

واقع الدلالة الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة
في رسائل ماجستير الموهبة والإبداع - جامعة البلقاء التطبيقية /الأردن

سناء نمر محمد الظاهر

عمر محمد احمد عبيدات

عروبة احمد صالح الزعبي

(جامعة البلقاء التطبيقية /الأردن ,كلية العلوم - قسم الرياضيات)

الملخص

استهدفت هذه الدراسة معرفة واقع الدلالة الإحصائية، وارتباطها بالدلالة العملية من جهة، وقوة الاختبار الإحصائي من جهة أخرى للفرضيات الإحصائية المستخدمة في رسائل ماجستير الموهبة والإبداع. تكونت عينة الدراسة من (119) فرضية صفرية تم فحصها باستخدام الإحصائي (ت)، و(100) فرضية تم فحصها باستخدام الإحصائي (ف)، والتي تم استخدامها في (24) رسالة من رسائل ماجستير الموهبة والإبداع في جامعة البلقاء التطبيقية.

تم حساب قيم (d, η^2) باعتبارها مؤشرات للدلالة العملية، كما تم حساب قوة الاختبارات الإحصائية باستخدام جداول كوهن، أشارت نتائج الدراسة أن حوالي (94%) من الفرضيات الصفرية فحصت عند مستوى دلالة ($\alpha = 0.05$)، وأن (91%) من الباحثين حددوا مستوى الدلالة الإحصائية بعد تحليل البيانات. كما أظهرت النتائج عدم تحيز الباحثين للنتائج الدالة إحصائياً، حيث كانت نسبة الفرضيات غير الدالة إحصائياً (75%) من مجموع الفرضيات. كما أظهرت النتائج أن (85%) من الفرضيات المفحوصة على الاختبارين كانت لها دلالات عملية ضعيفة، وأشارت النتائج أن (16%) من الفرضيات الدالة إحصائياً كانت دلالاتها العملية ضعيفة. كما أظهرت النتائج أن (13%) من الفرضيات غير الدالة إحصائياً كانت تتمتع بدلالات عملية تراوحت بين المتوسطة والعالية. وأظهرت أيضاً أن (58%) من الفرضيات كانت لها قوة اختبار ضعيفة، وأن (48%) من الفرضيات غير الدالة إحصائياً كانت لها قوة اختبار ضعيفة.

الكلمات المفتاحية: الدلالة العلمية، الدلالة الإحصائية، قوة الاختبار.

المقدمة

يعتمد البحث التربوي اختبارات الدلالة الإحصائية كأسلوب إحصائي علمي، لتقرير ما إذا كانت الفروق أو الارتباطات أو العلاقات بين المتغيرات دالة إحصائياً. (بابطين، 2001).
وتتم هذه الاختبارات بعدة خطوات، تبدأ بتحديد نوعين من الفرضيات التي يتم في ضوءها تفسير النتائج الخاصة بالبحث وتوضيحها، ويسمى النوع الأول بالفرضية الصفرية (null hypothesis) ويرمز لها (H_0)، والنوع الثاني من الفرضيات بالفرض البديل (alternative hypothesis) ويرمز لها (H_a).

وتشير الفرضية الصفرية إلى عدم وجود فروق معنوية في النتائج، أو أن المتغير المستقل لا يؤثر في المتغير التابع، أو أنه لا يوجد فرق بين خصائص العينة وخصائص المجتمع الذي سحبت منه. وينبغي أن يكون الفرض الصفرى مبنياً على مفاهيم معرفة تعريفاً إجرائياً، بحيث يمكن جمع بيانات تتعلق بها عن طريق الملاحظة والتجربة. والفرض الصفرى هو الذي يتم اختباره إحصائياً وهو الذي يخضع للفرض والقبول.

في حين تشير الفرضية البديلة إلى وجود فروقاً معنوية في النتائج، تعزى لتأثير المتغير المستقل على المتغير التابع، أو أن خصائص العينة التي يدرسها الباحث لا تتفق وخصائص المجتمع الذي سحبت منه. وقد تأخذ الفرضية البديلة أحد ثلاثة صور:

أ- الصورة الأولى: وتكون على النحو التالي $H_a : \mu \neq \mu_0$

ويدل ذلك على أن قيمة معلمة المجتمع لا تساوي قيمة محددة، وهذا يعني أن الفرضية البديلة عديمة الاتجاه (non-directional) أي لا يحدد الفرض اتجاه التأثير ولا يهتم بنوعيته.

ب- الصورة الثانية: وتكون على النحو $H_a : \mu > \mu_0$

وتشير إلى أن معلمة المجتمع (μ) أكبر من فرض معين (أي يتوقع أثر إيجابي) ويطلق على هذه الفرضية البديلة بالفرض ذي الاتجاه الواحد (directional).

ج- الصورة الثالثة: وتكون على النحو $H_a : \mu < \mu_0$

وتشير إلى أن معلمة المجتمع (μ) أصغر من فرض معين (أي يتوقع أثر سلبي) ويطلق على هذا الفرض البديل أيضاً ذي الاتجاه الواحد (خاطر وآخرون، 1998).

وبعد تحديد الفرضيات البديلة والصفية يختار الباحث الاختبار الإحصائي الملائم لفحصها، ويشكل توزيع المعاينة (Sampling Distribution) للإحصائي الإطار المرجعي لفحص الفرضية الصفية. وفي ضوء توزيع المعاينة ومستوى الدلالة الإحصائية المستخدمة، تحدد القيم الحرجة التي تمثل الحد الفاصل تحت التوزيع بين المساحة الممثلة لمنطقة القبول، والمساحة الممثلة لمنطقة الرفض.

ويختار الباحث بعد ذلك عينة بحجم مناسب من المجتمع الإحصائي، ويجري المعالجات على أفراد هذه العينة. وتقارن القيمة الناتجة مع القيم الحرجة على توزيع المعاينة، وفي ضوء هذه المقارنة يتخذ القرار المتعلق بقبول أو رفض الفرضية الصفية.

إن اتخاذ القرار باستخدام هذه الآلية لا يعني بأنه صحيح، بل أن هناك احتمالاً لنوعين من الخطأ يمكن أن يقع الباحث فيهما أثناء عملية اتخاذ القرار هما: الخطأ من النوع الأول (Type I Error)، والخطأ من النوع الثاني (Type II Error)، ويبين الجدول رقم (1) هذين الخطأين.

جدول رقم (1)

الأخطاء الممكنة الوقوع فيها نتيجة لعملية اتخاذ القرار

H_1 صحيحة	H_0 صحيحة	القرار المتخذ الفرضية الصفية في واقع الأمر
القرار صحيح	الخطأ من النوع الأول	رفض H_0
الخطأ من النوع الثاني	القرار صحيح	قبول H_0

يتضح من الجدول (1)

1. إذا رفض الباحث الفرضية الصفية وهي صحيحة، فإنه يرتكب الخطأ من النوع الأول.
2. إذا قبل الباحث الفرضية الصفية وهي غير صحيحة، فإنه يرتكب الخطأ من النوع الثاني.
3. إذا قبل الباحث الفرضية الصفية وهي صحيحة، فإن القرار المتخذ يكون صحيحاً.
4. إذا رفض الباحث الفرضية الصفية وهي خاطئة، فإن القرار المتخذ يكون صحيحاً، وهذا ما يعرف بقوة الاختبار (حجمات وعليان، 1997).

ويحدد الباحث مقدار الخطأ من النوع الأول الذي يمكن ارتكابه في فحص الفرضية الصفية، ويكون احتمال الوقوع في مثل هذا الخطأ مساوياً لقيمة مستوى الدلالة الإحصائية (α)، وغالباً ما يستخدم الباحثون مستويات للدلالة الإحصائية مساوية (0.05) أو (0.01)، ويفترض في

الباحث أن يحدد مستوى الدلالة التي يستخدمها في فحص فرضياته قبل القيام بجمع البيانات وإجراء التحليل الإحصائي (Winer, 1988).
إن شيوع أي من مستويات الدلالة الإحصائية له أهمية، حيث يمكن مقارنة النتائج التي توصل إليها باحثون آخرون طالما أن أي منهم قد استخدم أحد هذه المقادير (خاطر وآخرون، 1998).

وعند اتخاذ الباحث قراراً برفض الفرضية الصفرية وقبول الفرضية البديلة عند مستوى دلالة معين فإن النتائج عندئذ تكون ذات دلالة إحصائية، وهذا يعني أن هناك أثراً للمتغير المستقل على المتغير التابع، ولكنه لا يتضمن إشارة إلى حجم هذا الأثر، أو أن الفروق على المتغير التابع والتي تعود إلى المتغير المستقل ليست صفراً ولم تأتي بالصدفة (Cohen, 1977).
تاريخياً ساعدت اختبارات الدلالة الإحصائية التقليدية الباحثين في فهم معنى بياناتهم، حيث اعتبرت حجر الزاوية في الأبحاث الاجتماعية والإنسانية والبيولوجية على مدى 70 عاماً (Kirk, 2007)، لكن ثبت أن الاعتماد عليها بمفردها ينطوي على أخطاء، فخلال عقود تبين أن هذه الاختبارات غالباً ما يساء استعمالها، وأن تلك الاختبارات محدودة في المساعدة.
فعلى سبيل المثال، أغلب التساؤلات البحثية الأساسية تتضمن أسئلة حول حجم تأثير المتغيرات في المجتمع الإحصائي. لكن اختبارات الدلالة الإحصائية تخفق في الإجابة عن تلك التساؤلات.

لقد تطورت النظرة إلى الدلالة الإحصائية، وأصبح ينظر إليها على أنها تنتج معلومات قليلة صالحة ومناسبة للتساؤلات البحثية، وأن النتيجة الثنائية للدلالة الإحصائية (قبول، رفض) لم تعد ترضي كثيراً من العلماء، الذين يستخدمون الفرضية الصفرية لمعرفة مقدار التأثير أكثر من مجرد التعرف على وجوده من عدم وجوده. (بابطين، 2000).

وفي ضوء ذلك يمكن القول أن المعلومة التي تقدمها الدلالة الإحصائية والمتعلقة بالقرار المتخذ برفض الفرضية الصفرية لم تعد كافية. لذلك اقترح العديد من الباحثين في السنوات العشرين الأخيرة، استخدام (بالإضافة إلى الدلالة الإحصائية) مؤشرات ومقاييس أخرى لقياس الأثر الناتج عن المتغيرات المستقلة، مثل قوة الاختبار الإحصائي (Statistical Power)، وهي قدرة الاختبار على رفض الفرضية الصفرية عندما تكون في حقيقة الأمر خاطئة (عودة والخيلي، 2000)، ومؤشرات الدلالة العلمية أو ما يعرف بحجم التأثير (Effect Size)، وهي الدرجة التي توجد فيها الظاهرة في المجتمع، أو مدى تفسير المتغير التابع بواسطة المتغير المستقل (Kellow, 1998).

وأشار دنج (Deng, 2005) إلى أن الأبحاث التي تستخدم قوة الاختبار وحجم التأثير في سياق البحث فإنه يتم تفسير نتائجها بصورة أفضل.

وقد لاحظ تمكو وفويل (Tomcho & Foels, 2009) أن هذا التوجه (استخدام قوة الاختبار) نادراً ما يستخدم من قبل الباحثين، وأن الاختبارات التي ليس لها قوة كافية، تعزز عدم الدقة في اتخاذ القرار بقبول أو رفض الفرضية الصفرية من جهة، وعدم الدقة في تفسير النتائج من جهة أخرى. ولاحظ ثومبسون (Thompson, 1995) أن القوة إذا كانت عالية فإن النتائج غير الدالة إحصائياً تبقى تسهم في أساس المعرفة.

ويرى نيكس وبارنيت (Nix & Barnette, 1998) كما أورد (بابطين، 2001) بأنه إذا كان لدينا اختبار منخفض القوة، فإنه لن يمكننا تحسس والتقاط الدلالة الإحصائية، بينما إذا كان لدينا اختبار عالي القوة، فإن حجم الفروقات بغض النظر عن صغرها سوف تكون دالة إحصائياً.

ولعل أهم العوامل التي تساهم في جعل الاختبار الإحصائي أكثر قوة هي التصميم الإحصائي للدراسة، فكلما كانت نسبة التباين غير المفسر قليلة في التصميم، كان الاختبار المستخدم أكثر قوة، كذلك صدق وثبات الاختبار: فكلما كان الاختبار المستخدم في قياس المتغير التابع أكثر صدقاً وثباتاً، كلما قلت التباينات على المتغير التابع والعائدة إلى الأخطاء في القياس، وبالتالي فإن مثل هذه

الإختبارات تؤدي إلى تحسين قوة الإختبار الإحصائي (إبراهيم، 2000). وتتأثر قوة الإختبار الإحصائي بالعوامل التالية:

1. مستوى الدلالة الإحصائية α ، حيث ترتفع قوة الإختبار بزيادة مستوى الدلالة الإحصائية.
 2. حجم التأثير المتوقع ويقصد به أثر المتغير المستقل الذي يرغب الباحث بوجوده (مقدار الفرق بين المتوسطات)، فكلما كان حجم التأثير المراد الكشف عنه كبيراً، فإن الإختبار الإحصائي يكون أكثر قوة مما لو كان الإختبار معد للكشف عن حجم التأثير الصغير جداً.
 3. حجم العينة فكلما زاد حجم العينة إزدادت قوة الإختبار.
 4. كون الإختبار بإتجاه واحد أو بإتجاهين، يكون الإختبار بإتجاه واحد أقوى من الإختبار بإتجاهين إذا كانت القيمة الحقيقية للمعلمة في نفس الجهة التي تفترضها الفرضية البديلة.
 5. علاقة القيمة الحقيقية للمعلمة بقيمته في الفرضية الصفرية، حيث تزداد قوة الإختبار كلما ابتعدت القيمة الحقيقية للمعلمة عن القيمة المفروضة (عودة والخليلي، 2000).
- أما فيما يتعلق بحجم التأثير فقد أدى الإنتقاد المتزايد لإختبارات الدلالة الإحصائية كما يرى (كيرك، 1996)، إلى إعادة تركيز الأبحاث في العلوم الاجتماعية والتربوية على قيم حجم التأثير بصورة أكبر من اعتمادها على قيمة مستوى الدلالة P (Parker & Brossart, 2003) ، وقد عبر شيفر (Shaver) كما أورد (بابطين، 2000) عن ذلك بقوله: أعتقد بأن الدراسات يجب أن تنشر بدون إختبارات الدلالة الإحصائية، ولكن ليس بدون أحجام التأثير.
- وكمحصلة لهذه الإنتقادات، فقد تزايدت الإشارة إلى حجوم التأثير في نتائج الأبحاث النفسية والمنشورة (Parker & Brossart, 2003).

وقد أكدت الجمعية النفسية الأمريكية هذه الحقائق في دليلها الذي نشر عام 2001: "من أجل أن يفهم القارئ أهمية نتائجك بشكل كامل فمن الضروري أن تتضمن نتائجك مؤشرات عن حجم التأثير أو قوة الإرتباط (Robey, 2004)، وقد ذكرت (فريال، 2003) خمس فوائد لمقاييس فحص الدلالة العملية فهي أولاً: تعطي تقديراً لدرجة وجود الظاهرة في المجتمع، وذلك من خلال قيمة على مقياس متصل يبدأ من صفر حقيقي (يعني غياب الظاهرة)، ويمتد ليقدّم قيمة لحجم الأثر لهذه الظاهرة.

وثانياً: تعطي قيمة محددة تشكل محكاً لوصف حجم الأثر (صغير، متوسط، كبير) وثالثاً: يمكن إستخدام نتائجها كمقياس كمي للمقارنة بين دراستين ورابعاً: يمكن إستخدامها في تحليل القوة لتقدير حجم العينة اللازمة لدراسة ما. وأخيراً: عدم تأثرها بحجم العينة وهذا ما يميزها عن الدلالة الإحصائية.

لذلك إزدادت القناعة لدى الباحثين بضرورة إستخدام مؤشرات حجم التأثير، وحيث أن هذه المؤشرات تختلف وتتنوع بإختلاف الإحصائي المستخدم في فحص الفرضيات الصفرية، فقد ذكر (كيرك، 1996) ما يزيد عن 40 مؤشر، حيث قسم هذه المؤشرات في مجموعتين رئيسيتين تمثل الأولى: مؤشرات حجم التأثير أو الفرق بين المتوسطات بدرجات الإنحراف المعياري، والثانية: تمثل مؤشرات قوة التلازم، كما ميز عدد كبير من المؤشرات لا يندرج تحت أي من المجموعتين، ومن أهم مؤشرات حجم التأثير:

1. حجم الأثر d للإختبار t لعينة واحدة.
2. حجم الأثر d للإختبار t للفرق بين الأوساط الحسابية للعينات المستقلة.
3. حجم الأثر d للإختبار t للفرق بين الأوساط الحسابية للعينات المرتبطة.
4. مربع إيتا (η^2) المستخدم مع تصميم تحليل التباين.
5. مربع أوميغا (ω^2) المستخدم مع تصميم تحليل التباين.
6. حجم الأثر (r^2) لمعامل الإرتباط بيرسون.

7. حجم الأثر (F^2) لتصميم الإنحدار المتعدد.
8. حجم الأثر (D^2) لإختبار تحليل التباين المتعدد (MANOVA).

ومن بين هذه المؤشرات اهتمت الدراسة بالمؤشرات التالية :

1. مؤشرات حجم التأثير d لكوهن (Cohen) في حالة إستخدام الإحصائي (ت)، لفحص دلالة الفروق بين متوسط مجتمعين مستقلين. ويساوي القيمة المطلقة للفرق بين متوسط المجتمعين مقسوماً على الانحراف المعياري للمجتمع، على إفتراض تساوي المجتمعين وبالرموز $D = |\mu_1 - \mu_2| / \sigma$

وفي حالة عدم تساوي تباين المجتمعين، فإننا نستخدم الانحراف المعياري المشترك

(Common Variance)، الذي يحسب بالمعادلة $\sigma_{com} = \sqrt{\sigma_1^2 + \sigma_2^2} / 2$ حيث :

σ_1^2 : التباين للعينة الأولى.

σ_2^2 : التباين للعينة الثانية.

وحسب معايير كوهن يعد حجم الأثر من حساب d صغيراً عندما: $0.49 \geq d \geq 0$ ومتوسطاً عندما $0.79 \geq d \geq 0.50$ ، وكبيراً عندما $d \geq 0.80$

2- حجم التأثير f لكوهن (Cohen) والمتعلق بإختبار تحليل التباين الأحادي، ويعرف بالجزر التربيعي لحاصل قسمة مربع (η^2) على نفس المقدار مطروحا من واحد صحيح $(f = \sqrt{\eta^2 / 1 - \eta^2})$.

ويعرف مربع إبتا لمتغير ما، بنسبة تباين المتغير التابع التي يمكن تفسيرها من خلال تباين المتغير المستقل، وبأخذ قيمة تتراوح بين (صفر -1)، ويمكن حسابه بالمعادلة :

$$\eta^2 = SS_a / SS_t \text{ حيث:}$$

SS_a : مجموع مربعات إنحرافات المتغير المستقل a

SS_t : مجموع مربعات الإنحرافات الكلية.

وفي حالة عدم ذكر قيم SS_a, SS_t فيمكن حساب η^2 من المعادلة :

$$\eta^2 = (F_a)(df_a) / (F_a)(df_a) + df_e$$

حيث:

df_a : درجات الحرية للمتغير المستقل a

df_e : درجات الحرية للخطأ

F_a : قيمة الإختبار الاحصائي F للمتغير المستقل a

ووفق معايير كوهن يعد حجم الأثر من حساب f صغيراً عندما $0.24 \geq f \geq 0$ ،

ومتوسطاً عندما: $0.39 \geq f \geq 0.25$ ، وكبيراً عندما $1 \geq f \geq 0.40$

أما بالنسبة لقوة الإختبار فيمكن حسابها بإستخدام جداول كوهن كما يلي:

تكمن أهمية البحث في أن إختبارات الدلالة الإحصائية والتي إعتبرت حجر الزاوية في الدراسات التربوية على مدى سبعين عاماً، وقد واجهت في السنوات الأخيرة ولا تزال معارضة مكثفة وكبيرة من قبل عدد كبير من الباحثين.

وإن استخدامها لوحدها لم يعد كافياً لدى كثير من الباحثين، وبالتالي أصبح البحث عن مفاهيم جديدة (قوة الإختبار الإحصائي وحجم التأثير) أمراً ضرورياً، يوصي به عدد كبير من الباحثين على المستوى الفردي، وكذلك من قبل مراكز نشر ومجلات بحثية عالمية. وبالتالي جاءت هذه الدراسة للتعرف على واقع الدلالة الإحصائية، والدلالة العملية، وقوة الإختبار في الرسائل الجامعية، لتضيف معلومات جديدة حول هذه المفاهيم، فهذا من جهة يساعد الباحثين في:

1. تقدير حجم العينات المناسب عند تصميم أبحاثهم.
 2. تدعيم المعلومات الناتجة من إختبارات الدلالة الإحصائية لزيادة الثقة بها.
 3. لحساب فيما إذا كانت نتائج الأبحاث لها دلالة عملية.
 4. إستخدام نتائجها كمقياس كمي للمقارنة بين عدة دراسات، مما يؤدي إلى تكامل نتائج الأبحاث المختلفة.
- وهي من جهة أخرى تسلط الضوء على جوانب القصور فيها، من أجل الإهتمام بمعالجتها مستقبلاً لتسير في ركب الأبحاث العالمية.

أهداف الدراسة:

1. التعرف على أهمية مؤشرات الدلالة العملية وقوة الإختبار الإحصائي في الأبحاث التربوية.
2. التعرف على مدى إستخدام رسائل الماجستير في تخصص الموهبة والإبداع لمؤشرات الدلالة العملية وقوة الإختبار.
3. التعرف على مدى العلاقة بين الدلالة الإحصائية والدلالة العملية من جهة، والدلالة الإحصائية وقوة الإختبار الإحصائي من جهة أخرى.
4. التعرف على مستويات الدلالة العملية وقوة الإختبار المستخدمة في هذه الرسائل.

الدراسات السابقة:

لقد إهتم عدد كبير من الباحثين بمؤشرات حجم التأثير، وقوة الإختبار الإحصائي، فقد أجرى تمكو وفويل (Tomcho & Foels, 2009) دراسة هدفت إلى فحص القوة الإحصائية للأبحاث المهمة بالنظريات والأنشطة التعليمية، والمنشورة في مجلة تدريس علم النفس (Teaching of Psychology) للأعوام من (1974-2006). تكونت عينة الدراسة من (497) إختباراً إحصائياً تمت مراجعتهم من (197) دراسة، تم حساب متوسط قوة الإختبار للإختبارات الدالة إحصائياً، وأظهرت النتائج أن متوسط قوة الإختبارات لكشف الفروقات الكبيرة كانت (0.81)، والوسط الحسابي لكشف الفروقات المتوسطة (0.53) بينما كان الوسط الحسابي لكشف الفروقات الصغيرة يساوي (0.14).

كما أظهرت النتائج أن (66%) من الدراسات كانت قوة الإختبار لها كافية لكشف الفروقات الكبيرة، وأظهرت النتائج أيضاً أن الباحثين يستخدمون حجم عينات قليل بالنسبة للحجم المثالي.

وأجرى دنج (Deng, 2005) دراسة هدفت إلى فحص مستويات قوة الإختبار للإختبارات الإحصائية المستخدمة في رسائل الدكتوراه الكمية، التي تم إنجازها بنجاح في تخصص القيادة التعليمية أو الإدارة التعليمية في خمس جامعات في ولاية تينيسي، في الفترة من (1996-1998). تكونت عينة الدراسة من (2629) إختباراً إحصائياً، تمت مراجعتهم من (80) رسالة دكتوراه تم إختيارها من أصل (221) رسالة، حيث قام الباحث بإستبعاد الرسائل المتحيزة نتيجة لكبر حجم العينة، تم حساب مستويات قوة الإختبار، وكذلك تم حساب متوسط قوة الإختبار لجميع الإختبارات الدالة إحصائياً، أظهرت النتائج أن الإختبارات الدالة إحصائياً في المتوسط لديها فرصة (29%)

للكشف عن الفروقات عندما تكون صغيرة، وفرصة (75%) للكشف عن الفروقات عندما تكون متوسطة، وفرصة (93%) للكشف عن هذه الفروقات عندما تكون عالية.

كذلك أظهرت النتائج أن (66%) من الإختبارات لها قوة إختبار أقل من (0.30) لإكتشاف الفرق الصغيرة، وأن (91%) من هذه الإختبارات كان مستوى القوة لها أقل من المستوى الموصى به من قبل كوهن (0.80) لإكتشاف الفرق الصغيرة.

كذلك أظهرت النتائج إلى أن أقل الإختبارات قدرة على إكتشاف الفرق الصغيرة هو إختبار (ت) وكاي تربيع، وأن أكثر الإختبارات الدالة إحصائياً هو إختبار (ت) بنسبة (36%) ثم إختبار تحليل التباين بنسبة (34%)، وقد أوصت الدراسة بضرورة تدريب طلاب الدكتوراه في مواد الإحصاء والتركيز على مفاهيم قوة الإختبار، وحجم التأثير، وحجم العينة، والعلاقات المتداخلة بينهما.

وفي دراسة قام بها الجودة (2004) وهدفت إلى التعرف على مستويات الدلالة الإحصائية، والدلالة العملية، وقوة الإختبار للأبحاث المنشورة في مجلة أبحاث اليرموك / سلسلة العلوم الإنسانية والإجتماعية في الأعوام (1985-2001).

إشتملت عينة الدراسة على (889) فرضية صفرية، وأشارت النتائج إلى أن مستوى الدلالة ($\alpha = 0.05$) هو الأكثر استخداماً من قبل الباحثين بنسبة (48.6%)، كما أشارت النتائج أن (8.1%) من الفرضيات الصفرية فقط تم فيها الإشارة إلى مؤشرات للدلالة العملية، كما بينت النتائج أن (67%) من الفرضيات كانت دالة إحصائياً. وأن (50.7%) من الفرضيات الدالة إحصائياً كانت دلالتها العملية ضعيفة، وهذا يثير الشكوك حول دقة ومصداقية القرارات المتخذة اعتماداً على الدلالة الإحصائية.

كما بينت النتائج أن مستويات الدلالة العملية في المتوسط كانت ضعيفة ومتوسطة. وأظهرت النتائج أن (14.3%) من الفرضيات الغير دالة إحصائياً تمتعت بدلالة عملية متوسطة وعالية وهذه أيضاً نتيجة يثار حولها الشكوك في إتخاذ القرار. وأشارت النتائج إلى أن متوسط قوة الإختبارات جميعها كانت عالية نسبياً وبلغت (0.657)، وقد أوصت الدراسة بضرورة إستخراج مؤشرات الدلالة العملية لإعتمادها إلى جانب الدلالة الإحصائية في تفسير نتائج الباحثين لإتخاذ قرارات سليمة ودقيقة.

وفي دراسة قامت بها فريال (2003) هدفت إلى التعرف على واقع الدلالة الإحصائية، والدلالة العملية، وقوة الإختبار، لرسائل الماجستير الصادرة عن كلية العلوم التربوية في جامعة مؤتة في الفترة من (1995-2000)، واشتملت عينة الدراسة على (1167) فرضية صفرية. وأظهرت النتائج إلى أن (68.8%) من الفرضيات استخدمت مستوى دلالة (0.05)، كما بينت النتائج أن (45.2%) من الفرضيات كانت ذات دلالة إحصائية.

كما أشارت النتائج إلى أن (71%) من الإختبارات الدالة إحصائياً كانت لها دلالة عملية ضعيفة، كما أظهرت النتائج إلى أن (5.6%) من الفرضيات كانت غير دالة إحصائياً وتتمتع بدلالة عملية متوسطة وعالية، كما أشارت النتائج إلى أن (10.8%) من الفرضيات ذات الدلالة العملية كانت لها قوة إختبار ضعيفة.

وأجرى بابطين (2000) دراسة هدفت إلى التعرف على مشكلات الدلالة الإحصائية في البحث التربوي، والتعرف على أهم المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لتلك المشكلات، والكشف عن واقع تلك المشكلات في الرسائل العلمية لكلية التربية بجامعة أم القرى.

تكونت عينة الدراسة من (99) رسالة من رسائل الماجستير، وأشارت النتائج إلى أن أهم مشكلات الدلالة الإحصائية هي إستخدام نتائج الدلالة الإحصائية كتفسير لأهمية النتائج، وإستخدام قيمة (p) لتقدير حجم التأثير، والتحيز للنتائج الدالة إحصائياً.

كما توصلت النتائج إلى أن أهم المفاهيم والأساليب الإحصائية البديلة للإختبارات الإحصائية هي تحليل القوة الإحصائية وتقدير حجم التأثير، كما أظهرت النتائج إلى أن (98%)

من الرسائل تعتمد على الدلالة الإحصائية لتفسير النتائج، وأن (7.1%) تستخدم قيمة (p) في تقدير حجم التأثير، وأن نسبة الفرضيات الدالة إحصائياً كانت (47.6%).

كما أظهرت النتائج أنه لا يوجد أي استخدام لتحليل القوة الإحصائية، أما مقاييس حجم التأثير فلم تستخدم إلا في (2%) من الرسائل، وقد أوصت الدراسة بالتركيز في مقررات الإحصاء الإستدلالي على مفاهيم الأساليب الإحصائية، وحدودها، وإفتراساتها، وتوجيه المقررات نحو استخدام مؤشرات الدلالة العملية وقوة الإختبار.

وفي دراسة قام بها إبراهيم (2000) هدفت إلى التعرف على واقع الدلالة الإحصائية، وقوة الاختبارات الإحصائية المستخدمة في فحص الفرضيات، في البحوث المنشورة في مجلة (الدراسات - العلوم التربوية) في عامي (1997-1998) في الجامعة الأردنية، إشمطت عينة الدراسة على (623) إختباراً إحصائياً منشورة في (140) بحثاً، وتوصلت الدراسة إلى أن (95%) من فرضيات الدراسة استخدمت مستوى دلالة (0.05) في فحصها وأن (77.2%) من مجموع الفرضيات تم تحديد مستوى الدلالة الإحصائية لها بعد جمع البيانات، وبينت النتائج أن (5.6%) فقط من الفرضيات الدالة إحصائياً والتي تشكل نسبة (45%) من الفرضيات استخدمت لها مؤشرات للدلالة العملية، وأن (52.7%) من الفرضيات الدالة إحصائياً إقتربت بدلالة عملية ضعيفة، مما يشير إلى دقة في التصميمات المستخدمة التي كانت كافية للكشف عن الفروق حتى الصغيرة، كما تشير النتائج إلى أن (88.6%) من الفرضيات غير الدالة إحصائياً إقتربت بدلالة عملية ضعيفة.

كما بينت النتائج أن (67.3%) من الفرضيات غير الدالة إقتربت بقوة إختبار ضعيفة، مما يشير إلى ضعف في التصميمات المستخدمة في مثل هذه الدراسات، الأمر الذي يثير الشكوك في نتائج الدلالة الإحصائية.

وقام ثومبسون (Thompson, 1999) بفحص مدى إشارة مجلة الأطفال الإستثنائيون (Exceptional Children) لحجم التأثير، شملت عينة الدراسة المقالات المنشورة في المجلدين (63 و 64) من الأعوام (1996-1998)، حيث كان مجموع المقالات المنشورة (23) مقالاً. وجد الباحث أن (20) مقالاً من (23) مقال لم تذكر أي مؤشر لحجم التأثير، وتوصل الباحث إلى ضرورة تضمين الدراسات لمؤشرات حجم التأثير وفقاً لتعليمات الجمعية النفسية الأمريكية.

وفي دراسة قام بها حجمات وعليان (1997) هدفت إلى التعرف على واقع الدلالة الإحصائية، وعلاقتها مع كل من الدلالة العملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير الإرشاد النفسي والتربوي في الجامعة الأردنية، إشمطت عينة الدراسة على (188) فرضية صفرية، تم فحصها باستخدام الإختبار (ف) في (28) رسالة تجريبية وشبه تجريبية، وأشارت النتائج إلى أن (85%) من الفرضيات استخدمت مستوى دلالة ($\alpha = 0.05$) تم تحديد معظمه أثناء التحليل أو بعد الإنتهاء منه، كما بينت النتائج أن الباحثين إكتفوا بالدلالة الإحصائية لإستنتاج الفروق دون أي إشارة للدلالة العملية، كما بينت النتائج أن (6.1%) من الفرضيات كانت دالة إحصائياً.

كما أشارت النتائج إلى أن (75%) من الفرضيات المفحوصة كانت دلالاتها العملية أقل من (25%)، وأن (38.9%) من الفرضيات الدالة إحصائياً كانت لها دلالة عملية ضعيفة.

وبينت النتائج أن (3.2%) من الفرضيات كانت لها دلالة عملية ولم تك دالة إحصائياً وهذا يشير إلى ضعف في تقييم هذه الدراسات، وضعف في قوة الإختبار الإحصائي المستخدم.

كما أشارت النتائج أن (25%) من الإختبارات الإحصائية لم تكن ذات دلالة إحصائية وكان لها قوة إختبار ضعيفة.

وفي دراسة قامت بها أطاف (1997) هدفت إلى التعرف على الدلالة الإحصائية، والعملية، وقوة الإختبارات الإحصائية المستخدمة في رسائل الماجستير في قسم علم النفس التربوي

في الجامعة الأردنية في الأعوام (1990-1995)، وإشتملت عينة الدراسة على (310) فرضية إحصائية في (33) رسالة بحثية.

بينت النتائج إلى أن (30.4%) من الفرضيات الدالة إحصائياً كانت لها دلالات عملية معدومة تقريباً، وأن (30.9%) من الفرضيات الدالة إحصائياً كانت دلالتها العملية ضعيفة، وأن (11.8%) من هذه الفرضيات تمتعت بدلالة عملية من النوع المتوسط والكبير على الترتيب، بينما إقترنت (10.4%) من الفرضيات غير الدالة إحصائياً بدلالة عملية متوسطة وكبيرة. كما بينت النتائج أن (56%) من الفرضيات الغير دالة إحصائياً كانت قوة الإختبار لها ضعيفة.

منهجية البحث

مجتمع الدراسة: تكون مجتمع الدراسة من الإختبارات الإحصائية المستخدمة في رسائل الماجستير، في مجال الموهبة والإبداع في كلية الأميرة عالية التابعة لجامعة البلقاء التطبيقية، والتي تمت مناقشتها في الفترة الواقعة من (2003-2009)، وبعد حصر الرسائل المنشورة في الفترة السابقة بلغ عددها (80) رسالة.

عينة الدراسة:

تكونت عينة الدراسة من (24) رسالة من رسائل الماجستير، وبمراجعة هذه الرسائل تبين أن أكثر الإختبارات المستخدمة فيها هو الإختبار الإحصائي (ت)، والإختبار الإحصائي (ف)، أما بقية الإختبارات فكانت نسبة ظهورها قليلة، وبالتالي شملت عينة الدراسة (119) إختباراً للإحصائي (ت)، و(100) إختباراً للإحصائي (ف).

إجراءات الدراسة:

1. تم حصر رسائل الماجستير المنشورة في الفترة من (2003-2009) من سجلات دائرة القبول والتسجيل في الجامعة. وذلك بعد إستثناء الطلاب الذين تقدموا لإمتحان شامل كمسار آخر بدل الرسالة.
2. تم إعداد قائمة بالرسائل تضمنت عنوان الرسالة، وإسم الباحث، وتاريخ المناقشة، بعد الرجوع إلى مكتبة كلية الأميرة عالية، والمكتبة الرئيسية في مركز الجامعة، ومركز إيداع الرسائل الجامعية في الجامعة الأردنية.
3. تم جمع البيانات اللازمة للإجابة عن أسئلة الدراسة، تضمنت مستوى الدلالة، وهل تم تحديدها مسبقاً، ومتوسط كل مجموعة، وكذلك مجموع المربعات، وحجم العينة وتم تفرغها في جداول خاصة.
4. من هذه الجداول تم حساب متوسط قوة الإختبار لكل من الإختبار الإحصائي (ت)، والإختبار الإحصائي (ف)، بإستخدام جداول كوهن (Cohen, 1977) المعدة لهذا الغرض. وكذلك تم حساب متوسط حجم التأثير لكل من الإختبارين من خلال المعادلات التي تم توضيحها سابقاً.

النتائج:

هدفت هذه الدراسة إلى التعرف على مستويات الدلالة الإحصائية، والدلالة العملية، وقوة الإختبارات الإحصائية المستخدمة في فحص الفرضيات الصفرية في رسائل ماجستير الموهبة والإبداع.

وللإجابة عن السؤال الأول والمتعلق بمستويات الدلالة الإحصائية المستخدمة في فحص الفرضيات، وهل تم تحديدها قبل إختبار الفرضية أم بعدها؟
تم حساب النسب المئوية والتكرارات ووضعت النتائج في الجدول رقم (1)

جدول رقم (1)

توزيع الفرضيات وفق مستوى الدلالة المستخدمة في فحصها وزمن تحديدها

الإحصائي	مستوى الدلالة المستخدم	عدد الفرضيات	النسبة المئوية	الإحصائي	زمن التحديد	العدد	النسبة
ت	0.05	111	%93	ت	قبل التحليل	9	%8
	0.01	8	%7		بعد التحليل	110	%92
ف	0.05	95	%95	ف	قبل التحليل	10	%10
	0.01	5	%5		بعد التحليل	90	%90
المجموع الكلي	0.05	206	%94	المجموع الكلي	قبل التحليل	19	%9
	0.01	13	%6		بعد التحليل	200	%91

يتبين من الجدول (1) أن معظم الباحثين يستخدمون مستوى دلالة $\alpha = 0.05$ في فحص فرضياتهم الإحصائية، حيث كانت نسبتهم على الإختبارين (%94)، وأن (%91) من الباحثين يقومون بتحديد مستويات الدلالة الإحصائية بعد تحليل البيانات.

وللإجابة عن السؤال الثاني والمتعلق بنسبة الفرضيات الدالة إحصائياً، فقد تم حساب التكرارات والنسب المئوية للإختبارات الدالة إحصائياً مجتمعة ولكل إختبار على إنفراد، لاحظ جدول رقم (2).

جدول رقم (2)

أعداد ونسب الفرضيات الصفرية الدالة وغير الدالة إحصائياً

الإحصائي	الفرضيات الدالة إحصائياً		الفرضيات غير الدالة إحصائياً	
	العدد	النسبة	العدد	النسبة
ت	38	%32	81	%68
ف	15	%15	85	%85
المجموع الكلي	53	%25	166	%75

يتضح من جدول (2) أن عدد الفرضيات الدالة إحصائياً على الإختبارين كانت (%25)، أما على الإختبار (ت) فقد كانت الإختبارات الدالة إحصائياً (%32) وعلى الإختبار (ف) فكانت (%15).

السؤال الثالث: ما واقع الدلالة العملية للإختبارات الإحصائية المستخدمة في فحص الفرضيات؟

للإجابة عن سؤال الدراسة تم حساب التكرارات والنسب المئوية للفرضيات المفحوصة في كل فئة من فئات حجم التأثير، لكل من الإختبار الإحصائي (ف) والإختبار الإحصائي (ت) ووضعت النتائج في الجدول (3) والجدول (4).

جدول رقم (3)

التكرارات والنسب المئوية للفرضيات المفحوصة في كل فئة من فئات مربع إيتا على إختبار (ف)

النسبة المئوية	التكرار	مربع إيتا
56%	56	صفر - 0.10
30%	30	0.10 - 0.24
6%	6	0.25 - 0.39
8%	8	0.40 - فما فوق

يتضح من الجدول أن (56%) من الفرضيات المفحوصة على الإختبار (ف) كانت تتمتع بدلالات عملية أقل من (0.10)، وأن حوالي (86%) من الفرضيات المفحوصة على الإختبار (ف) كانت لها دلالات عملية ضعيفة (قيمة مربع إيتا أقل من (0.25)، كما يشير الجدول إلى أن (6%) من الفرضيات المفحوصة إحصائياً على الإختبار (ف) كانت دلالاتها العملية متوسطة (0.25-0.39)، في حين بلغت نسبة الفرضيات المفحوصة والتي لها دلالات عملية عالية (8%).

جدول (4)

التكرارات والنسب المئوية للفرضيات المفحوصة على الإختبار (ت) تبعاً للدلالة العملية

النسبة المئوية	التكرار	مستوى الدلالة
84%	100	صفر - 0.49
8.5%	10	0.50 - 0.79
7.5%	9	0.80 - 1

يتضح من الجدول (4) أن (84%) من الفرضيات المفحوصة على الإختبار (ت) كانت تتمتع بدلالات عملية ضعيفة (أقل من 0.50)، وأن حوالي (8.5%) من الفرضيات المفحوصة على الإختبار (ت) كانت لها دلالات عملية متوسطة (0.50 - 0.79)، كما يشير الجدول إلى أن نسبة الفرضيات المفحوصة على الإختبار (ت) والتي لها دلالات عالية كانت (7.5%). وبحساب المتوسطات الحسابية لأحجام التأثير كمؤشرات على الدلالة العملية (لاحظ جدول (5)، يتضح أن متوسط حجم التأثير للإختبارات الإحصائية على الإحصائي (ف) كانت (0.14)، وهذا يدل على أن الدلالة العملية للإختبار (ف) كانت ضعيفة في دلالاتها. في حين كان متوسط حجم التأثير للإختبارات الإحصائية على الإحصائي (ت) تساوي (0.26)، وهذا يدل على أن الدلالة العملية للإختبار (ت) كانت (ضعيفة) في دلالاتها حسب تصنيف كوهن.

جدول رقم (5)

متوسط الدلالة العملية للفرضيات الصفرية المستخدمة في الرسائل موزعة حسب نوع الإختبار

إختبار ت	إختبار ف	متوسط الدلالة
0.26	0.14	

كما لوحظ من تحليل البيانات أن أياً من الباحثين لم يستخدم مؤشرات إضافية لحساب حجم التأثير، أثناء أو بعد فحصهم للفرضيات، بل إعتبروا أن الدلالة الإحصائية كافية لتفسير النتائج.
السؤال الرابع: ما نسبة الفرضيات الإحصائية التي إقترنت فيها الدلالة الإحصائية بدلالة عملية للإختبارات؟

للإجابة عن السؤال تم حساب مربع كاي (χ^2) لفحص ما إذا كانت الدلالة العملية مستقلة عن الدلالة الإحصائية لكل إختبار.

كانت قيمة (χ^2) للإختبار ف (7.819) بدرجات حرية (2)، وهذه القيمة ذات دلالة إحصائية عند مستوى دلالة (0.01)، أي أن هناك علاقة ما بين الدلالة العملية والدلالة الإحصائية، والجدول رقم (6) يبين إقتران الدلالة الإحصائية بالدلالة العملية للإختبار (ف).

جدول رقم (6)

إقتران الدلالة العملية بالدلالة الإحصائية للإختبار الإحصائي (ف) في كل فئة من فئات مربع إيتا

الفرضيات غير الدالة إحصائياً		الفرضيات الدالة إحصائياً		الدلالة العملية / الدلالة الإحصائية
النسبة	العدد	النسبة	العدد	
72%	72	14%	14	صفر - 0.24
6%	6	-	-	0.39 - 0.25
7%	7	1%	1	1 - 0.40
85%	85	15%	15	المجموع

يتضح من الجدول (6) أن (14%) من الفرضيات الدالة إحصائياً كانت لها دلالة عملية (ضعيفة)، وأن (1%) من الفرضيات الدالة إحصائياً كانت لها دلالة عملية (قوية)، فيما كانت (72%) من الفرضيات غير الدالة إحصائياً لها دلالة عملية (ضعيفة). كما يتضح من الجدول أن (13%) من الفرضيات الغير دالة إحصائياً إقترنت بدلالة عملية (متوسطة وقوية)، وهذا يعني أن قوة الإختبار كانت ضعيفة، مما يثير شكوك حول دقة التصميم المستخدم. أما بالنسبة للإختبار الإحصائي (ت) فقد كانت قيمة (χ^2) تساوي (32.47) وهي قيمة ذات دلالة إحصائية عند مستوى دلالة (0.01) بدرجات حرية (2) أي أن الدلالة الإحصائية غير مستقلة عن الدلالة العملية، والجدول رقم (7) يبين إقتران الدلالة العملية بالدلالة الإحصائية.

جدول رقم (7)

إقتران الدلالة الإحصائية بالدلالة العملية ضمن كل فئة من فئات حجم التأثير (d)

الفرضيات غير الدالة إحصائياً		الفرضيات الدالة إحصائياً		الدلالة العملية / الدلالة الإحصائية
النسبة	العدد	النسبة	العدد	
66%	79	18%	22	صفر - 0.19
2%	2	7%	8	0.79 - 0.50
-	-	7%	8	1.00 - 0.80
-	81	-	38	المجموع

يتضح من الجدول (7) أن (18%) من الفرضيات الدالة إحصائياً كانت لها دلالة عملية (ضعيفة)، وأن (7%) من الفرضيات الدالة إحصائياً كانت لها دلالة عملية (متوسطة)، كما يبين الجدول أن (7%) من الفرضيات الدالة إحصائياً لها دلالة عملية (عالية). كما يتضح من الجدول أن (66%) من الفرضيات الغير دلالة إحصائياً إقترنت بدلالة عملية (ضعيفة)، وأن (2%) من الفرضيات الغير دالة إحصائياً كانت دلالتها العملية (متوسطة).

السؤال الخامس: ما واقع الإختبارات الإحصائية المستخدمة في فحص الفرضيات؟ للإجابة عن سؤال الدراسة تم حساب التكرارات والنسب المئوية للفرضيات المفحوصة في كل مستوى من مستويات قوة الإختبار الإحصائي (ف)، وفي كل مستوى من مستويات قوة الإختبار الإحصائي (ت)، ووضعت النتائج في الجداول (8) و (9).

جدول رقم (8)

التكرارات والنسب المئوية للفرضيات المفحوصة في كل مستوى من مستويات قوة الإختبار الإحصائي (ف)

النسبة المئوية	التكرار	مستوى القوة
65%	65	0.39 - صفر
20%	20	0.59 - 0.40
8%	8	1 - 0.60
7%	7	1.00 - 0.80

يتبين من الجدول (8) أن (65%) من الفرضيات كانت قوة الإختبار الإحصائي (ف) فيها (ضعيفة)، وحوالي (20%) من الفرضيات كانت قوة الإختبار الإحصائي فيها (متوسطة)، وأن (15%) من الفرضيات المفحوصة كانت قوة الإختبار الإحصائي فيها (عالية).

جدول رقم (9)

التكرارات والنسب المئوية للفرضيات المفحوصة في كل مستوى من مستويات قوة الإختبار الإحصائي (ت)

النسبة المئوية	التكرار	مستوى القوة
52%	62	0.39 - صفر
8%	9	0.59 - 0.40
40%	48	1.00 - 0.60

يتبين من الجدول (9) أن (52%) من الفرضيات المفحوصة كانت قوة الإختبار الإحصائي (ت) فيها (ضعيفة)، وأن (8%) من هذه الفرضيات المفحوصة كانت قوة الإختبار الإحصائي (ت) فيها (متوسطة)، بينما كان (40%) من هذه الفرضيات قوة الإختبار الإحصائي (ت) لها (عالية).

وبحساب المتوسطات الحسابية لقوة الإختبارات الإحصائية المستخدمة في فحص الفرضيات لكل إختبار إحصائي (جدول 10)، يتضح أن متوسط قوة الإختبار الإحصائي (ف) كانت (0.34)، وهي قوة إختبار (ضعيفة) في قوتها، في حين كان متوسط قوة الإختبار للإحصائي (ت) تساوي (0.47)، وهي قوة إختبار (متوسطة) في قوتها.

جدول رقم (10)

متوسط قوة الإختبار للفرضيات الصفرية المستخدمة في الرسائل موزعة حسب نوع الإختبار

متوسط القوة	إختبار ف	إختبارات
	0.34	0.47

السؤال السادس: ما نسبة الفرضيات الإحصائية التي إقترنت فيها الدلالة الإحصائية بقوة الإختبارات الإحصائية؟

للإجابة عن سؤال الدراسة تم حساب مربع كاي لفحص إستقلالية قوة الإختبارات الإحصائية عن الدلالة الإحصائية لكل إختبار، فكانت قيمة (χ^2) للإختبار (ف) تساوي (32.77) بدرجات حرية (2)، وهي ذات دلالة إحصائية عند مستوى دلالة (0.01)، أي أن الدلالة الإحصائية غير مستقلة عن قوة الإختبار الإحصائي. والجدول رقم (11) يبين إقتران الدلالة الإحصائية بقوة الإختبار الإحصائي.

جدول رقم (11)

إقتران الدلالة الإحصائية بقوة الإختبار الإحصائي (ف) ضمن كل مستوى من مستويات القوة

مستوى القوة / الدلالة الإحصائية	الفرضيات الدالة إحصائياً		الفرضيات غير الدالة إحصائياً	
	العدد	النسبة	العدد	النسبة
صفر - 0.39	-	-	65	65%
0.40 - 0.59	2	2%	18	18%
0.60 - 1.00	13	13%	2	2%

يشير الجدول (11) أن (2%) من الفرضيات الدالة إحصائياً والمفحوصة بالإختبار الإحصائي (ف) إقترنت بقوة إختبار (متوسطة)، وأن (13%) من هذه الفرضيات إقترنت بقوة إختبار (عالية)، كما يلاحظ من الجدول أن (65%) من الفرضيات الغير دالة إحصائياً إقترنت بقوة إختبار (ضعيفة)، وأن (18%) من هذه الفرضيات إقترنت بقوة إختبار (متوسطة)، في حين لم يقترن من هذه الفرضيات بقوة إختبار عالية سوى (2%).

أما بالنسبة للإختبار الإحصائي (ت) فقد كانت قيمة (χ^2) تساوي (54.39) بدرجات حرية (2)، وهي قيمة ذات دلالة إحصائية عند مستوى دلالة (0.01)، أي أن هناك علاقة ما بين الدلالة الإحصائية وقوة الإختبار، والجدول رقم (12) يبين إقتران الدلالة الإحصائية بقوة الإختبار الإحصائي (ت).

جدول رقم (12)

إقتران الدلالة الإحصائية بقوة الإختبار الإحصائي (ت) ضمن كل مستوى من مستويات القوة

مستوى القوة / الدلالة الإحصائية	الفرضيات الدالة إحصائياً		الفرضيات غير الدالة إحصائياً	
	العدد	النسبة	العدد	النسبة
صفر - 0.39	1	1%	62	52%
0.40 - 0.59	-	-	9	8%
0.60 - 1.00	36	30%	11	9%
المجموع	37	-	82	-

يتضح من الجدول (12) أن (1%) من الفرضيات الدالة إحصائياً والمفحوصة بالإختبار الإحصائي (ت) إقترنت بقوة إختبار (ضعيفة)، وأن (30%) من هذه الفرضيات إقترنت بقوة إختبار (عالية)، كما يلاحظ من الجدول أن (52%) من الفرضيات الغير دالة إحصائياً إقترنت بقوة إختبار (ضعيفة)، وأن (8%) من هذه الفرضيات إقترنت بقوة إختبار (متوسطة)، في حين لم يقترن من هذه الفرضيات بقوة إختبار عالية سوى (9%).

مناقشة النتائج:

توصلت الدراسة إلى أن (94%) من الفرضيات الصفرية تم فحصها عند مستوى دلالة $\alpha = 0.05$ ، وهذا يشير إلى وعي الباحثين لبعض المفاهيم الخاطئة والمتعلقة بمستوى الدلالة ومفادها "أن أهمية نتائج الدراسة تزداد بتقليل مستوى الدلالة من 0.05 إلى 0.01 أو 0.001"، وقد اتفقت هذه النتائج مع دراسة إبراهيم (2000).

كما أشارت النتائج إلى أن (91%) من الباحثين يقومون بتحديد مستويات الدلالة الإحصائية بعد تحليل البيانات. ولعل سبب ذلك يعود إلى اعتماد الباحثين على البرامج الإحصائية في استخراج النتائج، مما يوفر للباحث فرصة أفضل في تحديد مستوى الدلالة بعد تحليل البيانات.

كما أشارت النتائج أن (25%) من الفرضيات كانت دالة إحصائياً. حين كانت (75%) من الفرضيات غير دالة إحصائياً. وهذا يشير إلى عدم تحيز الباحثين للنتائج الدالة إحصائياً وقد اتفقت هذه النتائج مع دراسة بابطين (2001) وتبين من حساب الدلالة العملية للفرضيات المفحوصة أن (84%) من الفرضيات المفحوصة على الإختبار (ت) و(86%) من الفرضيات المفحوصة على الإختبار (ف) كانت لها دلالات عملية ضعيفة، كما نتضح أيضاً هذه الدلالات العملية الضعيفة من خلال المتوسطات العملية فقد كانت للإختبار (ف) تساوي (0.14) وللإختبار (ت) تساوي (0.26) وهي دلالات عملية ضعيفة حسب تصنيف كوهين. وقد اتفقت هذه النتائج مع دراسة حجيات وعليان (1997)، وهذا يشير إلى أن الباحثين لا يستخدمون أثناء تحليل البيانات أي من مؤشرات الدلالة العملية بل يكتفون بالدلالة الإحصائية وحدها في تفسير النتائج وقد اتفقت هذه النتائج مع ثومبسون (1999).

كما أظهرت النتائج أن (14%) من الفرضيات الدالة إحصائياً على الإختبار (ف)، وأن (18%) من الفرضيات الدالة إحصائياً على الإختبار (ت) كانت لها دلالة عملية ضعيفة وقد يرجع سبب وجود دلالة إحصائية لهذه الإختبارات بالرغم من أن دلالتها العملية ضعيفة إلى كبر حجم العينات لا إلى وجود تأثير للمتغير المستقل عن المتغير التابع وقد اختلفت هذه النتائج مع دراسة الجودة (2004). وكذلك أشارت نتائج الدراسة إلى أن (13%) من الفرضيات الغير دالة إحصائياً على الإختبار (ف)، وأن (2%) من الفرضيات الغير دالة إحصائياً على الإختبار (ت) تمتعت بدلالات عملية متوسطة وعالية. وهذه نتيجة يثار حولها الشكوك والتساؤلات في اتخاذ القرارات، فوجود مثل تلك الفرضيات يعني أن فروقا أو أثراً حقيقياً للمتغير المستقل على المتغير التابع، وأن تلك الإختبارات قد فشلت في الكشف عنها، وقد يعزى سبب ذلك إلى ضعف في التصاميم الإحصائية المستخدمة في فحص الفرضيات الصفرية، وقد يعود إلى أن الباحثين لا يهتمون بالشروط الواجب توفرها عند تطبيق الأساليب الإحصائية المختلفة. وقد اتفقت هذه النتيجة مع دراسة الجودة (2004). أما فيما يتعلق بقوة الإختبار فقد تبين أن (65%) من الفرضيات على الإختبار (ف)، وأن (52%) من الفرضيات على الإختبار (ت) كانت قوة الإختبار فيها ضعيفة، وهذا يشير إلى أن نسبة الوقوع في الخطأ من النوع الثاني كانت عالية مما يؤكد وجود ضعف واضح في تصميم البحث، ولعل ذلك يفسر نسبة وجود (75%) من الفرضيات الغير دالة إحصائياً، وقد اختلفت هذه الدراسة مع دراسة تمكو وفويل (2009)، وكذلك أشارت النتائج إلى أن الفرضيات الغير دالة إحصائياً والضعيفة في قوتها على الإختبار (ف) شكلت ما نسبته (50%)، فيما كانت هذه النسبة للإختبار (ت) تساوي (46%)، وقد يعود هذا إلى عدم وجود مثل هذه الفروق أو أن هذه الفروق موجودة

ولكن ضعف قوة الاختبار حال دون الكشف عنها. وقد اختلفت هذه النتائج مع دراسة حجيات وعليان (1997) حيث كانت النسب في تلك الدراسة تساوي (25%) فقط. كما أظهرت النتائج أن الاختبارات الدالة إحصائياً في المتوسط لديها فرصة (33%) للكشف عن الفروقات عندما تكون صغيرة، وفرصة (82%) للكشف عن الفروقات عندما تكون متوسطة، وفرصة (95%) للكشف عن هذه الفروقات عندما تكون عالية. وقد اتفقت هذه الدراسة مع دراسة دنج (2005).

التوصيات:

وقد أوصت الدراسة:

1. ضرورة تضمين الباحثين لمؤشرات عن حجم التأثير وقوة الاختبار في أبحاثهم.
2. التركيز على مفاهيم الدلالة العملية وقوة الاختبار في تدريس مواد الإحصاء.
3. ضرورة تضمين المجالات العلمية والدوريات وجود مؤشرات لحجم التأثير وقوة الاختبار في الأبحاث كمعيار لجودة البحث.

المراجع

أولاً: المراجع العربية

1. بابطين، عادل (1422هـ)، مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة، رسالة ماجستير، كلية التربية، جامعة أم القرى، مكة المكرمة.
2. خاطر، أحمد، محمد كشك، جمعة سلمى (1997)، التحليل الإحصائي للبحوث في الخدمة الاجتماعية، المكتب الجامعي الحديث، الإسكندرية.
3. حجيات، تحسين، عليان، خليل (1997)، واقع الدلالة الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير الإرشاد النفسي والتربوي بالجامعة الأردنية، مجلة دراسات العلوم التربوية، الجامعة الأردنية، عمان، الأردن، المجلد (24)، العدد الثاني.
4. عودة، أحمد، الخليلي، خليل (2000)، الإحصاء للباحث في التربية والعلوم الإنسانية، الطبعة الثانية، دار الأمل للنشر والتوزيع، إربد - الأردن.
5. إبراهيم، إبراهيم (2000)، واقع الدلالة الإحصائية والدلالة العملية وقوة الاختبارات الإحصائية المستخدمة في بحوث مجلة دراسات، العلوم التربوية في الأعوام (1997-1998) في الجامعة الأردنية، رسالة ماجستير، كلية التربية، الجامعة الأردنية، عمان.
6. محمود، فريال (2003)، الدلالة الإحصائية والدلالة العملية وقوة الاختبار للاختبارات الإحصائية المستخدمة في رسائل الماجستير الصادرة عن كلية العلوم التربوية في جامعة مؤتة، رسالة ماجستير، كلية التربية، جامعة مؤتة، مؤتة، الكرك.
7. الجودة، ماجد (2004)، الدلالة الإحصائية والدلالة العملية وقوة الاختبار للأبحاث المنشورة في مجلة أبحاث اليرموك، سلسلة العلوم الإنسانية والاجتماعية (1985 - 2001)، دراسة وصفية، رسالة ماجستير، جامعة اليرموك، إربد، الأردن.
8. جعفر، أطفاف رمضان (1997). واقع الدلالة العملية والإحصائية وقوة الاختبارات المستخدمة في رسائل الماجستير في علم النفس التربوي في الجامعة الأردنية بين عامي (1990-1995)، رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.

ثانياً: المراجع الأجنبية :

9. Cohen-J 1977 - Statistical power Analysis for The Behavioural Sciences. 2nd edition .Academic Pressinc. New York.

10. Kirk-Roger (2007) Effect magnitude; a different focus -Journal of statistical planning and inference-volume 137 -issue 5-pp 1634-1646.
11. R.I. Parker and D. F. Brossart (2003). Evaluating Single - case research data; A Comparison of Seven Statistical methods ;Behavior Therapy, vo 34-pp 189-211.
12. Roby. Randall (2004) Reporting Point and Interval estimates of effect-size for planned Contrasts; fixed within effect analyses of variance/Journal of Fluency disorders /Vol 29/Issue 4/pp 307-341.
13. Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines And explanations. American Psychologist, 54, 594- 604.
14. Tomcho, Thomas & Foels, Rob (2009). The Power of Teaching Activities: Statistical and Methodological Recommendations. Teaching of Psychology, Vol. 36, No. 2, pp 96-101.
15. Deng, Heping (2005). Does It Matter If Non-Powerful Significance Tests Are Used in Dissertation Research, Practical Assessment Research & Evaluation, Vol. 10, No 16.
16. Kellow, J. Thomas (1998). Beyond Statistical Significant Tests: The Importance of Using Other Estimates of Treatment Effects to Interpret Evaluation Results. American Journal of Evaluation. V19, n1, P 123-34.
17. Thompson, Bruce (1995b). Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms. Paper Presented at the Annual Meeting of the 'Mid-South Educational Research Association (Biloxi, MS).
18. Kirk, Roger E. (1996). Practical Significance: A concept Whose Time Has Come, Educational and Psychological Measurement. V56, n5, 746-759.
19. Winer-Russells (2000) comment on The Historical Growth of Statistical Significance Testing IN Psychology and Its Future prospects Educational and Psychological measurement .v60-n5-pp 663-696.