

## The effect of sample size on the statistical test power

Mohamed Ibrahim Ahmed Al shardi

Faculty of Education || Umm Al-Qura University || KSA

**Abstract:** This study aimed to reveal the effect of size of the sample and the number of independent variables on the strength of the statistical test in some of the statistical methods commonly used in educational studies. To achieve the aims of the study, the researcher used the descriptive approach, and the study tool was represented by a table consisting of a group of rows representing individuals and a group of columns representing variables. The study sample included a random sample of statistical vocabulary that ranged between (10) items and (330) items, depending on the nature of the statistical test used. Through the results it was concluded that with increasing sample size, noted that increasing of test (T) strength for one sample, when the sample size was (n=30) the test power was (0.877) while the value of (Beta) declined to (0.12), as the result of the test (T) for two interrelated samples the size of the sample (n=30) was very appropriate for high- power test (0.852) and the value of (Beta) declined to (0.148), in the test (T) for two independent samples the power of the test was high (0.773 – 0.856) and the value of (Beta) is low (0.228- 0.144) when the sample size between 40 – 50. After reaching these results, some recommendations has been proposed including increasing the size of the sample to the number that gives the strength to the suitable test estimated at about (0.8) in the field of educational sciences, and decrease the value of the error of the second type which is estimated at about (0.2) in the field of educational sciences.

**Keywords:** Sample size- Statistical test power – Effect size- PASS11.

### تأثير حجم العينة على قوة الاختبار الإحصائي

محمد إبراهيم أحمد الشاردي

كلية التربية || جامعة أم القرى || المملكة العربية السعودية

المستخلص: هدفت الدراسة إلى الكشف عن تأثير حجم العينة وعدد المتغيرات المستقلة على قوة الاختبار الإحصائي في بعض الأساليب الإحصائية شائعة الاستخدام في الدراسات التربوية. ولتحقيق أهداف الدراسة استخدم الباحث المنهج الوصفي، وتمثلت أداة الدراسة بجدول يتكون من مجموعة من الصفوف تمثل الأفراد ومجموعة من الأعمدة تمثل المتغيرات، وشملت عينة الدراسة عينات عشوائية من المفردات الإحصائية تراوحت من (10) مفردات إلى (330) مفردة، حسب طبيعة الاختبار الإحصائي. ومن خلال النتائج تم التوصل إلى أنه مع زيادة حجم العينة نلاحظ ازدياد قوة اختبار (ت) لعينة واحدة فعندما كان حجم العينة (n=30) كانت قوة الاختبار تساوي (0.877) بينما انخفضت قيمة (Beta) إلى (0.12)، أما في نتائج اختبار (ت) لعينتين مترابطتين كان حجم العينة (n=30) مناسباً جداً للحصول على قوة اختبار عالية (0.852) وقيمة Beta منخفضة (0.148)، وفي اختبار (ت) لعينتين مستقلتين كانت قوة الاختبار عالية من (0.773 – 0.856) وقيمة (Beta) منخفضة (0.228 - 0.144) عندما كان حجم العينة من (40 - 50). واستناداً للنتائج تم تقديم بعض التوصيات ومنها زيادة حجم العينة إلى العدد الذي يعطي قوة اختبار مناسبة والتي تقدر بحوالي (0.8) في مجال العلوم التربوية، ونقص في قيمة الخطأ من النوع الثاني والذي يقدر بحوالي (0.2) في مجال العلوم التربوية.

الكلمات المفتاحية: حجم العينة- قوة الاختبار الإحصائي – حجم التأثير- PASS11.

## المقدمة.

إن ما نشاهده من الأعداد الضخمة للدراسات المنشورة- عربياً وأجنبياً- وازديادها المطرد، ونظراً لهذا الازدياد والتوسع فإن التوجه لمراجعة الأبحاث أصبح محبباً في كثير من المجالات التربوية، ومن أهم التوجهات تلك التي تركزت على طرق البحث والمعالجات الإحصائية المستخدمة؛ وذلك لأن مجال الإحصاء مجال دائم التغير والنمو والتطور.

وقد تطور علم الإحصاء كنتيجة طبيعية لتطور المعارف الأخرى بسبب زيادة الاهتمام بالبحث وعملية اتخاذ القرارات، وعليه ظهرت العديد من النظريات والأساليب الحديثة (تشاو، 1990) التي من أبرزها أساليب فحص الفرضيات التي تشكل الأساس لاتخاذ القرارات باستخدام نوعين من الدلالات أحدهما إحصائية، والثانية عملية. ويشير الصياد (1988) إلى أن الدلالة الإحصائية شرط ضروري لصناعة قرار تربوي أو نفسي ولكن ليس كافياً، فالكفاية تتحقق بحساب قوة العلاقة بين المتغير المستقل والمتغير التابع، وقوة العلاقة هي ما يسمى بالدلالة العملية (Hays, 1963)، (Cohen, 1977).

ولعل أصعب مرحلة تواجه الباحث في بحثه هي مرحلة التحليل الإحصائي، وذلك لكثرة تعداد أنواع الطرق الإحصائية، ويلاحظ أن بعض الباحثين يعتمدون في تقدير نتائجهم على الدلالة الإحصائية دون محاولة الكشف عن مقدار العلاقة القائمة بين المتغيرين وتصبح هناك مغالاة في تفسير النتائج اعتماداً على مستوى الدلالة على الرغم من أنه ربما لا تكون لها قيمة من الناحية التطبيقية أو العملية (الشريبي، 2007).

وبناءً عليه فإن إهمال حجم التأثير يقلل من أهمية نتائج الدراسة وقد يؤدي إلى التضليل، فالأساس هو تقدير الدلالة العملية في البحث التربوي، خصوصاً إذا علمنا أن قوة الاختبار الإحصائية تكمن في علاقتها بالدلالة الإحصائية والدلالة العملية.

## مشكلة الدراسة:

في الأبحاث التربوية عادةً يتم اللجوء إلى تقنيات الإحصاء الاستدلالي وذلك من أجل تحديد حجم الأثر أو الفرق أو العلاقة بين المتغيرات والتأكد من دلالتها الإحصائية والتي ترضخ لعلامات كثيرة كحجم العينة وعدد المتغيرات وقوة الاختبار ومستوى الدلالة، ولقد ناقشت العديد من الدراسات ضرورة الاهتمام بتحديد الدلالة العلمية (حجم التأثير) وحجم العينة، كدراسة (البابطين، 2002)، و(الصانع، 1997)، و(نور، 1993)، و(النجار، 1991)، وقد أكد الصياد على أهمية قوة الاختبار كعامل مؤثر في كل من حجم العينة والدلالة العملية! (حجم الأثر) (الصياد، 1989)، كما أكدت دراسة (Huston, 1993) التي أكدت على أن الباحثين الذين يلجؤون إلى حساب قوة الاختبارات الإحصائية وحجم التأثير هم الذين يساهمون في تحسين نوعية البحوث.

ومن هذا المنطلق تتحدد مشكلة الدراسة في قصور ملحوظ وغموض يكتنف معظم الدراسات التربوية وإغفال كثير منها لحجم الأثر، وتأثير حجم العينة على مخرجات الأساليب الإحصائية شائعة الاستخدام في الدراسات التربوية، مع الحاجة الماسة لدى معظم الباحثين للتعرف على أهمية حجم التأثير لمتغيراتهم المتعددة.

## أسئلة الدراسة:

ومما سبق يمكن تحديد مشكلة الدراسة في التساؤلات التالية:

- 1- ما تأثير حجم العينة على قوة اختبار (ت) لعينة واحدة؟
- 2- ما تأثير حجم العينة على قوة اختبار (ت) لعينتين مترابطتين؟

- 3- ما تأثير حجم العينة على قوة اختبار (ت) لعينتين مستقلتين؟
- 4- ما تأثير حجم العينة على قوة اختبار تحليل التباين الأحادي (ف) للمقارنة بين المتوسطات الحسابية؟

#### أهداف الدراسة:

تسعى الدراسة الحالية للتعرف على:

1. أهم الطرق الإحصائية لحساب تأثير حجم العينة وعدد المتغيرات على قوة الاختبار الإحصائي في بعض الأساليب الإحصائية شائعة الاستخدام في الدراسات التربوية.
2. واقع تأثير حجم العينة على قوة اختبار (ت) لعينة واحد.
3. واقع تأثير حجم العينة على قوة اختبار (ت) لعينتين مترابطتين، وعينتين مستقلتين.
4. واقع تأثير حجم العينة على قوة اختبار تحليل التباين الأحادي (ف) للمقارنة بين المتوسطات الحسابية.

#### أهمية الدراسة:

يمكن تحديد أهمية البحث في التالي:

- يمكن للبحث الحالي أن يقدم للباحثين المعلومات اللازمة للتعامل مع ظاهرة أغفلها الباحثون وهي الاهتمام بحجم العينة وعدد المتغيرات.
- قد تفيد الباحثين في الاهتمام بالعوامل المؤثرة على قوة الاختبار الإحصائي والتي من المتوقع وجودها في الرسائل العلمية المنشورة واستخدام المناسب في كل منها مع طبيعة البحث وحجم العينة.

#### حدود الدراسة:

تقتصر الدراسة على الحدود الآتية:

- الحدود الموضوعية: تأثير حجم العينة وعدد المتغيرات المستقلة على قوة الاختبار الإحصائي في بعض الأساليب الإحصائية شائعة الاستخدام في الدراسات التربوية.
- الحدود البشرية: عينات عشوائية من المفردات الإحصائية التي تم توليدها تراوحت من (10) مفردات إلى (330) مفردة، حسب طبيعة الاختبار الإحصائي المستخدم.

#### مصطلحات الدراسة:

- حجم العينة: تعرف العينة بأنها: "جزء من مجتمع الدراسة من حيث الخصائص والصفات ويتم اللجوء إليها عندما تغني الباحث عن دراسة كافة وحدات المجتمع" (زويلف، 1998).
- ويعرفها الباحث في الدراسة الحالية على أنها: عامل هام ومؤثر بصورة طردية في قوة الاختبار فزيادة قوة الاختبار تنتج من ازدياد حجم العينة.
- قوة الاختبار الإحصائي: هي "قدرة الاختبار على رفض الفرض الصفري، عندما يكون في حقيقة الأمر خاطئاً، وتكون تلك القوة في صورة احتمال تعتمد على احتمال ارتكاب الخطأ نمط اثنين" (الشريبي، 2007).
- حجم التأثير: هو عبارة عن "قيمة تشير إلى درجة العلاقة بين متغيرات الدراسة ضمن مجتمع محدد مسبقاً هو مجتمع الدراسة بغض النظر عن الأسلوب الإحصائي المستخدم للتعبير عن ذلك الأثر" (نصار، 2002).

## 2- الإطار النظري والدراسات السابقة.

أولاً- الإطار النظري.

قوة الاختبار.

قوة الاختبار تعني رفض الفرض الصفري عندما يكون خاطئاً بالفعل، وهو يدل على المساحة الواقعة تحت المنحنى وبتحديد في منطقة الرفض. ويذكر الشربيني (2007) إن قوة الاختبار في البحوث الإنسانية بخاصة تكون مقبولة إذا كانت تتراوح ما بين (0، 40، 60، 0). كما أشار علام (2010) إلى أن قوة الاختبار تتراوح ما بين (صفر- 1) نظراً لأنها احتمالية. ويرى الباحث زيادة حجم العينة إلى العدد الذي يعطي قوة اختبار مناسبة والتي تقدر بحوالي (0.8) في مجال العلوم التربوية، ونقص في قيمة الخطأ من النوع الثاني والذي يقدر بحوالي (0.2) في مجال العلوم التربوية.

العوامل المؤثرة على قوة الاختبار:

لقد اتفق كلٌّ من (كوهن، 1977)، و(الشربيني، 2007)، و(عودة والخليلي، 2000)، و(علام، 2010) و(الصائغ، 1997) على أن العوامل التالية كعوامل مؤثرة على قوة الاختبار.

1. حجم العينة:

يرى كلٌّ من (Hays,1963) و(Cohen,1977) أن العلاقة بين حجم العينة وقوة الاختبار علاقة مباشرة فالزيادة في حجم العينة تزيد من القوة الإحصائية مع ثبات العوامل الأخرى عندما يكون الفرض الصفري غير صحيح، ويكون القرار رفض الفرض الصفري. ويتأثر احتمال الخطأ من النوع الثاني ( $\beta$ ) وقوة الاختبار باختلاف أو تباين المجتمع وحجم العينة، حيث أنه كلما قل تباين المجتمع أو زاد حجم العينة قل احتمال الخطأ من النوع الثاني ( $\beta$ ) وزادت قوة الاختبار.

2. مستوى الدلالة  $\alpha$ :

يتناسب مستوى الدلالة تناسباً طردياً مع قوة الاختبار، فإذا قل مستوى الدلالة من (0.05) على سبيل المثال إلى (0.01) فإن حدود الرفض (المنطقة الحرجة) تتغير من: ( $Z=1.969$  إلى  $Z=2.58$ ) وهذا يعني ازدياد صعوبة رفض الفرضية الصفرية، بمعنى أن احتمال ارتكاب الخطأ من النوع الأول يقل، لكن في الوقت نفسه يزداد احتمال ارتكاب الخطأ من النوع الثاني آخر الذي يخفض من قوة الاختبار (ابوزينة وآخرون، 2007).

3. علاقة القيمة الحقيقية للمجتمع بقيمته في الفرضية الصفرية:

اتفق كلٌّ من علام (2010)، وعودة والخليلي (2000) على أنه كلما زاد الاختلاف بين المتوسط الحقيقي (الأصلي للمجتمع) والمتوسط الفرضي زادت قوة الاختبار وقل الخطأ ( $\beta$ ).

4. الاختبار بديل والاختبار بديلين:

قوة الاختبار تزداد في حالة تحقق الباحث من صحة الفرض الصفري باختبار إحصائي ذي طرف واحد. فالباحث عندما يحدد المتوسط في ضوء الفرض الصفري، وكان الافتراض صحيحاً، فإن الخطأ من النوع الثاني أقل مما لو استخدم اختبار بطرفين.

5. الاختبارات المعلمية والاختبارات اللامعلمية:

إن الاختبارات المعلمية تعتبر أكثر قوة من الاختبارات اللامعلمية وذلك لأي قيمة معينة من قيم حجم العينة فاحتمالية الوقوع في الخطأ من النوع الثاني  $\beta$  يكون أقل في حالة الاختبارات الإحصائية للفرض الصفري والتي تفترض

اعتدالية التوزيع للمجتمعات وتساوي التباين حيث إن الأساليب المعلمية تراعي القيم الأصلية للدرجات وهذا يجعلها أكثر دقة في التوصل للاحتمالية عندما تتحقق الفروض.

#### 6. حجم التأثير:

قوة الاختبار تزداد كلما ازداد حجم الأثر، وتقل قوة الاختبار كلما قل حجم الأثر، أبو زينة وآخرون (2007). وتضيف الصائغ (1997) أنه كلما زاد حجم التأثير قل حجم العينة اللازم للوصول لقيمة معينة دالة إحصائياً، ويمكن إيجاد قوة الاختبار بالاعتماد على حجم التأثير.

#### ثانياً- حجم التأثير (الدلالة العملية):

تتطلب العديد من المجالات البارزة في مجال العلوم التربوية والنفسية والإنسانية على الباحثين تضمين قيمة حجم الأثر في أبحاثهم (الصائغ، 1997).

فحجم التأثير مرتبط بدلالة الفرق الإحصائي وتكمن في أهميتها العملية والتطبيقية والتفسيرية للقرار المتخذ، فمهمتها مناقشة نتائج البحث ثم اتخاذ قرار عملي يطبق واقعياً دون تكلفة أو جهد للموازنة بين المكاسب والخسارة بصورة عملية، لذلك فحجم التأثير يعتبر الوجه الآخر للدلالة لأنها تتعلق بالمزايا الفعلية الناتجة من تقييم الفرق الإحصائي المحسوب في اختبار دلالة الفرق.

ويذكر الصياد (1988) عن حجم التأثير أنه مؤشر لمدى قدرتنا على استخدام النتائج تفسيرياً أو تطبيقياً، أنه كم التباين الذي أمكن تفسيره للمتغير التابع حينما اعتبرنا المتغير المستقل في علاقة معه أو مؤثر عليه. وقد أشار هس (Haase, 1982) إلى أن مشكلة تقييم الدلالة العملية والدلالة الإحصائية لها تاريخ طويل بعض الشيء.

ويذكر نصار (2002) أن حجم الأثر هو عبارة عن أسلوب إحصائي يشير إلى الدلالة العملية للنتائج الإحصائية. ويؤكد علام (2010) أن على الباحث أن يضع في اعتباره أهمية الدلالة العملية أو التطبيقية والدلالة التفسيرية، لأنها مترابطة ويكمل بعضها الآخر.

فحجم الأثر يعتبر مكمل للدلالة الإحصائية ومقياس إضافي يجب أن يُسلم به الباحث التربوي، ويستعين به بعد الاستنتاج الإحصائي لاختبار دلالة الفرق لاتخاذ قرار صحيح.

#### • حجم التأثير وعلاقته بحجم العينة:

يرى هيس (Hays, 1963) بأن الدلالة الإحصائية لا تشير بالضرورة إلى قوة علاقه بين متغيرات الدراسة بل هي في أحسن الأحوال تشير إلى بعض الارتباط بين هذه المتغيرات دون تحديد مدى أهمية ذلك الارتباط. وأضاف هيس في هذا الخصوص أن العامل الأهم في الحصول على نتائج دالة إحصائياً هو حجم العينه وأنه من المهم التحديد الدقيق لمفهوم قوة علاقه الإحصائية.

#### المؤشرات الإحصائية المستخدمة للدلالة على حجم الأثر في الأساليب الإحصائية المختلفة:

يرى كلٌّ من أبو هاشم (2005)، وماكسويل ودليني (2004)، ونصار (2002)، أنه يمكن تقسيم المؤشرات التي تدل على مقدار حجم الأثر في الدراسات المختلفة إلى نوعين هما:

1. المؤشرات التي تدل على مقدار الأثر الذي يُحدثه المتغير المستقل في المتغير التابع: (Measures of effect size)

وتُعرف هذه المؤشرات أيضاً بأنها الفرق بين المتوسطات (متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة في الدراسات التجريبية) بوحدات معيارية (عوده وخليلي، 2000).

2. المؤشرات التي تدل على قوة العلاقة أو الارتباط بين متغيرات الدراسة: وتستخدم هذه المؤشرات في الأساليب

الإحصائية التي تعتمد على فحص التباين حيث تشير إلى نسبة التباين الذي يمكن تفسيره في المتغير التابع عن

طريق تباين المتغير أو المتغيرات المستقلة. ومن أمثلتها مؤشر مربع إيتا لبيرسون  $\eta^2$  (Pearson, 1905)، ومؤشر مربع أوميغا لهيس  $\omega^2$  (Hays, 1963). وسيقوم الباحث في هذا المبحث باستعراض بعض الأساليب الإحصائية الشائعة الاستخدام وبالتحديد اختبار "ت" (T-test) واختبار تحليل التباين (ANOVA).

#### أولاً- اختبار (ت) T-test.

يعد هذا الاختبار من أكثر الأساليب الإحصائية شيوعاً واستخداماً في مجال البحوث التجريبية التربوية والنفسية، حيث أشار كلٌّ من فرج (1996)، و(الأعسر وآخرون، 2002) إلى أن الفضل يعود في اشتقاق هذا الاختبار إلى العالم الإيرلندي وليلم كوسيت (W.S Gosset) عام 1908م، حيث نشر بحثنا باسم مستعار هو Student وعُرف هذا التوزيع بالاسم ((Students T- Distibution)) ويختصر باسم توزيع (T)، والذي توصل من خلاله إلى اشتقاق معادلة للتوزيع الاحتمالي (ت) حيث قيمته هي:

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

#### ثانياً- تحليل التباين ANOVA.

يذكر علام (2010) أن فيشر Fisher عالم الإحصاء الانجليزي قد صمم هذا الأسلوب لتيسير تحليل البيانات المستمدة من التجارب الميدانية والمختبرية في مجال البحوث الزراعية والبيولوجية، وتفسير نتائجها. ويمثل اليوم أحد أدوات البحث المهمة ليس فقط في هذه المجالات، وإنما في مختلف مجالات العلوم النفسية والتربوية والاجتماعية والطبيعية وغيرها.

كما يذكر أبو حطب وآخرون (1991) بأن تحليل التباين " يعد أسلوباً إحصائياً لازماً لفهم طبيعة المنهج التجريبي وشبه التجريبي".

في حين أشار عدس (1997) إلى أن تحليل التباين " يقدم لنا أسلوباً إحصائياً مناسباً لمقارنة عدة متوسطات مع بعضها البعض في نفس الوقت".

وتعتمد الطريقة الإحصائية لتحليل التباين على حساب تباين الخطأ وذلك بحساب المربعات داخل المجموعات، وحساب التباين المفسر وذلك بحساب المربعات بين المجموعات، وحساب درجات الحرية لتحويل تلك المربعات إلى التباين المقابل لها، وللكشف عن الدلالة الإحصائية للنسبة الفائية، حساب النسبة الفائية، والكشف عن دلالتها الإحصائية (السيد، 2005)

تقدير تباين المجتمع باستخدام العينة الأولى	=	حيث بين علام (2010) بأن النسبة الفائية الملاحظة
تقدير تباين المجتمع باستخدام العينة الثانية		
متوسط مجموع المربعات بين المجموعات	=	والنسبة الفائية
متوسط مجموع المربعات بين المجموعات		

أسباب استخدام تحليل التباين بدلاً من استخدام اختبار (ت):

يفضل استخدام تحليل التباين (F) بدلاً من اختبار (T) وذلك للأسباب التالية:

$$\frac{(1-n)n}{2}$$

- الجهد المبذول في عمل المقارنات، عدد المقارنات =  $\frac{(1-n)n}{2}$
- حيث أنه كلما زاد عدد العينات زاد عدد مرات إجراء اختبار النسبة زيادة ملحوظة.
- ضعف عملية المقارنة، في اختبار (T) يتم المقارنة بين كل متوسطين لمجموعتين على حده وبالتالي تهمل بقية المعلومات عن المجموعات الأخرى مؤقتاً والتي من الواجب أخذها بعين الاعتبار لأنها جزء لا ينفصل وبالتالي فهي تؤثر على قوة المقارنة.
- مخاطرة الوقوع في خطأ من النوع الأول، ذكر الشربيني (2007) بأن " عدد المقارنات ومستوى الدلالة يرتبطان باحتمالية الوقوع أو ارتكاب خطأ أو أكثر من النوع الأول طبقاً للعلاقة التالية:  
احتمالية الوقوع في خطأ من النوع الأول =  $n - (1 - \alpha)$   
عدد المقارنات =  $n$  مستوى الدلالة المستخدم في هذه المقارنات =  $\alpha$

#### ثانياً- الدراسات السابقة

- أ- دراسات بالعربية:
- دراسة بريمة (2016): التي هدفت إلى تسليط الضوء على الطرائق التي تؤثر في قوة الاختبار الإحصائي (حجم العينة، الانحراف المعياري، توزيع المجتمع) واختيار أفضل الطرائق التي تزيد من قوة الاختبار. وقد تم اتباع منهج وصفي واستدلالي. تم توليد (10000) مفردة عشوائياً باستخدام أسلوب المحاكاة من برنامج PASS12 بحيث اختيرت عينات عشوائية بالأحجام المختلفة. وكشفت نتائج الدراسة أنه بزيادة حجم العينة فإن قوة اختبار ANOVA تزداد وفي الحالات التي تماثل بيانات البحث فإن حجم العينة من 40 - 50 كان مناسباً للحصول على قوة اختبار عالية.
- دراسة حمادنة (2015) التي هدفت إلى تقييم القوة الإحصائية التي تم استخدامها في اختبار فرضية العدم وحجم التأثير في البحوث التربوية والنفسية المنشورة في مجلة المنارة للبحوث والدراسات في جامعة آل البيت من 2010 - 2014. اتبع الباحث المنهج الوصفي الاستدلالي وشملت الدراسة جميع الأبحاث التي تستخدم اختبائي t، (F)، وكان عدد هذه المقالات والبحوث (87)، وشملت (452) من الاختبارات الإحصائية، حيث استخدمت (231) منها T-test، وكشفت النتائج أن 63% من الفرضيات المحددة كان قوة الاختبار الإحصائي لها ضعيف، وأن استخدام قوة الاختبار الإحصائي أفضل من الاعتماد على الدلالة الإحصائية فقط.
- دراسة العبد القادر (2008)، التي هدفت إلى التعرف على أهم الطرق الإحصائية لحساب حجم التأثير للاختبارات المعلمية واللامعلمية المستخدمة في رسائل الماجستير بكلية التربية بجامعة الملك سعود، وتضمنت عينة الدراسة (192) رسالة ماجستير وذلك بواقع (20%) من حجم الرسائل الصادرة بكلية التربية بجامعة الملك سعود. وتوصلت الدراسة إلى أن أكثر الأساليب الإحصائية استخداماً في كلية التربية بجامعة الملك سعود هي الأساليب المعلمية،
- دراسة النجار (2005)، التي هدفت إلى تقويم استخدام الباحثين في الأبحاث الإدارية المنشورة في مجلة جامعة الملك سعود- العلوم الإدارية للدلالة العملية في اختبار  $\chi^2$ ، وقد أظهرت النتائج أن واقع الدلالة العملية المصاحبة للدلالة الإحصائية لاختبار  $(\chi^2)$  كانت عند حجم التأثير المنخفض، بالإضافة إلى ذلك أظهرت النتائج

أن زيادة حجم العينة قد يكون له تأثير سالب على الدلالة العملية، وأن زيادة حجم العينة في اختبار  $\chi^2$  فوق (200) أمر مكلف دون فائدة تذكر على مستوى الدلالة العملية المتحصل عليها، لذا فأن التعرف على حجم التأثير المناسب عند إجراء الدراسة سيساعد الباحثين على تخفيض التكاليف لمجمل أبحاثهم.

- دراسة بابطين (2002)، التي هدفت الدراسة إلى الكشف عن أهم مشكلات الدلالة الإحصائية في البحث التربوي والتعرف على أهم المفاهيم والأساليب الإحصائية التي يمكن أن تقدم حلولاً لتلك المشكلات، وكان أهم ما توصلت إليه الدراسة: أن أهم مشكلات اختبارات الدلالة الإحصائية في البحث التربوي هي: استخدام نتائج الدلالة الإحصائية كتفسير لأهمية النتائج، وكتفسير لتأثير المعالجة، واستخدام قيمة الاحتمالية (P) لتقدير حجم التأثير، وكتفسير لاحتمالية الفرضية الصفرية نفسها، ولتقويم إعادة النتيجة، إضافة إلى التحيز للنتائج الدالة إحصائياً، وصياغة التساؤلات البحثية والفرضيات، والطبيعة الثنائية لنتائج اختبار الدلالة الإحصائية.
- دراسة نصار (2002)، التي هدفت إلى تقديم مفهوم "حجم الأثر" كأسلوب إحصائي مكمل لفحص الفرضيات الإحصائية. وقد تضمنت الدراسة تحليلاً رقمياً لتبرهن أن حجم الأثر أقل تأثيراً بحجم العينة من الأساليب الإحصائية التي تستخدم لفحص الفرضيات. بالإضافة إلى ذلك قدمت الدراسة بعض المؤشرات الإحصائية التي تستخدم للدلالة على قيمة حجم الأثر في حالة استخدام بعض الأساليب الإحصائية وبالتحديد اختبار "ت" واختبار "ف" ومعامل ارتباط بيرسون وأخيراً الانحدار البسيط.
- دراسة حجيمات وعليان (1997)، والتي هدفت إلى تناول واقع الدلالة الإحصائية وعلاقتها بكل من الدلالة العملية وقوة وتكونت العينة من (188) فرضية صفرية، استخدم في فحصها الإحصائي (ف) وذلك في (28) رسالة من رسائل الماجستير التي استخدمت تصاميم تجريبية وشبه تجريبية في مجال الإرشاد التربوي والنفسي. جمعت البيانات اللازمة لحساب مربع آيتا وقوة الاختبار الإحصائي، وأشارت النتائج إلى أن حوالي 85% من الفرضيات الصفرية أُستخدم في فحصها مستوى دلالة إحصائية 0، 05 وأما الباقي فاستخدم في فحصها مستوى دلالة 0، 01 وأن حوالي 63% من الفرضيات التي كانت دالة إحصائية لم تكن ذات دلالة عملية.
- دراسة العجلان (1990)، التي هدفت إلى التعرف على مدى ملائمة الأساليب الإحصائية في الدراسات والبحوث التي أجريت في جامعة أم القرى بكلية التربية لما ينبغي أن يكون من حيث (نوع المتغيرات، ونوع التصميم، وحجم العينة) في كل منها وتقويم هذه الأساليب في ضوء المعايير التي يجب أن تؤخذ بعين الاعتبار وتوصلت الدراسة إلى أنه قد استخدم اختبار (كاي تربيع) 82 مرة منها 31 مرة استخدام مناسب أي بنسبة 37.8% و 51 مرة غير مناسب، و 24 مرة يرجع السبب فيها إلى عدم ملائمة نوع المتغيرات للأسلوب الإحصائي، بينما 27 يرجع السبب إلى عدم ملائمة كل من حجم العينة ونوع المتغيرات معاً.

#### ب- دراسات بالإنجليزية:

- دراسة (Hedges and Rhoads, 2010)، التي قدمت دليلاً لحساب القوة الإحصائية للتصاميم متعددة المستويات المعقدة التي تستخدم في معظم الدراسات الميدانية في مجال البحوث والتعميم وضحت الدراسة كيفية تصميم عينة عشوائية بدافع الاختلاف في الكميات التي لوحظت في تجربة عشوائية. كما تبين أن القوة الإحصائية في التصاميم المعقدة التي تنطوي على أخذ عينات عنقودية يمكن حسابها ببساطة من الجداول القياسية باستخدام القوة يتم تطبيق هذه المفاهيم على توفير طرائق لحساب القوة لكل من تصميمات البحوث الأكثر استخداماً في مجال البحوث والتعميم.

- دراسة (Thompson.2000)، التي هدفت الدراسة إلى فحص مدى إشارة مجلة (الأطفال الاستثنائيين) لحجم التأثير، وقد اشتملت عينة الدراسة المقالات المنشورة من المجلد رقم (63) في عام 1669م إلى مجلد (64) في عام 1998م، حيث بلغ عدد المقالات المنشورة (23) مقالاً. وتوصل الباحث إلى أن (20) مقالاً من (23) لم تذكر أي مؤشر لحجم التأثير. كما أشار إلى ضرورة تضمين الدراسات لمؤشرات عن حجم التأثير وفقاً لتعليمات (APA) الجمعية النفسية الأمريكية.
- دراسة (Lane.1999)، التي أشارت إلى وجود بعض الأخطاء الشائعة بشأن تفسير اختبار دلالة الفرض الصفري، حيث إن هناك الكثير من الباحثين الذين قاموا بتضخيم نتائج اختبار تلك الدلالة، وفي نفس الوقت أهملوا تماماً مؤشرات حجم التأثير، بينما البعض الآخر منهم يرى أن الدلالة الإحصائية تعكس حجم التأثير الكلي لنتائج الدراسة. وتعتبر لأن أن كل من الدلالة الإحصائية وحجم التأثير مكملان لبعضهما من حيث الإشارة إلى القوة الإحصائية لنتائج الدراسة.
- دراسة (Plucker, 1997)، التي قامت الدراسة بدراسة ثلاثة مجالات علمية ربع سنوية وأربعين مقالاً في مجالات أخرى نشرت خلال الخمس سنوات الأخيرة من الدراسة. وتوصلت الدراسة إلى أن تلك الأبحاث لا تتضمن في العادة أي مقاييس لحجوم التأثير، كما وجد أن عبارة (عالية الدلالة) موجودة في أغلب تلك المجالات والمقالات المنشورة. وكذلك توصل الباحث إلى أهمية إدراك الباحثين لمفهوم الدلالة الإحصائية وذلك لتحسين تفسير النتائج.
- دراسة (Snyder & Lawson: 1992) التي هدفت إلى تناول مقاييس حجم التأثير وكذلك تناولت الدراسة أسباب تشجيع المختصين في البحوث التربوية على استخدام حجم التأثير كمساعد تفسيري، والأنواع المختلفة من قياسات حجم التأثير مع تصنيفها إلى مجموعات حسب مؤشرات المجتمعات والعينات، والتحيز وعدم التحيز، وكذلك مؤشرات نماذج التصاميم الثابتة مقابل العشوائية. توصلت الدراسة إلى أن مقاييس حجم التأثير إذا استخدمت بطريقة صحيحة فأنها تعتبر مساعدة للباحثين الذين لا يرغبون في الاعتماد فقط على اختبارات الدلالة الإحصائية في تفسير النتائج المهمة.

#### التعليق على الدراسات السابقة:

من خلال استعراض الدراسات السابقة (العربية والأجنبية) تبين أن الدلالة الإحصائية وحجم الأثر مكملان لبعضهما من حيث الإشارة إلى قوة الاختبار الإحصائي لنتائج الدراسة، وأن قوة الاختبار في الأساليب الإحصائية المستخدمة كانت ما بين متوسطة إلى عالية مثل دراسة حمادنة (2015)، ودراسة نصار (2002)، ودراسة إبراهيم (1998)، ودراسة (Lane، 1999). كما تبين بأن قيمة مؤشرات الدلالة العملية (حجم التأثير) لاختبارات الدلالة الإحصائية كانت منخفضة وتحديداً في رسائل الماجستير في الجامعات السعودية وهذا في دراسة كل من العبد القادر (2008)، والنجار (2005) وباطين (2002)، والعجلان (1990).

وفي دراسة (Thompson، 2000) يشار إلى ضعف في ذكر المؤشرات لحجم التأثير في المقالات المنشورة حيث أنه من عام 1669م إلى 1998م، وهذا ما توصل إليه أيضاً (Plucker 1997)، حيث وجد أنه خلال خمس سنوات من الدراسة لم تتضمن ثلاث مجلات علمية ربع سنوية وأربعين مقالاً أي مقاييس لحجم التأثير، وقد أوصت دراسة العبد القادر (2008) بضرورة تضمين تدريس طرق حساب حجم التأثير في المقررات الدراسية.

واتفقت الدراسة الحالية مع الدراسات السابقة مثل دراسة بريمة (2016)، وحمادنة (2015)، نصار (2002)، وإبراهيم (1998)، ودراسة (Snyder & Lawson.1992) في تناول بعض الاختبارات الإحصائية المعلمية

الشائعة الاستخدام. حيث ركزت الدراسة الحالية على أثر حجم العينة على قوة الاختبار الإحصائية وهذا ما أوصت به بعض الدراسات السابقة.

### 3- منهجية الدراسة وإجراءاتها.

#### منهجية الدراسة:

بناء على مشكلة الدراسة، حدد الباحث المنهج الملائم للدراسة الحالية وهو المنهج الوصفي التحليلي والذي أشار العساف (2006) إلى أنه " لا يقتصر على جمع البيانات وتبويبها وإنما يمضي إلى ما هو أبعد من ذلك لأنه يتضمن قدراً من التفسير لهذه البيانات".

#### مجتمع الدراسة وعينتها:

تكون مجتمع الدراسة الحالية من عدد من المشاهدات الإحصائية تم توليدها بواسطة برنامج (PASS11) وتراوح مجتمع الدراسة من (100) مفردة إحصائية إلى (330) مفردة إحصائية، حسب طبيعة الأسلوب الإحصائي المستخدم، وتراوحت العينة من (10) مفردات إلى (330) مفردة وروعي في الأوامر المعطاة لبرنامج (PASS11)، أن تتحقق في هذه البيانات الافتراضات اللازمة لإجراء التحليلات الإحصائية المطلوبة.

#### أداة الدراسة:

كما هو الحال في دراسات المحاكاة، فإن أداة الدراسة عبارة عن جدول يتكون من مجموعة من الصفوف تمثل الحالات (الأفراد) ومجموعة من الأعمدة تمثل المتغيرات.

#### صدق وثبات الأداة:

أداة الدراسة الحالية، لا تحتاج إلى حساب مؤشرات الصدق والثبات، حيث إنها ليست استبانة أو اختبار، وإنما تم التأكد من طبيعة البيانات وملاءمتها للاختبارات الإحصائية في الدراسة الحالية؛ مما يعطي مصداقية للنتائج المتحصل عليها.

#### الأساليب الإحصائية:

للإجابة عن تساؤلات الدراسة تم استخدام الأساليب الإحصائية الآتية:

1. اختبار (ت) لعينة واحدة، ولعینتين مترابطتين، وعینتين مستقلتين مع تساوي حجم العينات، ولعینتين مترابطتين مع عدم تساوي حجوم العينات.
2. اختبار تحليل التباين الأحادي (ف) مع تساوي حجوم العينات، ومع عدم تساوي حجوم العينات.
3. قوة الاختبار الإحصائي.
4. الخطأ من النوع الثاني.

#### 4- عرض النتائج ومناقشتها.

● إجابة التساؤل الأول: "ما تأثير حجم العينة على قوة اختبار (ت) لعينة واحدة؟

وللإجابة على هذا التساؤل، تم توليد مجتمع من البيانات الإحصائية بحجم (N=100)، ثم أخذ عينات بأحجام من (n=10) إلى (n=100) وبمعدل زيادة (n=10)، بمتوسط عينة (M<sub>0</sub>=80) وبتوسط فرضي (M<sub>1</sub>=90)، وانحراف معياري (S=20)، ومجتمع البيانات الإحصائية، واستخدم اختبار (ت) لعينة واحدة لاختبار الفرض الصفري بعدم وجود فروق ذات دلالة إحصائية بين متوسط العينة والمتوسط الفرضي ((H<sub>0</sub>: μ<sub>0</sub> - μ<sub>1</sub> = 0) عند مستوى (Alpha) تساوي (0.05) ضد الفرض البديل وجود فروق ذات دلالة إحصائية بين متوسط العينة والمتوسط الفرضي ((H<sub>1</sub>: μ<sub>0</sub> - μ<sub>1</sub> ≠ 0)، ثم حساب قوة الاختبار الإحصائي (Power)، وفيما يلي عرض للنتائج:

جدول رقم (1) تأثير حجم العينة على قوة اختبار (ت) لعينة واحدة

Power	S	Mean1	Mean0	Beta	Alpha	n	Test
0.349	20	90	80	0.65	0.05	10	One-Sample T- Test Power Analysis
0.678	20	90	80	0.32	0.05	20	
0.877	20	90	80	0.12	0.05	30	
0.975	20	90	80	0.03	0.05	40	
0.997	20	90	80	0	0.05	50	
0.999	20	90	80	0	0.05	60	
1.000	20	90	80	0	0.05	70	
1.000	20	90	80	0	0.05	80	
1.000	20	90	80	0	0.05	90	
1.000	20	90	80	0	0.05	100	

تشير النتائج السابقة إلى وجود تأثير كبير لحجم العينة على قوة اختبار (ت) لعينة واحدة، حيث لوحظ أنه عندما كان حجم العينة (n=10) كانت قوة الاختبار تساوي (0.349) وقيمة (Beta) تساوي (0.65). ومما سبق يمكن استنتاج أن هناك تأثير كبير لحجم العينة على قوة اختبار (ت) لعينة واحدة، ومع زيادة حجم العينة تزداد قوة اختبار (ت) لعينة واحدة، وفي الحالات التي تماثل بيانات البحث الحالي فإن حجم العينة (30) كان مناسباً جداً للحصول على قوة اختبار عالية (0.877) وقيمة (Beta) منخفضة (0.12).

● إجابة التساؤل الثاني: ونصه "ما تأثير حجم العينة على قوة اختبار (ت) لعينتين مترابطتين؟

للإجابة على هذا التساؤل، تم توليد مجتمع من البيانات الإحصائية بحجم (N=100)، ثم أخذ عينات بأحجام من (n=10) إلى (n=100) وبمعدل زيادة (n=10) بمتوسط حسابي للعينة الأولى (M<sub>1</sub>=80) وبتوسط حسابي للعينة الثانية (M<sub>2</sub>=90)، وانحراف معياري (S=20)، ومجتمع البيانات الإحصائية يتوزع توزيعاً معتدلاً، وهناك استقلالية في اختبار العينات. واستخدم اختبار (ت) لعينتين مترابطتين لاختبار الفرض الصفري بعدم وجود فروق ذات دلالة إحصائية بين متوسطي العينتين ((H<sub>0</sub>: μ<sub>1</sub> - μ<sub>2</sub> = 0) عند مستوى دلالة (0.05) ضد الفرض البديل وجود فروق ذات دلالة إحصائية بين متوسطي العينتين ((H<sub>1</sub>: μ<sub>1</sub> - μ<sub>2</sub> ≠ 0)، ومن ثم حساب قوة الاختبار الإحصائي وفيما يلي عرض للنتائج:

جدول رقم (2) تأثير حجم العينة على قوة اختبار (ت) لعينتين مترابطتين

Power	S	Mean <sub>2</sub>	Mean <sub>1</sub>	Beta	Alpha	n	Test
0.284	20	90	80	0.716	0.05	10	Paired Means Power Analysis
0.630	20	90	80	0.370	0.05	20	
0.852	20	90	80	0.148	0.05	30	
0.968	20	90	80	0.032	0.05	40	
0.996	20	90	80	0.004	0.05	50	
1.000	20	90	80	0.000	0.05	60	
1.000	20	90	80	0.000	0.05	70	
1.000	20	90	80	0.000	0.05	80	
1.000	20	90	80	0.000	0.05	90	
1.000	20	90	80	0.000	0.05	100	

تشير النتائج السابقة إلى وجود تأثير كبير لحجم العينة على قوة اختبار (ت) لعينتين مترابطتين، ومما سبق يمكن استنتاج أن هناك تأثير كبير لحجم العينة على قوة اختبار (ت) لعينتين مترابطتين، ومع زيادة حجم العينة تزداد قوة اختبار (ت) لعينتين مترابطتين، وفي الحالات التي تماثل بيانات البحث الحالي فإن حجم العينة (30) كان مناسباً جداً للحصول على قوة اختبار عالية (0.852) وقيمة (Beta) منخفضة (0.148).

• إجابة التساؤل الثالث: " ما تأثير حجم العينة على قوة اختبار (ت) لعينتين مستقلتين؟

وللإجابة على هذا التساؤل، تم توليد مجتمع من البيانات الإحصائية بحجم ((N=200، ثم أخذ جميع العينات الممكنة بأحجام من (n=10) إلى (n=100) وبمعدل زيادة (n=10)، بمتوسط حسابي للعينة الأولى (M<sub>1</sub>=80) ومتوسط حسابي للعينة الثانية (M<sub>2</sub>=90)، وانحراف معياري للعينة الأولى (S<sub>1</sub>=10)، وانحراف معياري للعينة الثانية (S<sub>2</sub>=20)، ومجتمع البيانات الإحصائية يتوزع توزيعاً معتدلاً، وهناك استقلالية في اختيار العينات، وتجانس للتباين. واستخدم اختبار (ت) لعينتين مستقلتين لاختبار الفرض الصفري بعدم وجود فروق ذات دلالة إحصائية بين متوسطي العينتين (H<sub>0</sub>: μ<sub>1</sub>- μ<sub>2</sub> = 0) عند مستوى دلالة (0.05) ضد الفرض البديل وجود فروق ذات دلالة إحصائية بين متوسطي العينتين (H<sub>1</sub>: μ<sub>1</sub>- μ<sub>2</sub> ≠ 0)، ومن ثم حساب قوة الاختبار الإحصائي (Power) وفيما يلي عرض للنتائج:

أولاً- في حالة تساوي أحجام العينات (n<sub>1</sub> = n<sub>2</sub>)

جدول (3) تأثير حجم العينة على قوة اختبار (ت) لعينتين مستقلتين في حالة تساوي أحجام العينات

Power	S <sub>2</sub>	S <sub>1</sub>	Mean <sub>2</sub>	Mean <sub>1</sub>	Beta	Alpha	n <sub>2</sub>	n <sub>1</sub>	Test
0.237	20	10	90	80	0.763	0.05	10	10	Two Ind. Sample Means Power Analysis
0.469	20	10	90	80	0.531	0.05	20	20	
0.637	20	10	90	80	0.363	0.05	30	30	
0.773	20	10	90	80	0.227	0.05	40	40	

Power	S <sub>2</sub>	S <sub>1</sub>	Mean <sub>2</sub>	Mean <sub>1</sub>	Beta	Alpha	n <sub>2</sub>	n <sub>1</sub>	Test
0.856	20	10	90	80	0.144	0.05	50	50	
0.916	20	10	90	80	0.084	0.05	60	60	
0.949	20	10	90	80	0.051	0.05	70	70	
0.972	20	10	90	80	0.028	0.05	80	80	
0.983	20	10	90	80	0.017	0.05	90	90	
0.991	20	10	90	80	0.009	0.05	100	100	

تشير النتائج السابقة إلى وجود تأثير كبير لحجم العينة على قوة اختبار (ت) لعينتين مستقلتين، ومما سبق يمكن استنتاج أن هناك تأثير كبير لحجم العينة على قوة اختبار (ت) لعينتين مستقلتين، ومع زيادة حجم العينة تزداد قوة اختبار (ت) لعينتين مستقلتين، وفي الحالات التي تماثل بيانات البحث الحالي فإن حجم العينة من (40-50) كان مناسباً جداً للحصول على قوة اختبار عالية من (0.773 - 0.856) وقيمة (Beta) منخفضة (0.228 - 0.144).

ثانياً- في حالة عدم تساوي حجوم العينات ( $n_1 \neq n_2$ )

الحالة الأولى: ( $n_1 = 10$ ), ( $100 \geq n_2 \geq 10$ )

جدول (4) تأثير حجم العينة على قوة اختبار (ت) لعينتين مستقلتين

في حالة عدم تساوي حجوم العينات ( $n_1 = 10$ ), ( $100 \geq n_2 \geq 10$ )

Power	S <sub>2</sub>	S <sub>1</sub>	Mean <sub>2</sub>	Mean <sub>1</sub>	Beta	Alpha	n <sub>2</sub>	N <sub>1</sub>	Test
0.237	20	10	90	80	0.763	0.05	10	10	Two Ind. Sample Means Power Analysis
0.399	20	10	90	80	0.601	0.05	20	10	
0.485	20	10	90	80	0.515	0.05	30	10	
0.547	20	10	90	80	0.453	0.05	40	10	
0.584	20	10	90	80	0.416	0.05	50	10	
0.614	20	10	90	80	0.386	0.05	60	10	
0.634	20	10	90	80	0.366	0.05	70	10	
0.650	20	10	90	80	0.350	0.05	80	10	
0.662	20	10	90	80	0.338	0.05	90	10	
0.672	20	10	90	80	0.328	0.05	100	10	

تشير النتائج السابقة إلى وجود تأثير كبير لحجم العينة على قوة اختبار (ت) لعينتين مستقلتين، ومما سبق يمكن استنتاج أن هناك تأثير كبير لحجم العينة على قوة اختبار (ت) لعينتين مستقلتين في حالة عدم تساوي حجوم العينات، ومع زيادة حجم العينة تزداد قوة اختبار (ت) لعينتين مستقلتين، وفي الحالات التي تماثل بيانات البحث الحالي فإن النتائج تشير إلى أهمية أن حجوم العينات كبيرة في كل العينتين.

الحالة الثانية ( $n_1 = 20, n_2 \geq 10, 100 \geq n_2$ ).

جدول (5) تأثير حجم العينة على قوة اختبار (ت) لعينتين مستقلتين

في حالة عدم تساوي حجوم العينات ( $n_1 = 20, n_2 \geq 10, 100 \geq n_2$ )

Power	$S_2$	$S_1$	Mean <sub>2</sub>	Mean <sub>1</sub>	Beta	Alpha	$n_2$	$n_1$	Test
0.252	20	10	90	80	0.748	0.05	10	20	Two Ind. Sample Means Power Analysis
0.469	20	10	90	80	0.531	0.05	20	20	
0.599	20	10	90	80	0.401	0.05	30	20	
0.696	20	10	90	80	0.304	0.05	40	20	
0.756	20	10	90	80	0.244	0.05	50	20	
0.803	20	10	90	80	0.197	0.05	60	20	
0.833	20	10	90	80	0.167	0.05	70	20	
0.859	20	10	90	80	0.141	0.05	80	20	
0.876	20	10	90	80	0.124	0.05	90	20	
0.891	20	10	90	80	0.109	0.05	100	20	

مما سبق يمكن استنتاج أن هناك تأثيراً كبيراً لحجم العينة على قوة اختبار (ت) لعينتين مستقلتين، ومع زيادة حجم العينة تزداد قوة اختبار (ت) لعينتين مستقلتين، وفي الحالات التي تماثل بيانات البحث الحالي فإن حجم العينة الأولى ( $n_1=20$ ) و ( $n_2=60$ ) كان مناسباً للحصول على قوة اختبار عالية (0.803) وقيمة (Beta) منخفضة (0.197).

- إجابة التساؤل الرابع: ما تأثير حجم العينة على قوة اختبار تحليل التباين الأحادي (ف) للمقارنة بين المتوسطات الحسابية؟

أولاً- حجوم العينات متساوية ( $n_1 = n_2 = n_3$ )

تم توليد مجتمع من البيانات الإحصائية بحجم ( $N=300$ )، ثم أخذ ثلاثة عينات، كل عينة بأحجام من ( $n=10$ ) إلى ( $n=100$ ) وبمعدل زيادة ( $n=10$ )، بمتوسط حسابي للعينة الأولى ( $M_1=60$ ) والعينة الثانية ( $M_2=70$ ) والعينة الثالثة ( $M_2=90$ )، ومجتمع البيانات الإحصائية يتوزع توزيعاً معتدلاً، وهناك استقلالية في اختيار العينات، وتجانس للتباين بين المجموعات. واستخدام اختبار (ف) لاختبار الفرض الصفري بعدم وجود فروق ذات دلالة إحصائية بين متوسطات العينات الثلاث، (Alpha) تساوي (0.05) ضد الفرض البديل وجود فروق ذات دلالة إحصائية بين متوسطات العينات الثلاث ( $H_0: \mu_1 = \mu_2 = \mu_3$ )، ومن ثم حساب قوة الاختبار الإحصائي (Power)، وفيما يلي عرض للنتائج:

أولاً- في حالة تساوي أحجام العينات ( $n_1 = n_2 = n_3$ )

جدول (6) تأثير حجم العينة على قوة اختبار (ف) لثلاث عينات مستقلة في حالة تساوي حجوم العينات

Power	Beta	Alpha	Mean <sub>3</sub>	Mean <sub>2</sub>	Mean <sub>1</sub>	N	n <sub>3</sub>	n <sub>2</sub>	n <sub>1</sub>	Test
0.225	0.775	0.05	90	70	60	30	10	10	10	One Way ANOVA Power Analysis
0.435	0.565	0.05	90	70	60	60	20	20	20	
0.617	0.383	0.05	90	70	60	90	30	30	30	
0.754	0.246	0.05	90	70	60	120	40	40	40	
0.849	0.151	0.05	90	70	60	150	50	50	50	
0.910	0.090	0.05	90	70	60	180	60	60	60	
0.949	0.051	0.05	90	70	60	210	70	70	70	
0.971	0.029	0.05	90	70	60	240	80	80	80	
0.984	0.016	0.05	90	70	60	270	90	90	90	
0.992	0.008	0.05	90	70	60	300	100	100	100	

تشير النتائج السابقة إلى وجود تأثير كبير لحجم العينة على قوة اختبار (ف) للمقارنة بين مجموعة من المتوسطات الحسابية عند تساوي حجوم العينات لكل مجموعة، ومما سبق يمكن استنتاج أن هناك تأثير كبير لحجم العينة على قوة اختبار (ف) للمقارنة بين مجموعة من المتوسطات الحسابية، ومع زيادة حجم العينة تزداد قوة الاختبار.

#### ثانيا- في حالة عدم تساوي حجوم العينات ( $n_1 \neq n_2 \neq n_3$ )

تم توليد مجتمع من البيانات الإحصائية بحجم (N=330)، ثم أخذ ثلاثة عينات، كل عينة بأحجام من (n=10) إلى (n=100) وبمعدل زيادة (n=10)، مع مراعاة عدم تساوي حجوم العينات، وبمتوسط حسابي للعينة الأولى (M<sub>1</sub>=60) والعينة الثانية (M<sub>2</sub>=70) والعينة الثالثة (M<sub>2</sub>=90)، ومجتمع البيانات الإحصائية يتوزع توزيعاً معتدلاً، وهناك استقلالية في اختيار العينات، وتجانس للتباين بين المجموعات. واستخدام اختبار (ف) لاختبار الفرض الصفري بعدم وجود فروق ذات دلالة إحصائية بين متوسطات العينات الثلاث (H<sub>0</sub>:  $\mu_1 = \mu_2 = \mu_3$ ) عند مستوى (Alpha) تساوي (0.05) ضد الفرض البديل وجود فروق ذات دلالة إحصائية، ومن ثم حساب قوة الاختبار الإحصائي (Power)، وفيما يلي عرض للنتائج:

جدول (7) تأثير حجم العينة على قوة اختبار (ف) لثلاث عينات مستقلة في حالة عدم تساوي حجوم العينات

Power	Beta	Alpha	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	N	n <sub>3</sub>	n <sub>2</sub>	n <sub>1</sub>	Test
0.352	0.648	0.05	90	70	60	60	30	20	10	One Way ANOVA Power Analysis
0.537	0.463	0.05	90	70	60	90	40	30	20	
0.671	0.329	0.05	90	70	60	120	50	40	30	
0.784	0.216	0.05	90	70	60	150	60	50	40	
0.856	0.144	0.05	90	70	60	180	70	60	50	
0.914	0.086	0.05	90	70	60	210	80	70	60	
0.944	0.056	0.05	90	70	60	230	90	80	70	

Power	Beta	Alpha	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	N	n <sub>3</sub>	n <sub>2</sub>	n <sub>1</sub>	Test
0.963	0.037	0.05	90	70	60	270	100	90	80	
0.974	0.026	0.05	90	70	60	300	110	100	90	
0.985	0.015	0.05	90	70	60	330	120	110	100	

تشير النتائج السابقة إلى وجود تأثير كبير لحجم العينة على قوة اختبار (ف) للمقارنة بين مجموعة من المتوسطات الحسابية عند عدم تساوي حجوم العينات لكل مجموعة، ومما سبق يمكن استنتاج التأثير الكبير لحجم العينة على قوة اختبار (ف) للمقارنة بين مجموعة من المتوسطات مع عدم تساوي حجوم العينات، ومع زيادة حجم العينة تزداد قوة الاختبار، وفي الحالات التي تماثل بيانات البحث الحالي فحجم العينة (150 ≤ n ≤ 180) كان مناسباً للحصول على قوة اختبار عالية تراوحت من (0.784) إلى (0.856) وقيمة (Beta) منخفضة تراوحت من (0.329) إلى (0.216).

### التوصيات والمقترحات.

استناداً لنتائج الدراسة، يوصي الباحث ويقترح ما يلي:

- 1- زيادة حجم العينة إلى العدد الذي يعطي قوة اختبار مناسبة والتي تقدر بحوالي (0.8) في مجال العلوم التربوية، ونقص في قيمة الخطأ من النوع الثاني والذي يقدر بحوالي (0.2) في مجال العلوم التربوية.
- 2- عدم المبالغة في زيادة حجم العينة عن الحد الذي يعطي قوة اختبار مناسبة.
- 3- مراعاة أن يكون حجوم العينات مناسبة لنوع الاختبار الإحصائي المستخدم.
- 4- إجراء دراسة عن تأثير حجم العينة على حجم الأثر في بعض الأساليب الإحصائية.
- 5- إجراء دراسة عن العلاقة بين حجم الأثر وقوة الاختبار.

### قائمة المراجع.

أولاً- المراجع بالعربية:

- إبراهيم، أطفاف (1998). واقع الدلالات الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير علم النفس التربوي في الجامعة الأردنية. رسالة ماجستير، الأردن، الجامعة الأردنية.
- أبو حطب، فؤاد؛ وصادق، أمال (1991). مناهج البحث وطرق التحليل الإحصائي في العلوم النفسية والتربوية والاجتماعية. القاهرة، الأنجلو المصرية.
- ابوزينه، فريد والشايب، عبد الحافظ وعبابنة، عماد والنعيمي، محمد عبدالعال (2007): مناهج البحث العلمي الإحصاء في البحث العلمي. عمان، دار المسيرة.
- أبو هاشم، السيد (2005). مؤشرات التحليل البعدي لبحوث فعالية الذات في ضوء نظرية باندورا. مركز بحوث كلية التربية بجامعة الملك سعود، العدد 238. الرياض.
- الأعرس، عبد المنعم (2002). الإحصاء لعلوم الحياة والبيئة، جدة، الدار السعودية للنشر والتوزيع.
- البابطين، عادل (2002). مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة، رسالة ماجستير، جامعة أم القرى، مكة المكرمة.
- تشاو، لنكولن (1990). الإحصاء في الإدارة. ترجمة: عبد المرضي عزام، الرياض، دار المريخ للنشر.

- حجيمات، تحسين؛ وعليان، خليل (1997) واقع الدلالات الإحصائية والعملية وقوة الاختبارات الإحصائية المستخدمة في رسائل ماجستير الإرشاد النفسي والتربوي بالجامعة الأردنية. مجلة العلوم التربوية. المجلد (24)، العدد (2)، 398-408.
- زويلف، مهدي حسن. (1998). منهجية البحث العلمي. ط1. دار الفكر. عمان. الأردن.
- السيد، فؤاد (2005). علم النفس الإحصائي وقياس العقل البشري، القاهرة، دار الفكر العربي.
- الشربيني، زكريا (2007) الإحصاء وتصميم التجارب في البحوث النفسية والتربوية والاجتماعية، القاهرة، الأنجلو.
- الشربيني، زكريا (2007ب). الإحصاء اللابارامترية في العلوم النفسية والتربوية، القاهرة، الأنجلو المصرية.
- الصائغ، ابتسام (1997). الدلالة الإحصائية والدلالة العملية لاختبار (ت) و(ف) دراسة تحليلية تقويمية من خلال رسائل الماجستير التي قدمت في كلية التربية بجامعة أم القرى بمكة المكرمة، ماجستير أم القرى، مكة المكرمة.
- الصياد، عبد العاطي (1988). الدلالة العملية وحجم العينة المصاحبتين للدلالة الإحصائية لاختبار (ت) في البحث التربوي والنفسية العربي - بحوث مؤتمر البحث التربوي والواقع والمستقبل، المجلد الثاني، القاهرة، ص 199-230.
- الصياد، عبد العاطي أحمد (1989). جدول تحديد حجم العينة في البحث السلوكي. مجلة رابطة التربية الحديثة القاهرة، العدد الأول، يناير.
- العبد القادر، فيصل بن احمد (2008). حجم تأثير الاختبارات الإحصائية المعلميه واللا معلميه المستخدمة في رسائل الماجستير بكلية التربية بجامعة الملك سعود، رسالة ماجستير، الرياض، كلية التربية جامعة الملك سعود.
- العجلان، فتحية محمد (1990). دراسة تقويمية للأساليب الإحصائية المستخدمة في رسائل الماجستير بكلية التربية بجامعة أم القرى، مكة المكرمة، جامعة أم القرى.
- عدس، عبد الرحمن (1997). مبادئ الإحصاء في التربية وعلم النفس. الجزء الثاني: عمان: دار الفكر.
- العساف، صالح (2006). المدخل إلى البحث في العلوم السلوكية، الرياض، العبيكان.
- علام، صلاح الدين محمود (2010). الأساليب الإحصائية الاستدلالي البارامترية واللابارامترية في تحليل بيانات البحوث النفسية والتربوية. القاهرة: دار الفكر العربي.
- عودة، أحمد؛ والخليلي، خليل (2000). الإحصاء للباحث في التربية والعلوم الإنسانية. الأردن: دار الأمل.
- فرج، صفوت (1996). الإحصاء في علم النفس. القاهرة: الأنجلو المصرية.
- النجار، عبد الله عمر (1991). دراسة تقويمية للأساليب الإحصائية التي استخدمت في تحليل البيانات في رسائل الماجستير في كل من كلية التربية بجامعة أم القرى بمكة المكرمة وكلية التربية بجامعة الملك سعود بالرياض، رسالة ماجستير غير منشورة، بمكة المكرمة كلية التربية، جامعة أم القرى.
- النجار، عبد الله عمر (2005). الدلالة الإحصائية والدلالة العلمية لاختبار كا2 في البحوث الإدارية المنشورة (دراسة تقويمية). المجلة العربية للعلوم الإدارية. (12)، (2)، 169-193.
- نصار، يحيى (2002). حجم الأثر كأسلوب إحصائي مكمل لفحص الفرضيات الإحصائية، مركز بحوث كلية التربية - الرياض جامعة الملك سعود، العدد 176.

- نور، رجاء محمد (1993). تقويم استخدامات اختبار كاي تربيع في رسائل الماجستير بكلية التربية - جامعة أم القرى. رسالة ماجستير غير منشورة. مكة المكرمة: كلية التربية، جامعة أم القرى.

#### ثانياً- المراجع بالإنجليزية:

- Average Effect Sizes of Research in Counseling Psychology. Journal of Counseling Psychology. 29. 58-65.
- Cohen. J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.
- Haase. R. waechter. D. and Solomon. G. (1982). How Significant is a Significant Difference?
- Hays. W.L. (1963). Statistics for Psychologists. New York: Holt. Rinehart. and Winston.
- Huston. H (1993). Meaning fullness. Statistical Significance. Effect Size and power Analysis: A General Discussion with Implications for MANOVA. Paper Presented at the annual meeting of the mid- south Educational Research Association. 22<sup>nd</sup>. New Orleans. ERIC Document Reproductive Service. NO ED 364608.
- Lane. G (1999). Show me the Magnitude! The Consequences of Overemphasis on Null Hypothesis Significance Testing. Paper Presented at the annual meeting of the mid- south Educational Research Association (al. November 16- 19. 1999). (Eric document Reproductive service no. ED 436557).
- Pluker. Jontahan A. (1997). Debunking The Myth of ((Highly significant)) Result: Effult size in Gifted Education Research. Review. V20 n2. pp122- 26. 1997.
- Snyder. P. and Lawson. S (1992). Evaluating Significance using Corrected and Uncorrected of Effect Size Estimates. Paper Presented at Annual Meeting of the American Experimental Research Association (San Francisco. CA. Aril 20- 24. 1992).
- Thompson. Bruce. (2000). Improving Research Clarity and Usefulness with Effect Size Indices as to Statistical Significance Tests. Exceptional Children V65. n3. pp.329- 337. 1999.