

دراسة تحليلية

لقوة الاختبار الإحصائي في البحوث الإدارية المنشورة

عبدالله بن عمر النجار

كلية العلوم الإدارية والتخطيط - جامعة الملك فيصل

الأحساء - المملكة العربية السعودية

المخلص :

يعتبر مؤشر قوة الاختبار الإحصائي موضوع ذو أهمية لأي باحث يستخدم في دراسته منطبق الاستدلالات الإحصائية، فالدراسات التي تكون فيها الاختبارات الإحصائية ذات مستويات منخفضة من قوة الاختبار عادة تؤدي إلى نتائج غير حاسمة على الرغم من إنفاق الباحثين للكثير من الوقت والجهد لجمع بيانات أبحاثهم وتحليلها. لهذا الغرض قام الباحث باستعراض مؤشر قوة الاختبار الإحصائي لعدد من الأساليب الإحصائية المستخدمة في عدد من الأبحاث الإدارية المنشورة في مجلة الإدارة العامة الصادرة عن معهد الإدارة العامة.

وقد أشارت نتائج الدراسة الحالية إلى أن احتمال وجود الخطأ من النوع الثاني (أي احتمال وجود قوة الاختبار) منخفضة مع حجم التأثير المنخفض أكبر من احتمال وجود هذا الاقتران مع حجم التأثير المتوسط أو العالي، وأن قوة الاختبار الإحصائي في المتوسط عالية في البحوث الإدارية المنشورة، وأن السبب الرئيس لارتفاع قوة الاختبار الإحصائي ناتجة عن زيادة حجم العينة وليس لأن الباحثين تقصدوا الاهتمام بقوة الاختبار، وأن هذه الزيادة ربما كانت وسيلة تخفي وراءها الرغبة في الحصول على دلالة إحصائية لدلالة عملية ضعيفة، كما أن زيادة حجم العينة يُنقص من قيمة الخطأ من النوع الأول α وبذلك تتأثر قوة الاختبار نظراً للعلاقة الطردية بينهما.

- ومن خلال النتائج التي أسفرت عنها الدراسة الحالية فإن الباحث يوصي بالآتي :
١. بما أن نتائج الدراسة أشارت إلى حاجة الباحثين في مجال الإدارة للمعرفة بأهمية قوة الاختبار، لذا ينبغي على المتخصصين في مجال الإحصاء ضرورة توعية الباحثين بأهمية هذا المؤشر الإحصائي ودوره في نتائج البحث .
 ٢. لزيادة قوة الاختبار يقترح الباحث :
 - إجراء دراسات استطلاعية مصغرة للاختبارات الإحصائية لتقييم فاعلية الاختبارات حسب متغيرات البحث ، وذلك من خلال تحديد قيمة محددة لـ α ومثبتة على الاختبارات جميعاً ، ثم ومن خلال استخدام القيمة الاحتمالية لـ β يتم التعرف على التأثير النسبي لكل اختبار لمعرفة أيهما أكثر موائمة لظاهرة البحث .
 - تحديد قوة إحصائية بحيث لا تقل عن ٠,٨٠ مع مستوى دلالة وحجم تأثير عالي، ثم البحث عن أدنى حجم لعينة الدراسة .

المقدمة :

إن إجراء البحث ليس أمراً يسيراً ، وإنما يتطلب ذلك الملاحظة والوصف والتحليل والتفسير المناسب ، ويستخدم الباحثون فيه عادة أدوات قياس عديدة وأشكال وصف دقيقة ، كما يقومون باختيار الوسائل المناسبة لتجميع البيانات ، والقيام بإجراءات ملائمة في تحليلها وتفسيرها بهدف الوصول إلى قرارات إحصائية لرفض الفرض الصفري بأقل قدر من احتمالات الخطأ . وإن ما يجب التنويه عليه هنا أن جمع البيانات في حد ذاته ليس غاية يسعى إليها ، بل هي وسيلة إلى تحقيق أهداف معينة ، أي أن عمل الباحث لا ينتهي عند جمع البيانات ، بل على العكس من ذلك ، إن عمله يبدأ بها ، لذلك فالهدف النهائي من اختيار الوسائل المناسبة لجمع البيانات هو الحصول على البيانات التي تخدم في تحقيق أهداف البحث أو في دراسة المشكلة وإيجاد الحلول المناسبة لها . فقد تكون جميع مراحل جمع البيانات جيدة وصحيحة ، ولا يتخللها

ثغرات يمكن أن تؤثر في صحة البيانات، ولكن هذه البيانات تظل مادة خام لا تعطي ثمارها ما لم يوجد الباحث الجيد الذي يستطيع استثمارها على الوجه الأمثل.

ومما لاشك فيه أن تقويم الجانب الإحصائي في البحوث الإدارية يعطي تصورا صحيحا لواقع هذه البحوث، وجوانب ضعفها، وذلك لأن كثيرا من الباحثين في مجال العلوم الإنسانية ومنها الدراسات في مجال الإدارة أثناء قيامهم بالبحث يكون جزءا من مهمهم مركزا على مدى رفض أو قبول فروض الدراسة موضع البحث، وإن هذا الرفض أو القبول للفروض مرتبط بما يسمى بقوة الاختبار *Statistical Power*. فقوة الاختبار هي مدى قدرة الاختبار الإحصائي على رفض الفرضية موضع البحث عندما تكون في حقيقة الأمر خاطئة، لذلك فمن المهم أن نتذكر أن قوة الاختبار لا تعني رفض الفرضية الصفرية بشكل مطلق، بل تعني في حقيقة الأمر مدى قدرة البيانات المجمعة لدراسة ما ومدى كفاءة تصميمها على رفض فرضية الدراسة.

ومن الملاحظ أن الكثير من الباحثين لا يولون هذا المؤشر الإحصائي الاهتمام المناسب مما أثر سلبا على نتائج البحوث العلمية كما أشارت إلى ذلك العديد من الدراسات والأبحاث. فـ قوة الاختبار الإحصائي ضعيفة في الكثير من الدراسات المنشورة في مجال الإدارة والتسويق ونظم المعلومات الإدارية والتربية والاجتماع (الصياد، 1988؛ Mazen, et al 1987؛ Sawyer and Ball 1981؛ Chase and Katzer and Sodt 1973).

إن عدم قدرة الباحث للحصول على قدر كاف من قوة الاختبار سيكون له عواقب وخيمة على البحث ونتائجه، وفي هذا الصدد يقول (Stevens, 1980) بأنه في حال قيام الباحثين باستثمار وقتهم وجهدهم ومالهم في دراسة ما، فإنهم يطمحون للحصول على نتائج جيدة وذات دلالة عملية، فعندما تكون قوة الاختبار ضعيفة فإن هذه الاحتمالية (الحصول على نتائج جيدة وذات دلالة عملية) ستكون منخفضة، وإن نتائج الدراسة

ستكون غير مناسبة وسيؤدي ذلك إلى سوء فهم لنتائج البحث المتوصل إليها (Wilcox, 2003)، لذلك فإن تقييم قوة الاختبار الإحصائي في البحوث ضرورة مهمة لاتخاذ القرار المناسب وتعزيزا لصدق النتائج في البحث، ومن هذا المنطلق فإن الدراسة الحالية سوف تحاول الوقوف على طبيعة هذه المشكلة .

مشكلة الدراسة :

تعتبر البيانات مادة أساسية في البحوث العلمية، وعملية جمعها وتصنيفها وتحليلها إن لم تتم بطريقة مناسبة وجيدة أدى ذلك إلى انهيار البحث برمته وفقدانه قيمته العلمية والعملية . والمتتبع للبحوث في مجال الإدارة يجد أن عدم الاهتمام بقوة الاختبار الإحصائي مشكلة واضحة فيها، لذا فإنه مهما بلغت كمية البيانات ودقة أدوات جمعها، فإنها تبقى بيانات خام لا تعطي للباحث دلائل كمية وكيفية كافية حول الظاهرة المدروسة، وأن أي تقصير في إجراءات التحليل الإحصائي سيؤثر سلباً على المستوى العلمي للبحث ودرجة الثقة في نتائجه.

ويجب أن نؤكد هنا انه طالما نتعامل مع الإنسان، وينعكس اثر بحوثنا عليه - حيث أن كثيراً من القرارات التي تؤخذ تعتمد بالدرجة الأولى على نتائج الأبحاث - فإنه من الواجب أن نكون أكثر حرصاً وحذراً وأن نكون أكثر وعياً بالهدف وقوة الأسلوب الإحصائي الذي نستخدمه . وعليه فإن تقويم قوة الاختبارات الإحصائية في البحوث الإدارية سوف يساعد على إعطاء صورة واضحة عن الأخطاء والتجاوزات التي قد يقع فيها الباحثون عند تصميمهم لدراساتهم واستخدامهم للأساليب الإحصائية في تحليل بيانات أبحاثهم، وفي عام ١٩٦٢ أطلق Cohen أول استفسار عن قوة الاختبار الإحصائي في البحوث المنشورة، وقد أكد على أهمية ذلك (الصياد، ١٩٨٨) واعتبر أن قوة الاختبار الإحصائي من المشكلات التي تستحق اهتمام الباحثين. فقوة الاختبار الإحصائي تعتمد على الاختبار وطبيعته، وهذا الأمر يتطلب من

الباحث أن يحدد بشكل مناسب تصميم دراسته وفروضها والاختبار الإحصائي ومستوى الدلالة المطلوب كجزء من المواصفات الرئيسية للدراسة وبالعموم، فإن هذه الدراسة تحاول الإجابة عن السؤال الرئيس التالي : ما واقع استخدام الباحثون في مجال الإدارة لقوة الاختبار الإحصائي في أبحاثهم المنشورة ؟

أهداف الدراسة :

تهدف هذه الدراسة إلى إلقاء الضوء على واقع استخدام الباحثون لمؤشر قوة الاختبار الإحصائي في الأبحاث الإدارية وتوضيح كيفية تطبيق مؤشر قوة الاختبار والاستفادة منه في تقويم نتائج الأبحاث بصورة منطقية يمكن الاعتماد عليها في اتخاذ قرارات مناسبة حيال الظاهرة المدروسة وذلك من خلال عينة من البحوث الإدارية المنشورة في مجلة الإدارة العامة والصادرة عن معهد الإدارة العامة.

أهمية الدراسة وحدودها :

تأتي أهمية دراسة هذا الموضوع من كونه إضافة لأدبيات البحث في الإحصاء تحذر الباحثين من سوء استخدام الإحصاء كوسيلة، وتوجههم إلى ضرورة استخدام أفضل الأساليب الإحصائية وانسبها والتي يمكن الاعتماد عليها لزيادة قوة الاختبار الإحصائي وقدرته على رفض الفرضية موضع البحث عندما تكون في حقيقة الأمر خاطئة، فالإحصاء وسيلة أساسية لتحليل البيانات البحثية.

ولقد أكد (Thompson, 1992) إلى أن أصعب مرحلة تواجه الباحث في بحثه هي مرحلة التحليل الإحصائي، وذلك لتعدد الطرق الإحصائية المتاحة. فالباحث وهو يستعرض مجموعة كبيرة من مختلف الطرق الإحصائية سيقف حائراً ولا يدري أيها منها يختار ما لم يكن لديه معايير مسبقة يستتير بها في هذا الاختيار، وإن عدم الأخذ بهذه

المعايير تجعل الباحث في النهاية يستخدم اختبارات إحصائية ذات قوة منخفضة مما يؤدي بالتالي إلى استخدام طرقاً لا تتناسب مع طبيعة بيانات بحثه.

ولقد أشار كثيرا من الإحصائيين في مؤلفاتهم إلى ضرورة القيام بدراسات عن قوة الاختبار الإحصائي وبيان أهمية ذلك (Glenberg, 1996 ؛ Cohen, 1988 ؛ الصياد ؛ ١٩٨٨ ؛ Mazen, et al, 1987 ؛ Sawyer and Ball, 1981 ؛ Chase and Chase, 1976 ؛ Kater and Sordt, 1973)، وقد أكد معظم الباحثين على أنه لا يكفي بحساب هذا المؤشر الإحصائي نظرياً، ولكن أكدوا على ضرورة تطبيقه في الدراسات والبحوث مما يتطلب الحرص عند اختيار الأسلوب الملائم لطبيعة البحث، وهذا لا يتأتى إلا بالتدريب على إجراء البحوث ودراساتها دراسة عملية تتبعية.

وبناء على ما تقدم فالدراسة الحالية محاولة لتقييم الأبحاث الإدارية المنشورة في مجلة الإدارة العامة والصادرة عن معهد الإدارة العامة بالرياض، حيث تناول الباحث بالدراسة جميع الأبحاث العملية (التي أستخدم الإحصاء الاستدلالي في تحليل بياناتها) والمنشورة في هذه المجلة العلمية في الفترة من عام ١٤١٩هـ إلى عام ١٤٢٤هـ وذلك من خلال مراجعة هذه البحوث والتأكد من قوة الاختبار الإحصائي المستخدم بها.

أهمية قوة الاختبار الإحصائي :

تكمن قوة الاختبار الإحصائي في قدرته على رفض الفرضية موضع البحث عندما تكون في حقيقة الأمر خاطئة، وتعتمد قيمتها (قوة الاختبار) بشكل مباشر على احتمال ارتكاب الخطأ من النوع الثاني، حيث أن قوة الاختبار $= (1 - \beta)$ ، لذلك فمن المهم أن نتذكر أن قوة الاختبار لا تعني رفض الفرضية الصفرية بشكل مطلق، بل تعني في حقيقة الأمر مدى قدرة البيانات المجمعة لدراسة ما ومدى كفاءة تصميمها على رفض فرضية الدراسة.

ففي عام ١٩٦٢ أطلق Cohen أول استفسار عن قوة الاختبار الإحصائي في البحوث المنشورة، وبرز ما يُعرف باسم الدلالة العملية Practical Significance، فمنذ ذلك الحين أصبح الجدل بخصوص الدلالة الإحصائية والدلالة العملية يُمثلُ نموذجاً منهجياً في أبحاث العلوم الإنسانية (Brewer, 1972؛ Fern and Monroe, 1996؛ Thompson؛ and Snyder, 1997؛ Daniel, 1998؛ Wilkerson and Olson, 1997؛ Thompson, 1999).

فالدلالة العملية (حجم التأثير Effect Size) تعني مقدار التباين في المتغير التابع الذي يمكن تفسيره من خلال متغير مستقل يرتبط به أو يؤثر عليه (Haase, 2001). لذا فالدلالة الإحصائية ليست كافية لصناعة قرار إداري صائب، فهي مؤشر إحصائي لإظهار الفرق بين مجموعتين أو أكثر أو أن العلاقة بين متغيرين أو أكثر راجعة إلى فروق أو علاقة حقيقية وليست نتيجة للصدفة، ولكن هذا لا يعني الاعتماد على هذا الفرق واتخاذ القرارات بناء عليه (الصيد، ١٩٨٨)، كذلك زيادة حجم العينة قد يؤدي في بعض الأحيان إلى رفض الفرض الصفري (يعني وجود دلالة إحصائية)، وهنا يختلط الأمر على الباحث، هل الدلالة الإحصائية المتحصل عليها تعني أن العامل أو المعالجة تحت الدراسة لها تأثير على الناتج، أم أن هذه الدلالة الإحصائية جاءت نتيجة لكبير حجم العينة؟ (Kirk, 1996). وبالتالي فالدلالة الإحصائية المتحصل عليها قد تُوقع الباحث في حيرة مع نفسه، لأنه يُلاحظ أنه على الرغم من رفض الفرضية الصفريّة، إلا أنّ هذه النتيجة ليس لها أهمية عملية تُذكر، لذا فالكفاية تتحقق إذا ما حسبنا الدلالة العملية للاختبار (قوة العلاقة بين المتغير المستقل والمتغير التابع)، وهذه الدلالة العملية تُعتبر مؤشراً إحصائياً لمُدَى قُدْرَتنا على استخدام النتائج تفسيراً أو تطبيقاً.

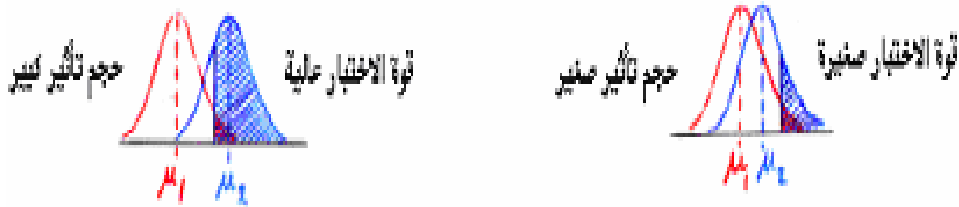
ومن الواضح كذلك أن قوة الاختبار الإحصائي تعتمد على الاختبار وطبيعته، وهذا الأمر يتطلب من الباحث أن يحدد بشكل مناسب تصميم دراسته وفروضها والاختبار

الإحصائي ومستوى الدلالة المطلوب كجزء من المواصفات الرئيسية للدراسة، واستعراض قوة الاختبار من خلال حساب الدلالة العملية المتوقعة والتي من الممكن الحصول عليها من خلال أحجام عينات مختلفة، فقوة الاختبار لا تُقدَّر أو تختبر، بل هي تُحسب من خلال التعرف على مؤشر حجم التأثير Effect Size للاختبار الإحصائي المستخدم. وهنا ينبغي على الباحث أن يطلع على تقديرات حجم التأثير من الدراسات السابقة للموضوع الذي يقوم بدراسته إذا كانت متوفرة، أو القيام بحساب حجم التأثير للبيانات الحالية.

أهم العوامل المؤثرة في قوة الاختبار الإحصائي :

لقد حدد الاحصائيون (Glenberg, 1996 ؛ الثبتي، ١٩٩٣؛ علام، ١٩٩٣ ؛ الشربيني، ١٩٩٠ ؛ Cohen, 1988 ؛ عودة والخليلي، ١٩٨٨) عددا من العوامل تتأثر قوة الاختبار بها، وهذه العوامل هي :

١. **حجم التأثير Effect Size (الدلالة العملية) :** يعتبر حجم التأثير أحد العوامل الأساسية التي تؤثر على قوة الاختبار، فحجم التأثير (الدلالة العملية) هو تلك القيمة الرقمية للفروق بين معالم المجتمع الحقيقية والغير معروفة (وهي تلك القيمة التي نفترضها عند صياغتنا للفرضية البديلة لمعالم المجتمع H_s) والقيم المحددة من خلال الفرضية الصفرية H_0 ، هذا الفرق ($\mu_s - \mu_0$) يمثل أحد أهم عناصر ما يسمى بمؤشر حجم التأثير، وهذا المؤشر يرتبط بشكل وثيق مع مؤشر قوة الاختبار، فكلما زاد حجم التأثير زادت قوة الاختبار، والشكل (١) يوضح ذلك :

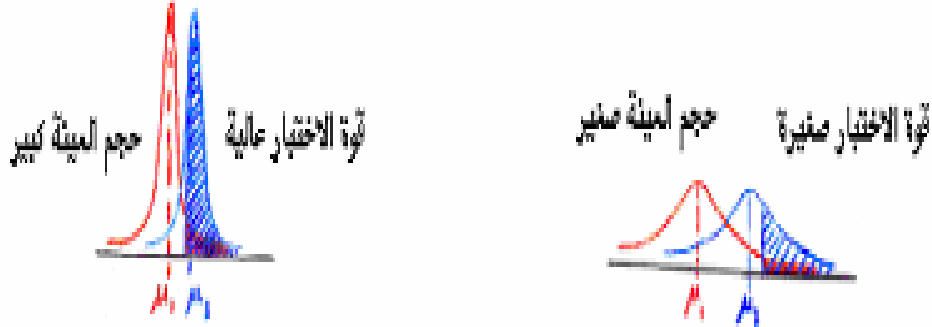


شكل (١) قوة الاختبار معتمدة على مؤشر حجم التأثير

وحجم التأثير (الدلالة العملية) مرتبطة بدلالة الفروق الإحصائية، وله أهميته العملية والتطبيقية للقرار المتخذ، فحجم التأثير يمثل الوجه الآخر للدلالة الإحصائية لأنه يتعلق بالمزايا الفعلية الناتجة من تقييم الفروق الإحصائية المحسوبة في اختبار دلالة الفروق، فهو مؤشر لمدى قدرتنا على استخدام النتائج تفسيرياً أو تطبيقياً (الصياد، ١٩٨٨)، لذا فإن الدلالة الإحصائية لا تكفي وحدها لاتخاذ قرار سليم، مما يتطلب حساب الدلالة العملية وذلك من خلال حساب قوة العلاقة بين المتغير التابع والمتغير المستقل للبيانات موضع البحث.

٢. **حجم العينة Sample Size** : يعتبر حجم العينة من أهم القرارات التي يتخذها الباحث عند اعداده لبحثه وذلك حتى تعطي أكبر قدر من الدقة بأقل تكلفه، وتزداد قوة الاختبار لقيمة معينة للمعلم الذي يراد اختبارها بازدياد حجم العينة، وهنا يجب ملاحظة إلى أن العينات الكبيرة تجعل أي فرق بسيط بين الإحصائي والمعلم المناظر له فرقاً ذا دلالة عملية، في حين أن الاختبار الإحصائي يشير إلى أنه يوجد فرق جوهري، وبذلك يصبح الاختبار الإحصائي الأقوى كفاءة هو الاختبار

الإحصائي الأكثر حساسية، لأنه يصل إلى مستويات الدلالة الإحصائية والعملية بعينة أصغر من غيره، والشكل التالي يوضح ذلك :



شكل (٢) قوة الاختبار معتمدة على حجم العينة

وقد رأى (Cohen، ١٩٨٨) أن العلاقة بين حجم العينة وقوة الاختبار علاقة مباشرة، فالزيادة في حجم العينة تزيد من القوة الإحصائية مع ثبات العوامل الأخرى وذلك عندما يكون الفرض الصفري غير صحيح، كما أن الزيادة في حجم العينة تقلل من حدوث الخطأ من النوع الأول والثاني، وقد وضع Cohen جداول لتحديد حجم العينة المناسب وذلك بمعرفة العوامل آنفة الذكر.

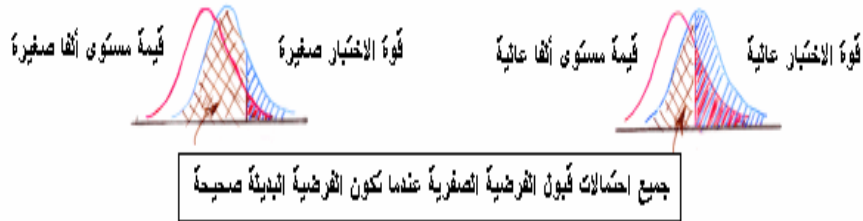
٣. مستوى الدلالة (α) *Alpha Level* : تزداد قوة الاختبار لقيمة معينة للمعلم الذي يراد اختباره بازدياد قيمة مستوى الدلالة، فزيادة قيمة ألفا (α) يقابلها نقصان لقيمة بيتا (β)، وبالتالي زيادة قوة الاختبار = ($1 - \beta$)، والجدول رقم (١) يوضح ذلك:

جدول (١)

يوضح احتمالية الخطأ في اختبار الدلالة الإحصائية لمجموعتين تجريبية وضابطة

| المجموعتين التجريبية والضابطة غير مختلفتين | المجموعتين التجريبية والضابطة مختلفتين | النتيجة النهائية المتحصل عليها من اخبارات العينة |
|---|---|---|
| الخطأ من النوع الأول الاحتمالية = (α) | قرار صائب الاحتمالية = $(1 - \beta)$ قوة الاختبار Power | توجد فروق ذات دلالة إحصائية (رفض الفرضية الصفرية) |
| قرار صائب الاحتمالية = $(1 - \alpha)$ | الخطأ من النوع الثاني الاحتمالية = (β) | لا توجد فروق ذات دلالة إحصائية (قبول الفرضية الصفرية) |

ويمكن توضيح هذه العلاقة من خلال الشكل التالي :



شكل (٣) قوة الاختبار معتمدة على مستوى الدلالة الإحصائية

وقد أشار (الصياد، ١٩٨٥) إلى أن المعيار الأساسي لاختبار الفروض الإحصائية من حيث اختبار مستوى الدلالة هو قوة الاختبار لموضوع البحث لذا ينبغي أن تكون قيمته على الأقل أكبر من حد الصدفة (٠,٥٠)، ورأى (Cohen، ١٩٨٨) أن تكون (٠,٨٠) كحد أعلى لتعامل الباحث مع المتغيرات المتغيرة دائماً ويكون الخطأ β مساوياً (٠,٢٠).

٤. تشتت أو تغير البيانات **Variability** : يمثل تشتت البيانات أحد أهم العوامل المؤثرة في قوة الاختبار، فكلما قل التشتت والتغير في بيانات العينة موضع الدراسة فإن قوة الاختبار تزداد، ويمكن التقليل من تشتت البيانات وتغيرها من خلال ثلاث طرق (١) تقنين طرق جمع البيانات للتخلص من تأثير أخطاء القياس، (٢) زيادة حجم العينة، (٣) جمع البيانات من خلال طريقة الأزواج *Paired Sampling* أو المجموعات الزوجية *Matched Sampling* لجعل الدرجات في العينات المختلفة مرتبطة ببعضها البعض.

٥. كون الاختبار بذيل واحد أو بذيلين : يكون الاختبار بذيل واحد أقوى من الاختبار بذيلين، إذا كانت القيمة الحقيقية للمعلم في نفس الجهة التي تفترضها الفرضية البديلة، وإن قوة الاختبار في الاختبار بذيل واحد تختلف بحسب قيمة المتوسط بالفرضية من حيث كونها أكبر أو أقل من قيمة المتوسط في الحقيقة، فعندما تكون قيمة μ_s الحقيقية عكس قيمة μ_0 في الفرضية البديلة تكون قوة الاختبار صفر و ذلك في الاختبار بذيل واحد، في حين نرى أن الاختبار ذو الذيلين يحافظ على التماثل في قوته في كلتا الجهتين.

٦. طبيعة الاختبار معلمية **Parametric** أو غير معلمية **Non Parametric** : إن طبيعة الاختبارات المعلمية أقوى كفاءة من الاختبارات اللامعلمية، فاحتمال وقوع الخطأ من النوع الثاني يكون أقل في حالة الاختبارات الإحصائية للفرض الصفري التي تفترض إعتدالية توزيع المجتمعات وتجانس التباين (الاختبارات المعلمية). ولكن في حالة عدم الوفاء بالافتراضات الأساسية للاختبار، تكون الاختبارات الإحصائية اللامعلمية أقوى كفاءة.

ومما ينبغي الإشارة إليه بأنه ستكون البعض من هذه العوامل قابلة للتطبيق أكثر من الأخرى فهي تعتمد على ظروف مشروع الدراسة موضع البحث، لذا ينبغي على الباحث تحقيق التوازن بين هذه العوامل والتي تساعد للوصول لأقصى قوة ممكنة للمختبر الإحصائي واكتشاف أي تأثير إن وجد للدراسة موضع البحث.

كيفية حساب قوة الاختبار الإحصائي :

هناك عدة طرق لحساب قيمة قوة الاختبار الإحصائي وتحديد معالمها (١) يمكن حساب قوة الاختبار الإحصائي من خلال المعادلات الرياضية، (٢) أو يمكن حسابها من خلال استخدام إحدى البرامج الإحصائية الجاهزة ومن ثم الرجوع إلى الجداول الإحصائية المعدة لهذا الغرض (٣) وكذلك يمكن حسابها من خلال البرامج الإحصائية مثل *GPOWER*، *Stat Power*، *Statistical Power Analysis*، *nQuery Advisor* وغيرها من البرامج الإحصائية المتخصصة والتي تهتم بحساب قوة الاختبار الإحصائي ببسر وسهولة <http://www.interchg.ubc.ca/cacb/power/>.

أولاً : حساب قوة الاختبار من خلال المعادلات الرياضية

إحدى المعادلات الرياضية لحساب قوة الاختبار هي حساب الدرجة المعيارية -Z Score للبيانات موضع الدراسة وذلك من خلال المعادلة التالية :

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$$

حيث أن :

\bar{X} = متوسط العينة موضع الدراسة ويمكن حساب \bar{X} من خلال معادلة Z-Score عن طريق ضرب طرفي المعادلة في $\sigma_{\bar{X}}$ فنحصل على المعادلة $\bar{X} - \mu_0 = (Z)(\sigma_{\bar{X}})$ ثم بعد ذلك إضافة μ_0 لطرفي المعادلة، وبعد الاختصار تكون معادلة \bar{X} هي

$$\bar{X} = \mu_0 + (Z)(\sigma_{\bar{X}})$$

$$\mu_0 = \text{متوسط المجتمع المحدد في الفرضية الصفرية } H_0 .$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \text{ = الخطأ المعياري للعينة والمحسوب من خلال المعادلة}$$

$$\sigma = \text{الإنحراف المعياري للمجتمع} .$$

ويتطلب حساب قوة الاختبار من خلال الدرجة المعيارية ثلاث خطوات هي:

- صياغة وتحديد الفرضية البديلة H_s (من خلال تحديد المتوسط في المجتمع).
- معرفة القيمة المتوسطة المحددة لمنطقة الرفض .
- حساب احتمالية الحصول على إحدى القيم المتوسطة والمعتمدة على توزيع العينة للفرضية البديلة H_s . والنتيجة المتحصل عليها من هذه الخطوة تمثل قوة الاختبار، فهي احتمالية رفض الفرضية الصفرية (وذلك لكون المتوسط \bar{X} الناتج من الدرجة المعيارية سيقع في منطقة الرفض Rejection Region) عندما تكون الفرضية البديلة H_s صحيحة .

ثانيا : حساب قوة الاختبار من خلال الجداول الإحصائية لقوة الاختبار

نظرا لصعوبة حساب قوة الاختبار الإحصائي من خلال المعادلات الرياضية (وذلك لصعوبة التعرف على بعض المعالم الأساسية للمعادلات الإحصائية المستخدمة لحساب قوة الاختبار)، لذا فمن الأسهل استخدام طريقة الجداول الإحصائية لقوة الاختبار والمعدة من قبل مجموعة من المتخصصين (Hoenig and Heisey, 2001؛ Cohen, 1988؛ Kraemer and Thiemann, 1987)، حيث قاموا بحساب مؤشر قوة الاختبار لمجموعة من الاختبارات الإحصائية الشائعة الاستخدام وعرضها في جداول إحصائية، وتم تحديد مجموعة من الشروط والافتراضات لاستخدام كل جدول.

فبعض هذه الجداول حدد حجم العينة Sample Size وقيمة مستوى الدلالة α مع تحديد وضع المنطقة الحرجة Critical Region موجهة أو غير موجهة وتحديد ثلاث مستويات لحجم التأثير وهي (منخفض، متوسط، كبير) كافتراضات لاستخدام

الجدول لأي إختبار محدد، ويمكن الرجوع إلى هذه الجداول والتي تساعد على التعرف على قوة الاختبار في (Cohen, 1988) وكذلك (Hays, 1981)؛ (Kraemer and Thiemann, 1987 ؛ Welkowitz, et al, 1982)

ثالثا : حساب قوة الاختبار من خلال البرامج الإحصائية

هناك العديد من البرامج الإحصائية المتخصصة لحساب قوة الاختبار لأي نوع من الاختبارات الإحصائية ومن هذه البرامج GPOWER ، Statistical Power Analysis ، nQuery Advisor ، Stat Power . الخ، ولقد أشار (Krebs & Thomas, 1997) إلى أهمية هذه البرامج وقارنوا بينها من حيث سهولة الاستخدام والتعلم في حساب قوة الاختبار الإحصائي ويمكن تلخيص ذلك من خلال الجدول رقم (٢).

جدول (٢)

يقارن بين البرامج الإحصائية المتاحة لحساب قوة الاختبار الإحصائي من حيث سهولة الاستخدام والتعلم

| سهولة التعلم | سهولة الاستخدام | البرنامج |
|--------------|-----------------|----------------------------|
| ❖❖ | ❖❖❖ | GPOWER |
| ❖❖❖ | ❖❖ | N |
| ❖❖❖❖ | ❖❖❖ | nQuery Advisor |
| ❖❖❖❖ | ❖❖❖❖ | PASS |
| ❖❖❖ | ❖ | PC-SIZE |
| ❖❖❖❖ | ❖❖❖ | Power and Precision |
| ❖❖ | ❖ | Powpal |
| ❖❖ | ❖❖ | Power Pack |
| ❖ | ❖❖ | Power Plant |
| ❖❖ | ❖ | PS |
| ❖❖ | ❖ | STPLAN |
| ❖❖❖❖ | ❖❖ | Statistical Power Analysis |
| ❖❖❖❖ | ❖❖❖ | Stat Power |

❖❖❖❖ ممتاز، ❖❖❖❖ جيد جدا، ❖❖❖❖ جيد، ❖❖❖❖ مقبول، ❖❖❖❖ ضعيف

استخدام مؤشر قوة الاختبار الإحصائي :

أوضحنا فيما سبق أهمية مؤشر قوة الاختبار الإحصائي وكيفية حسابه، وقد تبين لنا أن مؤشر قوة الاختبار مهم في جميع مراحل البحث (قبل، واثاء، وبعد)، فالتعرف على قوة الاختبار الإحصائي تتيح للباحث الإرتقاء ببحثه من خلال زيادة احتمالية الحصول على نتائج ذات دلالة وزيادة كفاءتها مما يُمكن الباحث من تفسير أفضل للنتائج المتحصل عليها، وذلك لأن قوة الاختبار هي المحك الأساسي لاختبار الفروض الإحصائية، مما يجعلنا نأخذها بعين الاعتبار عند اجراء البحوث الإدارية.

فالتحديد القبلي *Priori Determination* لقوة الاختبار هو أن يقوم الباحث بتحديد قيمة معينة لقوة الاختبار وفي ضوءها يتم استكمال اجراءات البحث والدراسة، وهذا التحديد يتيح للباحث استخدام مؤشر قوة الاختبار كأداة للتخطيط السليم لامكانية الحصول على نتائج ذات دلالة، فعند تحديد قوة اختبار عالية (٠.٨٠ على سبيل المثال) فهذا يستلزم الباحث دراسة حجم مناسب للعينة مع مراعاة العوامل المؤثرة في قوة الاختبار أنفة الذكر، وقد أشار (Cohen، ١٩٨٨) إلى عدد كبير من الجداول الإحصائية لعدد من الاختبارات الإحصائية الشائعة الاستخدام تربط بين قوة الاختبار الإحصائي وحجم العينة المناسب عند مستوى دلالة إحصائية α وحجم تأثير محدد.

كذلك يمكن للباحث استخدام تحليل قوة الاختبار الإحصائي لإجراء تقييم بعدي لقوة الاختبار *Post Hoc* لبعض الاختبارات الإحصائية، وقد اقترح (Medler, et al, 1981) استخدام تحليل قوة الاختبار البعدية في المراحل التخطيطية للدراسة وعند تحديد مستوى دلالة مناسب لنتائج البحث وذلك من خلال الاستفادة من بيانات أخرى متوفرة، فعلى سبيل المثال عندما يكون لدى الباحث بيانات محددة ويرغب في معرفة مدى مناسبة هذه البيانات لاستخدامها لبحث ظاهرة ما، في هذه الأثناء سيقوم الباحث بتقدير حجم تأثير متوقع وتحديد مستوى دلالة α مناسب

واختيار أسلوب إحصائي للتحليل وتحديد حجم عينة مناسب كل هذا سيساعد على تحديد هل لدينا قوة إحصائية لإجراء الدراسة أم لا 5.

تصميم وإجراءات الدراسة :

منهج الدراسة :

اتبع الباحث منهجا علميا تكامليا يجمع بين التحليل والتقويم، وإن لهذا النوع من البحوث أهمية في إثارة الوعي بقضايا ومشكلات البحث العلمي، والعمل على تزويد الباحثين بالمعلومات والحقائق المتعلقة بموضوع قوة الاختبار الإحصائي والتي قد تفيدهم في التغلب على هذه المشكلة ومعالجتها والوصول إلى تفسيرات واستنتاجات واقعية قابلة للتطبيق والتطوير.

مجتمع الدراسة وعينتها :

يمثل المجتمع الأصلي للدراسة الحالية البحوث المنشورة في مجلة الإدارة العامة منذ بداية إصدارها وحتى العدد الثاني من المجلد الثالث والأربعون والصادر في شهر ربيع الآخر من عام ١٤٢٤هـ يونيه ٢٠٠٣م، وقد بلغ عددها ١٧٢ عددا بواقع أربعة أعداد في المجلد الواحد وبمتوسط أربعة أبحاث في العدد الواحد. ومجلة الإدارة العامة هي عبارة عن دورية علمية متخصصة ومحكمة يصدرها معهد الإدارة العامة بالرياض - المملكة العربية السعودية وذلك كل ثلاثة أشهر.

ولغرض إلقاء الضوء على مدى الاهتمام بقوة الاختبار الإحصائي في البحوث الإدارية المنشورة قام الباحث بدراسة تقويمية لعينة من الأبحاث الإدارية المنشورة في مجلة الإدارة العامة خلال الست سنوات الأخيرة من ١٤١٩هـ - ١٤٢٤هـ، وقد بلغ عدد الأعداد التي تمت دراستها ٢٢ عددا وذلك بواقع ١٣٪ من مجمل الإصدارات الصادرة من هذه المجلة حتى

تاريخ هذه الدراسة، والسبب في تحديد هذه الفترة هو الرغبة في التعرف على الواقع الحالي لمؤشر قوة الاختبار في البحوث الإدارية المنشورة .
وقد بلغ تعداد البحوث التي تم تحليلها ودراستها (٨٨) بحثاً، إلا أنه وبعد مراجعة البيانات في هذه البحوث مراجعة دقيقة تم استبعاد عدداً من الدراسات بالكامل وذلك لعدم استخدام بعضها الإحصاء الاستدلالي في تحليل بياناتها أو لاستخدامها لبعض النماذج والمعادلات الرياضية والمؤشرات الاقتصادية والمالية لتحقيق أهدافها، وقد بقي بعد الاستبعاد (٤٦) بحثاً.

أسلوب جمع وتحليل بيانات الدراسة:

لقد صممت هذه الدراسة بهدف وصف واقع استخدام مؤشر قوة الاختبار الإحصائي في البحوث والدراسات الإدارية المنشورة، وقد تم اختيار مجموعة من الدراسات والأبحاث العلمية المنشورة في مجلة الإدارة العامة الصادرة من معهد الإدارة العامة بالرياض في المملكة العربية السعودية لغرض تحليلها والكشف من خلالها عن واقع قوة الاختبار الإحصائي.

ولقد تكونت عينة الدراسة من ٤٦ بحثاً منشوراً مركزة على ستة أساليب إحصائية ومجموع تكرارها ١٠٢٣، وقد تم استخدامها للإجابة على التساؤلات والفروض الرئيسية في هذه الدراسات، ولقد تم جمع المعلومات اللازمة للدراسة الحالية وفق التالي:

- تم حصر الأعداد المختارة لعينة الدراسة من مجلة الإدارة العامة.
- اعتمد الباحث في جمع بيانات الدراسة على تحليل الأبحاث ذات الصبغة العملية (الإمبيريقية) Empirical Studies وخاصة تلك التي استخدمت أساليب إحصائية استدلالية Inferential Statistics والمنشورة في مجلة الإدارة العامة وقد بلغ تعدادها (٤٦) بحثاً موزعين على سنوات نشرها كالتالي :

جدول (٣)

توزيع البحوث والدراسات العملية المنشورة في مجلة الإدارة العامة موزعة حسب سنة النشر

| سنة النشر | العدد | النسبة |
|-----------|-------|--------|
| ١٤١٩هـ | ١٣ | ٢٨,٤% |
| ١٤٢٠هـ | ١١ | ٢٣,٩% |
| ١٤٢١هـ | ٩ | ١٩,٦% |
| ١٤٢٢هـ | ٥ | ١٠,٩% |
| ١٤٢٣هـ | ٦ | ١٣,١% |
| ١٤٢٤هـ | ٢ | ٤,١% |
| المجموع | ٤٦ | ١٠٠% |

- ونظرا لقيام بعض الباحثين باستخدام عددا كبيرا من الأساليب الإحصائية لتحليل بيانات أبحاثهم، فقد تمّ تضمين الأساليب الإحصائية ذات العلاقة المباشرة بالفرضيات الأساسية في البحث، والجدول (٤) يحوي توزيع الأساليب الإحصائية التي تم حصرها من خلال عينة الدراسة، وقد تم تصنيف جميع الأساليب الإحصائية الغير معلمية (اللابارامترية) Non-Parametric تحت مسمى اختبارات غير معلمية وتحوي الاختبارات التالية (مان وتني Mann-Whitney، كروسكال والنز Kruskal-Wallis، ويلكوكسن Wilcoxon)، وعلى الرغم من كون اختبار مربع كاي (كا^٢) احد الاختبارات اللامعلمية Non-Parametric Tests إلا أنه تم التعامل معه بصورة منفصلة وذلك لشهرة هذا الأسلوب الإحصائي وكثرة استخدامه في البحوث والدراسات.

جدول (٤)

يوضح توزيع الأساليب الإحصائية المستخدمة في عينة الدراسة

| النسبة | التكرار | الاختبار الإحصائي |
|--------|---------|--|
| ٪١,٧ | ١٧ | تحليل الانحدار Regression Analysis |
| ٪٥٣,٣ | ٥٤٥ | معامل الارتباط Correlation |
| ٪١٥,١ | ١٥٥ | تحليل التباين الأحادي ANOVA |
| ٪١٨,٢ | ١٨٦ | اختبار (ت) للعينات المستقلة Independent T-Test |
| ٪٤,٣ | ٤٤ | اختبار مربع كاي (كا ^٢) Chi-Square |
| ٪٧,٤ | ٧٦ | الاختبارات اللامعلمية Non-Parametric |
| ٪١٠٠ | ١٠٢٣ | المجموع الكلي |

ويتضح لنا من الجدول (٤) أن عدد الأساليب الإحصائية المستخدمة في الأبحاث الإدارية المنشورة في مجلة الإدارة العامة في الفترة من عام ١٤١٩ - ١٤٢٤ هـ بلغ ستة أساليب إحصائية تكررت بمقدار ١٠٢٣ مرة، وهذه الأساليب الإحصائية اقتصرت فقط على تلك التي تجيب على التساؤلات الرئيسة فقط، وقد كان معامل بيرسون للارتباط هو أكثر الأساليب الإحصائية استخداماً حيث بلغ تكراره ٥٤٥ وبنسبة ٥٣,٣٪ من مجموع الأساليب الإحصائية المستخدمة في الدراسات الإدارية المنشورة في مجلة الإدارة العامة، وقد كان أقل الأساليب الإحصائية استخداماً هو تحليل الانحدار حيث بلغ تكراره ١٧ وبنسبة ١,٧٪ من مجموع الأساليب المستخدمة في المجلة.

- تم تحديد قوة الاختبار للأساليب الإحصائية الغير معلمية من خلال الرجوع إلى الجداول الإحصائية لقوة الاختبار للأساليب الإحصائية المعلمية المقابلة لها (Hays, 1981 ؛ Welkowitz, et al, 1982 ؛ Kraemer and Thiemann, 1987)،

- فعلى سبيل المثال اختبار مان وتني Mann-Whitney (اختبار غير معلمي) يقابله اختبار "ت" (اختبار معلمي)، وكذلك اختبار كروسكال والز Kruskal-Wallis (اختبار غير معلمي) يقابله اختبار تحليل التباين ANOVA وهكذا.
- لغرض تحديد حجم العينة المستخدم في هذه الدراسة تم الإطلاع على الجزء الخاص بمجتمع وعينة الدراسة في كل بحث من البحوث الإدارية المنشورة وتم من خلاله حصر حجم العينة المستخدم وفق متغيرات الدراسة .
 - تحديد مستوى الدلالة المستخدم بـ ٠,٠٥ لجميع الدراسات.
 - اعتبار جميع الأساليب الإحصائية المستخدمة بأنها غير موجهة Non-Directional (الاختبار بذيئين).
 - تم تحديد حجم الدلالة العملية (حجم التأثير) Effect Size من خلال إتباع الجداول المحددة من قبل (Cohen، ١٩٨٨) وذلك لصعوبة حساب حجم التأثير لكل أسلوب إحصائي على حده، حيث تم تحديد ثلاث مستويات لحجم التأثير وهي (منخفض، متوسط، كبير) لكل أسلوب إحصائي، هذه المستويات من حجم التأثير ساعدت على مقارنة قوة الاختبار الإحصائي بين كل دراسة وأخرى.
- تحليل النتائج وتفسيرها :**
- بعد جمع البيانات كان مجموع الأساليب الإحصائية المستخدمة في الأبحاث الإدارية المنشورة في مجلة الإدارة العامة ستة أساليب إحصائية تكررت بمقدار ١٠٢٣ مرة، عمل الباحث في الدراسة الحالية على تحليلها وتقويمها لغرض تحقيق أهداف الدراسة والإجابة على تساؤلاتها والتعرف على الواقع الفعلي لقوة الاختبار الإحصائي في الأبحاث الإدارية المنشورة .

ونظرا لاستخدام حجم العينة كأحد المعايير المستخدمة للحكم على قوة الاختبار الإحصائي في الأبحاث الإدارية المنشورة قام الباحث بحصر أحجام العينات المستخدمة لكل أسلوب من الأساليب الإحصائية المستخدمة في الدراسات والأبحاث الإدارية والجدول (5) يحوي توزيع متوسط حجم العينة وفقا لكل أسلوب إحصائي مستخدم في المجلة موضع الدراسة .

جدول (5)

يوضح توزيع حجم العينة في الأساليب الإحصائية المستخدمة في أبحاث عينة الدراسة

| الانحراف المعياري | المتوسط | الاختبار الإحصائي |
|-------------------|---------|--|
| ١٧٣,٠٨ | ٢٥٣ | تحليل الانحدار Regression Analysis |
| ١٧١,٧٣ | ٢١٣ | معامل الارتباط Correlation |
| ١٧٧,٨٣ | ٣٣٩ | تحليل التباين الأحادي ANOVA |
| ٨٠٢,٣٥ | ٤٦٣ | اختبار (ت) للعينات المستقلة Independent T-Test |
| ٤٨,٠٧ | ١٢١ | اختبار مربع كاي (كا ^٢) Chi-Square |
| ٩١,٠٤ | ١٠٧ | الاختبارات اللامعلمية Non-Parametric |
| - - | ٢٧٤ | المتوسط الكلي |

يتضح لنا من الجدول (5) أن معظم الأساليب الإحصائية المستخدمة في الأبحاث الإدارية المنشورة (تحليل التباين الأحادي، معامل الارتباط، اختبار "ت" للعينات المستقلة، تحليل الانحدار) استخدمت في المتوسط حجم عينة يزيد على الـ ٢٠٠، وهذا أمر مكلف لم يؤثر على حجم التأثير بشكل كبير كما أظهرت ذلك نتائج الدراسة

الحالية وأيدتها العديد من الدراسات السابقة (الصياد، ١٩٨٨؛ Kirk, 1996)، ذلك أن زيادة حجم العينة ربما كان وسيلة تخفي وراءها الرغبة في الحصول على دلالة إحصائية لدلالة عملية Effect Size ضعيفة، كما أن الزيادة في حجم العينة ينقص من قيمة الخطأ من النوع الأول α وبذلك تتأثر قوة الاختبار نظراً لعلاقتها الطردية بكل من حجم العينة وبالخطأ من النوع الأول. لذا فزيادة حجم العينة قد يؤدي في بعض الأحيان إلى رفض الفرض الصفري (يعني وجود دلالة إحصائية)، وهنا يختلط الأمر على الباحث، هل الدلالة الإحصائية المتحصل عليها تعني أن العامل أو المعالجة تحت الدراسة لها تأثير على الناتج، أم أن هذه الدلالة الإحصائية جاءت نتيجة لكبر حجم العينة؟ (Kirk, 1996). وبالتالي فالدلالة الإحصائية المتحصل عليها قد تُوقع الباحث في حيرة مع نفسه، لأنه يُلاحظ أنه على الرغم من رفض الفرضية الصفرية، إلا أن هذه النتيجة ليس لها أهمية عملية تُذكر، لذا فالكفاية تتحقق إذا ما حسينا الدلالة العملية للاختبار (قوة العلاقة بين المتغير المستقل والمتغير التابع)، وهذه الدلالة العملية تُعتبر مؤشراً إحصائياً لمدى قدرتنا على استخدام النتائج تفسيراً أو تطبيقاً.

ولقد تم تحديد قوة الاختبار الإحصائي لكل من الأساليب الإحصائية الستة وفقاً لتكرارها ١٠٢٣ مرة والتي تم استخدامها في الأبحاث الإدارية المنشورة وذلك وفقاً لجدول كوهن لقوة الاختبار، والجدول (٦) يوضح توزيع قوة الاختبار الإحصائي لهذه الأساليب الإحصائية أخذاً في الاعتبار مستويات حجم التأثير Effect Size الثلاثة (منخفض، متوسط، كبير).

الجدول (٦)

يوضح التكرار والنسبة التراكمية لقوة الاختبار الإحصائي للأساليب الإحصائية المستخدمة في الأبحاث الإدارية المنشورة في مجلة الإدارة العامة

| مستويات حجم التأثير | | | | | | فئات قوة الاختبار |
|---------------------|---------|------------------|---------|------------------|---------|--------------------|
| حجم تأثير كبير | | حجم تأثير متوسط | | حجم تأثير منخفض | | |
| النسبة التراكمية | التكرار | النسبة التراكمية | التكرار | النسبة التراكمية | التكرار | |
| %١٠٠ | ٨٦٨ | %١٠٠ | ٢٤٤ | - | - | ٠,٩٩ - ٠,٩١ |
| %١٥,١٥ | ٩ | %٧٦,١٥ | ٤٩٦ | - | - | ٠,٩٠ - ٠,٨١ |
| %١٤,٢٧ | ٧٤ | %٢٧,٦٧ | ٣٣ | - | - | ٠,٨٠ - ٠,٧١ |
| - | - | %٢٤,٤٤ | ٤٣ | - | - | ٠,٧٠ - ٠,٦١ |
| %٧,٠٤ | ٢٠ | %٢٠,٢٤ | ١٢٣ | - | - | ٠,٦٠ - ٠,٥١ |
| - | - | %٨,٢٢ | ٢ | - | - | ٠,٥٠ - ٠,٤١ |
| %٥,٠٨ | ٥٢ | %٨,٠٢ | ١٠ | %١٠٠ | ١١٠ | ٠,٤٠ - ٠,٣١ |
| - | - | %٧,٠٤ | ٢٠ | %٨٩,٢٥ | ١٤١ | ٠,٣٠ - ٠,٢١ |
| - | - | %٥,٠٨ | ٥٢ | %٧٥,٤٧ | ٦٩٠ | ٠,٢٠ - ٠,١١ |
| - | - | - | - | %٨,٠٢ | ٨٢ | ٠,١٠ - ٠,٠٠ |
| - - | ١٠٢٣ | - - | ١٠٢٣ | - - | ١٠٢٣ | المجموع |
| ٠,٩٣ | | ٠,٧٩ | | ٠,١٩ | | متوسط قوة الاختبار |

يتضح لنا من الجدول (٦) أن قوة الاختبار الإحصائي تختلف وفقا لاختلاف حجم التأثير، ففي حالة حجم التأثير المنخفض Small Effect Size فإن متوسط قوة الاختبار الإحصائي ٠,١٩ مما يدل على أنه في حالة افتراض حجم تأثير منخفض لظاهرة ما فإن قوة الاختبار الإحصائي Statistical Power لن تتجاوز في المتوسط ٠,١٩ أي أن فرصة رفض الفرضية موضع البحث عندما تكون في حقيقة الأمر خاطئة لن تتجاوز ١ من ٥،

كذلك نلاحظ من الجدول (٦) بأن قوة الاختبار الإحصائي لجميع الاختبارات الإحصائية المستخدمة في الأبحاث الإدارية المنشورة في حالة افتراض حجم تأثير منخفض أقل من الحد الأدنى وهو حد الصدفة لكوهن (٠,٥٠)، مما يعني أن قوة الاختبارات الإحصائية المستخدمة تمتلك فرصة أقل من ٥٠٪ للحصول على نتائج ذات دلالة إحصائية.

أما في حالة افتراضنا حجم تأثير متوسط Medium Effect Size فإن متوسط قوة الاختبار سيرتفع ويصل إلى ٠,٧٩ مقارنة بـ ٠,١٩ 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 ٢ إلى ٣. كذلك نلاحظ أن ٧٢٪ من الاختبارات الإحصائية المستخدمة في الأبحاث الإدارية المنشورة حصلت على قوة اختبار أعلى من ٠,٨٠، وأن أكثر من ٩١٪ من الاختبارات الإحصائية في حالة افتراض حجم تأثير متوسط حصلت على قوة اختبار أكبر من ٠,٥٠ أي أنها تمتلك فرصة أكبر من ٥٠٪ للحصول على نتائج ذات دلالة إحصائية.

أما في حالة افتراضنا حجم تأثير كبير Large Effect Size فإن متوسط قوة الاختبار سترتفع بشكل واضح وتصل إلى ٠,٩٣، ونلاحظ كذلك أن أكثر من ٨٤٪ من الاختبارات الإحصائية المستخدمة في البحوث الإدارية المنشورة حصلت على قوة اختبار ٠,٨٠ أو أكثر، وأن قرابة الـ ٦٪ حصلت على قوة اختبار أقل من الحد الأدنى وهو حد الصدفة لكوهن (٠,٥٠).

نلاحظ كذلك من الجدول السابق 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 وانخفاض لنسبة التباين غير المفسر (تباين الخطأ)، وهذه النتيجة تتفق مع ما

توصل إليه (Brewer, 1972) من أن قوة الاختبار الإحصائي مرتبطة ارتباطاً وثيقاً بحجم التأثير، وكذلك تتفق مع ما توصل إليه (Ottenbacher, 1982) والذي أشار أن الاختبارات التي يكون فيها حجم التأثير منخفضاً يكون احتمال شمولها على خطأ من النوع الثاني (β) كبيراً، وأن هذا الاقتران يكون أكثر حدوثاً مع حجم التأثير المنخفض من حدوثه مع حجم التأثير المتوسط أو الكبير.

ويلاحظ من الجدول أن تكرار الأساليب الإحصائية المستخدمة لحجم التأثير المتوسط عند قوة الاختبار ٠.٨١ فأكثر بلغ ٧٤٠ بنسبة ٧٢,٣٪ من مجموع تكرار الأساليب الإحصائية المستخدمة في الأبحاث الإدارية المنشورة، وهذا يؤكد توقع (الصيد، ١٩٨٨) من أن البحوث العربية لها قوة اختبار أعلى من البحوث الأجنبية وربما يرجع ذلك إلى أن الباحثين ارتضوا زيادة حجم العينة في سبيل الوصول لدلالة إحصائية وذلك برفض الفرض الصفري، وبالتالي الوصول إلى حجم تأثير ما دون إدراك منهم لأهمية قوة الاختبار، فكما هو معروف أنه ليس كل ما هو دال إحصائياً دال عملياً، وإن زاد حجم العينة، مع الأخذ بعين الاعتبار أنه بزيادة حجم العينة يمكن الوصول لقوة اختبار عالية وقد يكون بدلالة عملية ضعيفة أو مرتفعة حسب طبيعة الدراسة.

ولغرض التعرف على مؤشر قوة الاختبار لكل اختبار إحصائي قام الباحث بتحديد قوة الاختبار الإحصائي لكل اختبار تم استخدامه في البحوث الإدارية المنشورة وذلك من خلال الرجوع إلى الجداول الإحصائية المعدة لذلك (Cohen, ١٩٨٨) وكذلك (Kraemer and Thiemann, 1987؛ Welkowitz, et al, 1982؛ Hays, 1981)، والجدول (٧) يوضح ذلك.

الجدول (٧)

يوضح توزيع قيم قوة الاختبار الإحصائي في الأبحاث الإدارية المنشورة في مجلة الإدارة العامة وفقاً للأسلوب الإحصائي المستخدم

| مستويات حجم التأثير | | | | | | الأسلوب الإحصائي المستخدم |
|---------------------|---------|-------------------|---------|-------------------|---------|-----------------------------|
| حجم تأثير كبير | | حجم تأثير متوسط | | حجم تأثير منخفض | | |
| الانحراف المعياري | المتوسط | الانحراف المعياري | المتوسط | الانحراف المعياري | المتوسط | |
| ٠,٠٠٤ | ٠,٩٨ | ٠,٠١ | ٠,٨٥ | ٠,٠٢ | ٠,١٦ | تحليل الانحدار |
| ٠,١٦ | ٠,٩١ | ٠,١٨ | ٠,٧٨ | ٠,٠٣ | ٠,١٥ | معامل الارتباط |
| ٠,٠٢ | ٠,٩٨ | ٠,١٠ | ٠,٩٣ | ٠,٠٧ | ٠,٢٩ | تحليل التباين الأحادي |
| ٠,٠٦ | ٠,٩٦ | ٠,١٦ | ٠,٨٣ | ٠,٠٧ | ٠,٢٤ | اختبار (ت) للعينات المستقلة |
| ٠,٠٦ | ٠,٩٦ | ٠,١٢ | ٠,٦٣ | ٠,٠١ | ٠,١٠ | اختبار كاً |
| ٠,٢٦ | ٠,٦٩ | ٠,٢٩ | ٠,٤٦ | ٠,٠٦ | ٠,١٢ | الاختبارات اللامعلمية |

يتضح لنا من الجدول (٧) أن قوة الاختبار الإحصائي كانت أعلى من ٠,٨٠ في كل من تحليل الانحدار (٠,٨٥) وتحليل التباين الأحادي (٠,٩٣) واختبار "ت" للعينات المستقلة (٠,٨٣) وذلك عند افتراض حجم تأثير متوسط . وعلى الرغم من أن معامل الارتباط يمثل أكثر الأساليب الإحصائية استخداماً في البحوث الإدارية المنشورة حيث بلغ استخدامه ٥٤٥ بنسبة ٥٣,٣٪ إلا أن قوة الاختبار لهذا الأسلوب الإحصائي كانت أقل من ٠,٨٠ وذلك عند افتراض حجم تأثير متوسط.

كذلك نلاحظ أنه في حالة افتراض حجم تأثير كبير فإن قوة الاختبار الإحصائي لجميع الأساليب الإحصائية (ماعدا الاختبارات اللامعلمية) تكون أعلى من ٠,٩٠ ، أما الاختبارات اللامعلمية فإن حجم قوة الاختبار لها كانت ٠,٦٩ وهذا أمر طبيعي لكون الاختبارات اللامعلمية أقل قوة من الاختبارات المعلمية.

الخاتمة التوصيات والمقترحات :

هدفت الدراسة الحالية إلى إلقاء الضوء على واقع استخدام الباحثون لمؤشر قوة الاختبار الإحصائي في الأبحاث الإدارية وتوضيح كيفية تطبيق مؤشر قوة الاختبار والاستفادة منه في تقويم نتائج الأبحاث بصورة منطقية يمكن الاعتماد عليها في اتخاذ قرارات مناسبة حيال الظاهرة المدروسة وذلك من خلال عينة من البحوث الإدارية المنشورة في مجلة الإدارة العامة والصادرة عن معهد الإدارة العامة، ولقد أشارت نتائج هذه الدراسة إلى أن احتمال وجود الخطأ من النوع الثاني (أي احتمال وجود قوة الاختبار) منخفضة مع حجم التأثير المنخفض أكبر من احتمال وجود هذا الاقتران مع حجم التأثير المتوسط أو العالي، وأن قوة الاختبار الإحصائي في المتوسط عالية في البحوث الإدارية المنشورة، وأن السبب الرئيس لارتفاع قوة الاختبار الإحصائي ناتجة عن زيادة حجم العينة وليس لأن الباحثين تقصدوا الاهتمام بقوة الاختبار، وأن هذه الزيادة ربما كانت وسيلة تخفي وراءها الرغبة في الحصول على دلالة إحصائية لدلالة عملية ضعيفة، كما أن زيادة حجم العينة يُنقص من قيمة الخطأ من النوع الأول α وبذلك تتأثر قوة الاختبار نظرا للعلاقة الطردية بينهما.

ومن خلال النتائج التي أسفرت عنها الدراسة الحالية فإن الباحث يوصي بالآتي :

١. بما أن نتائج الدراسة أشارت إلى حاجة الباحثين في مجال الإدارة للمعرفة بأهمية قوة الاختبار، لذا ينبغي على المتخصصين في مجال الإحصاء ضرورة توعية الباحثين بأهمية هذا المؤشر الإحصائي ودوره في نتائج البحث.
٢. لزيادة قوة الاختبار يقترح الباحث :

- إجراء دراسات استطلاعية مصغرة للاختبارات الإحصائية لتقييم فاعلية الاختبارات حسب متغيرات البحث، وذلك من خلال تحديد قيمة محددة لـ α ومثبتة على الاختبارات جميعا، ثم ومن خلال استخدام القيمة الاحتمالية لـ β

يتم التعرف على التأثير النسبي لكل اختبار لمعرفة أيهما أكثر موثمة لظاهرة البحث.

- تحديد قوة إحصائية بحيث لا تقل عن ٠,٨٠ مع مستوى دلالة وحجم تأثير عالي، ثم البحث عن أدنى حجم لعينة الدراسة.

٣. في حالة اضطرار الباحث إلى التقييد بحجم عينة محدد لأسباب مالية أو عملية لا بد من تثبيت قيمة α ثم حساب قوة الاختبار بالنسبة لمستويات حجم التأثير، فإذا كانت قوة الاختبار أقل من ٠,٨٠ لجميع أحجام التأثير فلا بد أن يتجنب الباحث إجراء الاختبار الإحصائي إلا إذا قام بزيادة حجم العينة ولكن دون إفراط.

٤. إذا تساوت قوة الاختبار لأكثر من إحصائي لرفض الفرضية الصفرية فمن الأفضل للباحث حساب الدلالة العملية للاختبارات، والاختبار الذي له أهمية عملية وتفسيرية أكبر هو بالتالي الاختبار المناسب.

المراجع

أولا : المراجع العربية :

١. تشاور، لنكولن (١٩٩٠) "الإحصاء في الإدارة"، الرياض، دار المريخ للنشر .
٢. الثبيتي، علي حامد (١٤١٥) "أخطاء شائعة بين تصاميم البحوث التربوية والنفسية وعلاقتها ذلك بالصدق الإحصائي للنتائج وتصميمها"، رسالة الخليج العربي، مكتب التربية العربي لدول الخليج، الرياض - العدد ٥٢، السنة ١٣، ص ص ٥١ - ٧٤ .
٣. الشربيني، زكريا (١٩٩٠) "الإحصاء اللابارامتري في العلوم النفسية والتربوية والاجتماعية"، جدة، دار عكاظ للطباعة والنشر .
٤. الصياد، عبدالعاطي أحمد (١٩٨٥) "النماذج الإحصائية في البحث التربوي والنفسي والعربي بين ماهو قائم وما يجب أن يكون - دراسة تقويمية للواقع الإحصائي للنماذج الإحصائية"، رسالة الخليج العربي، مكتب التربية العربي لدول الخليج، الرياض - العدد ١٦، السنة ٥ .
٥. الصياد، عبدالعاطي أحمد (١٩٨٨) "الدلالة العملية وحجم العينة المصاحبتين للدلالة الإحصائية لأختبار (ت) في البحث التربوي والنفسي العربي - دراسة تقويمية"، بحوث مؤتمر البحث التربوي الواقع والمستقبل - المجلد الثاني، القاهرة .
٦. الصياد، عبدالعاطي أحمد (١٩٨٩) "جداول تحديد حجم العينة في البحث السلوكي"، مجلة رابطة التربية الحديثة، القاهرة، العدد الأول .
٧. علام، صلاح الدين محمود (١٩٩٣) "الأساليب الإحصائية الإستدلالية البارامترية واللابارامترية في تحليل بيانات البحوث النفسية والتربوية"، القاهرة، دار الفكر العربي .
٨. عودة، أحمد و أحمد الخطيب (١٩٩٤) "التحليل الإحصائي في البحوث التربوية - دراسة وصفية تحليلية"، مجلة اتحاد الجامعات العربية - العدد ٢٩، ص ص ٢٢٤ - ٢٤٢ .
٩. عودة، أحمد و خليل الخليلي (١٩٨٨) "الإحصاء للباحث في التربية والعلوم الإنسانية"، عمان، دار الفكر للنشر والتوزيع .

ثانيا : المراجع الأجنبية

1. Bradley, R. and Russell, L. (1996). Statistical Power in Complex Experimental Designs, Behavior Research Methods – Instruments and Computers, 28, 319-326.
2. Brewer, J. (1972). On the Power of Statistical Tests in the American Educational Research Journal, American Educational Research Journal, 9 (3), 391-401.
3. Chase, J. and Chase, B. (1976). A Statistical Power Analysis of Applied Psychological Research, Journal of Applied Psychology, 61 (2), 234-237.
4. Cohen, J. (1998). Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Hillsdale, NJ USA .
5. Daniel, L. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implication for the editorial policies of education journals, Research in the Schools, 5 (2), 23-32.
6. Fern, E. & Monroe, K. (1996). Effect-size estimates: issues and problems in interpretation, Journal of Vocational Education Research, 22 (4), 219-232.
7. Glenberg, A. (1996). Learning from Data – An Introduction to Statistical Reasoning, Lawrence Erlbaum Associates, Mahwah, NJ USA .
8. Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed, Educational and Psychological Measurement, 61 (2), 219-224.
9. Hays, L. (1981). Statistics, Holt, Rinehart and Winston Inc, New York, NY .
10. Hoening, M. and Heisey, M. (2001). The Abuse of Power – The Pervasive fallacy of Power Calculations for Data Analysis, The American Statistician, 55, 19-24.
11. Katzer, J. and Sodt, J. (1973). An Analysis of the Use of Statistical Theory in Communications Research, Journal of Communication, 23 (3), 251-265.
12. Kraemer, C. and Thiemann, S. (1996). How Many Subjects? – Statistical Power Analysis in Research, Sage Publications, Beverly Hills, CA .
13. Kirk, R. (1996). Practical significance: A concept whose time has come, Educational and Psychological Measurement, 56, 746-759.
14. Mazen, A. and Graf, L. (1987). Statistical Power In Contemporary Management Research, Academy of Management Journal, 30 (2), 369-380.
15. Medler, J. , Schneider, P. and Schneider, A. (1981). Statistical Power Analysis and Experimental Field Research, Evaluation Review, 5 (6), 834-850 .
16. Nickerson, S. (2000). Null Hypothesis Significance Testing – A Review of an Old and Continuing Controversy, Psychological Methods, (5), 241-301.
17. Ottenbacher, K. (1982). Statistical Power and Research in Occupational Therapy, The Occupational Journal of Research, 2 (1).

18. Sawyer, A. and Ball, D. (1981). Statistical Power and Effect Size in Marketing Research, *Journal of Marketing Research*, 18 (3), 275-290.
 19. Stevens, P. (1980). Power of the Multivariate Analysis of Variance Tests, *Psychological Bulletin*, 88 (3), 728-737.
 20. Thompson, B. (1992). The Use of Statistical Significance Tests in Research – Some Criticisms and Alternatives, Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, 20-24, ERIC Document Reproductive Service No. ED366654.
 21. Thompson, B. (1999). Why encouraging effect size reporting is not working: The etiology of researcher resistance to changing practices, *The Journal of Psychology*, 133 (2), 133-141.
 22. Thomas, L and Krebs, C. (1997). A Review of Statistical Power Analysis Software, *Bulletin of the Ecological Society of America*, 78 (2), 126-139.
 23. Thompson, B. & Synder, P. (1997). Statistical significance testing practices in The Journal of Experimental Education, *The Journal of Experimental Education*, 66 (1), 75-79.
 24. Welkowitz, J. , Ewen, R. and Cohen, J. (1982). *Statistics for the Behavioral Sciences*, Academic Press, New York, NY.
 25. Wilcox, R. (2003). Power – Basics Practical Problem and Possible Solutions, *Handbook of Psychology*, Hoboken, Wiley, NJ, (2) .
 26. Wilkerson, M. & Olson, M. (1997). Misconceptions about sample size, statistical significance, and treatment effect, *The Journal of Psychology*, 131 (6), 627-632.
-
-

An Analytical Study of Statistical Power in Published Administrative Research

Abdullah O. Alnajjar

Department of Statistics & Quantitative Studies
College of Administration and Planning, King Faisal University
Al-Hassa, Saudi Arabia

Abstract:

Statistical power is a significant topic for any researcher using statistical inference logic. Studies which use low levels of statistical power, obtain indecisive results, wasting the researchers' time and the effort put in collecting and analyzing data. To investigate this problem, this paper studies the statistical power of a number of statistical methods used in Administration and Management research works in the Journal of Public Administration published by the Institute of Public Administration.

The results obtained in the present paper show that the probability of having low type II error (i.e. the probability of achieving test power) in case of low effect, is higher than the probability of having this association in medium and high size effect. It is also found out that the average statistical power is high in the present corpus. But this high statistical test power results from the size of the sample, and not from the researchers' deliberate interest in statistical power. The large size of the data may have probably been used as a method which hides the desire to obtain statistical significance to support a weak practical significance. The increase in the sample size decreases type I error α , which affect the power of the test because of co-occurrence relationship between the two.

The above results should enable us to make the following recommendations:

1. Since the results have shown that researchers in Administration Sciences lack background knowledge in the area of statistical power, they need to be directed to the importance of this aspect of statistical operations and its role in their results.
 2. To increase statistical power, the paper suggests the following:
 - Undertaking a pilot study for statistical test, to evaluate the test effectiveness in relation to research variables. This can be achieved by defining the value of α in all the tests, and then by using the probable value of β , the relative effect of each test can be determined. This will help in choosing the most appropriate test for each research work.
 - Establishing the statistical power at a level no less than 0,80 together with a high level of significance. After this the smallest sample size should be used.
-