مجلة دراسات إقتصادية ISSN 1112 -7988 / E-ISSN 8627-2710
المجلد: 21 العدد: 01 السنة: 2021 ص ص: 503-522

أثر حجم العينة على القيمة الاحتمالية P-value وأهمية تضمين حجم الأثر في الاختبارات الاحصائية

The Effect of Sample Size on the Probability Value and The Importance Of Including The Effect Size In The Statistical Tests

محمد روابة*، جامعة ابن خلدون تيارت، mohammed.rouaba@univ-tiaret.dz

تاريخ الارسال: 2021/05/07	تاريخ القبول: 2021/06/02	تاريخ النشر: 2021/06/30	المؤلف المرسل: محمد روابة

الملخص:

هدفت هذه الدراسة إلى تقديم مفهوم "حجم الأثر" كأسلوب احصائي مكمل لإختبار الفرضيات الاحصائية. وتضمنت الدراسة تحليلا لبرهنة أن حجم الأثر أقل تأثرا بحجم العينة من الأساليب الاحصائية الأخرى التي تستخدم لإختبار الفرضيات الاحصائية، بالإضافة إلى ذلك قدمت الدراسة بعض المؤشرات الاحصائية التي تستخدم للدلالة على قيمة حجم الأثر في حالة استخدام بعض الأساليب الاحصائية. حيث تم التركيز من خلال هذه الورقة البحثية على دراسة تأثير حجم العينة على القيمة الاحتمالية P-value التي على أساسها يتم اتخاذ القرار بعدم رفض الفرض الصفري أو قبوله بحيث تم توليد 400 مفردة عشوائيا من خلال برنامج (PASS V16) واختيرت عينات عشوائية بأحجام مختلفة، وتوصلت الدراسة إلى أن هناك تأثير كبير لحجم العينة على القيمة الاحتمالية P-value، حيث عند نفس المستوى من الفروق بين متوسطي العينتين تكون الفروق غير دالة احصائيا عند حجم العينات الصغير بينما تكون نفس هذه الفروق دالة احصائيا عندما يكون حجم العينة كبيرا.

الكلمات المفتاحية: الدلالة الاحصائية، الدلالة العملية، حجم الأثر، قوة الاختبار الاحصائي.

Abstract :

The aim of this study is to present the concept of "Effect size" as a statistical method to supplement the testing of statistical hypotheses. The study included an analysis to demonstrate that the effect size is less affected by the size of the sample than the other statistical methods used to test the statistical hypotheses. The focus of this research paper is on the study of the effect of sample size on the probability value

* **المؤلف المرسل:** محمد روابة

on which the decision is made not reject the null hypothesis or accept it. 400 words have randomly been created using Pass v16 software program. Random samples of different sizes have been used, and the researcher has concluded that the sample size affecting the probability value, where at the same level of differences between the two averages the differences are not statistically significant at the size of the small samples while the same differences are statistically significant when the size of the sample is large.

Keywords: statistical significance, practical significance, effect size, power of statistical test.

1. مقدمة:

بعد أن يتم إخضاع الفرضية الصفرية للإختبار يتخذ الباحث قرار بحق الفرضية الصفرية، وهذا القرار يكون إما برفضها أو بعدم رفضها، ويتم اتخاذ القرار في ضوء الاختبار الاحصائي المستخدم، بإعتباره قانونا أو قاعدة تربط إحصائية العينة مع معلمة المجتمع، ومن أمثلة الاختبارات الاحصائية المستخدمة اختبارz ، اختبار t، اختبار χ^2 ...إلخ، وفي الواقع لا يستطيع الباحث التأكد فيما إذا كان القرار الذي اتخذه صائبا أم غير صائب، وكل ما يعرفه هو احتمال صواب هذا القرار، فإذا كان الفرض صحيحا، و لم يتمكن من رفضه (مستوى الثقة) أو كان خاطئا بالفعل ورفضه الباحث (قوة الاختبار)، فإنه في كلتا الحالتين يكون قد اتخذ قرار صائبا، وإذا رفض القرار وهو في الواقع صحيح (الخطأ من النوع الأول)، أو لم يتمكن من رفضه، في حين أنه في الواقع خاطئ (الخطأ من النوع الثاني)، عندئذ يكون الباحث قد وقع في خطأ، لذا فإنه أيا كان القرار هناك درجة من الخطأ، وينبغي ألا ينظر إلى اختبار الفرضيات باعتباره برهنة لها.

قد تشير نتائج اختبار الفروض إلى وجود فروق ذات دلالة إحصائية بين متوسطات المعالجة مثلا. ولا تعزى هذه الفروق إلى الصدفة أو خطأ المعاينة، أي أن المتغير المستقل له أثر في المتغير التابع، ولكن يبقى السؤال، ماهو حجم هذا التأثير؟ هل هو قوي أو صغير؟ فأحيانا إذا كانت قيمة الاختبار الاحصائي ذات دلالة عند احتمال 0.05 توصف بأنها ذات دلالة فقط، ولكن إذا كانت ذا دلالة عند 0.01 توصف بأنها ذات دلالة قوية، وقد يوحي ذلك بأن تأثير المعالجة أكبر، إلا أن هذا التفسير غير صحيح، فإختبار الدلالة الاحصائية ليس مقياسا ملائما لقياس تأثير المعالجة لأن الدلالة الاحصائية لها علاقة بحجم العينة، إذ أنه بزيادة حجم العينة بشكل كبير تصبح إمكانية رفض الفرضية الصفرية أكبر، لذا فإنه ومن أجل إيجاد قوة العلاقة بين المتغير المستقل والمتغير التابع أو من أجل إيجاد حجم الفروق في المتغير التابع الناتجة عن اختلاف مستويات المتغير المستقل، أو مدى مساهمة المتغير المستقل في التباين في المتغير التابع يتم اللجوء إلى ما يسمى بالدلالة العملية.

وبناء على ذلك فإن هذه الدراسة تحاول الاجابة على التساؤل التالي:

هل يؤثر حجم العينة على مستوى القيمة الاحتمالية P-value لإختبار t ؟

1.1. فرضيات البحث:

- يؤثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينة واحدة.
- يؤثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينتين مستقلتين.
- يؤثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينتين مرتبطتين.

2.1. أهداف الدراسة: تهدف هذه الدراسة إلى:

- الكشف عن أهم مشاكل الدلالة الاحصائية في البحوث.
- التعرف على أهم المفاهيم التي يمكن أن تقدم حلولا لمشاكل الدلالة الاحصائية في البحوث.
- دراسة أثر حجم العينة على مستوى القيمة الاحتمالية P-value لإختبار t لعينة واحدة، ولعينتين مستقلتين وكذلك لعينتين مرتبطتين.

2.1. أهمية الدراسة: من الملاحظ أن كثيرا من الباحثين يعتمدون في تقرير نتائج دراساتهم على الدلالة الاحصائية دون محاولة تحديد مقدار العلاقة القائمة بين المتغيرات (المستقلة والتابعة)، الأمر الذي يجعلهم في بعض الأحيان تفسير النتائج الدالة احصائيا بطريقة خاطئة على الرغم من أنها قد لا تكون لها قيمة من الناحية العملية، لذلك تتضح أهمية البحث الحالي من خلال إلقاء المزيد من الضوء على أهمية حجم الأثر كمحك عملي للدلالة وعدم الاكتفاء بالدلالة الاحصائية فقط، وهو من الموضوعات المهمة للوقوف على أهمية النتائج المستخلصة من البحوث.

3.1. مجتمع الدراسة: يتكون مجتمع الدراسة الحالية من عدد N=400 مشاهدة، حيث تم توليدها بإستخدام برنامج PASS V16، وتم التركيز على حجم العينة باعتباره عاملا مؤثرا في مستوى القيمة الاحتمالية P-value، كما تم توليد البيانات عشوائيا بوسط حسابي وانحراف معياري ثابت، وهذه البيانات تتوافر فيها مجموع الافتراضات اللازمة لإجراء التحليلات الاحصائية المطلوبة، سواء فيما يتعلق بالاختيار العشوائي، مستوى القياس الفئوي، التوزيع الطبيعي، تجانس التباين.

4.1. الدراسات السابقة: هناك العديد من الدراسات التي أشارت إلى أهمية استخدام حجم الأثر، ومن بين هذه الدراسات دراسة Thompson,1999 بعنوان "تحسين وتوضيح نتائج الأبحاث، وفائدة مؤشرات حجم الأثر كاختبارات مكملة لاختبارات الدلالة الاحصائية"، وتوصل من خلالها إلى ضرورة تضمين الدراسات لمؤشرات حجم الأثر وفقا لتعليمات الجمعية النفسية الأمريكية APA. 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 الاحصائية وهي التفسير الخاطئ للدلالة الاحصائية وما يتعلق بها من فشل وتفسير حجم الأثر الممثل في جميع التحاليل الكمية.

أما في دراسة قام بها Portillo,2001 بعنوان "اختبارات الدلالة الاحصائية وقياسات حجم الأثر" ، أوصى من خلالها بأن الأبحاث يجب أن تتضمن الدلالة الاحصائية وقياسات حجم الأثر، وأوصت الدراسة أيضا بأن يشترك الباحثين من أجل الحوار حول كلا من الدلالة الاحصائية والدلالة العملية والمقارنة بينهما.

وفي دراسة أجراها King,2002 بعنوان "الدلالة الاحصائية ضد الدلالة العملية (آثار لتحليل بيانات التعليم الطبي المستمر)" فقد أوصت بتطبيق اختبارات الدلالة الاحصائية بشكل مدروس، وليس بطريقة ميكانيكية، وبفهم كامل لقوة كل منها وضعفها، وببيان قياس حجم الأثر مع فترة الثقة المناسبة له، ومحاولة تقييم الدلالة العملية للحصول على النتائج المناسبة (King, 2002, p. 1).

2. الدلالة الاحصائية:

لقد ظهر مفهوم الدلالة الاحصائية منذ ثلاثمائة عام تقريبا ولكن تم استخدامه في بداية القرن العشرين بصورة أكثر كثافة، وخاصة عند استخدام الاختبارات الاحصائية مثل اختبار χ^2 واختبار t ، واختبار تحليل التباين ANOVA. ولقد انتشر استخدام أسلوب فحص الدلالة الاحصائية للفرضيات الصفرية في الآونة الأخيرة بصورة متزايدة، وكان الهدف من ذلك هو اتخاذ قرار فيما إذا كانت النتائج التي تم التوصل إليها دالة احصائيا أم غير دالة، ويعتمد هذا الأسلوب على صياغة فرضية صفرية تبنى حول مجتمع الدراسة ويتم اختبارها من خلال بيانات عينة من المفترض أنها تتشابه في خصائصها مع خصائص ذلك المجتمع. ويستخدم مفهوم مستوى الدلالة الاحصائية لإختبار فيما إذا كانت الفروق بين متوسطات مجموعتين دالة احصائيا أم لا (نصار، 2017، صفحة 356).

تستخدم اختبارات الدلالة الإحصائية للاستدلال على خصائص المجتمع من خلال العينة المأخوذة منه، وتسمح تلك الاختبارات للفرد بحساب احتمالية الحصول على نتائج العينة بافتراض أن الفرض الصفري صحيحا (أي افتراض أن المجتمع يتصف بما هو مفترض في الفرض الصفري) .

واختبارات الدلالة الإحصائية تقدر قيمة احتمالية P-value انحراف نتائج العينة عن المحدد بالفرض الصفري للمجتمع، عند حجم عينة معطى، وبعبارة أخرى هذه الاختبارات لا تعمل على تقدير احتمالية أن نتائج العينة تصف بدقة المجتمع الاحصائي، وبدلا من ذلك تفترض الاختبارات الاحصائية أن الفرضية الصفرية تصف بدقة المجتمع الاحصائي ومن ثم تختبر احتمال العينة، بالطبع هذا المنطق معقد بعض الشيء لأنه لا يخبرنا بما نريد معرفته عن معالم المجتمع واحتمالية تكرار النتائج للعينات المستقبلية المسحوبة من نفس المجتمع الاحصائي، كما تؤكد الدراسات التجريبية بأن العديد من مستخدمي اختبارات الدلالة الإحصائية في الواقع لا يفهمون ما الذي تقوم به هذه الاختبارات الاحصائية فعلا، فبعض مستخدميها يقلصون دورها إلى مجرد قول (عند مستوى سطحي)، أن هذه الاختبارات تقيم ما إذا كانت النتائج ناشئة عن الصدفة أم لا (Thompson, 2002, p. 65).

1.2. حدود استخدام الدلالة الاحصائية:

تاريخيا ساعدت اختبارات الدلالة الاحصائية التقليدية الباحثين في فهم معنى بياناتهم، حيث اعتبرت حجر الزاوية في البحوث الاجتماعية والانسانية والبيولوجية، لكن ثبت أن الاعتماد عليها بمفردها ينطوي على أخطاء،

فخلال عقود تبين أن هذه الاختبارات غالبا ما يساء استعمالها، وأن تلك الاختبارات محدودة المساعدة، فعلى سبيل المثال أغلب التساؤلات البحثية الأساسية تتضمن أسئلة حجم تأثير المتغير المستقل في المتغير التابع، لكن اختبارات الدلالة الاحصائية تخفق في الاجابة عن تلك التساؤلات.

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

ناقش Hopkins, (2011) مفهوم مستويات الدلالة الإحصائية وتطورها بقوله: ما هو الخاص في القيمة P-value=0.05 ؟ في الحقيقة لا شيئ. شخص ما قرر أن هذه القيمة معقولة، ولذا فإننا الآن جامدون عندها. كما أصبحت P-value<0.01 أيضا تقليدا لتحديد الدلالة الاحصائية، فكلا القيمتين من الآثار القديمة للأيام التي سبقت استخدام الحاسب الآلي. عندما كان من الصعب حساب القيمة الدقيقة لـــلقيمة الاحتمالية P-value لإحصاءة الاختبار. وبدلا من ذلك استخدم الباحثين الجداول لقيم احصاءة الاختبار المقابلة لقيم قليلة وعشوائية لـــلقيمة الاحتمالية P-value وهي 0.1، 0.05، 0.01 وفي بعض الأوقات 0.001 (حمدي، 2013، الصفحات 353-354).

لقد أصبحت هذه القيم معتمدة كعتبة حدية للقيم التي تحدد الدلالة الاحصائية ثم اختبار الدلالة الاحصائية. وعادة ما يكون على الباحث أن ينص على أي واحدة يستخدم. فعلى سبيل المثال إذا نص الباحث على أن مستوى الدلالة هو 0.05 إذا يسمح له بإعتبار أي نتيجة لها قيمة P-value تساوي أو أقل من 0.05 بأنها دالة إحصائيا (حمدي، 2013، صفحة 354).

كما اعتبر Mclean & Ernest, (1997) أن مستوى الدلالة أصبح الأساس الذي من خلاله تبنى المعرفة العلمية، بحيث أن النتائج التي لا تقابل المعيار ينظر إليها على أنها غير مهمة، وأن هذه الآلية في تحديد مستوى الدلالة الاحصائية ثم اختبار الدلالة الاحصائية للوصول إلى القرار الاحصائي، تبدو محددة لصنع القرار مما يعطي شعورا بعدم الموضوعية العلمية في اتخاذ القرار (حمدي، 2013، صفحة 354).

وباختصار اختبارات الدلالة الإحصائية لا تخبر الباحثين بأهمية أو قيمة نتائجهم، وإنما تخبرهم فقط باحتمالية النتيجة، وما إذا كانت نتائج عيناتهم ناشئة عن الصدفة أو خطأ المعاينة. فنحن نفترض أولا أن الفرض الصفري صحيح في المجتمع، وأن أي عينة يتم سحبها من هذا المجتمع سوف تعطي نتائج مماثلة تقريبا للفرض الصفري، ثم نحدد احتمال الحصول على نتائج العينة لحجم العينة المُعطي، وكلما ابتعدت نتائج العينة عن المحدد بالفرض الصفري انخفض احتمال حصولنا على النتائج عن طريق الصدفة (أي انخفضت قيم P-value المحسوبة). ولسوء الحظ أن اختبارات الدلالة الإحصائية لا تخبرنا بما إذا كانت العينة تصف المجتمع (وهو ما نريد أن نعرفه) وإنما تخبرنا فقط باحتمالية نتائج العينة إذا افترضنا أن الفرض الصفري صحيحا. فدلالة أو عدم دلالة النتيجة لا يشير إلى أهميتها عمليا، فقد تكون النتيجة دالة إحصائيا نتيجة حجم العينة الكبير إلا أنها غير مهمة عمليا (مثل الفرق الضعيف جدا

الدال إحصائيا بين متوسطي عينتين حجمهما كبير جدا)، وعلى العكس من ذلك قد تكون النتيجة غير دالة إحصائية نتيجة حجم العينة الصغير إلا أنها مهمة عمليا (مثل الفرق الكبير غير الدال إحصائيا بين متوسطي عينتين حجمهما صغير)، ومن هنا فاختبارات الدلالة الإحصائية غير كافية لتحديد أهمية نتائج البحث، ولذا نحتاج إلى أحجام التأثير لتقييم الدلالة العملية للنتائج حيث تساعد مؤشرات حجم الأثر القراء على التمييز بين النتائج الدالة عمليا وتلك النتائج الدالة إحصائيا.

ويلخص King, (2002) ما لا تفعله اختبارات الدلالة الإحصائية في النقاط التالية (عزت، 2016، صفحة 27):

-احتمالية أن يكون الفرض الصفري صحيحا: رغم أن هذا مرغوب جدا، إلا أنه لا يمكن حساب مثل هذا الاحتمال.

-الاحتمال بأن هذه النتائج تم الحصول عليها بالصدفة: فهذا الاحتمال مجهول سواء كانت العينة تمثل المجتمع تمثيلا تاما، أو كانت النتائج شاذة.

-احتمال وجود هذه النتائج بالدراسات المستقبلية.

-احتمال وجود تأثير حقيقي في المجتمع: فاختبارات الدلالة الإحصائية لا تقيس حجم الأثر بطريقة مباشرة (مثل: مقدار الفروق المشاهدة)، وبدلا من ذلك تكون قيمة P-value موجودة وتدمج حجم الأثر مع حجم العينة ولهذا تظهر النتائج على أنها دالة إحصائيا نتيجة حجم الأثر الكبير أو حجم العينة الكبير أو الاثنان معا.ولذلك قد تكون النتائج دالة إحصائيا بسبب حجم العينة الكبير، إلا أنها ليس لها دلالة عملية بسبب حجم الأثر الصغير والعكس صحيح.

3. الدلالة العملية وحجم الأثر:

تعد الدلالة العملية من المفاهيم التي تستخدم في الكثير من الاختبارات الاحصائية التي تعتمد الأبحاث فيها على الفرض الصفري، ويعتبر مقياس لدرجة خطأ الفرض الصفري في المجتمع الاحصائي. فإذا كان الفرض الصفري يعبر عن غياب الظاهرة في مجتمع الدراسة، فإن الدلالة العملية تعبر عن درجة وجود هذه الظاهرة في ذلك المجتمع، وكلما كانت قيمة الدلالة العملية كبيرة دل ذلك على وجود الظاهرة بدرجة أكبر في ذلك المجتمع (Cohen, 1988, pp. 9-10).

إن الدلالة العملية مرتبطة بدلالة الفرق الاحصائي، ولها أهميتها العملية والتطبيقية فهي تعتبر الوجه الآخر للدلالة لأنها تقوم بتقييم الفرق الاحصائي المحسوب في اختبار دلالة الفروق، أما الدلالة الاحصائية فهي محدودة في تفسير النتائج العملية، وقد أوصى بعض الباحثين أمثال Hays ، Cohen، Thompson، McNamara بضرورة حساب قوة العلاقة بين المتغير التابع والمتغير المستقل، وبهذا فإن الدلالة الاحصائية شرط ضروري من شروط إتخاذ القرار ولكنها ليست كافية لإتخاذ القرار لأن القرار المتخذ يتحقق فقط عندما نحسب قوة العلاقة بين المتغير المستقل والمتغير التابع.

وبهذا يشير مفهوم الدلالة العملية للنتائج إلى استكشاف قوة العلاقة بين المتغيرات قيد الدراسة من خلال استخدام أساليب احصائية أقل تأثرا بحجم العينة من تلك الأساليب الاحصائية التي تعتمد على فحص الدلالة الاحصائية للفروق بين المتوسطات أو حتى عند فحص العلاقة الارتباطية بين المتغيرات (نصار، 2017، صفحة 356).

كما أكد Kirk,(1996) على أن اختبارات الدلالة الإحصائية تقيم فقط "علاقات ترتيبية" Ordinal relationships (على سبيل المثال إذا ما كان متوسطي مجموعتين مختلفان، أو أن أحدهما أكبر من الآخر) ومن خلال ما سبق هل يمكن اعتبار أن هذ المنطق يعمل على تطوير العلوم؟، ما نريد أن نعرفه هو حجم الفرق بين (A)، (B) والخطأ المرتبط بتقديرنا، وتعتبر معرفتنا بأن (A) أكبر من (B) ليست كافية. هذا ما يجعلنا نولي أهمية كبيرة للدلالة العملية والتي ليست بالمفهوم الجديد، حيث يعود جذورها إلى Fisher (1925) الذي دعا من خلال تحليل التباين ANOVA إلى حساب مؤشر يعرف بمربع إيتا، وبالمثل اقترح Kelly (1935) مؤشر ANOVA آخر للدلالة العملية "مربع إبسيلون"، هذه المؤشرات سميت بشكل عام بإسم "أحجام التأثير"، حيث نجد العديد من الخيارات المتاحة لحساب الدلالة العملية لكل اختبار احصائي، ويعتبر حجم الأثر مهم بشكل خاص لأن الاختبارات الإحصائية تتأثر بشدة بحجم العينة، وهذا يعتبر أحد الأسباب التي دفعت إلى تشجيع استخدام الدلالة العملية كمكمل للاختبارات الإحصائية التقليدية (Thompson, 2002, p. 65). وقد طرح Thompson (1993) المثال التالي ليوضح الفرق بين الدلالة الإحصائية والدلالة العملية: إذا قرر باحث ما أن يقارن بين متوسطي درجات الذكاء لمجموعتين كبيرتين وهم يمثلون طلاب مدرسة ما بياناتهما كالتالي: $(\bar{x}_1 = 100{,}15;\ s_1 = 15;\ n_1 = 12000)$ $(\bar{x}_2 = 99{,}85;\ s_2 = 15;\ n_2 = 188000)$ حيث تمثل المجموعة الأول مجموعة من الطلاب تابعين لمنطقة ما بينما المجموعة الثانية تمثل مجموعة الطلاب من باقي المناطق ، حيث: $\bar{x}_1;\bar{x}_2$ تمثل على التوالي متوسط درجات الذكاء للمجموعة الأولى والثانية و $s_1;s_2$ تمثل الانحراف المعياري للمجموعتين و $n_1;n_2$ تمثل حجم المجموعة الأولى والثانية، ومن خلال ما سبق سيجد الباحث أن الفرق بين متوسطي المجموعتين دال إحصائيا لأن

$$(Z_{Calc} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(100{,}15 - 99{,}85)}{\sqrt{\frac{15^2}{12000} + \frac{15^2}{188000}}} = 2{,}12 > Z_{Crit} = 1{,}96;\quad p \leq 0{,}05)$$

ويقرر أن المجموعة الأولى أعلى ذكاء من المجموعة الثانية، ومن خلال هذا قد يقترح الباحث الغير واعي لمعنى الدلالة الاحصائية على أعضاء مجلس إدارة هذه المدرسة إنشاء مدرسة خاصة للطلاب الموهبين في المنطقة التي يقطن فيها 12000 طالب، بحيث أنهم أكثر ذكاء بشكل ملحوظ على أقرانهم في المناطق الأخرى، بالرغم من أن الفرق في متوسط درجات الطلاب بلغ (100,15-99,85=0,3) وهو يقل عن درجة ذكاء واحدة وهو غير مهم عمليا. (Thompson, 2002, pp. 65-66).

إن الهـــدف الأســـاسي من جميع الأساليب السابقة هو الإجابة على السؤال التالي: هـــل أحدثت المعالجـــة فـــرقا عمليا بين المجموعـــات التجريبية والضابطة، وما مقـــــدار الفرق الذي حققته؟ ويعتبر أسلـــوب حجم الأثر مفيدا لأنه يهتم بتقرير أهمية النتائـــــج على المستوى العملي وهو بذلك يتعدى مفهوم الـــدلالة الاحصائية التي تهتم فقط فيما إذا كانت النتائج تعـــود إلى عـــامل الصدفة أم لا (نصار، 2017، صفحة 357).

1.3. تعريف حجم الأثر:

حجم الأثر هو مفهوم يقصد به الأساليب التي يتم من خلالها معرفة حجم الفرق أو حجم العلاقة بين متغيرين أو أكثر، ويسمى أحيانا الدلالة العملية، ويطلق عليه في أحيان ثالثة مقاييس قوة التأثير وفي أحيان رابعة قوة الترابط...وهكذا (رشدي، 1997، صفحة 57) .

كما يمكن تعريف حجم الأثر بأنه أي إحصاءة تحدد درجة اختلاف نتائج العينات عن التوقعات المحددة في الفرضية الصفرية، وهناك العشرات من مقاييس حجم الأثر المتاحة، ولكل منها نقاط القوة والضعف الخاصة بها (Sun, Wei , & Lihshing, 2010, p. 991).

إن مقاييس حجم الأثر تتناول حجم الفرق أو قوة الارتباط دون أن يكون دالة لحجم العينة (لا يعتمد على حجم العينة)، لكن حجم الأثر لا يعبر عن مدى الثقة التي نوليها لهذا الفرق بل يترك هذا لمستوى الدلالة الاحصائية. ومن هنا جاء مفهوم حجم الأثر ليكمل مفهوم الدلالة الاحصائية، ويتبين مما تقدم أن مفهوم الدلالة الاحصائية للنتائج يركز على مدى الثقة التي نضعها في النتائج بصرف النظر عن حجم الفرق أو قوة الارتباط، بينما يركز مفهوم حجم الأثر على الفرق أو حجم الارتباط بصرف النظر عن مدى الثقة التي نضعها في النتائج (رشدي، 1997، الصفحات 58-59).

إن فكرة حجم الأثر تقوم بكل بساطة على صياغة الفروق بين المتوسطات بإستخدام الانحراف المعياري كوحدة قياس لمقدار الفرق بين تلك المتوسطات (كما هو الحال في اختبار t للعينات المستقلة والمترابطة)، أو التعبير عن العلاقة بين المتغير المستقل أو المتغيرات المستقلة من جهة والمتغير التابع أو المتغيرات التابعة من جهة أخرى عن طريق استخراج حجم تباين المتغير التابع أو المتغيرات التابعة الذي يمكن تفسيره عن طريق المتغير المستقل أو المتغيرات المستقلة، كما هو الحال في تحليل التباين، أي أن قيمة حجم الأثر تبين إلى أي درجة يمكن التنبؤ بالمتغير التابع أو تفسيره من خلال المتغير المستقل.

ومن وجهة نظر Maxwell & Delaney,(1990) يمكن تقسيم المؤشرات التي تدل على مقدار حجم الأثر في الدراسات المختلفة إلى فئتين رئيسيتين هما (نصار، 2006، صفحة 39):

-المؤشرات التي تدل على مقدار التأثير الذي يحدثه المتغير المستقل في المتغير التابع. وتعرف هذه المؤشرات أيضا بأنها الفرق بين المتوسطات بوحدات معيارية، وتستخدم هذه المؤشرات في الدراسات التي تعتمد على الأساليب

الاحصائية التي تفحص الفروق بين المتوسطات كما هو الحال في اختبارt سواء للعينات المستقلة أم المترابطة. وأشهر هذه المؤشرات هو المؤشر الذي قدمه Cohen,(1988).

-المؤشرات التي تدل على قوة العلاقة أو الارتباط بين متغيرات الدراسة. وتستخدم هذه المؤشرات في الأساليب الاحصائية التي تعتمد على فحص التباين حيث تشير إلى نسبة التباين الذي يمكن تفسيره في المتغير التابع عن طريق تباين المتغير أو المتغيرات المستقلة، ومن أشهر هذه المؤشرات "مربع إيتا".

2.3. المؤشرات الاحصائية المستخدمة للدلالة على حجم الأثر في اختبار t:

تختلف مقاييس حجم الأثر بإختلاف الاختبار الإحصائي المستخدم لإختبار الفرضية، وفيما يأتي أبرز هذه المقاييس المستخدمة كمؤشرات لحجم الأثر في اختبار t.

أ- لإختبار الفرضية المتعلقة بالفرق بين متوسطي مجتمعين مستقلين فإن: $d = \frac{|\bar{X}_1 - \bar{X}_2|}{s_p}$ و $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$.

حيث تشير d إلى حجم الأثر، $\bar{X}_1$: متوسط العينة الأولى، $\bar{X}_2$: متوسط العينة الثانية، s_p: الانحراف المعياري المشترك للعينتين.

1. إذا كانت $d = 0$ فإن ذلك يعني تساوي متوسطي العينتين، وكلما زاد الفرق عن الصفر يزداد حجم الأثر.
2. إذا كانت قيمة حجم الأثر d تساوي $0,2$ أو أقل فإنه يعتبر حجم أثر صغير، وإذا كانت هذه القيمة أكبر من $0,2$ وأقل من $0,8$ فيعتبر حجم أثر متوسط، أما إذا كانت قيمته $0,8$ فأكبر فإنه يعتبر حجم أثر كبير.

أو يمكن استخدام مربع إيتا (η^2) كبديل لحجم الأثر d حيث أن: $\eta^2 = \frac{T^2}{T^2 + df}$

حيث T تمثل قيمة إحصاءة الاختبار، df تمثل درجات الحرية وتتراوح قيمة η^2 بين 0 و 1.

1. إذا كانت $\eta^2 = 0$ فمعنى هذا أن متوسط الفروق يبلغ صفرا. وفي هذه الحالة فإن القيمة المتوسطة لا تختلف داخل كل من المجموعتين (أي هناك تطابقا تاما بين متوسطي المجموعتين).
2. إذا كانت قيمة $\eta^2 = 1$ فمعنى هذا أن متوسطي المجموعتين مختلفان.
3. إذا كانت قيمة مربع إيتا تساوي $0,01$ أو أقل يعتبر حجم أثر صغير، وإذا كانت هذه القيمة أكبر من $0,01$ وأقل من $0,14$ فيعتبر حجم أثر متوسط، أما إذا كانت $0,14$ فأكبر فإنه يعتبر حجم أثر كبير.

ب- لإختبار الفرضية الاحصائية المتعلقة بمتوسطي مجتمعين مرتبطين فإن: $d = \frac{|\bar{X}_{x_a - x_b}|}{s_{x_a - x_b}}$ أو $d = \frac{T}{\sqrt{n}}$.

حيث تشير d إلى حجم الأثر، $\bar{X}_{x_a - x_b}$: تمثل متوسط الفروق بين قيم x_a وقيم x_b المتناظرتين أو بين الاختبار القبلي والاختبار البعدي، $s_{x_a - x_b}$: الانحراف المعياري للفروق.

أو يمكن استخدام مربع إيتا (η^2) كبديل لحجم الأثر d حيث أن: $\eta^2 = \frac{T^2}{T^2 + df}$

حيث $df = n - 1$ تمثل درجات الحرية.

جـــ- لإختبار الفرضية المتعلقة بمتوسط حسابي لمجتمع واحد فإن $d = \frac{|\bar{X} - \mu_0|}{s}$.

حيث تشير d إلى حجم الأثر، $\bar{X}$: متوسط العينة، μ_0: قيمة متوسط المجتمع المحددة في الفرضية الصفرية s: الانحراف المعياري للعينة.

3.3. أهمية حجم الأثر:

قد يكون حجم الأثر مفيدا في ثلاثة تطبيقات عملية على الأقل. أولا، قبل إجراء الدراسة، حيث يمكن استخدام تقديرات أحجام التأثير المتوقعة لإظهار حجم العينة الذي يكون مناسبا لكشف النتائج ذات الدلالة الإحصائية. ومن خلال هذا يمكن حساب الحد الأدنى لحجم العينة المناسب لكشف حجم تأثير معين بعد تقدير أو تحديد قيم حجم الأثر و مستوى الدلالة α وقوة الاختبار. وهذا يساعد في تقليل مخاطر النتائج الإحصائية المضللة بسبب عدم كفاية حجم العينة. ثانيا، تقدير حجم الأثر يساعد الباحثين على إصدار حكم حول الأهمية العملية للدراسة. ويرى Fan (2001) أن القيمة الاحتمالية P-value وحجم الأثر هما وجهان لعملة واحدة: فهما يكملان بعضها البعض، لكنهما لا يحلان محل بعضها البعض. لذلك ، يجب على الباحثين النظر في كلا الجانبين، بحيث يجب أن يكون الغرض من البحث هو قياس حجم الأثر بدلا من مجرد الدلالة الاحصائية؛ وبالتالي، فإن تقدير حجم الأثر وتفسيره أمر بالغ الأهمية. ثالثا بالاضافة إلى أن أحجام التأثير تهدف إلى قياس الأهمية العملية، فقد يتم استخدامها لمقارنة نتائج الدراسات المختلفة مع بعضها البعض وتقييم إمكانية تكرار النتائج . بمعنى أنها توفر أداة إحصائية للتحليل البعدي (Meta-analysis) والذي يعتبر تقنية إحصائية للجمع بين نتائج مجموعة من الدراسات التي تعالج الفرضيات البحثية المشتركة وبعبارة أخرى هو ذلك التحليل الاحصائي لمجموعة عريضة من نتائج التحليلات الفردية بغرض دمج النتائج حيث تصبح وحدة التحليل في هذه الحالة هي الدراسات نفسها (عيشوش، 2017، صفحة 211). كما أن استخدام أحجام التأثير في الدراسات يساعد الباحثين المهتمين بالتحليل البعدي على تجنب استخدام أحجام التأثير التقريبية التي تعتمد على افتراضات إحصائية ضعيفة في بعض الأحيان (Sun, Wei , & Lihshing, 2010, p. 991).

ولقد أوجز Huston, (1993) فوائد مقاييس حجم الأثر على النحو التالي (حمدي، 2013، صفحة 356):

-حجم الأثر يشير إلى درجة وجود الظاهرة في المجتمع بمستوى قياس نسبي، والذي يعني الصفر فيه عدم وجود الظاهرة.

-تزود الباحثين بمؤشرات للدلالة العملية بخلاف اختبارات الدلالة الإحصائية.

-يمكن استخدامها في المقارنة الكمية بين نتائج دراستين أو أكثر كما هو مستخدم في التحليل البعدي Meta Analysis.

-يمكن استخدامها في تحليل القوة الاحصائية لتحديد حجم العينة المطلوبة في دراسة معينة.

ويمكن بيان أهمية عمل حجم الأثر من خلال النقاط التالية:

- مقياس يوضح قوة العلاقة بين متغيرين أو أكثر، وذلك من خلال قيمة واحدة ومحددة تتبع الأسلوب الاحصائي المستخدم والدال احصائيا.
- مقياس يبين فيما إذا كان المتغير التابع مُفسرا ومتأثرا بالمتغير المستقل.
- مقياس لمؤشر الفروق الحقيقية والمعيارية بين متوسطين أو أكثر.
- مقياس للدلالة العملية ومدى أهمية النتائج المتحصل عليها، وكونها لا ترجع للمصادفة.

4.3. تفسير حجم الأثر:

إن الاكتفاء بعملية تقدير حجم الأثر فقط في البحوث لا يكفي بل يجب على الباحثين تفسير وتقييم حجم الأثر وفق الأهمية العملية. وبهذا فإن كيفية تفسير حجم الأثر هو أيضا بالغ الأهمية. حيث تتمثل الممارسة الشائعة في تفسير أحجام التأثير في استخدام المعايير الخاصة بالتأثيرات "الصغيرة" و "المتوسطة" و "الكبيرة" التي قدمها (1988) Cohen إلا أنه يجب التنبيه إلى أن تفسير نتائج أحجام التأثير والدلالة العملية هو أمر ذاتي Subjective، حيث يعتمد بشكل كبير على أهداف البحث والدراسة، وذلك لكي لا تتكرر مشكلة مستويات الدلالة الاحصائية (0.01 ، 0.05) التي قد تمنح الباحث شعورا بالموضوعية، ولذا فإن العرف الذي وضعه Cohen في تصنيف وتفسير حجم الأثر (صغير متوسط، كبير) ليس مناسبا، لأن Cohen قدم هذه المعايير كمبادئ توجيهية عامة للباحثين الذين يعملون في المجالات البحثية الغير مستكشفة. أما في حالة المجالات البحثية التي تتوفر على دراسات سابقة، من غير المناسب تطبيق إرشادات Cohen بصورة عمياء. لذلك يجب أن يركز التفسير المناسب لحجم الأثر على المقارنة الصريحة والمباشرة بين أحجام التأثير في نتائج الدراسة وأحجام التأثير في الدراسات السابقة، كما يجب تضمين كل من حجم وطبيعة التأثير في التفسير (بابطين، 2002، الصفحات 108-109).

كما تعتمد أهمية ومعنى حجم الأثر على عوامل عديدة، مثل سياق الدراسة، وأهمية النتائج، وحجم وطبيعة التأثير الذي تم الحصول عليه في الدراسات السابقة .لذلك، ينبغي على الباحثين تفسير حجم الأثر ضمن إطار الدراسة مع الأخذ بعين الاعتبار نتائج أحجام التأثير للدراسات السابقة. كما أن هناك أربعة نتائج محتملة في الاختبارات الاحصائية: (أ) وجود دلالة احصائية مع حجم أثر مهم عمليا، (ب) عدم وجود دلالة احصائية مع حجم أثر غير مهم عمليا، (ج) وجود دلالة احصائية مع حجم أثر غير مهم عمليا، و (د) عدم وجود دلالة احصائية مع حجم أثر مهم عمليا .فيما يخص الحالتين الأولى والثانية لا يوجد مشكلة في تفسير النتائج وذلك بسبب وجود اتساق بين نتائج الدلالة الاحصائية والدلالة العملية. بينما في الحالة الثالثة نجد أن هناك دلالة احصائية بالرغم من أن حجم الأثر غير

مهم عمليا وفي الحالة الأخيرة تفشل الدلالة الاحصائية في تحديد حجم الأثر المهم. وبهذا في هاتين الحالتين الأخيرتين يكون هناك عدم اتساق بين الدلالة الاحصائية والدلالة العملية (حجم الأثر)، مما يوحي بوجود تهديدات محتملة لصحة تصميم الدراسة (Sun, Wei , & Lihshing, 2010, p. 991).

كما أشار Zientek, Capraro, and Capraro (2008) إلى أن عدم تضمين حجم الأثر في الدراسات يمكن أن يكون له العديد من السلبيات بسبب عدم الاتساق المحتمل بين نتائج الدلالة الاحصائية وحجم الأثر وبهذا يكون تضمين حجم الأثر في الدراسات مهما للغاية للاختبارات الإحصائية. لأن القيمة الاحتمالية P-value الصغيرة لا تشير بالضرورة إلى حجم تأثير مهم عمليا؛ ومن ناحية أخرى، قد تكون القيمة الاحتمالية P-value كبيرة بينما حجم الأثر مهم عمليا بسبب القدرة الإحصائية المحدودة للكشف عن مثل هذا التأثير، وبالتالي، يجب على الباحثين تضمين حجم الأثر في الدراسات للاختبارات ذات الدلالة الاحصائية والاختبارات التي ليست لها دلالة احصائية، ومن جهة أخرى عدم تضمين حجم الأثر في الدراسات يلزم الباحثين المهتمين بإجراء تحليل ثانوي أو تحليل بعدي استخدام صيغة تحويل تقريبية لتقدير حجم الأثر بسبب عدم الوصول إلى البيانات الخام، لذلك يجب على الباحثين تضمين حجم الأثر الدقيق عند إجراء دراساتهم من خلال البيانات الأولية (الخام) لتحسين دقة التقدير للتحليل الثانوي. وفي الأخير يمكن القول بأن عدم تضمين حجم الأثر ليس له سلبيات على نتائج دراسة واحدة فقط وإنما تشمل هذه السلبيات أيضا تراكم المعرفة على المدى الطويل (Sun, Wei , & Lihshing, 2010, pp. 991-992).

وتجدر الإشارة إلى أن مقاييس حجم الأثر لها حدودها وليست علاجا لجميع مشاكل الدلالة الاحصائية وتفسير النتائج، أولا، يعتمد تقدير حجم الأثر على المتوسط الحسابي والانحراف المعياري للعينة، وبالتالي، سوف يختلف من عينة إلى أخرى، لذلك يجب تفسير أي حجم تأثير بحذر، ثانيا، يعتبر حجم الأثر المحسوب من خلال العينة كمقدر نقطي لحجم الأثر في المجتمع ، كما أنه لا يشير إلى دقة التقدير. وقد اقترح Henson,(2006) طريقتين يمكن من خلالهما التغلب على حدود استخدام أحجام التأثير: (أ) الاعتماد على أحجام التأثير الخاصة بالتحليل البعدي (Meta-analysis)، (ب) بالاضافة إلى ذلك القيام ببناء فترات الثقة لمتوسط المجتمع، وبناء فترات الثقة لحجم الأثر في الدراسات الفردية التي تشمل على جميع القيم التي من المتوقع أن تكون داخل هذه الفترات بمستوى ثقة معين بدلا من استخدام المقدر النقطي لحجم الأثر الحقيقي في المجتمع .كما أوصى كل من Kline (2004) and Thompson (2008) بهاتين الطريقتين (Sun, Wei , & Lihshing, 2010, p. 992).

حاولنا من خلال ما سبق أن نبين مدى أهمية تضمين حجم الأثر بالإضافة إلى الدلالة الاحصائية في الاختبارات الاحصائية لما في هذه الأخيرة من عيوب، حيث يعمل حجم الأثر على تدعيم الدلالة الاحصائية من أجل تفادي البعض من هذه العيوب، وفيما يلي سوف نتطرق إلى الجزء التطبيقي الذي سنحاول من خلاله إبراز أحد أهم عيوب الدلالة الاحصائية والمتمثل في تأثر القيمة الاحتمالية P-value بحجم العينة وهذا عند مستوى ثابت من حجم الأثر، وسنعتمد على ثلاثة أنواع من الاختبارات الاحصائية وهي اختبار t لعينة واحدة، اختبارt لعينتين مسقلتين واختبارt لعينتين مرتبطتين.

4. الدراسة التطبيقية لأثر حجم العينة على مستوى القيمة الاحتمالية P-value:

1.4. أثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينة واحدة: للإجابة على هذا التساؤل تم استخدام أسلوب المحاكاة بطريقة مونتي كارلو لتوليد البيانات المتعلقة بالعينات محل الدراسة من خلال برنامج PASS V16 حيث تم توليد مجتمع من البيانات الاحصائية بحجم N=400، ثم أخذ عينات من n=10 إلى n=200 وبمعدل زيادة n=10، بمتوسط عينة ($\bar{X}=100$) ومتوسط فرضي ($\mu_1=105$)، وإنحراف معياري ($S=20$)، ومجتمع البيانات يتوزع توزيع طبيعيا، واستخدام اختبار t لعينة واحدة لإختبار الفرض الصفري ($H_0:\mu_0-\mu_1=0$) ضد الفرض البديل وجود فروق ذات دلالة احصائية بين متوسط المجتمع والمتوسط الفرضي ($H_1:\mu_0-\mu_1\neq 0$)، ثم حساب مستوى القيمة الاحتمالية P-value، وفيما يلي عرض النتائج:

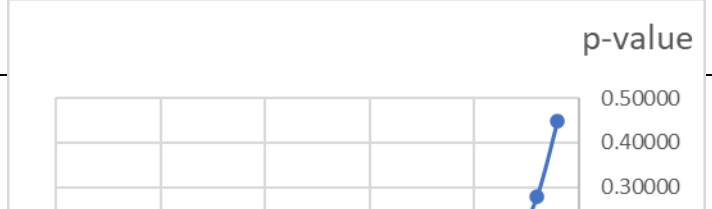
الجدول رقم 01: يوضح أثر حجم العينة على القيمة الاحتمالية لإختبار t لعينة واحدة

حجم الأثر	القيمة الاحتمالية	T*	S	μ_1	$\bar{X}$	N
0,250	0,44952	0,79057	20	105	100	10
0,250	0,27749	1,11803	20	105	100	20
0,250	0,18142	1,36931	20	105	100	30
0,250	0,12192	1,58114	20	105	100	40
0,250	0,08333	1,76777	20	105	100	50
0,250	0,05760	1,93649	20	105	100	60
0,250	0,04015	2,09165	20	105	100	70
0,250	0,02817	2,23607	20	105	100	80
0,250	0,01986	2,37171	20	105	100	90
0,250	0,01406	2,50000	20	105	100	100
0,250	0,00999	2,62202	20	105	100	110
0,250	0,00712	2,73861	20	105	100	120
0,250	0,00508	2,85044	20	105	100	130
0,250	0,00364	2,95804	20	105	100	140
0,250	0,00261	3,06186	20	105	100	150
0,250	0,00188	3,16228	20	105	100	160
0,250	0,00135	3,25960	20	105	100	170
0,250	0,00097	3,35410	20	105	100	180
0,250	0,00070	3,44601	20	105	100	190
0,250	0,00051	3,53553	20	105	100	200

المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج PASS V16

والتمثيل البياني للعلاقة بين حجم العينة ومستوى القيمة الاحتمالية P-value موضح في الشكل الموالي:

الشكل رقم 01: يوضح العلاقة بين حجم العينة ومستوى القيمة الاحتمالية



المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج PASS V16

في الجدول أعلاه تم حساب القيمة الاحتمالية P-value بناءا على المعطيات وبالأخذ بعين الاعتبار أن الفرضية البديلة هي: $\mu_0 \neq \mu_1$، كما يلي: $P\left(|T| > T^*\right) = 2P\left(T > T^*\right)$، حيث: $T \sim t_{n-1}$ و $T^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ وهذه القيم محسوبة في الجدول أعلاه، ومن خلال الجدول نلاحظ أيضا أن حجم الأثر $d = \frac{|\bar{X} - \mu_0|}{s} = \frac{|100-105|}{20} = 0.25$ وهو ثابت في جميع الحالات إلا أننا نلاحظ وجود علاقة تأثير عكسية لحجم العينة على مستوى القيمة الاحتمالية P-value لإختبارt لعينة واحدة، حيث تبين أنه عندما كان حجم العينة n=10 كانتP-value =0,44952 وعند هذا المستوى لا نرفض الفرضية الصفرية أي لا توجود فروق ذات دلالة احصائية بين متوسط العينة والمتوسط الفرضي، وعند زيادة حجم العينة إلى أن وصلت n=70 كانت P-value =0,04015 وهي أقل من مستوى الدلالة 0,05 وبهذا نرفض الفرضية الصفرية أي توجود فروق ذات دلالة احصائية بين متوسط العينة والمتوسط الفرضي.

مما سبق يمكن استنتاج أن حجم العينة يعتبر من بين أهم العوامل التي تؤثر على مستوى القيمة الاحتمالية P-value لإختبارt لعينة واحدة، فإذا كان حجم العينة كبيرا بما فيه الكفاية فإننا نصل إلى فروق دالة احصائيا حتى ولو كانت هذه الفروق صغيرة، وفي المقابل إذا كان حجم العينة صغير فإننا لا نصل إلى فروقات دالة احصائيا حتى ولو كانت حجم الفروقات كبيرة. ومن هنا تحديدا ظهر مفهوم حجم الأثر ليعوض أو ليكمل هذا النقص.

2.4. أثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينتين مستقلتين: للإجابة على هذا التساؤل تم استخدام مجموعة البيانات الاحصائية المتاحة من خلال برنامج PASS V16 حيث تم توليد مجتمع من البيانات الاحصائية بحجم n=400 ، ثم أخذ عينات من n=10 إلى n=200 ، وبمعدل زيادة n=10، بمتوسط حسابي للعينة الأولى $(\bar{X}_1 = 100)$ ومتوسط حسابي للعينة الثانية $(\bar{X}_2 = 110)$،و إنحراف معياري $(S = 20)$، ومجتمع البيانات يتوزع توزيع طبيعيا، واستخدام اختبارt لعينتين مستقلتين لإختبار الفرض الصفري بعدم وجود فروق ذات دلالة احصائية بين متوسطي المجتمعين $(H_0 : \mu_1 = \mu_2)$ ضد الفرض البديل وجود فروق ذات دلالة احصائية بين متوسطي المجتمعين $(H_1 : \mu_1 \neq \mu_2)$، ثم حساب مستوى القيمة الاحتمالية P-value في حالة تساوي حجم العينتين، وفيما يلي عرض النتائج:

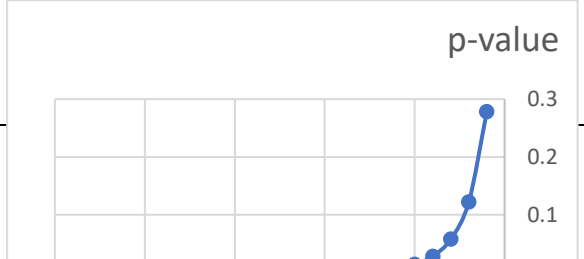
الجدول رقم 02: يوضح أثر حجم العينة على القيمة الاحتمالية لإختبار t لعينتين مستقلتين

حجم الأثر	القيمة الاحتمالية	T^*	S_p	δ	$\bar{X}_2$	$\bar{X}_1$	N	N2	N1
0,50	0,27825	-1,11803	20	-10	110	100	20	10	10
0,50	0,12214	-1,58114	20	-10	110	100	40	20	20
0,50	0,05768	-1,93649	20	-10	110	100	60	30	30
0,50	0,02821	-2,23607	20	-10	110	100	80	40	40
0,50	0,01408	-2,50000	20	-10	110	100	100	50	50
0,50	0,00713	-2,73861	20	-10	110	100	120	60	60
0,50	0,00364	-2,95804	20	-10	110	100	140	70	70
0,50	0,00188	-3,16228	20	-10	110	100	160	80	80
0,50	0,00097	-3,35410	20	-10	110	100	180	90	90
0,50	0,00051	-3,53553	20	-10	110	100	200	100	100
0,50	0,00026	-3,70810	20	-10	110	100	220	110	110
0,50	0,00014	-3,87298	20	-10	110	100	240	120	120
0,50	0,00007	-4,03113	20	-10	110	100	260	130	130
0,50	0,00004	-4,18330	20	-10	110	100	280	140	140
0,50	0,00002	-4,33013	20	-10	110	100	300	150	150
0,50	0,00001	-4,47214	20	-10	110	100	320	160	160
0,50	0,00001	-4,60977	20	-10	110	100	340	170	170
0,50	0,00000	-4,74342	20	-10	110	100	360	180	180
0,50	0,00000	-4,87340	20	-10	110	100	380	190	190
0,50	0,00000	-5,00000	20	-10	110	100	400	200	200

المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج PASS V16

والتمثيل البياني للعلاقة بين حجم العينة ومستوى القيمة الاحتمالية P-value موضح في الشكل الموالي:

الشكل رقم 02: يوضح العلاقة بين حجم العينة ومستوى القيمة الاحتمالية



المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج PASS V16

في الجدول أعلاه تم حساب القيمة الاحتمالية P-value بناءا على المعطيات وبالأخذ بعين الاعتبار أن الفرضية البديلة هي: $H_1: \delta = \mu_1 - \mu_2 \neq 0$، كما يلي: $P\left(|T| > T^*\right) = 2P\left(T > T^*\right)$، حيث: $T \sim t_{n_1+n_2-1}$ و $T^* = \sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}} . \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{s_p}$ و $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$ وهذه القيم محسوبة في الجدول أعلاه ومن خلال الجدول نلاحظ أيضا أن حجم الأثر $d = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{s_p} = \frac{|100-110|}{20} = 0.50$ وهو ثابت في جميع الحالات إلا أننا نلاحظ وجود علاقة تأثير عكسية لحجم العينة على مستوى القيمة الاحتمالية P-value لإختبارt لعينتين مستقلتين، حيث تبين أنه عندما كان حجم العينة n=10 كانت P-value=0,27825 وعند هذا المستوى لا نرفض الفرضية الصفرية أي لا توجود فروق ذات دلالة احصائية بين متوسطي العينتين، وعند زيادة حجم العينة إلى أن وصلت n=40 كانت P-value=0,02821 وهي أقل من مستوى الدلالة 0,05 وبهذا نرفض الفرضية الصفرية أي توجود فروق ذات دلالة احصائية بين متوسطي العينتين.

مما سبق يمكن استنتاج أن حجم العينة يعتبر من بين أهم العوامل التي تؤثر على مستوى القيمة الاحتمالية لإختبارt لعينتين مستقلتين، فإذا كان حجم العينة كبيرا بما فيه الكفاية فإننا نصل إلى فروق دالة احصائيا حتى ولو كانت هذه الفروق صغيرة، وفي المقابل إذا كان حجم العينة صغير فإننا لا نصل إلى فروقات دالة احصائيا حتى ولو كانت حجم الفروقات كبيرة.

3.4. أثر حجم العينة على مستوى القيمة الاحتمالية لإختبارt لعينتين مرتبطتين: للإجابة على هذا التساؤل تم استخدام مجموعة البيانات الاحصائية المتاحة من خلال برنامج PASS V16 حيث تم توليد مجتمع من البيانات الاحصائية بحجم N=400، ثم أخذ عينات من n=10 إلى n=200 وبمعدل زيادة n=10، بمتوسط (*Mean of Paired Differences* = 5)،و إنحراف معياري s=20، ومجتمع البيانات يتوزع توزيع طبيعيا، واستخدام اختبارt لعينتين مرتبطتين لإختبار الفرض الصفري بعدم وجود فروق ذات دلالة احصائية بين متوسطي العينتين ضد الفرض البديل وجود فروق ذات دلالة احصائية بين متوسطي العينتين، ثم حساب مستوى القيمة الاحتمالية، وفيما يلي عرض النتائج:

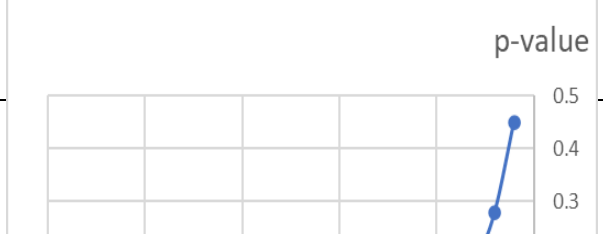
الجدول رقم 03: يوضح أثر حجم العينة على القيمة الاحتمالية لإختبار t لعينتين مرتبطتين

حجم الأثر	القيمة الاحتمالية	T*	S	Mean of Paired Differences	N
0,25	0,44952	0,79057	20	5,00	10
0,25	0,27749	1,11803	20	5,00	20
0,25	0,18142	1,36931	20	5,00	30
0,25	0,12192	1,58114	20	5,00	40
0,25	0,08333	1,76777	20	5,00	50
0,25	0,05760	1,93649	20	5,00	60
0,25	0,04015	2,09165	20	5,00	70
0,25	0,02817	2,23607	20	5,00	80
0,25	0,01986	2,37171	20	5,00	90
0,25	0,01406	2,50000	20	5,00	100
0,25	0,00999	2,62202	20	5,00	110
0,25	0,00712	2,73861	20	5,00	120
0,25	0,00508	2,85044	20	5,00	130
0,25	0,00364	2,95804	20	5,00	140
0,25	0,00261	3,06186	20	5,00	150
0,25	0,00188	3,16228	20	5,00	160
0,25	0,00135	3,25960	20	5,00	170
0,25	0,00097	3,35410	20	5,00	180
0,25	0,00070	3,44601	20	5,00	190
0,25	0,00051	3,53553	20	5,00	200

المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج PASS V16

والتمثيل البياني للعلاقة بين حجم العينة ومستوى القيمة الاحتمالية P-value موضح في الشكل الموالي:

الشكل رقم 03: يوضح العلاقة بين حجم العينة و القيمة الاحتمالية



المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج PASS V16

في الجدول أعلاه تم حساب القيمة الاحتمالية P-value بناءا على المعطيات وبالأخذ بعين الاعتبار أن الفرضية البديلة هي: (Mean of Paired Differences ≠0)، كما يلي: $P\left(|T|>T^*\right)=2P\left(T>T^*\right)$، حيث: $T \sim t_{n-1}$ و $T^*=\frac{\bar{X}_{x_a-x_b}}{s_{x_a-x_b}/\sqrt{n}}$ وهذه القيم محسوبة في الجدول أعلاه، ومن خلال الجدول نلاحظ أيضا أن حجم الأثر $d=\frac{|\bar{X}_{x_a-x_b}|}{s_{x_a-x_b}}=\frac{|8|}{20}=0.40$ وهو ثابت في جميع الحالات إلا أننا نلاحظ وجود علاقة تأثير عكسية لحجم العينة على مستوى القيمة الاحتمالية P-value لإختبارt لعينتين مرتبطتين، حيث تبين أنه عندما كان حجم العينة n=10 كانتP-value =0.44952 وعند هذا المستوى لا نرفض الفرضية الصفرية أي لا توجود فروق ذات دلالة احصائية بين متوسطي العينتين، وعند زيادة حجم العينة إلى أن وصلت n=70 كانت P-value =0.04015 وهي أقل من مستوى الدلالة 0.05 وبهذا نرفض الفرضية الصفرية أي توجود فروق ذات دلالة احصائية بين متوسطي العينتين.

مما سبق يمكن استنتاج أن حجم العينة يعتبر من بين أهم العوامل التي تؤثر على مستوى القيمة الاحتمالية لإختبار t لعينتين مرتبطتين، فإذا كان حجم العينة كبيرا بما فيه الكفاية فإننا نصل إلى فروق دالة احصائيا حتى ولو كانت هذه الفروق صغيرة، وفي المقابل إذا كان حجم العينة صغير فإننا لا نصل إلى فروقات دالة احصائيا حتى ولو كانت حجم الفروقات كبيرة.

5. خاتمة:

من خلال هذه الورقة البحثية تم التطرق إلى مفهوم الدلالة العملية والدلالة الاحصائية وأهميتهما مع تبيان جوانب قصور هذه الأخيرة (الدلالة الإحصائية) وذلك بتأثرها بحجم العينة، وكذا أهمية اقتران الدلالة العملية أو ما يطلق عليها بحجم الأثر بالدلالة الاحصائية في تفسير مخرجات البحوث، من أجل الحصول على قرارات أكثر دقة، الأمر الذي يجعلنا نثق في النتائج المتحصل عليها في البحوث.

1.5. اختبار فرضيات الدراسة: من خلال هذه الورقة البحثية تم تأكيد فرضيات البحث والتي هي:

- يؤثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينة واحدة.
- يؤثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينتين مستقلتين.

– يؤثر حجم العينة على مستوى القيمة الاحتمالية لإختبار t لعينتين مرتبطتين.

2.5. نتائج الدراسة:

إن تقدير حجم الأثر يقدم فائدتين كبيرتين أولهما أنه يقدم تقريرا حول مدى العلاقة بين المتغير المستقل والمتغير التابع، أي أنه لا يكتفي بالإجابة عن الفرضية الصفرية بـ "نعم" أو "لا" بل إنه يحدد مقدار أثر المتغير المستقل على المتغير التابع، حيث أنه في بعض الأحيان قد يكون الفرق بين المجموعتين كبيرا ولكنه غير دال إحصائيا بسبب أحد العوامل التي تم ذكرها سابقا. ومن جهة أخرى قد يكون هذا الفرق صغيرا جدا ولكنه دال بسبب كبر حجم العينة أو نتيجة استخدام مستوى دلالة مرتفع. أما الفائدة الثانية لتقرير "مقدار حجم الأثر" فتتمثل في أنه يزود الباحثين الآخرين بفكرة عن قوة العلاقة أو حجم تأثير المتغير المستقل على المتغير التابع، ومن هنا فإن حساب مقدار حجم الأثر يقدم معلومات وفائدة لا تقل أهمية عن عملية اختبار الفرضيات الاحصائية، بل إنه يفصل ويوضح ويكمل المعلومات المستخلصة من خلال اختبار الفرضيات.

لقد تبين أنه حتى تكون للنتائج قابلية للتعميم وموثوقق بمصداقيتها، لا بد للباحثين من الإهتمام بالدلالة العملية إضافة إلى الدلالة الاحصائية، لأن الدراسات التي تتمتع بدلالة عملية يمكن أن تساعد في تفسير نتائج الدراسات التي تبحث في أثر المتغيرات التابعة، وبالتالي يمكن أن يتم تعميم نتائجها، بحيث تكون قابلة للتطبيق. لذا ينبغي على الباحثين حساب الدلالة العملية لنتائج مجهوداتهم البحثية ذات الدلالة الاحصائية، لإبراز مدى الاتفاق بين الدلالتين الاحصائية والعملية من عدمه.

3.5. التوصيات:

هناك مجموعة من التوصيات المقترحة لتفسير نتائج اختبارات الدلالة الاحصائية موضحة فيما يلي:

1- يجب حساب حجم الأثر (مؤشرات الدلالة العملية) بشكل مرافق لكل قيمة من القيم الاحتمالية المحسوبة في البحث، بحيث لا تعرض أي نتيجة من نتائج اختبارات الدلالة الاحصائية بدون ارفاقها بمؤشر واحد على الأقل من مؤشرات حجم الأثر.

2- يجب عدم الاكتفاء بعرض نتائج مؤشرات حجوم التأثير المحسوبة فقط حسب الاستراتيجية السابقة، ولكن يجب أن تفسر في جميع التحاليل سواء كانت مقاييس فروق معيارية أو نسب التباين المفسر، كما يجب إدراك أن تفسير نتائج حجوم التأثير والدلالة العملية هو أمر ذاتي، حيث يعتمد بشكل كبير على أهداف البحث والدراسة، وذلك لكي لا تتكرر مشكلة مستويات الدلالة الاحصائية (0.01 ، 0.05) التي قد تمنح الباحث شعورا بالموضوعية ولذا فإن العرف الذي وضعه Cohen في تصنيف وتفسير حجم الأثر (صغير ، متوسط، كبير) ليس مناسبا حيث أن وجود قيمة لمؤشر حجم الأثر أصغر من 0.2 والذي يصنفه Cohen بأنه الحد الأدنى لحجم الأثر الصغير قد يعني تأثيرا كبيرا عند دراسة عقار على مرض عضال.

3- إذا كانت نتائج الدراسة غير دالة إحصائيا يفترض أن يقوم الباحث بتفسير تلك النتيجة وعدم الاكتفاء بالإشارة إلى عدم دلالتها إحصائيا، كما يفترض أن يقوم الباحث بعرض نتيجة اختبار القوة الاحصائية وذلك للتأكد

من أن الاختبار الاحصائي المنفذ لا يعاني من انخفاض القوة، وأن عدم الحصول على الدلالة لم يكن بسبب انخفاض القوة الاحصائية للاختبار المستخدم.

6. قائمة المراجع:

1.6. قائمة المراجع باللغة العربية:

1. رياض عيشوش، استخدام تحليل الميتا في البحوث الادارية –المقاربة النظرية والتطبيقات العملية-. مجلة البحوث الاقتصادية والمالية،2017.
2. عادل بابطين، مشكلات الدلالة الاحصائية في البحث التربوي وحلول بديلة. بحث تكميلي لنيل درجة الماجستير في علم النفس تخصص إحصاء وبحوث، جامعة أم القرى، السعودية،2002.
3. عزت عبد الحميد محمد حسن، الدلالة الاحصائية والدلالة العملية في البحوث، مجلة كلية التربية النوعية للدراسات التربوية والنوعية العدد 1، 2016.
4. منصور فام رشدي، حجم الأثر الوجه المكمل للدلالة الاحصائية. المجلة المصرية للدراسات النفسية، 7، 1997.
5. يحي نصار، استخدام حجم الأثر لفحص الدلالة العملية للنتائج في الدراسات الكمية. مجلة العلوم التربوية والنفسية، 2006.
6. يحي نصار، الدلالة الإكلينيكية للبحوث النفسية والتربوية المستخدمة للتصاميم التجريبية. مجلة الدراسات التربوية والنفسية، 11 .2017
7. يونس أبوجراد حمدي، قوة الاختبارات الاحصائية وحجم الأثر في البحوث التربوية المنشورة في مجلة جامعة القدس المفتوحة للأبحاث والدراسات. مجلة العلوم التربوية والنفسية، 14، 2013.

2.6. قائمة المراجع باللغة الأجنبية:

8. Cohen J., Statistical Power Analysis For the Behavioral Sciences. America: Lawrence Erlbaum Associates,1988.
9. King, J. E. Statistical Significance Versus Practical Significance: Implications for Analyzing CME Data. Poster presented at the annual meeting of the Alliance for Continuing Medical Education, Orlando, FL, February, 2002.
10. Sun S., Wei P., & Lihshing L., A Comprehensive Review of Effect Size Reporting and Interpreting Practices in Academic Journals in Education and Psychology. Journal of Educational Psychology, 102, 2010.
11. Thompson B., "Statistical," "Practical," and "Clinical": How Many Kinds of Significance Do Counselors Need to Consider? JOURNAL OF COUNSELING & DEVELOPMENT, 80,2002.