكنة زيادة فرصة إضافة بيانات لهذه الخلية الخالية، وذلك بزيادة حجم العينة الكلي بعدد ثابت. وهناك حل آخر شائع وهو إضافة قيمة ثابتة ضئيلة جداً إلى جميع الخلايا وليس الخلية الخالية فقط. وهذه الثوابت المقترحة بلغ أدناها في البحوث السابقة ١٠-٨ وأقصاها ٠,٥ (Grissom & Kim, 2005, p.190).

ويمكن اختبار الفرض الصفري فـ$_0$: ن أمجتمع = ١ في مقابل الفرض البديل أن فـ$_1$: ن أمجتمع ≠ ١. وذلك باستخدام اختبار مربع كاي المصحح (أي طرح ٠,٥ من البسط قبل عملية التربيع. وتزداد دقة عندما تكون التكرارات المتوقعة كبيرة. ولا يجب استخدام هذه الطريقة إذا كان أي من التكرارات المتوقعة يقل عن ٥. والاختبار الإحصائي هو:

$$\chi^2 = \text{مجـ} \frac{[|ك_{رص} - (ن_{ر} ن_{ص} \backslash ن)| - ٠,٥]^2}{ن_{ر} ن_{ص} \backslash ن}$$ معادلة ٣-١١

حيث:

تجرى عملية الجمع على الخلايا الأربع للجداول.

$ك_{رص}$ التكرار الملاحظ في كل خلية.

$ن_{رن_{ص}}$ المجموع الكلي لعمود معين توجد به الخلية المقصودة.

$ن_{رن_{ص}} \backslash ن$ التكرار المتوقع في خلية معينة تحت الفرض الصفري.

(يلاحظ أنه على عكس الحال في مربع كاي المستخدم في حساب فاي يجب استخدام المعادلة ٣-١١ التي عدل فيها البسط).

وبتطبيق بيانات جدول ٣-٢ على المعادلة ٣-١١ فإننا نجد:

$$\chi^2 = [|١٠ - (٢٣ \times ١١ \backslash ٤٧)| - ٠,٥]^2 \backslash (٢٣ \times ١١ \backslash ٤٧) + [|١٣ - (٢٣ \times ٣٦ \backslash ٤٧)| - ٠,٥]^2 \backslash (٢٣ \times ٣٦ \backslash ٤٧) + [|(٢٤ \times ١١ \backslash ٤٧) - ١|]^2 \backslash (٢٤ \times ١١ \backslash ٤٧) + [|(٢٤ \times ٣٦ \backslash ٤٧) - ٢٣|]^2 \backslash (٢٤ \times ٣٦ \backslash ٤٧) = ٨,٠٥.$$

وبالرجوع إلى أي جدول للقيم الحرجة لمربع كاي في أي كتاب من كتب الإحصاء فإننا نجد أنه بالنسبة لدرجات حرية (ص-١) (ع-١) = (٢-١) (٢-١) = ١ أن القيمة ٨,٠٥ دالة إحصائياً عند مستوى يقل عن ٠٠٥, وبذلك يكون لدينا دليل كاف على أن ن أمجتمع لا تساوي ١.

حدود الثقة لقيمة ن أمجتمع

يمكن بناء حدود ثقة تقريبية غير مباشرة للقيمة ن أمجتمع على أساس المنحنى الاعتدالي. وكبر حجم العينة يساعد على دقة التقريب. وسوف نستخدم هنا أيضاً أبسط طريقة.

ويمكن بناء حدود الثقة للوغاريتم الطبيعي للقيمة ن أمجتمع، أي لـو ن أمجتمع لأنه مع زيادة حجم العينة يحدث تقريب أسرع للمنحنى الاعتدالي للتوزيع العيني لـ لـو ن أ أسرع من التوزيع العيني لقيمة ن أ. ثم يعطينا المقابل اللوغاريتمي لحدود هذه المسافة حدود الثقة للقيمة لـو ن أمجتمع ذاتها. وبإضافة القيمة الثابتة ٠,٥ لتكرار كل خلية نقلل من التحيز في تقدير لـو ن أمجتمع، ولذلك نستخدم هذا التعديل. وحدود الثقة (١–α) ح ث للقيمة لـو ن أمجتمع التقريبية هي:

لـو ن أمجتمع ± زα\٢ ع لو ن أ معادلة ٣–١٢

حيث:

زα\٢ قيمة ز (الانحراف المعياري الاعتدالي) والتي بعدها تقع نسبة α\٢ من المساحة تحت المنحنى الاعتدالي.

ع لو ن أ الانحراف المعياري (الخطأ المعياري) للتوزيع العيني للقيمة لـو ن أ:

$$ع_{\text{لو ن أ}} = \left(\frac{١}{ك_{٢١} + ٠,٥} + \frac{١}{ك_{٢١} + ٠,٥} + \frac{١}{ك_{١٢} + ٠,٥} + \frac{١}{ك_{٢٢} + ٠,٥}\right)^{\frac{١}{٢}}$$ معادلة ٣–١٣

ورغم أن هناك مشكلة بدقة هذه المعادلة وذلك بسبب صغر حجم العينات بالنسبة لما جار في البحوث السلوكية فسوف نطبق بيانات جدول ٣–٢ على هذه المعادلة:

$$ع_{\text{لو ن أ}} = \left(\frac{١}{١٠ + ٠,٥} + \frac{١}{١٣ + ٠,٥} + \frac{١}{١ + ٠,٥} + \frac{١}{٢٣ + ٠,٥}\right)^{\frac{١}{٢}} = ٠,٩٣٧$$

وبعمل حدود الثقة المعتادة (١–α) = (١–٠٥,٠) = ٠٩٥,٠ ح ث، فإننا نستخدم زα\٢ = ١,٩٦ لأن ما مجموعه ٠٥,٠ (أي ٠٢٥,٠ + ٠٢٥,٠) من المساحة تحت المنحنى

الاعتدالي تقع في الذيلين خارج ز = ± ١,٩٦. ولذلك فإن ٩٥, حدود ثقة للقيمة ن أمجتمع على أ ن أمعدل = ١٢,١٩ من القسم السابق هي لو (١٢,١٩) ± ١,٩٦ (,٩٣٧) التي هي ٦٦٤, و ٤,٣٣٧. واللوغاريتم المقابل للقيمتين ٦٦٤, و٤,٣٣٧ تعطي ٩٥, حدود ثقة للقيمة ن أمجتمع ذاتها ١,٩٤ و ٧٦,٤٨.

ويلاحظ أن حدود الثقة في هذا المثال واسعة للغاية (١,٩٤ و ٧٦,٤٨). ومع ذلك فمن المحتمل أنه يعطي نتائج حرة، بمعنى أن الحدود الناتجة عنه أضيق من الحدود الفعلية. ومن البرامج الإحصائية التي تعطي حدود ثقة SAS (الإصدار التاسع) وStatXact اللذين يعطيان حدود ثقة دقيقة لقيمة ن أمجتمع للمستوى ١–α أي ٩٥, .

احتمال التفوق: الفروق بين عينتين مستقلتين

لنفرض أننا نريد حساب حجم أثر يعكس ما يمكن أن يحدث إذا كنا قادرين على أخذ كل درجة من درجات المجتمع (أ) ومقارنتها مع كل درجة من المجتمع (ب)، كل على حدة لمعرفة أي الدرجتين أكبر، مع تكرار هذه المقارنات حتى نستغرق كل درجة في المجتمعين بحيث نقارن كل درجة من المجتمع (أ) بكل درجة من درجات المجتمع (ب). وإذا حدث أنه في معظم المقارنات كانت الدرجات في المجتمع (أ) أكبر من نظيراتها في المجتمع (ب)، فإن ذلك يشير إلى نزعة لتفوق المجتمع (أ) عن المجتمع (ب). ونحصل نتيجة لهذه الطريقة على حجم أثر لا يتضمن مقارنة مقاييس النزعة المركزية للتوزيعين مثل المتوسط أو الوسيط. وهذا النوع من حجم الأثر يعرف بأنه احتمال أن عضواً مختاراً اختياراً عشوائياً من المجتمع (أ) سوف يحصل على درجة (صأ) تكون أعلى من نظيرتها في المجتمع ب (صب) التي حصل عليها فرد مختار اختياراً عشوائياً من المجتمع (ب).

والتعبير المستخدم لهذا النوع من حجم الأثر هو ل(صأ > صب) حيث (ل) تشير إلى الاحتمال. وقد أطلق جريسوم وكيم (Grissom & Kim, 2005) على هذا الاحتمال احتمال التفوق (أت) (Probability of superiority, PS). ويمكن تعرف (أت) على النحو التالي:

أت = ل (صأ > صب) معادلة ٣-١٤

وتقيس أت احتمال تفوق درجات مجموعة من الأفراد على مجموعة أخرى. ونظراً لأن أت احتمال ولأن الاحتمالات تتراوح بين صفر و ١، فإن أت تتراوح بين صفر و١. ولذلك فإن أقصى نتيجتين عندما نقارن المجتمع (أ) بالمجتمع (ب) يمكن أن تكون:

* أت = ٠، حيث يتفوق كل فرد في المجموعة (ب) على كل فرد في المجموعة (ب).

* أت = ١، حيث يتفوق كل فرد في المجموعة (أ) على كل فرد في المجموعة (ب).

* أت = ٠,٥، حيث لا يوجد أي أثر لمجموعة على المجموعة الأخرى، وهذه هي أقل النتائج تطرفاً، لأن أفراد المجتمعين أ و ب يتفوقان على بعضهما البعض بدرجة متساوية.

وتقدر نسبة العينة الاحتمال في المجتمع. فإذا أحصينا مثلاً ٥٢ صورة في عينة قدرها ١٠٠ (رمية عشوائية للعملة) فإن نسبة الصور في نتائج العينة تبلغ ٥٢\١٠٠ = ٠,٥٢، ويكون تقديرنا لاحتمال الحصول على الصور في مجتمع الرميات العشوائية لتلك العملة بالذات ٠,٥٢. وبالمثل يمكن تقدير أت من نسبة عدد المرات التي تفوق فيها المشاركون في العينة نأ على المشاركين في العينة نب وذلك بمقارنة رمية كل عملة نحصل فيها على صورة من بين جميع المقارنات الزوجية الممكنة لدرجات كل عضو في العينة (أ) مع درجات كل عضو في العينة (ب). والعدد الكلي لهذه المقارنات الممكنة نحصل عليه من ضرب حجم العينتين في بعضهما البعض، نأنب. ولذلك إذا قلنا مثلاً أن نأ = نب = ١٠ (وليس شرطاً تساوي حجم العينتين) وحصلنا في ٧٠ رمية من مقارنات الدرجات التي عددها = نأنب = ١٠٠ على درجات أعلى للعينة (أ) من العينة (ب) فإن تقدير أت يكون ٧٠\١٠٠ = ٠,٧٠ (Grissom & Kim, 2005).

ويعطينا جريسوم وكيم مثالاً بسيطاً ولكنه أكثر تفصيلاً إذا افترضنا أن العينة (أ) التي تتكون من ثلاثة أشخاص أ، ب، جـ وأن العينة (ب) التي تتكون من ثلاثة أشخاص أيضاً هم د، هـ، و. في هذه الحالة يكون عدد المقارنات الممكنة نأنب = ٣ × ٣ = ٩ مقارنات زوجية نلاحظ فيها أي الأشخاص يحصل على درجة أعلى:

أ في مقابل د
أ في مقابل هـ
أ في مقابل و
ب في مقابل د
ب في مقابل هـ
ب في مقابل و
جـ في مقابل د
جـ في مقابل و
جـ في مقابل هـ

ولنفرض أنه في خمس من هذه المقارنات الزوجية التسع تفوقت درجات الأشخاص أ، ب، جـ (أعضاء العينة أ) على درجات الأشخاص د، هـ، و (أعضاء العينة ب)، ويفوز أفراد العينة (ب) في أربع مقارنات زوجية. في هذا المثال يمكن تقدير أت = ٥\٩ = ٥٦,. وبالطبع لا يمكن أن نبني تقديراتنا في البحوث الحقيقية على عينة صغيرة تتكون من ثلاثة أفراد فقط.

وتزيد التقديرات على ٥, عندما يتفوق أعضاء العينة (أ) على أعضاء العينة (ب) في أكثر من نصف المقارنات الزوجية. وتقل التقديرات عن ٥, عندما يتفوق أفراد العينة (ب) على أفراد العينة (أ). وإذا وجدت عقد (تساوي المقارنات) فإن أبسط حل هو تقسيم هذه العقد على العينتين بالتساوي. وفي هذا المثال إذا تفوق أفراد العينة (أ) أربع مرات وليس خمس مرات، وكان هناك عقدة واحدة، فإن نصف هذه العقدة يذهب إلى العينة (أ)، والنصف الثاني

يذهب إلى العينة (ب)، وفي هذه الحالة يكون عدد مرات تفوق العينة (أ) ٤,٥ من تسع مقارنات زوجية، ويكون تقدير أت = ٤,٥\٩ = ٠,٥ . وهناك مقياس آخر يرتبط بمقياس أت ولكنه يتجاهل العقد سوف نناقشه فيما بعد.

وعدد المرات التي تتفوق فيها درجات عينة محددة على درجات عينة أخرى تتم معها عمل المقارنات الزوجية (بسط نسبة العينة المستخدم في حساب أت) يطلق عليه إحصاءة U (نسبة إلى مان و ويتني). ونظراً لأن عدد المقارنات الممكنة سبق تحديده بأنه يساوي نأنب وباستخدام لَأ>ب لترمز إلى نسبة العينة التي تقدر أت يمكن الآن أن نعطي التعريف التالي (Grissom & Kim, 2005, p. 100):

$$لَأ>ب = \frac{U}{ن_أ ن_ب}$$ معادلة ٣-١٥

وفي المعادلة ٣-١٥ يمثل البسط عدد مرات الفوز لعينة محددة والمقام عدد فرص الفوز في مقارنات زوجية لدرجات كل فرد من أفراد العينة مع درجات كل فرد من أفراد العينة الأخرى. ويمكن حساب قيمة U يدوياً ولكن هذه العملية قد تكون مرهقة جداً إذا كان عدد أفراد العينة كبيراً. ورغم عدم توافر برامج إحصائية لحساب لَأ>ب إلا أن كثيراً من البرامج المتوفرة تحسب إحصاءة U لمان ويتني أو القيمة المناظرة وهي الإحصاءة W_m. وعند الحصول على قيمة U من برنامج إحصائي يمكن قسمتها على ن أ ن ب للحصول على لَأ>ب. وإذا كان البرنامج يعطي القيمة المناظرة لويلكوكسن وهي إحصاءة مجموع الرتب يمكن حساب U بحساب $U = U_m - ن_جـ[(ن_جـ + ١)]\٢$، حيث ن جـ عدد أفراد العينة الأصغر، وإذا تساوت العينتان يمكن استخدام حجم أي العينتين. وتعتبر لَأ>ب تقديراً لحجم الأثر كما ذكر جريسوم وكيم، ولذلك يمكننا إعادة صياغة المعادلة ٣-١٥ على النحو التالي:

$$ح = \frac{U}{ن_أ ن_ب}$$ معادلة ٣-١٦

وفي هذه الحالة نختبر الفرض الصفري بأن فـ$_0$: ح = ٠,٥ في مقابل الفرض البديل فـ$_ب$: ح≠ ٠,٥. ويمكن للباحثين الذين يفضلون استخدام المتوسط لتسليمهم باعتدال التوزيع وتساوي التباين، استخدام اختبار (ت) لمقارنة المتوسطين وفي هذه الحالة يمكنهم استخدام حجم الأثر الخاص باختبار (ت) كما جاء في الفصل الثاني من هذا الكتاب.

العينات غير المستقلة (المتكافئة)

إن احتمال التفوق الذي سبق الإشارة إليه لا ينطبق على تصميمات المجموعات غير المستقلة. وفي هذه الحالة يمكن تقدير حجم أثر مشابه يمكن أن نطلق عليه احتمال التفوق للمجموعات غير المستقلة أو المتكافئة، ويمكن تعريفه بالمعادلة رقم (٣-١٧).

أح$_{متكافئة}$ = ل (ص$_{س\ ب}$ > ص$_{س\ أ}$) معادلة ٣-١٧

حيث ص$_{س\ ب}$ هي درجة الفرد تحت ظروف الحالة (ب) و ص$_{س\ أ}$ هي درجة نفس الفرد تحت ظروف الحالة (أ). وسوف نستخدم حالة إعادة القياس (أي نفس الفرد) حتى نهاية هذا الجزء.

و أح$_{متكافئة}$ كما تم تعريفها في المعادلة (٣-١٧) هي احتمال أنه في معاينة عشوائية لزوجين متكافئين من الدرجات (أي درجتين من نفس المشارك في ظرفين مختلفين)، فإن الدرجة التي نحصل عليها تحت الظرف (ب) تكون أكبر من الرجة التي نحصل عليها تحت الظرف (أ). ويلاحظ أن هناك فرقاً بين تعريف أح السابق وتعريفه بمقتضى أح$_{متكافئة}$، ففي هذه الحالة الأخيرة فإننا نقدر حجم الأثر الذي ينجم عن الدرجة التي نحصل عليها لكل فرد من أفراد المجتمع الذي سحبنا منه العينة، وذلك بمقارنة درجتين عليهما من نفس الفرد تحت ظرفين مختلفين (أي الحالة أ، والحالة ب) لمقارنتهما بغرض معرفة أيهما أكبر.

وبمعنى أدق فإننا نبدأ بتقدير أح$_{متكافئة}$ وذلك بالحصول على درجتين من كل فرد ومقارنة الدرجتين. مثال ذلك الحصول من الطالب أحمد على درجة في

الحالة (أ)، وعلى درجة أخرى له في الحالة (ب)، ثم نقارن بين الدرجتين. وحجم الأثر في هذه الحالة هو نسبة جميع المقارنات الداخلية للأفراد، والتي تكون فيها الدرجة تحت ظروف الحالة (ب) أكبر من الدرجة تحت ظروف الحالة (أ). ويتم تجاهل العقد في مثل هذه الحالة. فإذا حصلنا مثلاً من مائة مشارك (ن = ١٠٠) على درجات كل منها ٦٠ درجة أكبر تحت ظروف الحالة (ب) منها تحت الحالة (أ)، فإن تقدير أحمتكافئة هو:

$$\text{ل متكافئة} = ٦٠ \backslash ١٠٠ = ٠,٦٠$$

تقدير حجم الأثر في الجداول الكبيرة

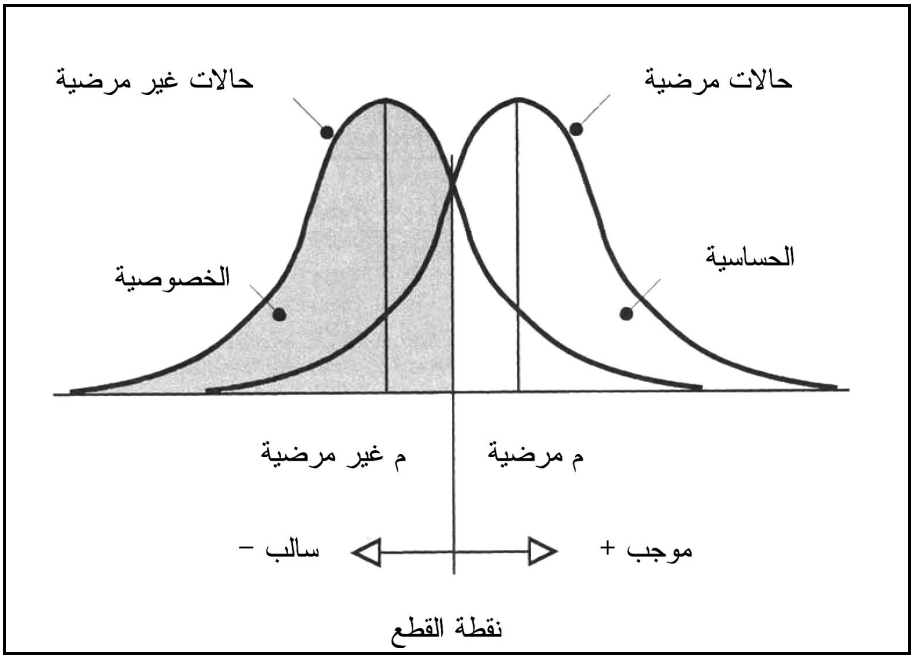
إذا زاد عدد المتغيرات القطعية عن متغيرين، أو كان هناك أكثر من مجموعتين، فإن الجدول الاحتمالي يزيد عن ٢ × ٢. فإنه يمكن مقارنة مقاييس المجازفة النسبية لمثل هذا الجدول (وهي نسبة الأرجحية «ن أ» ونسبة المجازفة «ن خ» والفروق في المجازفة «ف خ») إذا أمكن اختزاله إلى جدول ٢ × ٢ وذلك بحذف بعض الصفوف أو الأعمدة أو دمجها. وأفضل مقياس للارتباط للجداول الاحتمالية التي تزيد عن صفين أو عمودين هو معامل ارتباط كريمر Cramer's V، وهو امتداد لمعامل ارتباط فاي في العينة φ، ومعادلته:

$$V = \sqrt{\frac{\chi^{٢}(ص - ١)(ع - ١)}{\text{أقل } (ص - ١)(ع - ١) \times ن}} \qquad \text{معادلة ٣-١٨}$$

حيث البسط يتكون من إحصاءة $\chi^{٢}$ لدرجات حرية تساوي حاصل ضرب الصفوف (ص) ناقصا ١ في الأعمدة (ع) ناقصا ١. أما المقام فهو حاصل ضرب حجم العينة في أصغر بعد للجدول – ١. مثال ذلك إذا كان حجم الجدول ٣ × ٤ فإن (٣-١ ، ٤-١) = ٢. وبالنسبة للجداول ٢ × ٢ فإن معادلة كريمر تختزل إلى φ. وبالنسبة للجداول الأكبر لا تعتبر V معامل ارتباط من الناحية الفنية رغم أن مداها يتراوح بين ٠-١,٠٠. ولذلك لا يمكن تفسير مربع V كنسبة من التباين المفسر. ونظراً لأن Cramer's V تعميم لمعامل φ فهو مقيد بنفس قيود هذا المعامل الأخير (Klime, 2004).

قيمة الحساسية والخصوصية والتنبؤ

لنفرض أن لدينا معياراً تشخيصياً فردياً دقيقاً إلا أنه مرتفع التكاليف، وأن لدينا في نفس الوقت اختباراً جمعياً متصلاً لفرز الحالات أقل تكلفة، ولكنه يحتاج إلى تحديد نقطة قطع تصنف عندها الحالات بين حالات مَرَضية، وحالات سوية. ولكن عادة ما تتداخل الحالات عند نقاط القطع. ويوضح شكل (٣-١) الموقف الذي يكون فيه متوسط المجموعة غير السوية أعلى من متوسط المجموعة السوية. ويوضح الشكل كذلك نقطة القطع الذي تصنف الحالات على أساسها بين حالات سوية وحالات غير سوية. ويلاحظ أن بعض الحالات السوية تحصل على درجات تصنفها مع الحالات غير السوية، وأن بعض الحالات غير السوية حصلت على درجات تضعها ضمن الحالات السوية. وهذان الوضعان المذكوران يمثلان خطأ في التشخيص.



شكل (٣-١) - توزيع مجموعتين بالنسبة لدرجة القطع في اختبار تشخيصي متصل

ويبين الجدول الرباعي بأعلى الجدول (٣-٣) العلاقة بين نتائج الاختبار التشخيصي (+ أو -) والحالة الفعلية كما تم تشخيصها بالمعيار التشخيصي الفردي في حالات مرضية وحالات غير مرضية. وتمثل الحروف في الجدول تكرارات خلايا الجدول التي لوحظت. مثال ذلك أن الحرف «أ» يمثل الحالات المرضية التي تم تشخيصها كحالات موجبة بالاختبار التشخيصي، ويمثل الحرف «د» الحالات غير المرضية التي تم تشخيصها كحالات سالبة. وتمثل هاتان الخليتان التنبؤ الصحيح بالنسبة للاختبار التشخيصي. أما الخليتان الأخريان في الجدول (ب و جـ) فتمثلان الحالات الموجبة الزائفة (التي شخصت خطأ)، والحالات السالبة الزائفة في الاختبار التشخيصي.

ويوجد بالجزء الأسفل من جدول (٣-٣) قيم الحساسية والخصوصية والتنبؤ والمعدل الأساسي. وقد حسبت كل هذه القيم باستخدام الخلايا التكرارية الموجودة في الجزء الأعلى من الجدول. و**الحساسية** هي نسبة نتائج التشخيص المرضية (أ\(أ+جـ)). فإذا كانت قيمة الحساسية ٨٠, فمعنى ذلك أن ٨٠٪ من نتائج الاختبار في هذه المجموعة حالات موجبة صحيحة، أما الباقي وقدره ٢٠٪ فهي حالات سالبة زائفة. و**الخصوصية** هي نسبة نتائج الحالات غير المرضية الصحيحة وهذه تساوي (د\(ب+د)). وإذا كانت الخصوصية ٧٠, فمعنى هذا أن ٧٠٪ من النتائج في هذه المجموعة نتائج صحيحة، والباقي وقدره ٣٠٪ نتائج موجبة زائفة. والاختبار المثالي يجب أن يعطي ١٠٠٪ حساسية، و١٠٠٪ خصوصية. وبالنسبة للجدول (٣-٣) لم يحقق الاختبار هذه المثالية نتيجة للتداخل بين المجموعتين وتحدد نقطة القطع في الاختبار التشخيصي الحساسية والخصوصية. ولذلك لا بد من الأخذ في الاعتبار التكلفة النسبية لكلا النوعين من الخطأ في التشخيصُ السالبة الزائفة أو الموجبة الزائفة، عند تحديد نقطة القطع.

جدول (٣-٣)

تعريف قيم الحساسية والخصوصية والتنبؤ والمعدل الأساسي

نتائج الاختبار التشخيصي	التنبؤ	الحالة	
		حالات مرضية	حالات غير مرضية
+	مرضية	أ	ب
-	غير مرضية	جـ	د
العملية الإحصائية			التعريف
الحساسية			أ\(أ+جـ)
الخصوصية			د\(ب+د)
القيمة التنبؤية (موجبة +)			أ\(أ+ب)
القيمة التنبؤية (سالبة -)			د\(جـ+د)
المعدل الأساسي		(أ+جـ) \ (أ + ب + جـ + د)	

وتؤثر الحساسية والخصوصية على القيمة التنبؤية، وهي نسبة نتائج الاختبار التشخيصي الصحيحة وبهذا المعنى فهي تعكس الثقة التي يمكن أن يضعها المشخصون في نتائج الاختبار التشخيصي. وهناك نوعان من القيم التبؤية: الموجبة والسالبة. والقيمة التنبؤية الموجبة ويرمز لها بالرمز (ق ت+) هي نسبة جميع النتائج الموجبة الصحيحة، أي التي نحصل عليها من الحالات المرضية، أي أ\(أ+ب) في جدول (٣-٣). أما القيمة التنبؤية السالبة (ق ت-) فهي نسبة نتائج الاختبار السالبة الصحيحة، أي التي تنتمي للأفراد غير المرضى، أي د\(جـ+د) في الجدول. وبشكل عام تزداد قيمة (ق ت+) و (ق ت-) مع زيادة الحساسية والخصوصية.

وتتأثر القيمة التنبؤية كذلك بعامل آخر مهم جداً وهو المعدل الأساسي (م أ) وهو نسبة جميع الأفراد المرضى، أي أن (م أ = (أ+جـ)\ن)) في جدول (٣-

٣). ويتضح أثر المعدل الأساسي على القيمة التنبؤية في جدول (٣-٤). والجدولان الرباعيان المختلفان واللذان يكونان جدول (٣-٤) يشتملان على مجتمع فرضي يتكون من ١٠٠٠ حالة واختبار تشخيصي ونسبة الحساسية الناتجة تبلغ ٨٠, والخصوصية ٧٠,. وبالنسبة للجدول الأول تبلغ النسبة الأساسية ١٠, لأن عدد الحالات المرضية يبلغ ١٠٠\١٠٠٠ منهم ٨٠, من الحالات مشخصة تشخيصاً صحيحاً (حالات موجبة)، وما مجموعه ٩٠٪ من الحالات ليسوا من المرضى (٩٠٠)، منهم ٧٠٪ (٦٣٠) حصلوا على نتائج صحيحة (سالبة). إلا أنه من بين جميع النتائج الموجبة نجد أن ٨٠\٣٥٠ فقط (أي ٢٣٪) حالات صحيحة. وفي مقابل ذلك نجد أن معظم الحالات السالبة حالات صحيحة (٦٣٠\٦٥٠ = ٩٧٪) ولذلك فإن ق ت- = ٩٧,. وبالنسبة لهذه القيم التنبؤية يمكننا القول إن الاختبار التشخيصي اختبار دقيق جداً في استبعاد الحالات المرضية، ولكن ليس كذلك في اكتشافها.

وإذا لم تكن النسبة القاعدية ١٠٪ تتغير القيمتان التنبؤيتان. وهذا ما يظهر من الجزء الثاني في جدول (٣-٤) لنفس الاختبار التشخيصي. ولكن بالنسبة لمعدل قاعدي قدره ٧٥٪. ومن بين الحالات جميعها وقدرها ٦٧٥ حالة نتائجها موجبة، يوجد ٦٠٠ حالة مريضة، ولذلك فإن ق م+ = ٦٠٠\٦٧٥ = ٨٩,. وبالمثل من بين الحالات السالبة وقدرها ٣٢٥ حالة يبلغ عدد غير المرضى ١٧٥، ولذلك فإن ق س- ١٧٥\٣٢٥ = ٥٤,، ولذلك يمكن الوثوق أكثر في النتائج الموجبة للاختبار التشخيصي أكثر من النتائج السالبة، وهذا هو بالضبط عكس الموقف بالنسبة للمعدل القاعدي الذي يبلغ ١٠٪. وبشكل عام فإن ق م+ تنقص وتزداد ق م- مع اقتراب المعدل القاعدي من الصفر. ومعنى هذا أن الاختبارات التشخيصية تنزع إلى أن تكون أكثر فائدة في استبعاد وجود الاضطرابات النادرة نسبياً منها في التنبؤ الصحيح بوجود هذه الاضطرابات. والعكس صحيح للاضطرابات الأكثر شيوعاً، إذ عندما يتزايد المعدل القاعدي فإن ق م+ تتزايد وتتناقص ق م-. وفي هذه الحالة يكون الاختبار أكثر فائدة في الكشف عن المرض من استبعاده.

جدول (٣-٤)

القيم التنبؤية الموجبة والسالبة بالنسبة لمعدلين قاعديين مختلفين لاختبار تشخيصي ذي حساسية قدرها ٨٠٪ وخصوصية قدرها ٧٠٪

نتائج الاختبار التشخيصي	الحالة الفعلية		المجموع	القيم التنبؤية	
	مرضية	غير مرضية		ق م +	ث م –
المعدل القاعدي = ١٠,					
+	٨٠	٢٧٠	٣٥٠	,٢٣	,٩٧
–	٢٠	٦٣٠	٦٥٠		
المجموع	١٠٠	٩٠٠	١٠٠٠		
المعدل القاعدي = ٧٥,					
+	٦٠٠	٧٥	٦٧٥	,٨٩	,٥٤
–	١٥٠	١٧٥	٣٢٥		
المجموع	٧٥٠	٢٥٠	١٠٠٠		

وأثر المعدل القاعدي على القيم التنبؤية له أهميته، ولكن كثيراً ما يتم تجاهله حتى بين المشخصين ذوي الخبرة. وأكثر سوء الفهم يأتي من الخلط بين الحساسية والخصوصية، والتي علاقتها ثابتة مع المعدل القاعدي ق م+ و ق م- المختلفة. وهذا يعني أن المشخصين كثيراً ما يفشلوا في تعديل تقديراتهم لدقة التشخيص التي يحصلون عليها من الاختبارات التشخيصية وذلك بالنسبة للتغيرات في المعدل القاعدي.

مثال من البحوث:

تقدم فيما يلي مثالاً لحساب حجم الأثر للمتغيرات القطعية. وقد ورد هذا المثال لدى كلاين (Kline, 2004).

التدخين ومخاطر الإصابة بأمراض القلب

حصل كلاين على بيانات هذا المثال من (Glass & Hopkins, 1996) اللذين زاوجا بين ١٢٠ موظفاً مصابين بمرض القلب (الشريان التاجي) و١٢٠ موظفاً آخر في نفس الشركة ليسوا مصابين بهذ المرض. وقد صنف جميع الموظفين في أربع فئات بالنسبة للتدخين: غير المدخنين، والمدخنين الذين يدخنون أقل من علبة واحدة في اليوم، والمدخنين الذين يدخنون علبة واحدة في اليوم، والمدخنين الذي يدخنون أكثر من علبة في اليوم. وهذه فئات مرتبة، إلا أنها حللت بالطرق المستخدمة مع الفئات غير المرتبة. وتعتبر هذه الدراسة كذلك دراسة استعادية retrospective لأنه تم التعرف على الحالات بناء على وجود أو عدم وجود المرض قبل تقويم مستوى تعرضهم لعامل خطورة مفترض، وهو التدخين. ونظراً لأن الدراسة استعادية فإن أفضل طريقة لحساب حجم الأثر هي نسبة الأرجحية. وقد استخدمت بعض الوسائل الإحصائية اللامعلمية الأخرى لأغراض تعليمية.

ويبين جدول (٣-٥) فئات التدخين وحالة المرض التي تبلغ ٤ × ٢.

جدول (٣-٥)

مستوى التدخين وأمراض القلب التاجية

فئة التدخين	العدد	مرض القلب التاجي	
		مصاب	غير مصاب
أقل من علبة في اليوم	٤٢	١٩	٢٣
علبة في اليوم	٦٤	٣٩	٢٥
أكثر من علبة في اليوم	٣١	٢٠	١١
غير مدخن	١٠٣	٤٢	٦١

وتبلغ قيمة مربع كاي لهذا الجدول χ^2 (٣) = ٩,٥٦، كما تبلغ قيمة معامل كريمر Cramer's V (٩,٥٦\٢٤٠)$^{\frac{1}{2}}$ = ٠,٢٠، وتصف هذه القيمة حجم الارتباط الكلي بين هذين المتغيرين. ولتقدير المجازفة النسبية لاحتمال الإصابة بمرض الشريان

التاجي دمجت فئات التدخين الثلاث بجدول (٣-٣) في فئة واحدة، ونتج عن ذلك جدول رباعي هو الموضح في الجزء الأيمن من جدول (٣-٤). أما الجزء الأيسر من الجدول فيبين قيم حجم الأثر للإحصاء اللامعلمي وحدود الثقة ٩٥٪ بالتقريب.

ويلاحظ من جدول (٣-٤) أن ما مجموعه ٧٨ مدخناً من ١٣٧ مصابون بمرض القلب (الشريان التاجي)، ولذلك فإن احتمال المجازفة يبلغ ل$_{م}$=١٣٧\٧٨=٠,٥٧ . ومن بين غير المدخنين وعددهم ١٠٣ يبلغ مجموع المصابين بمرض القلب ٤٢. ولذلك فإن احتمال المجازفة ل$_{غ}$ = ٤٢\١٠٣ = ٠,٤١ . والفرق بين الاحتمالين (الفرق في المجازفة) يبلغ «ف م» = ٠,٥٧ - ٠,٤١ = ٠,١٦ . ولذلك يمكن القول إن المدخنين يزيد احتمال إصابتهم بمرض القلب بما يزيد على ١٦٪ من غير المدخنين. وحدود الثقة التقريبية ٩٥٪ للفرق في المجازفة في المجتمع يبلغ ٠,٠٣-٠,٢٩ ولذلك فإن ف م = ١٦-٠, وهو ما يعني أن خطورة الإصابة بمرض القلب في المجتمع تتراوح بين ٠٣, و ٢٩, وذلك بثقة قدرها ٩٥٪. ونسبة المجازفة «ن م» = ٠,٥٧\٠,٤١ = ١,٣٩، أي إن معدل الإصابة بالمرض بين المدخنين يبلغ ١,٤ مرة أكثر من غير المدخنين. وحدود الثقة التقريبية ٩٥٪ لنسبة المجازفة في المجتمع ١,٠٦ - ١,٨٤، ولذلك فإن «ن م» = ١,٣٩، أي إن نسبة المجازفة في أدنى حالاتها تبلغ ١,٠٦، وفي أقصى حالاتها تبلغ ١,٨٤. وتبلغ نسبة الأرجحية:

ن أ = ٠,٥٧ (١ - ٠,٥٧) \ ٠,٤١ (١ - ٠,٤١) = ١,٩٢

أي إن أرجحية الإصابة بمرض الشريان التاجي تبلغ بين المدخنين ضعفها بين غير المدخنين. وتتراوح حدود الثقة التقريبية ٩٥٪ لنسبة الأرجحية في المجتمع بين ١,١٤ - ٣,٢٢. أي إن ن أ = ١,٩٢ وهذا ما يؤيد نسبة لرجحية تصل إلى ١,١٤ في حدها الأدنى و٣,٢٢ في حدها الأعلى وذلك بثقة قدرها ٩٥٪. وتشير حدود الثقة الثلاثة التي ذكرناها الآن إلى مدى متسع من عدم الدقة، ويرجع هذا إلى صغر حجم المجموعات نسبياً. ويبلغ معامل الارتباط بين التدخين والإصابة بمرض القلب φ = ١٦,، وهو ما يدل على نسبة قدرها ٢,٥٪ من التباين في التدخين تفسر التباين في الإصابة بمرض القلب. وهذه نسبة نموذجية لحجم الأثر الذي يفسر التباين في هذا المجال ولا يمكن إغفالها باعتبارها قيمة غير مهمة.

جدول (٣-٦)

تصنيف رباعي للمدخنين وغير المدخنين بالنسبة للمصابين بمرض القلب وغير المصابين بالمرض وذلك بالنسبة لبيانات جدول (٣-٣)

		الإصابة بمرض القلب		حجم الأثر					
المجموعة	ن	نعم	لا	لم	لغ	ف م	ن م	ن أ	φ
مدخن	١٣٧	٧٨	٥٩	,٥٧	,٤١	,١٦	١,٣٩	١,٩٢	,١٦
غير مدخن	١٠٣	٤٢	٦١						

تلخيص:

ناقشنا في هذا الفصل مؤشرات حم الأثر لمقارنة المجموعات في المتغيرات القطعية. وبعض هذه المؤشرات يقيس المجازفة النسبية عبر مجموعتين مختلفتين بالنسبة لنتيجة غير مرغوبة في مقابل نتيجة أخرى مرغوبة مثل ناجح/ راسب. وتوضع هذه البيانات في جدول رباعي. ومؤشر المجازفة الذي يتميز بالخاصية السيكومترية الأفضل هو نسبة الأرجحية، أي نسبة الأرجحية داخل المجموعات لنتيجة معينة. ويسبب استخدام معامل ارتباط فاي φ للعينات مشكلة ترجع إلى الخواص الهامشية لعضوية المجموعات في المتغيرات القطعية.

كما ناقشنا المفاهيم المتعلقة بقيم الحساسية والخصوصية والتنبؤ. وهذه الطريقة تأخذ في اعتبارها بشكل صريح أثر المعدل القاعدي في المجتمع على دقة القرارات المتخذة بناء على نتائج الاختبار التشخيصي بوجود أو عدم وجود اضطراب أو حالة مرضية. وبشكل عام فإن الاختبارات التشخيصية ليست مفيدة في التعرف على وجود الحالات النادرة، ولكنها دقيقة في تحديد عدم وجود هذه الحالات. والعكس صحيح عندما يكون المعدل القاعدي أقرب إلى ١,٠ منه إلى الصفر.

Effect size of experimental treatments and the significance of statistical significance

Abstract

This booklet discusses a very important and recent topic in statistical analysis: Effect size of experimental treatments. Effect size has become an integral part of statistical analysis to the extent that a research topic that ignores using effect size as a basis for the interpretation of statistical significance is considered incomplete.

This booklet consists of three chapters. The first chapter deals with the controversial issue in Psychology about statistical tests and the events that led to them. Special mention was given to APA's fifth edition of the Publication Manual which discussed the statistical tests of the null hypothesis. This chapter pinpoints the importance of the problem of statistical significance and concludes that the problem is not in the statistical significance test itself. If we chooser a plausible problem, write reasonable hypotheses, select a random sample, observed the assumptions of sampling distribution, defined statistical power, and understand the real meaning of p-values, there would be no problem in using statistical significance tests as the main tool of statistical inference.

The best solution until now for this problem is to add effect size to the results of experimental treatment analysis. Chapter one presented several definitions of effect size. The basis of those definitions was the methods of computing effect sizes. The first known equation for effect size was the equation of Glass which is based on the difference of the means of experimental and control groups divided by the standard deviation of the control group because this latter group is not affected by the experimental treatment of the independent variable.

Chapter two discusses effect size in parametric statistics: Correlation, t-test, ANOVA, and Factorial analysis of variance. The discussion of those statistical tests shows that there are different types of effect size depending on the methods of calculation. These differences in the methods of calculation result sometimes in getting conflicting effect sizes. Accordingly researches should mention the method of calculation of effect size so that readers would be aware not mix between conflicting methods of estimating effect size.

The third chapter deals with effect size in non-parametric statistics and categorical variables. Some of the indices of effect size are based on ratio probabilities known as risk ratios. The risk index with the best psychometric properties is the odds ratio.

المراجع

١ – رجاء محمود أبو علام (١٩٨٩). **مدخل إلى مناهج البحث التربوي.** الكويت: مكتبة الفلاح.

٢ – رجاء محمود أبو علام (٢٠٠٣). **التحليل الإحصائي للبيانات باستخدام برنامج SPSS.** القاهرة: دار النشر للجامعات.

٣ – رجاء محمود ابو علام (٢٠٠٥). **تقويم التعلم.** عمان، الأردن: دار المسيرة للنشر والتوزيع.

٤ – رجاء محمود ابو علام (٢٠٠٦). **مناهج البحث في العلوم النفسية والتربوية.** (الطبعة الخامسة). القاهرة: دار النشر للجامعات.

٥ – رشدي فام منصور (يونيو، ١٩٩٧). حجم التأثير: الوجه المكمل للدلالة الإحصائية. **المجلة المصرية للدراسات النفسية**، ١٦ (٧)، ٥٧–٧٥.

٦ – صلاح أحمد مراد (٢٠٠٠). **الأساليب الإحصائية في العلوم النفسية والتربوية والاجتماعية.** القاهرة: مكتبة الأنجلو المصرية.

7 - American Psychological Association (1994) **Publication manual of the American Psychological Association,** (4th ed.). Washington, DC: Author.

8 - American Psychological Association (2001). **Publication manual of the American Psychological Association,** (5th ed.). Washington, DC: Author.

9 - Coe, R. (2000). **What is an 'Effect Size'?: A guide for users.**

10 - Cohen, (1994). The earth is round (p > .05). **American Psychologist. 49,** 997-1003.

11 - Cohen, J. Cohen, P., West, S. G., & Allen, L. S. (2003). **Applied multiple regression/correlation analysis for the behavioral sciences** (2nd ed.) Mahwah, NJ: Laurence Erlbaum Associates.

12 - Cortina, J. M., & Nouri, H. (2000). **Effect size for ANOVA designs.** Thousand Oaks, CA: Sage Publications.

13 - Cumming, G. & Finch, S. (2001). A primer on the understanding, use,

and calculation of confidence intervals that are based on central and noncentral distributions. **Educational and Psychological Measurement, 61,** 532-574.

14 - Furlong, N., Lovelace, E., & Lovelace, K. (2000). **Research methods and statistics: An integrated approach.** Fort Worth, TX: Harcourt College Publishers., (2000)

15 - Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. **Educational Researcher, 5,** 3-8.

16 - Glass, G. V., McGraw, B., & Smith, M.L. (1981). **Meta-analysis in social research.** Beverly Hills: Sage Publications.

17 - Grissom, R. J.(1996). The magical number: 7± 2. Meta-meta-analysis of the probability of the superior outcome in comparisons involving therapy, placebo, and control. **Journal of Consulting and Clinical Psychology, 64,** 973-982.

18 - Grissom, R. J. & Kim, J. J. (2005). **Effect sizes for research: A broad practical approach.** Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

19 - Hedges, L. & Olkin, I. (1985). **Statistical methods for meta-analysis.** New York: Academic Press.

20 - Hubbard, R. & Ryan, P. A. (2000). The Historical Growth of Statistical Significance Testing in Psychology - And Its Future Prospects". **Educational and Psychological Measurement,** 60. 661-681

21 - Hunter, J. E. (1997). Needed: A ban on the significance test. **Psychological Science, 8,** 1-7.

22 - Hunter, J.E. & Schmidt, F. L. (1990). **Methods of meta-analysis: Correcting error and bias in research findings.** Newbury Park, CA: Sage Publications.

23 - Hunter & Schmidt, (2004). **Methods of meta-analysis**, (2^{nd} ed.). Thousand Oaks, CA: Sage Publications.

24 - Kaufman, A. S., (1998). Introduction to the special issue on statistical significance. **Research in the Schools, 5,** 1.

25 - Kiess, H. O. (1989). **Statistical concepts for the behavioral sciences.** Boston: Allyn and Bacon.

26 - Kirk, R. E. (1996). Practical significance: A concept that whose time has come. **Educational and Psychological Measurement, 61,** 213-218.

27 - Kline, R. B. (2004). **Beyond significance testing: Reforming data analysis methods in behavioral research.** Washington, D. C.: American Psychological Association.

28 - Levine, T.R. & Hullett, C.R. (2002). Eta squared, partial eta squared and misreporting of effect size in communication research. **Human Communication Research, 28,** 612-625.

29 - Lipsey, M.W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatments: Confirmation from meta-analysis. **American Psychologist, 48,** 1181-1209.

30 - Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. **Current Directions in Psychological Science, 5,** 161-171.

31 - Lunneborg, C. E. (2001). Random assignments of available cases: Bootsrap standard errors and confidence intervals. **Psychological Methods, 6,** 402-412.

32 - Mittag, K. C., & Thompson, B. (2000). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. **Educational Researcher, 29,** 14-20.

33 - Mulaik. S. A., Raju, N. S., & Hasherman, R. A. (1997). There is a time and place significance testing. In L. L. Harlow, & J. H. Steiger (EDS.), **What if there were no significance tests?** (pp. 65-115). Mahwah, NJ: Erlbaum.

34 - Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretations of significance levels and effect sizes by psychological researchers. **American Psychologist, 41,** 1299-1301.

35 - Oakes, M. (1986). **Statistical inference.** New York: Wiley.

36 - Olejnik, S. & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. **Contemporary Educational Psychology, 25,** 241-286.

37 - Platt, J. R. (1964). Strong inference. **Science, 146 (October 16),** 347-353.

38 - Pollard, P. (1993). How significant is "significance"? In G. Keren & C. Lewis (Eds.), **A handbook for data analysis in the behavioral sciences: Vol 1. Methodological issues** (pp. 449-460). Hillsdale, NJ: Lawrence Erlbaum.

39 - Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. **Journal of Educational Psychology, 74,** 166-169.

40 - Rosenthal, R., Rosnow, R., & Rubin, D. B. (2000). **Contrasts and effect size in behavioral research: A correlational approach.** Cambride: Cambridge University Press.

41 - Rubin, D. B. (1992). Meta-analysis: Literature synthesis or effect-size surface estimation. **Journal of Educational Statistics, 17, 4,** 363-374.

42 - Schimdt, N. (1996). Statistical significance testing and cumulative knowledge in psychology: Implication for training of researchers. **Psychological Methods, 1,** 115-129.

43 - Shrout, P. E. (1997) Should significance tests be banned? Introduction to a special section exploring the pros and cons. **Psychological Science, 8,** 1-2.

44 - Thompson, B. (Ed.) (2003). **Score reliability: Contemporary thinking on reliability issues.** Thousand Oaks, CA: Sage.

45 - Thompson, W. L. (2001). 402 citations questioning the indiscriminate use of null hypothesis significance tests in observational studies. In R. B. Kline (ed), **Beyond significance testing: Reforming data analysis methods in behavioral research.** Washington, D. C.: American Psychological Association, 2004.

46 - Tversy, A., & Kahneman, D. (1971). Belief in the law of small numbers. **Psychological Bulletin, 76,** 105-110.

47 - Vincent, D. & Crumpler, M. (1997). An analysis of a standard spelling test in Britian. **American Educational Research Journal, 17,** 419-433.

48 - Welkowitz, J., Ewen, R. B., & Cohen, J. (1982). **Introductory statistics for the behavioral sciences.** (3rd ed.), New York: Academic Press.

49 - Wilcox, R. R. (2003). **Applying contemporary statistical techniques.** San Diego, CA: Academic Press.

50 - Wilkinson, L., & APA Task Force on statistical inference (1999). Statistical methods in psychology journals: Guidelines and explanations. **American Psychologist, 54,** 594-604.