الدلالة العملية ضرورة حتمية في البحوث النفسية والتربوية
مؤشر "كوهين" لحالات اختبار "ت"

د. محمد حسين سعيد حسين (*)

المُلخص:

يمثل الإحصاء التربوي أهمية كبيرة للباحثين في مجال العلوم النفسية والتربوية والاجتماعية، إذ يساعدهم في اتخاذ قرارات بشأن فروضهم البحثية، من جهة قبولها أو رفضها، اعتمادا على مستوى الدلالة الإحصائية للاختبارات المستخدمة في اختبار صحة هذه الفروض. وقد تعرضت مستويات الدلالة الإحصائية والاعتماد عليها إلى كثير من أوجه النقد، مثل ارتباطها بشروط الاختبار الإحصائي المستخدم، ومدى التزام الباحث بهذه الشروط، كذلك تأثرها بحجم العينات المستخدمة، فكلما زاد حجم العينة؛ كانت نتيجة الاختبار الإحصائي المستخدم ذات دلالة إحصائية. لذا دعت الكثير من الجمعيات النفسية المتخصصة إلى ضرورة استخدام مؤشرات الدلالة العملية بجانب الدلالة الإحصائية. ويعد اختبار «ت» بحالاته المختلفة؛ لعينة واحدة، ولعينتين مستقلتين، ولعينتين مرتبطتين من أكثر الاختبارات الإحصائية استخداما في البحوث النفسية والتربوية. ويتناول هذا المقال مؤشر كوهين بوصفه من أهم مؤشرات الدلالة العملية وطرائق استخدامه وتفسيره في حالات اختبار «ت» المختلفة للمقارنة بين المتوسطات.

الكلمات المفتاحية: الدلالة الإحصائية، الدلالة العملية، حجم التأثير، اختبار ت، مؤشر كوهين.

(*) أستاذ علم النفس التربوي كلية التربية جامعة بني سويف.mohamedhussein40@yahoo.com

Practical Significance is an Imperative in Psychological And Educational Research "Cohen's Index in Cases of T-Test"

Mohamed Hussein Said Hussein(*)

Abstract:

Educational statistics are of great importance to researchers in the field of psychological, educational and social sciences, as it helps them to make decisions about their research hypotheses in terms of accepting or rejecting them depending on the level of statistical significance of the tests used to test these hypotheses. The levels of statistical significance and dependence on them have been subjected to many aspects of criticism, such as their connection to the terms of the statistical test used and the extent of the researcher's commitment to these conditions, as well as being affected by the size of the samples used. Therefore, many specialized psychological societies called for the necessity of using indicators of practical significance in addition to statistical significance. This article deals with the Cohen index as one of the most important indicators of practical significance, and ways of using and interpreting it in the different cases of t-tests to compare means.

Keywords: statistical significance, practical significance, effect size, t-test, Cohen index.

(*) Professor of Educational Psychology, Faculty of Education, Beni-Suef University.

مقدمة:

يمثل البحث في المجالات النفسية والتربوية أساسًا للتقدم في هذه المجالات، وحجر الزاوية في تطويرها؛ فنتائج تلك البحوث تسهم في اتخاذ القرارات، وتعتمد في ذلك على الإحصاء؛ سواء كان إحصاءً وصفيًّا؛ يتناول عرض النتائج في صورةِ أشكال أو تكرارات أو تلخيص للبيانات في صورة مقاييس النزعة المركزية والتشتت، أو كان إحصاء استدلاليًّا يركز على اتخاذ القرارات الإحصائية التي قد يقع الباحث في أثنائها في خطأ من النوع الثاني، وهو قبول فرض وهو خاطئ ويرمز له بالرمز بيتا β، أو يقع في خطأ من النوع الأول وهو رفض فرض وهو صحيح ويرمز له بالرمز ألفا α. ويطلق عليه مستوى الدلالة الإحصائية[1]، التي قد تكون للطرف الواحد إذا كان الفرض صفريًا أو فرضًا بديلًا غير موجه، أو للطرفين إذا كان الفرض بديلًا موجهًا.

ويوجد مستويان للدلالة الإحصائية في البحوث النفسية والتربوية، وهما: مستوى الدلالة الإحصائية ٠.٠٥ ويقابله مستوى ثقة ٠.٩٥، ومستوى الدلالة الإحصائية ٠.٠١، ويقابله مستوى ثقة ٠.٩٩. ويعتمد الباحث في رفضه أو قبوله للفرض الصفري على مستوى الدلالة الإحصائية، فإذا كانت قيمة مستوى الدلالة الإحصائية ٠.٠٥ أو أقل يرفض الباحث فرضه الصفري، ويقبل الفرض البديل سواء كان موجها أو غير موجه. وإذا كانت قيمة مستوى الدلالة الإحصائية أكبر من ٠.٠٥ فإن الباحث يقبل الفرض الصفري.

وعلى الرغم من أهمية مستوى الدلالة الإحصائية في البحوث النفسية والتربوية، فإنه تعرض لكثيرٍ من أوجه النقد مثل: اعتماد مستوى الدلالة الإحصائية على مدى تحقق الباحث من شروط الاختبار الإحصائي الذي يستخدمه، فلكل أسلوب إحصائي شروط يجب التحقق منها قبل استخدامه.

1- Statistical Significance

فعلى سبيل المثال يستخدم اختبار «ت»[1] للمقارنة بين متوسطين مرتبطين أو غير مرتبطين. وتوجد مجموعة من الشروط التي يجب التأكد منها قبل استخدامه مثل: الاعتدالية وتجانس العينات والفرق بين حجمي العينتين التي يُقارن بينهما وكذلك حجم كل عينة، وإذا ما تم الإخلال بأحد هذه الشروط فقد تتأثر قيمة مستوى الدلالة الإحصائية لهذا الاختبار، كذلك من بين أوجه النقد التي تعرضت لها مستويات الدلالة الإحصائية ارتباطها بزيادة حجم العينة (Snyder & Lawson, 1993; Thompson, 2002). فعلى سبيل المثال عند المقارنة بين مجموعتين مستقلتين باستخدام اختبار "ت"، إذا ما ثُبِّتَ كل من متوسطي المجموعتين والانحراف المعياري للمجموعتين، فإن قيمة "ت" تصبح ذات دلالة إحصائية عند زيادة حجم العينتين، على الرغم من أن متوسطي المجموعتين وانحرافهما المعياري نفسه في الحالتين، ويوضح الجدول التالي ذلك.

جدول (١)

نتائج اختبار "ت" لدلالة الفروق بين المجموعة أ والمجوعة ب

الحالة الأولى المجموعة	العدد	المتوسط	الانحراف المعياري	قيمة "ت"	درجات الحرية	مستوى الدلالة
أ	٣٠	٢٠.٥	٢	٠.٩٦٨٢	٥٨	٠.٣٣٧
ب	٣٠	٢٠	٢			
الحالة الثانية المجموعة	العدد	المتوسط	الانحراف المعياري	قيمة "ت"	درجات الحرية	مستوى الدلالة
أ	٢٠٠	٢٠.٥	٢	٢.٥	٣٩٨	٠.٠١٣
ب	٢٠٠	٢٠	٢			

من جدول (١) السابق يتضح أن قيمة "ت" في الحالة الأولى (عدد أفراد كل مجموعة = ٣٠) غير دالة إحصائيًا فقيمة الدلالة الفعلية[2] كانت

1- T test
2- p value

٠.٣٣٧ وهي أكبر من ٠.٠٠٥. بينما كانت قيمة "ت" في الحالة الثانية (عدد أفراد كل مجموعة = ٢٠٠) دالة إحصائيًا فقيمة الدلالة الفعلية كانت ٠.٠١٣، وهي أقل من ٠.٠٠٥. أي إنه عند زيادة حجم كل عينة من ٣٠ إلى ٢٠٠ انخفضت قيمة الدلالة الفعلية لاختبار "ت"؛ ومن ثَمَّ يقبل الباحث الفرض الصفري في الحالة الأولى ويرفضه في الحالة الثانية، على الرغم من أن قيمة متوسط كل مجموعة في الحالتين لم يتغير، ومن ثَمَّ الفرق بين المجموعتين في الحالة الأولى هو نفسه الفرق بين المجموعتين في الحالة الثانية، وكذلك الانحراف المعياري للمجموعتين لم يتغير في كل حالة.

كذلك دعت الجمعية الأمريكية للإحصاء[1] الباحثين إلى ضرورة الاستخدام الأمثل لمفهوم الدلالة الإحصائية وعدم إساءة استخدامه بوصفه وسيلة وحيدة لاتخاذ القرارات الإحصائية (Wasserstein & Lazar, 2016). في ضوء ما سبق تُوجه إلى ضرورة استخدام الدلالة العملية في البحوث النفسية والتربوية، وعدم الاعتماد فقط على استخدام مستوى الدلالة الإحصائية، حتى لو لم تكن هناك دلالة إحصائية للاختبار الإحصائي المستخدم.

مفهوم الدلالة العملية[2]:

تعبر الدلالة العملية في حالة اختبار "ت" عن حجم الفروق بين المجموعات، بغض النظر عن مدى الثقة في النتائج، الذي يستدل عليه من مستوى الدلالة الإحصائية. أي إن كلًا من الدلالة الإحصائية والدلالة العملية يكمل كل منهما الآخر، ويؤدي استخدامهما معًا إلى إثراء نتائج البحوث النفسية والتربوية (منصور، ١٩٩٧). فالدلالة العملية هي كمية التباين في المتغير التابع، الذي يمكن تفسيره من المتغير المستقل، ويستخدم مفهوم حجم التأثير للتعبير أو الإشارة إلى الدلالة العملية في حالة اختبار "ت" التي تعني

1- American Statistical Association (ASA)
2- Practical Significance.

مقدار أو كمية الفروق بين المتغيرات في البحوث النفسية والتربوية (Mahadevan,2000). وتوجد ثلاث حالات لاختبار "ت" هي اختبار "ت" لعينة واحدة مثل: مقارنة متوسط مجموعة من الأطفال بمتوسط ذكاء معلوم أو متوسط طول معلوم أو متوسط وزن معلوم... إلخ. والحالة الثانية هي اختبار "ت" لعينتين مستقلتين مثل: المقارنة بين متوسط الدافعية للذكور بمتوسط الدافعية للإناث. والحالة الثالثة هي اختبار "ت" للعينات المرتبطة ويتم فيها المقارنة بين متوسط مجموعة واحدة قبل تعريضها لفنيات علاج أو تدريب معين بمتوسط المجموعة نفسها بعد الانتهاء من فترة العلاج أو التدريب. ولكل حالة من هذه الحالات يُستخدم حجم التأثير الذي لا تختلف طريقة حسابه بطريقة كبيرة من حالة إلى أخرى.

حساب حجم التأثير في حالة اختبار "ت" لعينة واحدة[1]:

يُحْسَب حجم التأثير في حالة المقارنة بين متوسط مجموعة m وانحرافها المعياري s بمتوسط معلوم μ بأكثر من صيغة على النحو التالي:

الصيغة الأولى (Curtis & Araki, 2002)**:**

$$d = \frac{m - \mu}{s}$$

حيث d حجم التأثير لكوهين، وm متوسط العينة، وμ متوسط المجتمع أو المتوسط المعلوم المراد مقارنة متوسط العينة به، وs الانحراف المعياري للعينة.

الصيغة الثانية (Rosenthal, 1991)**:**

$$d = \frac{t}{\sqrt{n}}$$

حيث d حجم التأثير لكوهين، وt قيمة اختبار "ت" لعينة واحدة، وn حجم العينة.

1- One Sample T Test

ويمكن حساب حجم التأثير لعينة واحدة Online من خلال الموقع التالي:

https://peterstatistics.com/CrashCourse/2-SingleVar/Scale/Scale-2b-Effect-Size.html

حساب حجم التأثير في حالة اختبار "ت" لعينتين مرتبطتين[1]:

يُحْسَب حجم التأثير في حالة المقارنة بين متوسط مجموعة$m1$ بمتوسط المجموعة نفسها، بعد تعريضها لتدخل معين $m2$، بأكثر من صيغة على النحو التالي:

الصيغة الأولى (Cohen, 1988):

$$d = \frac{m_1 - m_2}{s}$$

حيث d حجم التأثير لكوهين، وm_1 متوسط القياس القبلي، وm_2 متوسط القياس البعدي، وs الانحراف المعياري للفرق بين درجات القبلي ودرجات البعدي.

الصيغة الثانية (Rosenthal, 1991)**:**

$$d = \frac{t}{\sqrt{n}}$$

حيث d حجم التأثير لكوهين، وtقيمة اختبار "ت" لعينتين مرتبطتين، وn حجم العينة.

الصيغة الثالثة (Dunlap et al, 1996):

$$d = t\sqrt{\frac{2(1-r)}{n}}$$

حيث d حجم التأثير لكوهين، وtقيمة اختبار "ت" لعينتين مرتبطتين، وr قيمة معامل الارتباط بين درجات القياسين القبلي والبعدي، وn حجم العينة.

1- Paired Samples T Test

حساب حجم التأثير في حالة اختبار "ت" لعينتين مستقلتين[1]:

يُحْسَب حجم التأثير في حالة المقارنة بين متوسط مجموعة (الضابطة) $_1$ بمتوسط مجموعة أخرى (التجريبية) m_2 بأكثر من صيغة على النحو التالي:

الصيغة الأولى: في حالة تجانس تباين المجموعتين (Cohen, 1988):

$$d = \frac{m_1 - m_2}{s}$$

حيث d حجم التأثير لكوهين، وm_1 متوسط المجموعة الأولى (الضابطة)، وm_2 متوسط المجموعة الثانية (التجريبية)، وs الانحراف المعياري لأي من المجموعتين.

الصيغة الثانية: في حالة عدم تجانس تباين المجموعتين (Cohen, 1988):

$$d = \frac{m_1 - m_2}{s_p}$$

حيث d حجم التأثير لكوهين، وm_1 متوسط المجموعة الأولى (الضابطة)، وm_2 متوسط المجموعة الثانية (التجريبية)، وs_p الانحراف المعياري للمجموعتين ويُحْسَب كالتالي:

$$s_p = \sqrt{\frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}}$$

ويمكن حساب حجم التأثير لعينتين مستقلتين Online من خلال الموقع
https://peterstatistics.com/CrashCourse/3-TwoVarUnpair/BinScale/BinScale-ES-CohenDs.html

أو من الموقعhttps://www.socscistatistics.com/effectsize/default3.aspx

1- Independent Samples T Test

ويُحْسَب حجم التأثير في برنامج SPSS (27) ويعطي ثلاث قيم لحجم التأثير هي: حجم التأثير لكوهين[١]، وحجم التأثير لهيدجز[٢]، وحجم التأثير لجلاس دلتا[٣] (شكل ١).

Independent Samples Effect Sizes

			COHEN'S D	95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
anxi	Cohen's d	3.006	-.434	-.875	.010
	Hedges' correction	3.034	-.430	-.867	.010
	Glass's delta	3.488	-.374	-.818	.075

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

شكل (١) مخرجات برنامج SPSS (27) لحجم التأثير للعينات المستقلة

تتشابه قيم حجم التأثير لكل من كوهين وهيدجز للغاية، عندما يكون حجم العينة أكبر من ٣٠، وعندما تكون أحجام العينات أقل من ٢٠، فيفضل استخدام حجم التأثير لهيدجز؛ ولذلك يطلق عليه أحياناً اسم حجم التأثير المصحح[٤]. وإذا كانت الانحرافات المعيارية مختلفة بشكل كبير بين المجموعات، فيفضل استخدام جلاس دلتا الذي يستخدم فيه الانحراف المعياري للمجموعة الضابطة. ويُحْسَب حجم التأثير المصحح لهيدجز من الصيغة التالية(Cumming, 2012) :

$$\text{Hedges's } g_s = \text{Cohen's } d_s \times \left(1 - \frac{3}{4\,(n_1+n_2)-9}\right)$$

1- Cohen's d
2- Hedges 'g
3- Glass delta
4- Corrected effect size

كما يُحْسَب حجم التأثير لجلاس دلتا الذي يعرف بأنه متوسط الفرق بين المجموعة التجريبية والمجموعة الضابطة مقسومًا على الانحراف المعياري للمجموعة الضابطة من الصيغة التالية:

$$Glass's\ delta = \frac{m_1 - m_2}{s_c}$$

حيث: m_1 متوسط المجموعة الأولى (الضابطة)، m_2 متوسط المجموعة الثانية (التجريبية)، s_c الانحراف المعياري للمجموعة الضابطة.

كما يمكن حساب حجم التأثير بدلالة قيمة "ت" في حالة العينات المرتبطة والعينات غير المرتبطة باستخدام مربعا إيتا[1] η^2 كالتالي (Curtis & Araki, 2002):

$$\eta^2 = \frac{t^2}{t^2 + df}$$

حيث η^2 مربع إيتا، t^2 مربع قيمة اختبار "ت" لعينتين مرتبطتين أو مستقلتين، و df درجات الحرية وتساوي $n_1 + n_2 - 2$ في حالة العينات المستقلة حيث n_1 هو حجم العينة الأولى n_2 حجم العينة الثانية، بينما تساوي $n - 1$ في حالة العينات المرتبطة.

تفسير حجم التأثير:

يُفسر حجم التأثير لكل من كوهين وهيدجز وجلاس دلتا بالطريقة نفسها التي أشار إليها كوهين بالقاعدة التالية **(Cohen, 1988)**:

٠.٢ ≤ d ٠.٥ تأثير صغير[2]

٠.٥ ≤ d ٠.٨ تأثير متوسط[3]

d ≤ ٠.٨ تأثير كبير[4]

1- Eta squared
2- Small
3- Medium
4- Large

وأضاف (Sawilowsky, 2009) إلى التفسير السابق ما يلي:

١.٢٠ □ ≥ d ٢.٠٠ تأثير كبير جدًا[١]

≤ d ٢.٠٠ تأثير ضخم[٢]

وفي حالة استخدام مربع إيتا يكون التفسير على النحو التالي (أبو حطب وصادق، ١٩٩٦):

٠.٠١ ≤ η^2 ٠.٠٦ تأثير صغير

٠.٠٦ ≤ η^2 ٠.١٤ تأثير متوسط

η^2 ≤ ٠.١٤ تأثير كبير

في ضوء ما سبق يتضح أن الدلالة الإحصائية بمفردها غير كافية لتقييم نتائج البحوث النفسية والتربوية، بغض النظر عما إذا كانت هناك دلالة إحصائية، أو عدم وجود دلالة إحصائية؛ فقد تبين أنها تتأثر بحجم العينة، لذا فمن الضروري استخدام الدلالة العملية بجانب حساب الدلالة الإحصائية لتلك البحوث. وفي حالات اختبار "ت" المختلفة، يمكن استخدام مؤشر كوهين المناسب لكل حالة، حتى لو لم تكن هناك دلالة إحصائية لقيمة "ت" المحسوبة.

1- Very large
2- Huge

المراجع:

منصور، رشدي فام (١٩٩٧). حجم التأثير الوجه المكمل للدلالة الإحصائية. **المجلة المصرية للدراسات النفسية، ٧(١٦)**، ٥٧-٧٥.

أبو حطب، فؤاد، وصادق، آمال. (١٩٩٦). **مناهج البحث وطرق التحليل الإحصائي في العلوم النفسية والتربوية والاجتماعية**. القاهرة: الأنجلو المصرية.

Cohen, J. (1988). ***Statistical power analysis for the behavioral sciences***. New York, NY: Routledge Academic.

Cumming, G. (2012). ***Understanding the new statistics: Effect sizes, Confidence Intervals, and Meta-Analysis***. New York, NY: Routledge.

Curtis, D., & Araki, C. (2002). ***Effect size statistics: An analysis of statistics textbooks used in psychology and education***. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Dunlap, W., Cortina, J., Vaslow, J. & Burke, M. (1996). Meta-analysis of experiments with matched groups or repeated measured designs. ***Psychological Methods, 1***, 170–177.

Mahadevan, L.(2000).The effect size statistic: Overview of various choices. ***(Eric Document Reproduction Service No. ED 438308)***.

Rosenthal, R. (1991). ***Meta-analytic procedures for social research***. New bury Park, CA: SAGE Publications, Incorporated.

Sawilowsky, S. (2009). "New effect size rules of thumb". ***Journal of Modern Applied Statistical Methods, 8***(2), 467–474.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. ***Journal of Experimental Education,61***, 334- 349.

Thompson, B. (2002) "Statistical", "Practical", and "Clinical": How Many Kinds of Significance Do Counselors Need to Consider? ***Journal of Counseling and Development, 80***, 64-71.

Wasserstein, R. & Lazar, N. (2016). The ASA's statement on p-values: context, process, and purpose. ***The American Statistician, 70***(2), 129-133.