.

الدلالة الاحصائية وحجم التأثير و قوة الاختبار الاحصائي "ت" المستخدم في بعض بحوث الماجستير والدكتوراه
- دراسة تحليلية تقويمية -

أ. زيـــاد رشيـــد – جامعـــة وهـــران2–
Ziadpsy@gmail.com

ملخص :

هدفت الدراسة الراهنة إلى تناول ثلاثة مفاهيم إحصائية قد يخلط الكثير من غير المختصين بينهما ، وهما: الدلالة الإحصائية والدلالة العملية (حجم التأثير) وقوة الاختبار الإحصائي ، وبيان أهميتهما والعلاقة بينهما ، وتقديم إطارا نظرياً وتطبيقياً للباحثين وتقصي واقع استخدام أسلوب الدلالة العملية (حجم التأثير) إلى جانب الدلالة الإحصائية وقوة الاختبار الإحصائي في بعض بحوث الماجستير والدكتوراه الجزائرية مابين 2005– 2015 ، والتعرف على نسبة الفرضيات التي استخدمت الأسلوب الإحصائي " ت " T-test والتي اقترنت فيها الدلالة الإحصائية بالدلالة العملية (حجم التأثير) وقوة الاختبار الإحصائي.

الكلمات المفتاحية : الدلالة الإحصائية ، حجم التأثير ، قوة الاختبار الإحصائي ، اختبار " ت" T-test.

مقدمـــــة :

يعتمد البحث التربوي والنفسي اختبارات الدلالة الإحصائية كأسلوب إحصائي علمي لتقرير ما إذا كانت الفروق أو الارتباطات أو العلاقات بين المتغيرات دالة إحصائيا ، ومع رسوخ هذا الأسلوب أصبحت اختبارات الدلالة الإحصائية لاختبار الفرضيات في الأبحاث التربوية طريقة تقليدية ، وكما يقول Keaster(1998) فقد تحولت اختبارات الدلالة الإحصائية من أساليب إلى طقوس.

لقد أصبح استخدام وسوء استخدام اختبارات الدلالة الإحصائية القضية الرئيسية لطرق البحث التربوي والنفسي الحالي (kaufman,1998)، وكثفت المعارضة للدلالة الإحصائية في السنوات الأخيرة (Burdenski, 1998) ، وارتفعت دعوات عالمية إلى ضرورة أن تقاد اختبارات الدلالة الإحصائية إلى مكانها الصحيح في العملية العلمية (keaster,1998) . كما أن هناك حاجة علمية إلى أن يتجاوز التركيز في مناقشة قصور

الدلالة الإحصائية إلى ما بعد الإدانات الإضافية لتلك الاختبارات ، حيث الحاجة إلى مزيد من الاقتراحات لتحسين استخدام ممارسة الاختبارات الإحصائية في اختبار الفروض (عادل بابطين،1421: 4).

إذ من الملاحظ أن بعض الباحثين يعتمدون في تقرير نتائجهم على الدلالة الإحصائية دون محاولة الكشف عن مقدار العلاقة الفعلية القائمة بين المتغيرين ، وبالتالي تصبح هناك مغالاة في تفسير النتائج بالاعتماد على مستوى الدلالة الإحصائية فقط على الرغم من أنها ربما لا تكون لها قيمة من الناحية التطبيقية أو العملية ، وعليه فإنه عندما تشير النتائج إلى أن القيمة دالة إحصائيا ، فذلك يعني أن المتغير المستقل له تأثير غير صفري أو أن العلاقة بين المتغيرات ليست صفرية ، وهذا لا يدل على حجم التأثير أو درجة العلاقة بين المتغيرات (الشربيني،1995: 32).

ففكرة الدلالة العملية أو حجم التأثير Effect size تقوم على صياغة الفروق بين المتوسطات وذلك باستخدام الانحراف المعياري كوحدة قياس لمقدار الفرق بين تلك المتوسطات ، أو التعبير عن العلاقة بين المتغيرات المستقلة من جهة والمتغيرات التابعة من جهة أخرى عن طريق استخراج حجم تباين المتغير التابع الذي يمكن تفسيره عن طريق المتغير المستقل كما هو الحال في تحليل التباين (Cohen,1988) .

وأكد كل من ماكلين وارنست من خلال مراجعتهما لأدبيات البحث التربوي المتعلقة بمؤشرات حجم الأثر بقولهما :" لم يجد المؤلفان مقال يشكك في أهمية تضمين بعض النتائج لحجم الأثر أو تقدير الدلالة العملية في تقارير الأبحاث ، وعن أهمية مقاييس حجوم الأثر بأن الدراسات يجب أن تنشر بدون اختبارات الدلالة الإحصائية ، ولكن ليس بدون أحجام الأثر"(McClaien and Ernest ,1998 : 18)

وفي دراسة لثومبستون Thompson(1999) بعنوان " تحسين وتوضيح نتائج الأبحاث ، وفائدة مؤشرات حجم التأثير كاختبارات مكملة لاختبارات الدلالة الإحصائية" . قام بفحص مدى إشارة مجلة (الأطفال الأستثنائيون) لحجم التأثير ، شملت عينة الدراسة (23) مقالاً ، وجد الباحث منها (20) مقالاً لم تذكر أي مؤشر لحجم التأثير ، وتوصل الباحث إلى ضرورة تضمين الدراسات لمؤشرات عن حجم التأثير وفقاً لتعليمات الجمعية النفسية الأمريكية APA. وفي دراسة أخرى لـ(1997) Thompson and Snyder بعنوان " ممارسة اختبارات الدلالة الإحصائية في مجلة التربية التجريبية" . وقد شملت عينة الدراسة (22) مقالاً منشوراً استخدمت اختبارات الدلالة الإحصائية ، قام الباحثان بحساب حجم التأثير لكل الدراسات وذلك للتحقق من افتراضهما بان الباحثين يعتمدون على الاختبارات الإحصائية دون التحقق من أن تلك النتائج ربما لا تكون مفيدة أو جديرة بالاهتمام ، حيث أنهم يختبرون الدلالة الإحصائية فقط دون حساب حجوم التأثير ، وقد افترض الباحثان أيضاً أن بعض الباحثين يحصلون على نتائج دالة إحصائيا بسبب حجم العينة الكبير ، بينما نتائجهم تستلزم حجوم

تأثيرات صغيرة فقط. وقد وجد الباحثان أن بعض الباحثين في المقالات (22) أشاروا إلى اهتمام محدود بشكل واضح للتأثير الناتج عن الخطأ من النوع (I) ، كما أن اللغة المستخدمة في كل المقالات الـ(22) تضمنت وصف النتائج بأنها (دالة) وليست (دالة إحصائيا) ، كما أن بعضها تضمنت لغة (تقترب من الدلالة) ، (كانت تقريبا دالة) ، (لا تدنو من الدلالة) ، (تأثير المعالجة اقتربت من مستوى α المحدد). كما وجدا أن (8) مقالات من (22) مقال لم تتضمن الإشارة إلى حجوم التأثير بشكل مصاحب لاختبارات الدلالة الإحصائية.

وفي دراسة أخرى قام بها Thompson(1998) أيضا بعنوان " خمسة أخطاء منهجية تحدث في البحث التربوي " . توصلت الدراسة إلى أن أحد هذه الأخطاء تخص الدلالة الإحصائية وهي التفسير الخاطئ للدلالة الإحصائية وما يتعلق بها من فشل تقرير وتفسير حجم الأثر الممثل في جميع التحاليل الكمية.

مشكلة الدراسة :

بعد أكثر من 150 عام من اعتمادها الجمعية الأمريكية للإحصاء ASA تحدد الأسس العملية الصحيحة لاستخدام قيمة P-value (الدلالة الإحصائية) ، ففي حدث مفاجئ دعت هذه الجمعية في مقال نشر بتاريخ 08–03–2016 إلى وقف الاعتماد على قيمة P-value في الدراسات الإحصائية ، أو بشكل أدق دعت الباحثين إلى حسن استخدام هذه الأداة والوقف الفوري لاستخدامها كأداة شبه وحيدة لاتخاذ القرار ، كما حددت في هذه المقالة المبادئ الستة لاستخدام وتفسير القيمة P-value وهذه المبادئ الستة مختصر ،كالآتي :

1. قيمة P-value يمكن أن تشير إلى كيفية تعارض (عدم اتفاق) البيانات مع النموذج الإحصائي المستخدم.
2. قيمة P-value لا تحسب احتمالية أن الفرضية المدروسة صحيحة أو احتمالية وأن البيانات كانت مختارة عشوائياً فقط.
3. الاستنتاجات العلمية والاقتصادية ، بل وحتى السياسية لا يجب أن تبنى استناداً فقط على تجاوز قيمة P-value لعتبة معينة أم لا.
4. الاستدلال السليم يتطلب تقديم تقارير كاملة وشفافة.
5. قيمة P-value لا تقيس حجم تأثير أو أهمية نتيجة.
6. القيمة P-value بحد ذاتها ، لا توفر قدراً كبيراً من الأدلة فيما يخص نموذجاً أو فرضية معينة.

(Ronald,w and Nicole,L ,2016 :01-17)

ومن هذا المنطلق يرى الباحث ضرورة إلى أن تقاد اختبارات الدلالة الإحصائية في بحوثنا ككل وبحوث الماجستير والدكتوراه الجزائرية بشكل خاص إلى مكانها الصحيح في العملية العلمية ، فالعلوم الاجتماعية بشكل

عام والتربوية والنفسية بشكل خاص سوف تتقدم إلى الأمام بشكل أسرع عندما يصبح البحث بحثاً عن التأثيرات الثابتة الواضحة في الحجم في سياق البحث العلمي .

وبناءا على ذلك فإن الدراسة الحالية جاءت لتحاول الإجابة عن التساؤلات ، الآتية :

1. ما نسبة استخدام مفهوم حجم التأثير(الدلالة العملية) إلى جانب الدلالة الإحصائية بالنسبة للاختبار الإحصائي " ت" في بعض بحوث الماجستير والدكتوراه الجزائرية ما بين 2005-2015؟

2. ما نسبة الفرضيات التي استخدمت الاختبار الإحصائي " ت" والتي اتفقت فيها نتائج الدلالة الإحصائية مع حجم الأثر (الدلالة العملية) في بعض بحوث الماجستير والدكتوراه الجزائرية ما بين2005-2015؟

3. ما النسبة التراكمية لقوة الاختبار الإحصائي " ت" المستخدم في بعض بحوث الماجستير والدكتوراه الجزائرية ما بين 2005-2015؟

أهداف الدراسة :

تهدف الدراسة إلى إلقاء الضوء على واقع استخدام الباحثين في الدراسات العليا (ماجستير ودكتوراه) بالجزائر لمؤشرات قوة الاختبار الإحصائي وحجم التأثير أو الدلالة العملية إلى جانب الدلالة الإحصائية ، وتوضيح كيفية تطبيق هذه المفاهيم والاستفادة منهما في تقويم نتائج الأبحاث بصورة منطقية يمكن الاعتماد فيها في اتخاذ قرارات مناسبة حيال الظاهرة المدروسة .

أهمية الدراسة : تتضح أهمية الدراسة فيما يلي :

1. تناولها للمفاهيم التالية : الدلالة العملية (حجم التأثير) والدلالة الإحصائية ، وقوة الاختبار الإحصائي.

2. تكشف الدراسة عن جوانب القوة أو القصور في بحوث الماجستير والدكتوراه الجزائرية في البحث التربوي والنفسي ، وتوجه نظر الباحثين لجوانب القوة للاستفادة منها ، وجوانب القصور لتلافيها.

منهجية الدراسة : اتبع الباحث منهجاً علميا تكامليا يجمع بين التحليل والتقويم ، وان لهذا النوع من البحوث أهمية في إثارة الوعي بقضايا ومشكلات البحث العلمي.

الجانب النظري :

مفهوم الدلالة الإحصائية:

أشار البهي (1971) إلى أن الدلالة الإحصائية هي درجة اقتراب قمة مقاييس العينات من مقاييس المجتمع الأصلي. ويمكن اكتشافها بواسطة الاختبار الإحصائي وهو مجموعة من القواعد تمكن الباحث من رفض أو قبول الفرض الإحصائي و بموجبه يمكن الحكم على الفرض الإحصائي (الصياد،1990). كما أشار كاشن و جيجر(cachen and geiger,2004 :12) إلى أن الاستدلال الإحصائي الكلاسيكي يختبر الفرض

الصفري H_0(Null.hypothesis) أي أنه لا توجد أي فروق بين المعالم قيد الاختبار، أو أن الفروق يمكن أن تهمل، و التي يقارنها الباحثون بالفرض البديل H_1(Altermative.hypotheisis) وهي تعني أن الظاهرة قيد التحقيق حاضرة ، أو أنه توجد فروق بين المعالم قيد الاختبار ، ولأن الباحثون عادة ما يأملون في رفض الفرض العدمي ، فإنهم عادة ما يرصدون الفرض خاطئ. ومما سبق يلاحظ أن الدلالة الإحصائية يتم قياسها عن طريق العديد من اختبارات الفروض التي تقوم على اختبار الفرض العدمي H_0 ، ضد الفرض البديلH_1 ، والتي يعمم الباحث فيها رفض الفرض العدمي ، حيث أن ذلك دلالة على صحة توجهه ، وذلك طبقاً لمستوى الدلالة (α) ، وتوزيع المعاينة المناسب، وتحديد القيمة الحرجة(Critical Value) على ذلك التوزيع الاحتمالي ، والتي تفصل بين منطقتي الرفض والقبول ، ثم بمقارنة القيمة الحرجة مع القيمة المحسوبة بين الاختبار الإحصائي تأخذ القرار إما برفض الفرض العدمي وبالتالي قبول الفرض البديل ، أو العكس(باهي،2010: 14-15). ويمكن تلخيص صحة وخطأ القرارات في اختبارات الفروض الإحصائية بالجدول ، الآتي :

جدول (01) يوضح صحة وخطأ القرار الاختبار الفروض.

القرار الإحصائي / الفرض	قبول H_0	رفض H_0
H_0 صحيح	1−α قرار صحيح	خطأ من النوع الأول باحتمالα
H_1 خطأ	خطأ من الثاني باحتمالβ	β − 1 قرار صحيح

مفهوم حجم التأثير:

حجم التأثير هو الذي يقيس قوة العلاقة (التلازم) بين المتغيرات الموجودة في الدراسة ، وقد أوضح هايز أن مستوى الدلالة الإحصائية بمفردها لا تخبرنا بشيء عن قوة التلازم بين المتغيرات ، وكما قال ثومبستون بان حجم التأثير هو الذي يوجه التفسير نحو ما هو حقيقة في البحث وجلب الانتباه لقضية جدارة النتائج (wilkinson,1992) .ويرى كيلو kellow(1998) أن أول ما يميز حجوم التأثير هو إمكانية تفسيرها بشكل مستقل عن حجم العينة ، وهي مفيدة في تحديد المقدار الفعلي للفروق بين متوسطات المجموعات أو درجات التلازم بين المجموعات ، ولذا فإنها تزود الباحثين بمعلومات تتجاوز حدود اختبارات الدلالة الإحصائية ، وتمنحهم فهما واضحا للبيانات التي بين أيديهم . وقد أوضح ماكلين McClain(2000) فائدة حجم التأثير في منع سوء تفسير الاختبارات الإحصائية بقوله "حجوم التأثير تمنح صورة أوضح للعلاقات الموجودة في البيانات ، هذه المؤشرات ليست ضبابية ولا محجوبة بحجم العينة مثل الإحصاءات الناتجة من اختبارات الدلالة

الإحصائية... وبحق نشعر بان فهم واستخدام مؤشرات حجم التأثير يمنع الاختبارات الإحصائية من سوء التفسير لمؤشرات لها أهمية... وبالتأكيد البحث النفسي يمكن أن يفسر بطريقة أفضل بإستخدام حجوم التأثير "

ولقد أوجز هوستون(1993) فوائد مقاييس حجم التأثير على النحو التالي :

1. حجم التأثير يشير إلى درجة وجود الظاهرة في المجتمع بمقياس متصل ، بحيث يعني الصفر عدم وجود الظاهرة.
2. تزود الباحثين بمؤشرات للدلالة العملية بخلاف اختبارات الدلالة الإحصائية.
3. يمكن استخدامها في المقارنة الكمية بين نتائج دراستين أو أكثر كما هو مستخدم في التحليل البعدي.
4. يمكن استخدامها في تحليل القوة الإحصائية لتحديد كم عدد العناصر المطلوبة في دراسة معينة.

ولاستخدام مؤشرات الدلالة العملية (حجم التأثير) المرافقة لاختبارات الدلالة الإحصائية الشائعة يمكن الرجوع إلى كتابي كوهين Cohen(1988) والقريشي (1421) حيث عرض أهم المؤشرات ومنها :

1. مؤشرات حجم التأثير (d) لاختبار (ت) للفروق بين المتوسطات:

قدم كوهين Cohen(1988) مؤشر (d) لحجم التأثير لنتائج اختبار (ت) حسب المعادلة التالية :

$$d = \frac{|\mu 1+\mu 2|}{\sigma}$$

حيث : (d) هو مؤشر حجم التأثير.

$|\mu 1 + \mu 2|$ هو الفرق بين متوسطي العينتين.

σ هو الانحراف المعياري لإحدى العينتين (بافتراض تساويهما).

أما إذا كان الانحراف المعياري للعينتين غير متساويتين ، فإن قيمة الانحراف المعياري المستخدم في المعادلة السابقة يحسب كالآتي : $\sigma = \sqrt{\frac{(\sigma 1)^2+(\sigma 2)^2}{2}}$

حيث : $(\sigma 1)$ و $(\sigma 2)$ هما الانحراف المعياري للعينتين (1) و(2) .

ويمكن الحصول على المؤشر (d) من نتيجة اختبار (ت) على النحو التالي : $d = 2t\sqrt{df\sigma}$

2. مؤشر حجم التأثير (f) لاختبار تحليل التباين (ف): $f = \sqrt{\frac{\eta^2}{\eta^2-1}}$

حيث : (η^2) هي مربع إيتا ،وتحسب كالآتي : $\eta^2 = \frac{SSbetween}{SStotal}$

حيث : SSbetween هو مجموع المربعات بين المجموعات

SSbetween هو مجموع المربعات الكلي (Cohen,1988)

3. مؤشر حجم التأثير (ω) لاختبار مربع كاي (χ^2): $\omega = \sqrt{\frac{\chi 2}{n}}$

حيث : (n) هي حجم العينة (Cohen,1988)

4. مؤشر حجم التأثير الخاص بمعامل ارتباط بيرسون (r): مؤشر حجم التأثير الخاص بمعامل ارتباط بيرسون (r) وهو (r) نفسه ، ويستخدم للارتباط بين متغيرين متصلين ، كما يعبر مربع معامل ارتباط بيرسون (r^2) عن نسبة التباين المشترك الذي يمكن أن يرجع إلى أحد المتغيرين (القريشي،1421: 351).

كما أن كوهن قد وضع جداول خاصة لاستخراج جميع المؤشرات السابقة ومؤشرات أخرى أيضاً (Cohen,1988) ، وهي كالآتي :

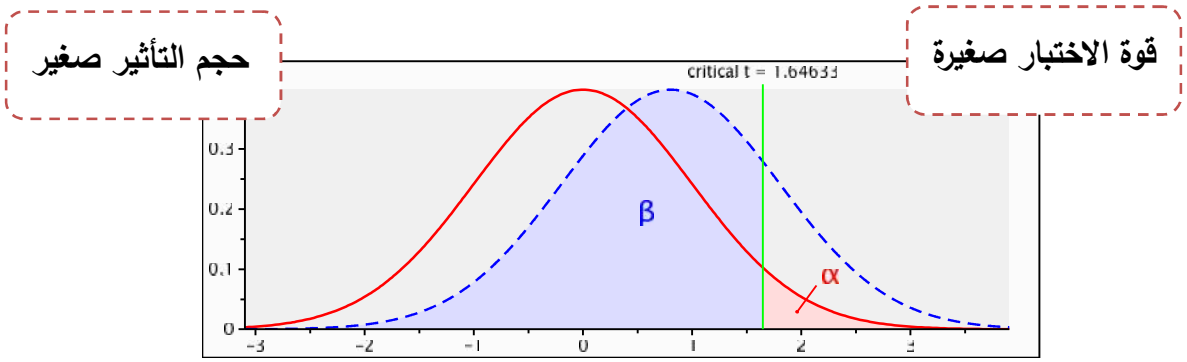
الاختبارات الإحصائية ومؤشرات حجوم التأثير ومدى كل مؤشر

اختبار (t) مؤشر d لكوهين	اختبار (f) مؤشر η^2 مربع إيتا لفريدمان	اختبار (χ^2) مؤشر ω أوميغا	اختبار (r) مؤشر r^2 مربع آر
0.2–0.49 ضعيف	0.059–0.01 ضعيف	0.01–0.29 ضعيف	0.01–0.29 ضعيف
0.5 –0.79 متوسط	0.06–0.139 متوسط	0.3 –0.49 متوسط	0.3–0.49 متوسط
0.8 فأكثر	0.14 فأكثر كبير	0.5 فأكثر كبير	0.5 فأكثر كبير

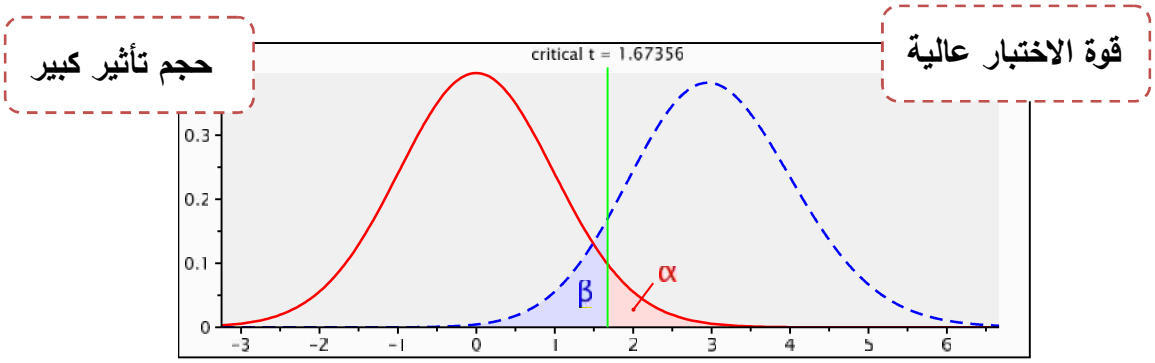
مفهوم القوة الإحصائية:

تحليل القوة الإحصائية هو احد الاختبارات المكملة لاختبارات الدلالة الإحصائية، وهو يقدر احتمالية الخطأ من النوع (II) عندما تكونH_0 (عدم رفض الفرضية الصفرية عندما تكون خاطئة): (wilkimsoom 1992) وقد عرفها هوستون بأنها (احتمال أن يرفض الاختبار الفرضية الصفرية بطريقة صحيحة). وقد لاحظ كوهين سنة(1988) و شيفر (1980) أن هذا التوجه نادراً ما يستخدم من قبل الباحثين وهو تحليل يتضمن التعامل مع α وβ، وحجم العينة n وقد لاحظ ثومبسون (1987) أن القوة إذا كانت عالية، فإن النتائج غير الدالة إحصائيا تسهم في أساس المعرفة، وكما يرى بلاي (1973) بأن أي باحث يقبل الفرضية الصفرية بدون أن يعرف قوة الاختبار الإحصائي فإنه يكون مسؤولاً عن حدوث نسبة خطا كبير من النوع (II) وأن الباحثين الذين يؤيدون وبطريقة غير صحيحة قبول الفرضية الصفرية إذا لم يلاحظوا أي دلالة إحصائية دون اعتبار للقوة الإحصائية يسيئون قيادة القراءة (wilkimsoom1992). ويرى بارنيت(1998) بأنه إذا كان لدينا اختبار

منخفض للقوة فإنه لن يمكننا تحسس والتقاط الدلالة الإحصائية، بينما إذا كان لدينا اختبار عالي القوة فإن حجم الفروق بغض النظر عن صغرها سوف تكون دالة إحصائيا، ويعتقد الكثيرون أن التركيز على التحكم في الخطأ من النوع(I) المستعمل في الأساليب الشائعة، ربما يكون السبب وراء رؤية العديد من الدراسات الإحصائية التي تعاني من نقص في القوة وعلى العموم فإن التحكم في الخطأ من النوع(I) ما يجعل الباحثين يدفعون الثمن وهو العجز في التحكم في الخطأ من النوع(II) أي النقص في قوة الاختبار ، مما يعني الفرصة الأقل للحصول على نتائج دالة إحصائيا (أشرف بن إدريس،2015: 128).



الشكل رقم (01): مستقطع من برنامج G*POWER يوضح العلاقة ما بين قوة الاختبار الصغيرة وحجم التأثير الصغير.



الشكل رقم (02): مستقطع من برنامج G*POWER يوضح العلاقة ما بين قوة الاختبار العالية وحجم التأثير الكبيرة.

الجانب الميداني :

النتائج :

السؤال الأول : ما نسبة استخدام مفهوم حجم التأثير(الدلالة العملية) إلى جانب الدلالة الإحصائية بالنسبة للاختبار الإحصائي " ت" في بعض بحوث الماجستير والدكتوراه الجزائرية ما بين 2005-2015؟

من خلال مراجعة البيانات الخاصة بالدراسة لوحظ انه تم حساب حجم الأثر (الدلالة العملية) إلى جانب الدلالة الإحصائية في عدد (16) فرضية مستخدمة في بحوث الماجستير والدكتوراه فقط بنسبة (11.34%) من إجمالي نسبة الفرضيات التي اعتمادها في الدراسة الحالية ، والجدول التالي يوضح البيانات الخاصة بالاختبار الإحصائي " ت ".

جدول رقم (02)

تكرار الدلالة الإحصائية وحجم الأثر (الدلالة العملية) بالنسبة للاختبار الإحصائي "ت" في الفرضيات المعتمدة في الدراسة الحالية.

العدد الكلي للفروض التي استخدمت الاختبار الإحصائي "ت".		عدد الفروض التي حسبت لها الدلالة الإحصائية فقط.		عدد الفروض التي حسبت لها الدلالة العملية (حجم الأثر) فقط.		عدد الفروض الإحصائية التي حسبت فيها الدلالة الإحصائية إلى جانب حجم الأثر (الدلالة العملية)	
التكرار	النسبة	التكرار	النسبة	التكرار	النسبة	التكرار	النسبة
141	%100	125	%88.65	0	%0	16	%11.34

من الجدول رقم (02) يتضح أنه ، من تكرار قدره (141) فرضية استخدمت اختبار "ت" ، وهو العدد الكلي للفرضيات المعتمدة في هذه الدراسة قام الباحثون بحساب الدلالة الإحصائية جنباً إلى جنب مع الدلالة العملية (حجم التأثير) بتكرار ضعيف جداً قدره (16) فرضية ونسبة مئوية قدرها (11.34%) ، بينما حسبت الدلالة الإحصائية فقط دون التطرق إلى الدلالة العملية (حجم التأثير) في بقية الفرضيات بتكرار (125) فرضية ونسبة (88.65%) ، كما لوحظ أنه لم تحسب الدلالة العملية (حجم التأثير) مفردة بأي تكرار.

السؤال الثاني : ما نسبة الفرضيات التي استخدمت الاختبار الإحصائي " ت" والتي اتفقت فيها نتائج الدلالة الإحصائية مع حجم الأثر (الدلالة العملية) في بعض بحوث الماجستير والدكتوراه الجزائرية ما بين2005-2015؟

وللإجابة على هذا التساؤل سنقوم بعرض الاتفاق والاختلاف بين نتائج الدلالة الإحصائية وحجم الأثر (الدلالة العملية) بالنسبة للاختبار الإحصائي "ت" المستخدم في بعض بحوث الماجستير والدكتوراه.

جدول رقم (03)

بيانات الدلالة الإحصائية ، وحجم الأثر (الدلالة العملية) الخاصة بالاختبار الإحصائي "ت" للفروق بين متوسطي عينتين مستقلتين.

دلالة إحصائية / حجم الأثر	دالة	النسبة	غير دالة	النسبة	إجمالي	النسبة
ضعيف	06	%8.57	64	%91.47	70	%49.64
متوسط	05	%12.19	36	%87.80	41	%29.07
قوي	22	%73.33	08	%26.66	30	%21.27
إجمالي	33	%23.40	108	%76.59	141	%100

من الجدول رقم (03) ليتضح أن الدلالة الإحصائية لاختبار "ت" للفرق بين متوسطي عينتين مستقلتين اتفقت مع الدلالة العملية (حجم التأثير) في (33) فرضية استخدمت اختبار "ت" بنسبة (23.40%) ، حيث كانت نتيجة (64) فرضية غير دالة إحصائيا مع الدلالة عملية (حجم التأثير) ضعيف ونتيجة (27) فرضية منها دالة إحصائيا مع دلالة عملية قوية أو متوسطة.

بينما اختلفت في (06) فرضيات بنسبة (8.57%) حيث كانت دالة إحصائيا ودلالتها العملية (حجم التأثير)ضعيف وفي (36) فرضية غير دالة إحصائيا مع دلالة عملية متوسطة ، وفي (8) فرضيات غير دالة إحصائيا مع دلالة عملية قوية.

السؤال الثالث : ما النسبة التراكمية لقوة الاختبار الإحصائي " ت" المستخدم في بعض بحوث الماجستير والدكتوراه الجزائرية ما بين 2005-2015؟

جدول رقم (04)

يوضح التكرار والنسبة التراكمية لقوة الاختبار الإحصائي "ت" المستخدم في بعض بحوث الماجستير والدكتوراه.

فئات قوة الاختبار	مستويات حجم التأثير					
	حجم أثر منخفض		حجم أثر متوسط		حجم أثر كبير	
	التكرار	النسبة التراكمية	التكرار	النسبة التراكمية	التكرار	النسبة التراكمية
صفر- 0.5	90	%80.35	02	%16.66	00	%00
0.51 - 0.79	13	%11.60	04	%33.33	01	%5.88
0.80- 1	9	%08.03	06	%50	16	%94.11
المجموع	112	%100	12	%100	17	%100
متوسط قوة الاختبار	0.18		0.64		0.93	

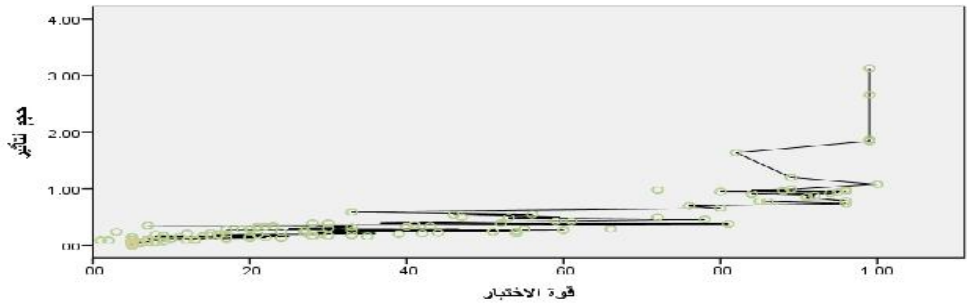
يتضح لنا من الجدول رقم (04) أن قوة الاختبار الإحصائي تختلف وفقا لاختلاف حجم التأثير ، ففي حالة حجم التأثير المنخفض Small Effect فإن متوسط قوة الاختبار الإحصائي 0.18 مما يدل على أنه في حالة افتراض حجم تأثير منخفض لظاهرة ما فإن قوة الاختبار الإحصائي لن تتجاوز 0.18 أي أن فرصة رفض الفرضية موضع البحث عندما تكون في حقيقة الأمر خاطئة لن تتجاوز 1 من 5.

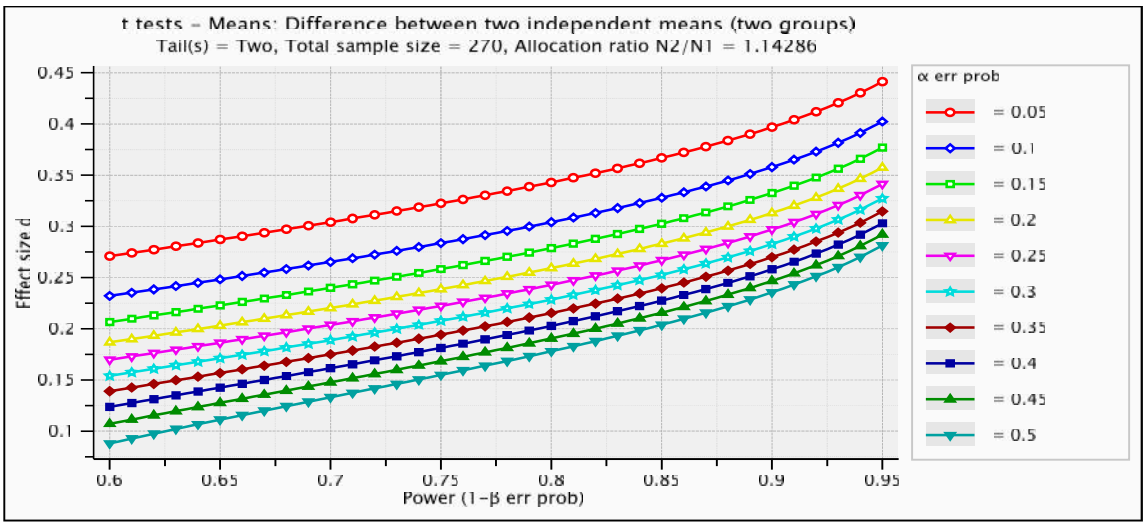
أما في حالة افتراضنا حجم تأثير متوسط Medium Effect فإن متوسط قوة الاختبار سيرتفع ويصل إلى 0.64 مقارنة بـ0.18 والتي تم الحصول عليها في حالة حجم التأثير المنخفض ، وهذا بالتالي يرفع احتمالية رفض الفرضية الصفرية عندما تكون في حقيقة الأمر خاطئة إلى أكثر من 2 إلى 3.

أما في حالة افتراضنا حجم تأثير كبير Large Effect فإن متوسط قوة الاختبار سيرتفع بشكل واضح وتصل إلى 0.93. نلاحظ من خلال هذه القيم أن هناك علاقة طردية بين قوة الاختبار وحجم التأثير فكلما ارتفعت قيمة مستوى التأثير ارتفعت قيمة قوة الاختبار وهذا يشير بطبيعة الحال إلى أن ارتفاع حجم التأثير (الدلالة العملية) عامل مؤثر وهام في ارتفاع قوة الاختبار و انخفاض لنسبة التباين غير المفسر (تباين الخطأ) ، وهذه النتيجة تتفق مع ما توصل إليه (Brewer,1972) من أن قوة الاختبار الإحصائي مرتبطة ارتباطا وثيقا بحجم التأثير ، وكذلك تتفق مع ما توصل إليه (Ottenbacher,1982) والذي أشار أن الاختبارات التي يكون فيها حجم التأثير منخفضا يكون احتمال شمولها على خطأ من النوع الثاني (β) كبيرا ، وأن هذا الاقتران يكون أكثر حدوثا مع حجم التأثير المنخفض من حدوثه مع حجم التأثير المتوسط أو الكبير.

ولمعرفة هل هناك علاقة ذات دلالة إحصائية بين الاتفاق من عدمه بين نتائج قوة الاختبار الإحصائي والدلالة العملية (حجم التأثير) لاختبار "ت" للفرق بين متوسطي عينتين مستقلتين في هذه الدراسة ، استخدم الباحث معامل الارتباط "بيرسون" حيث أشارت النتائج إلى وجود علاقة ارتباطية طردية موجبة بين نتائج قوة الاختبار الإحصائي والدلالة العملية (حجم التأثير) لاختبار "ت" للفرق بين عينتين مستقلتين بقيمة قدرها (0.80) وهي دالة إحصائيا عند مستوى دلالة (0.01) مع حجم تأثير قدره (0.64) حيث تدل هذه النتيجة على دلالة عملية (حجم تأثير) قوي. والشكلان 3 و4 التاليان يوضح هذه العلاقة :

شكل رقم (03)





شكل رقم (04): مستقطع من برنامج G*POWER يوضح العلاقة الطردية بين قوة الاختبار الإحصائي وحجم التأثير.

توصيات :

1. منح المزيد من الاهتمام بالدلالة العملية (حجم التأثير) وقوة الاختبارات في البحوث التربوية والنفسية المحلية والعربية.

2. تحديد قوة الاختبار الإحصائي بحيث لا تقل عن 0.80 مع مستوى دلالة وحجم تأثير عالي ثم البحث عن أدنى حجم لعينة الدراسة.

3. يجب على الباحثين الإشارة إلى مقدار حجم الأثر وقوة الاختبار وذلك لفهم نتائج دراساتهم بصورة أفضل والاعتماد على هذه النتائج لبناء أبحاث وفق أسس نظرية سليمة.

4. نشر الثقافة الإحصائية في أوساط الباحثين من خلال إعطاء دورات إلزامية للأساتذة في مجال التحليلات الإحصائية.

5. عرض البحوث ورسائل الماجستير والدكتوراه على متخصصين في الإحصاء أو القياس والتقويم.

خاتمة :

بعد تحليل وتقويم بعض بحوث الماجستير والدكتوراه الجزائرية2005-2015 من خلال فحص (141) فرضية استخدمت اختبار "ت" وتقصي مدى استخدام هذه البحوث للدلالة الإحصائية مقابل الدلالة العملية (حجم التأثير) وقوة الاختبار ، حيث تكمن قوة الدراسة الحالية كما يراها الباحث في الموضوع ذاته وماله من تطبيقات

وتأثيرات على واقع ومستقبل البحث النفسي والتربوي بالجزائر ، ويأمل الباحث أن تكون معالجته للموضوع وما توصل إليه من نتائج تتناسب مع خطورته وأهميته.

المراجع :

1. أشرف بن إدريس، و احمد محمد عبد الله .(2015). تأثير حجم العينة و توزيع المجتمع على قوة الاختبار (t) للعينتين المستقلتين و المرتبطتين .

SUST.journal of Natural and Medical Siences (jNMs)vol _16(2)
http://journal-sutech.edu/

2. باهي، مصطفى حسين.(2010). العلاقة بين الدلالة الإحصائية وحجم التأثير في البحوث التربوية النفسية ، القاهرة ، المؤتمر العلمي السادس عشر لإعداد المعلم وتنميته.

3. بابطين ، عادل أحمد .(2001). مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة ، رسالة ماجستير غير منشورة ، كلية التربية ، قسم علم النفس ، جامعة أم القرى.

4. الشربيني، زكريا.(1995). الإحصاء وتصميم التجارب في البحوث النفسية والتربوية الاجتماعية ، القاهرة : الأنجلو مصرية.

5. طلال هياز ، حسن الباقري.(2012). واقع الدلالة الإحصائية والدلالة العملية للبحوث المنشورة بمجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية في المدة من 1425– 1430هـ.

6. القريشي ، عبد الفتاح إبراهيم.(1421هـ). تصميم البحوث في العلوم السلوكية ، الكويت : دار القلم.

7. Cohen,jacob.(1988). Statistical Power Analysis for the Behavioral Sciences .New york :Academic Press.

8. Kirk ,Roger.E.(1996).Practical Significance Aconcepet whose Time hascome. Educational and Psychological Measurment.V56 ,n5,746-759.

9. McClaien,james et al .(2000).Are ALL Effect Sizes.Created Equal .paper presented at the Annual Meeting of Mid-South. Educational Research Association (28th , Bowl ing green ,Ky,Nevember.15-17)

10. McClaien and Ernest .(1998). The Role of statistical Significance Testing in Educati onal Research in the Schools .v5 n2,15-22.

11.Ronald,w and Nicole,L.(2016).The ASA’s ststement on P-value : context ,process ,and propose . The American Statisticion (online) journal .pp01-17.
hhtp://anstat.tandfonline.com/lai /utas20.