ASJP Algerian Scientific Journal Platform
مجلة دفاتر مخبر الشعرية الجزائرية
ISSN : 1112-9727 EISSN : 2676-1661
https://www.asjp.cerist.dz/en/PresentationRevue/459

المجلد:**08**، العدد: **01 (2023)** **ص:142 - 157**

اختبارات الدلالة الإحصائية (أوجه القصور والحلول البديلة)

Statistical Significance Tests (Deficiencies and Alternative Solutions)

د. عبد الحق بحاش*

جامعة الدكتور يحي فارس (المدية)

drabdelhak85@gmail.com

معلومات المقال

تاريخ الارسال:
2022/04/23
تاريخ القبول:
2022/10/16
تاريخ النشر:
2023/03/26

الكلمات المفتاحية:
- ✓ إختبارات الدلالة الاحصائية
- ✓ أوجه القصور
- ✓ الحلول البديلة

الملخص:

هدفت الدراسة الحالية إلى تقديم مفهوم شامل للدلالة الإحصائية وأساليبها المتنوعة في معالجة فرضيات مختلف البحوث النفسية والتربوية، كما ركزت الدراسة على تبيان أهم العوامل والعيوب التي تؤثر وتتأثر بها الدلالة الإحصائية جراء الاعتماد عليها لوحدها من طرف الباحثين على مختلف مستوياتهم مع إعطاء بعض الحلول البديلة التي من شأنها أن تضع الباحث في الطريق الصحيح عند صناعة أي قرار نفسي كان أم تربوي.

Article info

Received
23/04/2022
Accepted
16/10/2022

Keywords:
- ✓ *Statistical Significance Tests*
- ✓ *Deficiencies*
- ✓ *Alternative Solutions*

Abstract :

The present study aimed at comprehensive submission concept of statistical significance and its various methods by manipulating various assumptions psychological and educational research, the study also focused on the main factors and flaws that affect and influenced by the statistical significance relying on them by researchers at their different levels with alternative solutions that would put the researcher in the right direction in making any decision the psychological decision was an educational mother.

*عبد الحق بحاش

1. مقدمة

هناك اعتقاد سائد بأن العلاقة بين الإحصاء والبحث العلمي غالبا ما يساء فهمهما من قبل الباحثين الذين يتجاهلون الكثير من الاعتبارات، منها أن لغة الإحصاء لغة تواصل علمي وأداة جيدة من أدوات البحث العلمي، لكنه ليس هو نفسه البحث العلمي.

ويترتب عن هذه الحقيقة أن اختيار التحليل الإحصائي المناسب وتفسير البيانات بشكل صحيح واستخلاص النتائج هي مسؤولية تقع على عاتق الباحث، وأن التصميم الجيد للبحث بطريقة تلاءم مشكلة البحث المطروحة هي مسألة منهجية مختلفة عن التحليل الإحصائي وهي أيضا مسؤولية الباحث، فالتحليل الإحصائي الرفيع لا يمكن أن يكون تعويضا عن مشروع بحث لم يخطط له جيدا، أو عن أداة متواضعة لجمع البيانات، ذلك لأن الإحصاء لا يمكنه التكفل بإصلاح الأخطاء الناتجة عن إتباع منهجية غير ملائمة وموافاة الباحث بما يرغبه من بيانات واستنتاجات يستحيل أن توفرها المنهجية.

لكن المهمة لم تنته، فغالبا ما تنتج القياسات أرقاما كثيرة جدا، ولذا يجب تنظيمها وتلخيصها ورسم تمثيلها بيانيا، كما يجب القيام بحسابات الطبيعة العامة للقياسات، والعلاقات التي يجب أن توصف، وأخيرا هناك قرارات يجب أن تتخذ للإجابة على تساؤلات الدراسة، وهذه المهام كلها جزء من الإحصاء.

إن تطور علم الإحصاء ناجم عن طبيعة التطور في المعارف الأخرى ولهذا الغرض ظهرت العديد من النظريات والأساليب التي من أبرزها أساليب فحص الفرضيات التي تشكل الأساس في اتخاذ القرارات باستخدام نوعين من الدلالات إحداهما إحصائية والأخرى عملية. فالأولى تعني احتمال رفض الفرض الصفري باستخدام أساليب إحصائية، أما الثانية فيقصد بها أن تكون الفروق الإحصائية أو العلاقات بين المتغيرات كبيرة إلى درجة تبرر عملية الأخذ بنتائجها.

إذا فبلا شك أن الدلالة الإحصائية لها أهميتها في إظهار الفرق الإحصائي بين مجموعتين أو أكثر، ولكن هذا لا يعني الاعتماد على هذا الفرق لوحده واتخاذ القرار بناء عليه فقد أجمع التربويون على أنه لا يمكن لأي باحث إيجاد فرق الدلالة الإحصائية برفض الفرض الصفري.

صحيح أن الدلالة الإحصائية شرط ضروري لصناعة أي قرار تربوي أو نفسي ولكنه ليس كافيا، فالكفاية تتحقق بحساب قوة العلاقة بين متغيرين وقوة العلاقة هي ما يسمى بالدلالة العميلة التي أغفلها الكثير من الباحثين.

إن السبب الرئيسي الذي جعل الباحثين والتربويين والنفسيين يغفلون الدلالة العملية هو ناجم عن عدم فهمهم لها، وكذا الخلط في ذهن البعض منهم حول الدلالة العملية والإحصائية حيث وضح دانييل (1977) أن الخلط الظاهري في أذهان الكثير من الباحثين فيما يتعلق بالفرق بين الدلالة الاحصائية والدلالة العملية مهد

الطريق لإجراء الكثير من البحوث كمحاولة لإزالة هذا الخلط ومن هذه الدراسات دراسة هانسون وآخرون 1986 وماكنمار 1978 وبروتر وزملاؤه 1978 وكوهين 1983 وغيرهم.

ومن هذا المنطلق فإن إغفال الدلالة العملية وأهميتها يقلل من أهمية وقيمة النتائج في الدراسات وأحيانا قد يضللها فأساس البحث التربوي أو النفسي هو تقدير الدلالة العملية، خاصة إذا علمنا أن الدلالة الإحصائية والدلالة العملية يتأثران بالعديد من العوامل التي من بينها قوة الاختبار وحجم العينة ومستوى الدلالة وما إلى ذلك .

لذا فالباحث الجيد لابد أن يأخذ في الاعتبار عند تعامله مع الدلالة الإحصائية القضايا المتعلقة بالدلالة العملية، وقوة الاختبار وحجم العينة لمعرفة العلاقة الهادفة بين المتغييرات المستقلة والتابعة، ولوجود علاقة كبيرة بين المؤشرات الأربعة (الدلالة الإحصائية، الدلالة العملية، قوة الاختبار، وحجم العينة)، ومعرفة هذه العلاقات تمكن من اتخاذ القرار السليم حيال الظاهرة المدروسة .

ومن هذا المنطلق أمكننا صياغة التساؤلات التالية:

-ما هو المقصود بالدلالة الإحصائية؟

-ماهي العوامل تتأثر بها الدلالة الإحصائية؟

-ماهي الحلول البديلة لذلك؟

- **أهمية الدراسة:**

تكمن أهمية هذه الدراسة في كونها تقدم المعلومات اللازمة عن بعض المؤشرات التي تتأثر بها الدلالة الإحصائية والتي أغفلها العديد من الباحثين في مجال علم النفس وعلوم التربية وهذا من خلال توضيح مفهوم الدلالة الإحصائية والاعتماد عليها لوحدها في تفسير النتائج، فمن المعلوم أن ليس كل فرق دال إحصائيا هو ذا قيمة في أرض الواقع. وإضافة إلى إبراز قصور الباحثين في اعتمادهم في تفسير النتائج على الدلالة الإحصائية لوحدها فإن هاته الدراسة ستركز على إعطاء بعض المؤشرات التي تساهم في ترشيد القرارات المتخذة بشأن ظاهرة ما.

2.مفهوم الدلالة الإحصائية

عبر كارفر (1987) المشار إليه في دراسة (البارقي، 2012) بقوله :" إن الدلالة الإحصائية ببساطة تعني الندرة الإحصائية، فالنتائج قد تبدو ذات أهمية من وجهة النظر الإحصائية لأنها تظهر بشكل نادر في العينات العشوائية تحت شروط الفرض الصفري، وبالتالي فالدلالة الإحصائية تعني القليل أو لا شئ"

أي أن الدلالة الإحصائية تهتم بمستوي الثقة التي نوليها للنتائج فنقول ما دام الفرق دالا عند مستوي (0.05) فهذا يعني إن الفرق بين المجموعتين حقيقي وان مجتمع المجموعة الأولي يختلف عن مجتمع المجموعة الثانية، وأننا نثق في هذا الحكم بنسبة (95%). (منصور، 1997: 57- 58).

و يري ثومسون (1994) بأن قليلا جدا من الباحثين يفهمون ما تفعله اختبارات الدلالة الإحصائية وما لا تفعله، ولذلك نتائجهم عادة ما تكون سيئة التفسير، وحتي إذا فهم بعض الباحثين العناصر المكونة لاختبار الدالة الإحصائية، فإن مفهومهم عن الدالة الإحصائية غالبا لا يتكامل في البحث، فعلي سبيل المثال: تاثير حجم العينة علي الدلالة الإحصائية ربما يكون معروفا من قبل الباحث، لكن هذا الاستبصار لا يظهر عندما تفسر نتائج الدراسة التي تشمل عدة آلاف من أفراد العينة.

وقد أكد لان**lane(2000)**هذه المشكلة عندما راي أن هناك خلطا في المفهوم الدقيق للاحتمالية المحسوبة في اختبارات الدلالة الاحصائية، ويري هوستن **Huston (1993)** أنه كلما كان الباحثون والمحررون وطلاب الدراسات العليا والمجتمع اكثر إحاطة بفوائد وحدود الدلالة الإحصائية ... كان من المرجح, أن يكونوا اكثر قدرة علي عمل تفسيرات متقنة ومحنكة ومفيدة لمعني نتائج البحث.

ومما سبق يلاحظ أن الدلالة الإحصائية يتم قياسها عن طريق العديد من اختبارات الفروض التي تقوم علي اختبار الفرض العدمي (**NullHypothesis**)، ضد الفرض البديل (**Alternative Hypothesis**)، والتي يهم الباحث فيها رفض الفرض العلمي، حيث أن ذلك دلالة علي صحة توجهه، وذلك طبقا لمستوي الدلالة ، وتوزيع المعاينة (**Sampling Distribution**) المناسب، وتحديد القيمة الحرجة (**Critical Valeur**) علي ذلك التوزيع الاحتمالي. والتي تفصل بين منطقتي الرفض والقبول، ثم بمقارنة القيمة الحرجة مع القيمة المحسوبة من الاختبار الإحصائي نأخذ القرار إما برفض الفرض العدمي وبالتالي قبول الفرض البديل أو العكس.

1.2 أهم افتراضات اختبارات الدلالة الإحصائية:

يتطلب نموذج الدلالة الإحصائية البدء بمعاينة عشوائية، ولن ينتج اختبار الدلالة الإحصائية المستخدم بدون عشوائية معلومات صادرة حول احتمالية النتيجة عند صحة الفرضية الصفرية حيث يرى "وينر " **(Winer,2000:693)**أن انتهاك افتراض عشوائية المعاينة لعناصر من المجتمع وعشوائية التعيين للعناصر في المعالجات ربما يمنع الصدق كليا في الدراسات وذلك بسبب أن العشوائية تزودنا بالضمان أن الأخطاء موزعة بشكل مستقل بين وفي ظروف المعالجة **(Shaver.1992)** وقد أشار إلى ذلك أيضا "جلاس وهوبكينز" (1984) بقولهما أن الإحصاء الاستدلالي يعتمد على افتراض عشوائية المعاينة من المجتمعات، وأن العينات العشوائية واحدة من أسس بناء اختبار الفرضيات **(Shaver.1992)**

إن ضرورة والتزامية العشوائية يمكن أن تبرهن بشكل أكبر عندما يصاغ التساؤل التالي في اختبار الدلالة الإحصائية: "ما مدى تمثيل العينة (إحصاء العينة) للمجتمع (معلم المجتمع) كما هو محدد في الفرضية الصفرية؟ " بدون عشوائية لايمكن الإجابة على هذا السؤال بصدق باستخدام اختبارات الدلالة الإحصائية الشائعة **(shaver. 1992)** ، وكما أوضح ذلك "جلاس وهوبكينز" بأن طريقة الاختيار العشوائي للعينات

يؤكد- مع هامش معروف ومؤكد من الخطأ – تمثيل العينات، ويسمح بتأسيس حدود والتي من المتوقع أن تقع المعالم فيها باحتمالية معينة **(Shaver.1992) .**

إن القدرة على تقدير درجة الخطأ بسبب الصدفة (خطأ المعاينة) خاصية مهمة للعينة العشوائية وأنه من غير الممكن تقدير درجة الخطأ في عينة صدفية وفي العديد من استراتيجيات المعاينات الأخرى بسبب أنها تحتوي على أنواع ودرجات غير معروفة من الصدفة وخطأ المعاينة وللتوضيح ذلك ضرب "شيفر" المثال التالي:

افترض أن رجلا سار إلى غرفة ورأى (10) عملات على الطاولة، (8) منها كانت على وجهها الأمامي و(2) كانت على جانبها الخلفي تساءل الرجل عما إذا كانت تلك العملات متحيزة؟ وتساءل عن احتمالية أن يكون هذا الترتيب قد حدث بالصدفة، تناول الرجل مرجعا فوجد أن احتمالية الحصول على (8) واجهات أمامية هي ($\frac{45}{1024}$)أي (0.044) ولان هذه الاحتمالية هي أقل من مستوى التقليدية (0.05)، فانه استنتج أن النتيجة من غير المحتمل أن تحدث بالصدفة عند صحة الفرضية الصفرية (50%- 50%)، ولذا فقد اعتبر أن لديه برهان بأن العملات متحيزة في الحقيقة، مثل هذا الرجل ليس لديه إثبات على تحيز تلك المجموعة من العملات، والسبب ببساطة أنه لم يتأكد بأن ذلك الوضع للعملات أتى بعد رميها، أي أنه لم يعرف العملية التي من خلالها وصلت تلك العملات إلى ذلك الوضع وعليه فان التوزيع النظري الثنائي سيكون قليل التعلق بالموضوع لان البيانات لم تنتج بالطريقة التي تتوافق مع الافتراضي الأساسي لاستخدام التوزيع، تماما مثل التوزيع الثنائي الذي لم يتمكن من تزويدنا بمعلومات حول التحيز الممكن في العملات المشاهدة فان الباحثين التربويين الذين يستعملون مجموعات غير عشوائية لا يمكن الحصول على معلومات صادقة حول احتمالية فروق مجموعات- عند صحة الفرضية الصفرية – باستخدام الاختبارات الإحصائية المعروفة، أما المعاينة العشوائية فإنها لا تنتج ذلك، وكذلك مقارنة العينة غير العشوائية بتوزيع المعاينة المولدة عشوائيا لا يمكن أن تزودنا بجملة احتمالية صادقة للحدث**(Shaver.1992) .**

لقد وصف (**Shaver.1992**) انتهاك هذا الافتراض في أرض الواقع من قبل الباحثين حيث رأى بأنه في البحث التربوي عادة ما يذهب الباحث إلى مدرسة أو عدة مدارس ويحصل على مجموعات موجودة (بدون عشوائية المعالجة ، ولا عشوائية الاختيار) ، ثم يجمع البيانات أحيانا بمعالجات وأحيانا بدونها – وبعد ذلك يطبق اختبارات الدلالة الإحصائية. نتائج تلك الاختبارات الاستدلالية ليس لها بالضرورة معنى، الا أذا كان أحدهم مهتما بالمقارنة بمعيار منتقى بالاحتمالية مثل الموضح بالسؤال: "ما هي احتمالية الحصول على النتيجة (المحصلة) إذا كانت العينات العشوائية سحبت فعليا؟ وباختصار ، فان اختبارات الدلالة الإحصائية الشائعة

تزود الباحث بمعلومات محددة عن مقدار احتمالية النتيجة بافتراض أن الفرضية الصفرية صحيحة وعند عشوائية المعاينة والتعيين وحجم العينة (**n**)، وبدون عشوائية تكون نتيجة اختبار الدلالة الإحصائية عديمة المعنى، في أحسن أحوالها تكون صلتها بجملة احتمالية غير محددة أو معروفة.

3.العوامل المؤثرة في اختبارات الدلالة الإحصائية

ويتضح مما سبق أن اختبارات الدلالة الإحصائية تدل الباحث على احتمالية حدوث النتيجة فقط، وذلك بافتراض صحة الفرضية(H_0) وعند عدد (**n**)من حالات العينة العشوائية، غير أن هناك العديد من العوامل المؤثرة في هذه الاختبارات وعلى الباحث أن يعرف هذه العوامل المؤثرة وهذا ما يساعده في فهم أدق لمفهوم نتائج اختبارات الدلالة الإحصائية، وقد حدد شنيدر ودارسي(**Snyder & Darcy,1984**)المشار إليهما في دراسة (بابطين،2001) سبعة عوامل محددة لنتيجة اختبارات الدلالة الإحصائية وهي:

1-عدد الحالات المستعملة في الدراسة (حجم العينة)

2-مستوى الدلالة المختارة.

3-التشتت في المتغيرات ذات العلاقة.

4-القوة الفعلية للتأثير.

5-مناسبة القياس الإحصائي والاختبارات الإحصائية المستخدمة.

6-الفرضية المختبرة.

7-تعقيد التحليل.

وسيتم التفصيل في العوامل الثلاثة الأولى فقط وفقا لما تقتضيه هذه الدراسة:

1.3تأثير حجم العينة على نتيجة اختبارات الدلالة الاحصائية:

يرى ثومسون **(Thompson .1989)** و ويلكنسون **(Wilkirson .1992)** وسنايدر ولاوسن **(Snyder & Lawson, 1992)**بأن أكبر مؤثر على نتيجة اختبار الدلالة من العوامل المذكورة هو عدد الحالات المستعملة في الدراسة أي حجم العينة**(n)** ، وأن الوصول إلى الدلالة الإحصائية وان كانت دالة لسبعة خصائص في الدراسة على الأقل، إلا أن حجم العينة هو المؤثر الأساسي الذي جعل النتائج دالة أو غير دالة إحصائيا.

قام دانييل **(Daniel,1998)** بعرض طريقة لتوضيح أثر حجم العينات في نتائج اختبارات الدلالة الإحصائية كما يلي:

بافتراض أن باحث يجري تجربة تربوية والتي عين فيها مجموعتين بطريقتين تدريسيتين مختلفتين وبعد ذلك، أراد أن يقوم نتائج مقياس الانحراف، هذا الباحث ربما يستخدم تحليل التباين لتقويم نتيجة التجربة، حيث يفترض أولا الفرضية الصفرية ...والتي تنص على عدم وجود فروق بين الأشخاص في نفس الظروف

التجريبية، بعد ذلك يحسب قيمة الاحصاءة **(F)** والتي عن طريقها سيقوم الفرضية الصفرية **(H_0)**، وحيث أن **(F)**ببساطة عبارة عن إحصاءات نسبة، تعتمد على حاصل قسمة متوسط مربعات التأثيرات **Ms** ،على متوسط مربعات الخطأ، ونظرا لان متوسط المربعات هي نتيجة حاصل لقسمة مجموع المربعات لكل تأثير على درجات الحرية، وأن متوسط مربعات الخطأ سوف يصبح أصغر كلما زاد حجم العينة، فان هذا المقام القاسم لمتوسط مربعات التأثير سينتج قيمة أكبر لاحصاءة **.(F)** وكمثال على ذلك فان العينة المتكونة من (2002) حالة ستكون أكثر احتمالية بمقدار خمس مرات لنتائج الدلالة الإحصائية من عينة مكونة من (602) حالة فقط، وهذا يعني أنه مع عينة كبيرة مثل (2002)، حتى الفروق الضئيلة بين مجموعتين ستكون دالة إحصائيا**(Daniel, 1998: 63-65).**

وقد لخص ثومسون**(Thompson)** هذا بقوله: "قيمة **(P)** المحسوبة في دراسة معطاة، عبارة عن دالة لعدة خصائص للدراسة، ولكنها متأثرة بالتحديد بخليط التأثيرات لحجم العينة المستخدمة في الدراسة و حجوم التأثير للدراسة، وبسبب أن قيمة **(P)**مركبة، نجد أنه نظريا يمكن أن تكون لـ(100) دراسة بأحجام عينات مختلفة و(100) حجوم تأثير مختلفة قيمة **(P)**محسوبة واحدة، كما يمكن أن يكون لـ(100) دراسة بحجم تأثير واحد (100) قيمة محسوبة ومختلفة لـ**.(48-35 :Thompson1998) "(P)**

ان الاعتماد الواضح و الكبير لنتائج اختبارات الدلالة الإحصائية على حجم العينة جعل العديد من الباحثين يذهبون إلى أنه افتراضيا عندما نتعامل مع عينات كبيرة فان كل الفرضيات الصفرية سوف ترفض لان الفرضية الصفرية لعدم وجود فروق في الغالب ليست تماما صحيحة في المجتمع، وعبر عن ذلك أيضا (هايز) بقوله: "افتراضيا، أي دراسة يمكن أن تظهر نتائج دالة إحصائيا إذا استخدم الباحث مفردات كافية **."(Palomers, 1990).**

وذهب أيضا إلى ذلك كارفر **(Carver,1993)** بقوله: "أن من النادر أن يكون الارتباط أو الفرق بين المجموعات بالتمام صفرا (كما تفرضه الفرضية الصفرية)، ولذا فان الفرضية الصفرية **(H_0)** أو فرضية عدم وجود فرق نادرا ما تبقي **.(Carver,1993:287)**"هذا يعني عمليا مع حجم عينة كبير أي فرضية صفرية يمكن أن ترفض، وقد قرر هذا أيضا "ماكلين" (**Maclean**,1997)بقوله: "اختبار الدلالة الإحصائية انتقد بشكل متكرر بسبب اعتماده المفرط على حجم العينة، وعدم القدرة على توضيح معنى النتائج ....والباحثون الذين لديهم خبرة العمل مع عينات كبيرة، يدركون حالا بأن كل الفرضيات الصفرية سوف ترفض عند مستوى معين من حجم العينة.....، ذلك لان الفرضية الصفرية المأخوذة حرفيا، دائما خاطئة (مرفوضة) ...واقعيا، الدلالة الإحصائية يمكن تحققها عند حجم معين من العينة، أنها ببساطة تصبح مسألة الحصول على عدد كافي من العناصر.

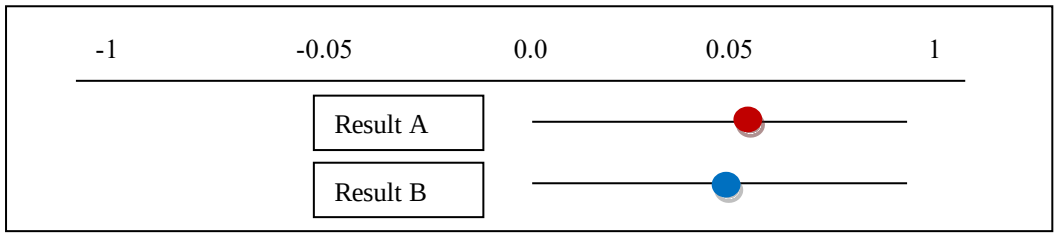
وكما رأى دانييل (**Daniel,1997**) بأن الباحث في الحقيقة يمكن أن يعرف حتى قبل البدء بالتجربة بأن اختيار الدلالة الإحصائية هو اختبار لمدى ما إذا كانت العينة كبيرة أم لا. وإذا جمعت بيانات كافية فان الفرضية الصفرية (H_0) عادة سوف ترفض، وستكون الفروق صفرية في تقديرات معالم المجتمع من العينات الكبيرة، بغض النظر عن عدم دلالتها العملية، نتيجة دالة إحصائيا .

2.3 مستوى الدلالة الإحصائية ألفا(α)

ذكر (عادل بابطين ،2001) في دراسته أن اختبار الدلالة الإحصائية ينفذ بعد تحديد مستوى الدلالة (**α**)، في احتمالية أن تؤخذ النتيجة المعطاة بسبب أخطاء المعاينة، بمعنى احتمالية ارتكاب الخطأ من النوع (**I**) وهو رفض الفرضية الصفرية وهي صحيحة، ومستوياتها(**α**) التقليدية (0.05) أو (0.01) والمستعملة من قبل الباحثين هي تقليد عشوائي (**Kellaw,1998**).

وقد ناقش هوبكينز **(Hopkins,1997)** مفهوم مستويات الدلالة الإحصائية وتطورها بقوله: "ما هو الخاص في القيمة 0.05 ؟ في الحقيقة لا شيء ، شخص ما قرر أن هذه القيمة معقولة، ولذا فإننا الآن جامدون عندها **(P)<0.01** أصبحت أيضا تقليدا لتحديد الدلالة الإحصائية، كلا القيمة الدقيقة لـ **(P)**لاحصاءة الاختبار بدلا من ذلك استخدام الناس الجداول لقيم احصاءة الاختبار المقابلة لقيم قليلة ومختارة وعشوائية لـ **(P)**وهي (0.05)، (0.01) وفي بعض الأوقات (0.001) هذه القيمة أصبحت مقدسة كعتبة حدية القيم التي تحدد الدلالة الإحصائية، وعادة ما يكون على الباحث أن ينص على أي واحدة يستخدم. فعلى سبيل المثال إذا نص الباحث على أن مستوى الدلالة هو (5%)، اذا يسمح له باعتبار أي نتيجة لها قيمة **(P)** تساوي أو أقل من (0.05) بأنها دالة إحصائيا.

وعلى هذا فقد عرض "هوبيكنز" مثالا توضيحيا لمشكلة هذا التقليد الذي يشبه التقديس لاستخدام قيم محددة لمستويات الدلالة حيث رأى بأن الباحثين والإحصائيين يدعون أن القرار يجب أن يعمل لمعرفة ما إذا كانت النتيجة دالة إحصائيا، وفقا لهذا المنطق إذا كانت (**P< 0.05**) فالباحث لديه نتيجة قابلة للنشر.[لاحظ مشكلة التحيز في البحث التربوي للنتائج الدالة إحصائية] وإذا كانت (**P> 0.05**) فنتيجة الباحث ليست كذلك. والشكل رقم (01) يوضح كما يرى هوبكينز، حيث أن إحدى النتيجتين دالة إحصائيا والأخرى غير دالة إحصائيا فأي النتيجتين صالحة للنشر؟ الإجابة كلاهما، لأن القيمتين (**P**) هما (0.049) و (0.051). (**Hopkins,1997**).



شكل رقم (01) مثال على مشكلة الاعتماد على مستوى الدلالة الإحصائية

وكما يرى ماكلين(**Mclain,1997**) بأن مستوى الدلالة (0.05) أصبح حارس بوابة عشوائي للمعرفة العملية، بحيث أن النتائج التي لا تقابل المعيار ينظر على أنها غير مهمة، وأن هذه الآلية في تحديد مستوى الدلالة الإحصائية للوصول إلى القرار الإحصائي تبدوا محددة لصنع القرار مما يعطي شعورا بالموضوعية العلمية لكنها في الواقع ليست لها معنى، لان تكلفة هذه الموضوعية هي المعنى.

ويرى الكثير من الباحثين ومنهم ماكلين أن فيشر دعم مستوى الدلالة (0.05) كمعيار للدلالة الإحصائية لكن من المثير الملاحظة أن فيشر (**Fisher**) لم يضع مدا ثابت لمستوى الدلالة، بل انه دعا الباحثين لتقويم الحالات الفردية بالرجوع للبيانات والنظريات.

3.3التشتت في المتغيرات ذات العلاقة:

ذكر (عادل بابطين، 2001) في دراسته أن نتيجة اختبار الدلالة الإحصائية تعتمد على قوة الاختبار في رفض الفرض الصفري بطريقة صحيحة والحصول على نتائج دالة إحصائيا، **(Wilkirson,1992)**هذه القوة الإحصائية للاختبارات بذاتها تعتمد من ضمن ما تعتمد عليه على ثبات البيانات، الذي بدوره يعتمد على التشتت أو التغاير للبيانات، وقد لاحظ العديد مثل، **(Huston,1993)**ونيكس وبارنيت **(Nix and Barnette,1998)**أن أكثر طريقة مشاعة لزيادة الثبات والقوة الإحصائية هي زيادة حجم العينة، وعلى أي حال، فان عامل حجم العينة الحاسم قد سبق مناقشته ولكن يلاحظ أن درجة تشتت الدرجات عند ثبات حجم العينة تؤثر في قيمة الاحتمالية المحسوبة **(P)**بحيث أن زيادة تقلب وتغاير وتشتت الدرجات ينتج قيمة أعلى للاحتمالية المحسوبة **(P)** ، بينما يؤدي انخفاض درجة التشتت إلى انخفاض قيمة **(P)** ومن ثم الحصول على نتائج دالة إحصائيا.

وهناك طريقة أخرى يلاحظ بها تأثير تشتت الدرجات هي قيمة بعض إحصاءات اختبار الدلالة الإحصائية مثل اختبار (**F**)، حيث أن قيمة الاحصاة هي:

$$F = \frac{S^2_{between}}{S^2_{within}}$$

ونظرا لان المقام (**withins²**) متأثر باختلاف ضمن المجموعات الذي يعتمد على التباين في الدرجات ضمن المجموعات، فان كلما قل ذلك التباين انخفضت قيمة المقام وبذلك ارتفعت قيمة (**f .cal**) مما يزيد احتمالية الحصول على نتائج إحصائية. (**Nix and Barnette. 1998**)

4.قوة الاختبار الإحصائي كحل بديل والعوامل المؤثرة فيه

1.4قوة الاختبار (Power of the test):

يعني هذا المصطلح قدرة الاختبار على رفض الفرض الصفري عندما يكون خاطئا بالفعل ، وهو يشير الى المساحة الواقعة تحت المنحنى في منطقة الرفض (الصائغ، 1996).

وتعتمد قوة الاختبار الإحصائي على مستوى الدلالة والخطأ من النوع الثاني بيطا (ß) وحجم العينة وبالتالي فان قوة الاختبار الإحصائي تساوي 1-ß ، وهي احتمال قرار رفض **(H_0)** عندما يكون الفرض البديل صحيحا وتزداد قوة الاختبار الإحصائي عن طريق زيادة مستوى الدلالة وتباين الدرجات وحجم العينة وتزداد قوة الاختبار الإحصائي أيضا كلما انخفضت قيمة (ß)، وتتراوح قوة الاختبار الإحصائي فيما بين الصفر كحد ادنى والواحد كحد اقصى ، وتكون قوة الاختبار مقبولة حينما تكون بين (0,40، 0,60) (الدردير، 2006: 56).

ان صيغة العلاقة لحسابه حسب ما أشار إليه البلداوي (2008)

$$\beta = \sqrt{n}(\mu - \mu_0)/_{\sigma}$$

فلو كان لدينا : $\square = 25,6$ ، و$\mu_0 = 25$، و$\square = 11$ ، و$n = 5000$

فان قوة الاختبار الاحصائي او احتمال رفض الفرضية الصفرية هو:

$$h_0 = \mu_0 = 3,85$$

يتم بعدها ايجاد قيمة متوسط العينة $(\bar{X})$ الذي يؤدي الى رفض (H_0) كالاتي :

فالخطأ المعياري ل $(\bar{X})$ يتمثل بالصيغة الاتية بحيث ان $(S_{\bar{x}})$ هو الخطا المعياري للعينة:

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11}{\sqrt{5000}} = 0,155$$

فعند مستوى دلالة (α=0.05) نرفض (H_0) إذا ما وقع $(\bar{X})$ خارج :

$$25 \pm (1.96)(0.155)$$

$$25 \pm 0,304$$

أي أن قرار رفض (H_0) هو:

$$\bar{X} < 24,696$$

$$\bar{X} > 25,304$$

فاحتمال وقوع $(\bar{X})$ عند اقل من (24,696) وباستخدام الصيغة التالية :

$$\square = \frac{\bar{X} - \mu}{\sigma/n} = \frac{24.696 - 25.6}{0.155} = -5.832$$

فاحتمال وقوع $(\bar{X})$ عند اكبر من (25,304) هو :

$$\square = \frac{25.304 - 25.6}{0.155} = -1.909$$

ومنه:

$$P(0\ to - 5.832) = 0$$

$$P(0\ to - 1.909) = 0,9712$$

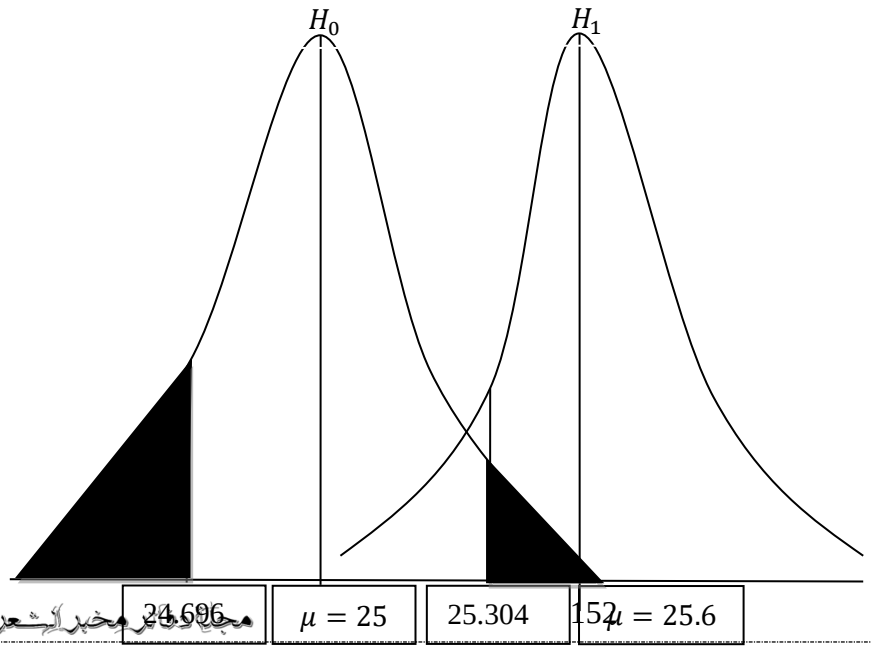
اذن قوة الاختبار هي :

$$0 + 0,09712 = 0,9712$$

بكلمة اخرى فان الوقوع في الخطأ من النوع الثاني ضئيل جدا عند احتمال :

$$P = 1 - \beta = 1 - 0.9712 = 0.0288$$

(البلداوي، 2008: 287)



الشكل رقم (02) يوضح قوة الاختبار الإحصائي في رفض الفرضية الصفرية. المصدر: البلداوي، 2008

2.4العوامل المؤثرة في قوة الاختبار:

تتأثر قوة الاختبار بعدد من العوامل ، حيث وضح كل من كوهين (Cohen;1977) و(الثبيتي،1993)، (وعلام ، 1993) ان العوامل المؤثر في قوة الاختبار هي:

أولا/ حجم العينة:

يعتبر حجم العينة من اهم القرارات التي يتخذها الباحث عند اختباره لها حتى تعطي اكبر قدر من الدرجة باقل تكلفة.ويجدر بالباحث قبل تكوين عينة البحث ان يحدد اولا مجتمعه الاصلي بدقة وان يعد قائمة كاملة بجميع مفردات ذلك المجتمع، ثم يأخذ من هذه المفردات عدد كبير نسبيا حيث يكون ممثلا لخصائص المجتمع الاصلي، وبوجه العموم فان حجم العينة الكافية يتحدد بثلاث عوامل اساسية :

-طبيعة المجتمع الاصلي

-طريقة اختبار العينة

-درجة الدقة المطلوبة

كما اضاف كوهين على هذه العوامل عوامل اخرى هي(Cohen,1977:14)

-حجم الاثر الادنى المحدد من قبل الباحث منسوبا للانحراف المعياري (SD)

-قوة الاختبار الإحصائي المستخدم في تحليل النتائج

-اتجاهية الاختبار بمعنى هل هو اختبار ذو اتجاه ام ذو اتجاهين.

ونظرا لاهميتها فان عدم قدرة الباحث في تحديد حجم العينة او اهمالها يؤدي الى مضاعفات جسمية توصله الى نتائج بعيدة عن الصحة، فالسؤال الذي يطرح نفسه هو : كيف يختار الباحث حجم العينة التي تتناسب مع اهداف بحثه ؟ وقد اتفق الاحصائيون ان الخطوة الاولى هي ان يقرر الباحث مستوى الثقة الذي يريده في نتائجه ويرمز له بالرمز (ث) ثم يقوم بتطبيق المعادلة الخاصة بحجم العينة وهي :

$$\text{عن} = \left\{ \frac{\text{المعياريالانحراف للمجتمع} \times \text{النسبة المئوية للثقة ممثلة معياريبدرجةانحراف}}{\text{الخطأ المعياري للمتوسط الحسابي}} \right\}^2$$

مثال : إحدى الشركات المنتجة لمسحوق الحليب تريد معرفة العينة التي يجب استخدامها لضبط الجودة في إنتاجها اليومي الذي يبلغ عدة مئات من العلب بسعة ثلاث كيلوغرامات للعلبة الواحدة، فإذا كانت نسبة الثقة المطلوبة 99%بالإضافة إلى المعلومات التالية:

•إن معدات التعبئة بتفاوت قدره ¾ أوقية للعلبة الواحدة

•إن درجة الخطأ المعيارية المسموح بها (درجة الخطأ للمتوسط الحسابي) هي اوقية، إذن عدد العلب في العينة

$$\text{ن ع} = \left\{\frac{0.75 \times (2.576)}{0.25}\right\}^2 = \left\{\frac{3/4 \times (1 \times 2.576) \text{ أي } 1\%}{1/4}\right\}^2 = 59{,}72 \approx 60 \text{ علبة}$$

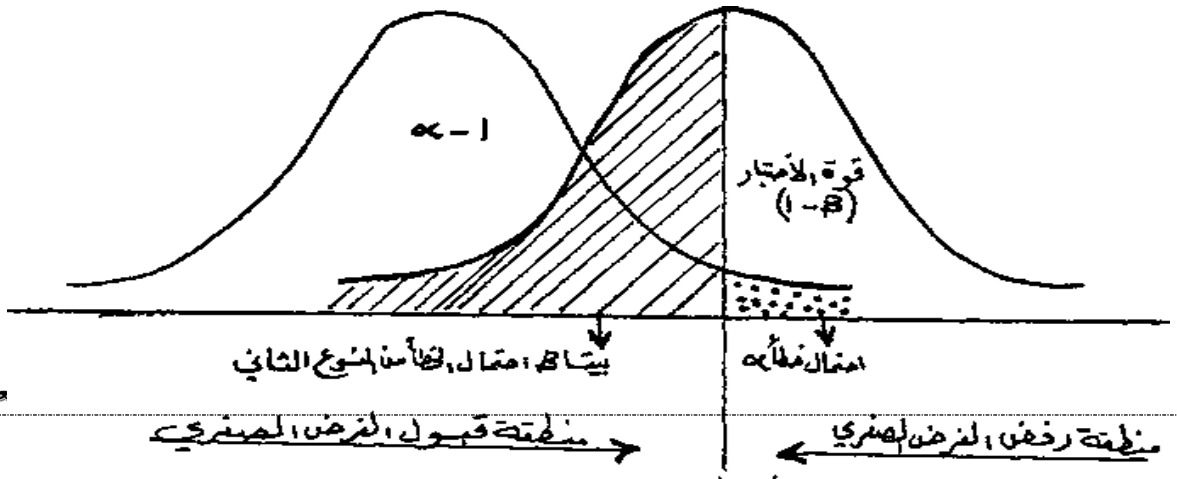
(المغربي ،2006: 140- 222)

فزيادة حجم العينة قد يفيد في حالات معينة وهي (الصائغ، 1996):

- وجود متغيرات غير مضبوطة بأي طريقة من طرق ضبط المتغيرات
- توقع إعادة تقسيم المجموعة الكلية إلى مجموعات جزئية حسب مستويات المتغيرات المستقلة
- عندما لا يكون المجتمع متجانس .
- عندما يكون ثبات المقياس للمتغير التابع منخفضا . مما يعني وجود أخطاء في القياس وقد رأى كوهين أن العلاقة بين حجم العينة وقوة الاختبار علاقة مباشرة فالزيادة في حجم العينة نزيد من القوة الإحصائية مع ثبات العوامل الأخرى عندما يكون الفرض (H_0) غير صحيح ويكون القرار رفض الفرض الصفري .

ثانيا/ مستوى الدلالة (α) (Level of Significance):

تزداد قوة الاختبار كما ذكرت (الصائغ ،1996) بازدياد قيمة مستوى الدلالة فزيادة (α) تنقص (β) وبالتالي تزداد قوة الاختبار باعتبار أن (P=1-β) ، ففي الشكل رقم (03) تكون الزيادة لقيمة (α) بتحريك الخط الرأسي إلى اليسار وبالتالي تنقص قيمة (β) ومن ثم ترتفع قوة الاختبار وقد أشار (الصياد،1985) إلى أن المعيار الأساسي لاختبار الفروض الإحصائية من حيث اختبار مستوى الدلالة هو قوة الاختبار (P) بحيث يجب أن تكون قيمته على الأقل اكبر من حد الصدفة (0,50) ، ورأى كوهين أن تكون (0,80) كحد أعلى لتعامل الباحث مع المتغيرات غير الثابتة دائما ويكون الخطأ (β) مساويا (0,20) ، ويمكن تحديد قوة الاختبار مسبقا قبل إجراء الدراسة لمعرفة أي الاختبارات الإحصائية أكثر قوة للتعامل معه



الشكل رقم (03) يوضح العلاقة بين كل من قوة الاختبار ومستوى الدلالة والخطأ المصدر : الصائغ،1996

ثالثا/ علاقة القيمة الحقيقية للمجتمع بقيمته في الفرضية الصفرية :

كلما زاد الاختلاف بين المتوسط الحقيقي (الأصلي للمجتمع) والمتوسط الفرضي زادت قوة الاختبار وقل الخطأ من النوع الثاني (β)

رابعا/ الاختبار بطرف والاختبار بطرفين :

تزداد قوة الاختبار عندما يتحقق الباحث من صحة الفرض الصفري (H_0) باختبار إحصائي ذي طرف واحد . فإذا حدد الباحث المتوسط في ضوء الفرض الصفري ، وكان هذا الافتراض صحيحا فالخطأ من النوع الثاني (β) اقل مما لو استخدم اختبار ذو طرفين .

خامسا/ الاختبارات البارامترية واللابارامترية :

تتحدد قوة الاختبار وفقا للاختياريين البارامتري، حيث ان الاختبار البارامترية اكثر قوة من نظارها اللابارامترية لاي قيمة معينة فاحتمال وقوع الخطا من النوع الثاني يكون اقل في حالة الاختبارات الاحصائية للفرض الصفري التي تفترض اعتدالية توزيع المجتمعات وتساوي التباين حيث الاساليب البارامترية تراعي القيم الاصلية للدرجات مما يجعلها اكثر دقة في التواصل للاحتمالية عند ما تتحقق الفروض .

وتختلف الاساليب الاحصائية المستخدمة لاختبار الفروض الصفرية في مدى قوتها فالاساليب التي تميل الى رفض الفرض الصفري تسمى اساليب اكثر قوة من غيرها ، اما الاساليب الاقل قوة فيفضل استخدامها في البحوث ذات العينات الكبيرة .ولهذا تعتبر قوة الاختبار الاحصائي من اهمم حكات اختبار النموذج الاحصائي المناسب .

سادسا/ الدلالة العملية (حجم التاثير) (Effect Size):

وهي من العوامل التي يرى (الثبيتي ،1993: 62) انها اضافة الى ما سبق ذات اثر في قوة الاختبار فهو مفهوم عام يتم تطبيقه في كثير من الاختبارات الاحصائية للفرض الصفري . ويعتبر مقياس لدرجة خطأ (H_0) في المجتمع الاحصائي ، كما يعتبر بديلا نوعيا للفرض (H_0) في مقابل (H_1) العام وهذا المفهوم يساعد في عملية تحليل القوة ، فكلما زاد حجم التاثير قل حجم العينة اللازم للوصول لقيمة معينة دالة احصائيا ويمكن ايجاد قوة الاختبار بالاعتماد على حجم التاثير.

5. خاتمة

إن إتباع الباحث للمؤشرات الاحصائية الملازمة والملائمة بنوعيها (دلالة إحصائية ودلالة عملية) من خلال البيانات يقدم له فائدتين كبيرتين، أولهما انه يقدم تقدير حول مدى العلاقة بين المتغيرين المستقل والتابع . أي انه لا يكتفي بالإجابة عن الفرضية الصفرية " بنعم أو" لا " بل انه يحدد مقدار اثر المتغير المستقل على المتغير التابع، حيث انه في بعض الأحيان قد يكون الفرق بين المجموعتين التجريبية والضابطة كبيرا ولكنه غير دال إحصائيا ومن جهة أخرى قد يكون هذا الفرق صغير جدا ولكنه دال بسبب كبر حجم العينة أو نتيجة استخدام مستوى دلالة مرتفع، أما الفائدة الثانية فتتمثل في انتلك المؤشرات تزود الباحثين الآخرين بفكرة عن قوة العلاقة أو حجم التأثير للمتغير المستقل على المتغير التابع من هنا فان حساب مقدار حجم الأثر يقدم معلومات وفائدة لا تقل أهمية عن عملية فحص الفرضيات الإحصائية، فالاهتمام بحساب الدلالة الإحصائية في البحوث التربوية ليس أمر كافيا لبيان أن هذه العلاقة الارتباطية أو مقدار الفرق بين المتوسطات له اثر وقيمة تربوية، ولهذا لابد من حساب الدلالة العملية إلى جانب الدلالة الإحصائية ذلك لتبيان نسبة تباين المتغير التابع التي ترجع إلى المتغير المستقل وليس إلى عوامل أخرى خارجية .

قائمة المراجع

1- ابتسام حسن مدني الصائغ (1995): الدلالة الإحصائية والدلالة العملية لاختبار (ت) و(ف): دراسة تقويمية مقارنة لرسائل الماجستير في كلية التربية بجامعة أم القرى، رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى، المملكة السعودية.

2- رشدي فام منصور : حجم التأثير الوجه المكمل للدلالة الاحصائية، المجلة المصرية للدراسات النفسية ، المجلد 7، العدد 16، ص.ص 50-67.

3- طلال هيازع حسن البارقي (2012): واقع الدلالة الإحصائية والدلالة العملية للبحوث المنشورة بمجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية، في المدة من 1425-1430هـ) ، جامعة أم القرى، المملكة العربية السعودية.

4- عادل أحمد بابطين (2001): مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة، رسالة ماجستير، كلية التربية، جامعة أم القرى، المملكة العربية السعودية.

5- عبد الحميد عبد المجيد البلداوي (2008): الأساليب الاحصائية التطبيقية، عمان، الأردن، درا الشروق.

6- عبد المنعم أحمد الدردير (2006): الإحصاء البارامتري واللابارامتري في اختبار الفروض النفسية والتربوية والاجتماعية، القاهرة ، عالم الكتاب.

7- Carver, R, P (1993): The Case Against Statistical Significance Testing, Revisited, The Journal of Experimental Education, 61 **.(4)**

8- Cohen, J (1977): Statistical Power Analysis for the Behavioral Sciences. New York :Acadimic Press **.**

9- Daniel, Larry G (1998): Statistical Significance Controversy Is Definitely Not Over A Rejoinder to Responses by Thompson, Knapp, and Levin, Research in the Schools. V5 n2 .

10- Hopkins, Will G (1997): A New View of Statistics. (Online Book) .

11- Shaver, James P (1992): What Statistical Significance Testing Is, and What It Is Not. Paper presented at the Annual Meeting of the American Educational Research Association, (San Francisco, CA, April 20-24) .

12- Thompson, Brus (1994): Common Methodology Mistakes in Dissertation, Revisited, Paper presented at the Annual Meeting of the American Educational Research Association, (New Orlens, LA, April)

13- Thompson, Brus (1998): Common Methodology Mistakes in Dissertation Quality, Paper presented at the Annual Meeting of the Mid-South Educational Research Association, (Louisvill, KA, Nouvember 8-11)

14- Winer, Russell S (2000): Comment on "The Historical Growth of Statistical Significance Testing in Psychology - and Its Future Prospects". Educational and Psychological Measurement. V60, n5, October.