الدلالة الإحصائية و الدلالة العملية في البحوث

إعـــداد

أ.د / عزت عبد الحميد محمد حسن

أستاذ علم النفس التربوي

مدير مركز التقويم والقياس وبنوك الأسئلة

كلية التربية جامعة الزقازيق

مقدمة

تشير **الدلالـة الإحصائية** Statistical Significance إلى إذا ما كانت النتيجة ناشئة عن الصدفة أو المتغيّرية Variability في العيّنة، بينما تشير **الدلالـة العمليـة** Practical Significance إلى إذا ما كانت النتيجة مفيدة في العالم الحقيقي، حيث من خلال الدلالة العملية يمكن عمل تفسيرات ذات مغزى لنتائج البحث وتطبيقها في العالم الحقيقي. فقيمة p (p-value) تساعد في إصدار الحكم الذي يتعلق بالدلالة الإحصائية للنتائج، إلاَّ أن قيمة p لا تكفي بمفردها في مساعدة القرَّاء لفهم الدلالة العملية للبحث، وذلك لعدة أسباب منها:

١) أن قيم p تتأثر بالعديد من العوامل ولذا لا يمكن استخدامها لاتخاذ قرار يتعلق بمقدار تأثير المعالجة.

٢) أن قيم p لا تدرس مباشرة القضية الحرجة لإعادة النتيجة؛ فقيم p الصغيرة لا تدلّ على ثقة كبيرة في استخلاص أن نتائج العينة تتكرر، ولهذا يوصي بأن يُتْبَع اختبار الدلالة بحساب حجم التأثير الذي يمكن أن يستخدم لإصدار الحكم حول الدلالة العملية للنتائج. **فالدلالـة الإحصـائية وحجـم التـأثير وجهـان لعملـة واحـدة**، يكمل كل منهما الآخر ولكنّ لا يحل أحدهما مكان الآخر، ولذا يجب على الباحثين أن يضعوا في اعتبارهم كلا الجانبين (Sun, 2008, 6).

٣) أن **الدلالــة الإحصـائية** قد تشير إلى وجود فرق دال إحصائيًا بين مجموعتي المقارنة (التجريبية والضابطة)، أو بين القياسين القبلي والبعدي، ولكن هذا الفرق الدال لا يشير إلى

تحسن حقيقي لدى المجموعة التجريبية في الظاهرة موضع القياس.

ويشير (رشدي منصور، ١٩٩٧، ٥٩) إلى أن مفهوم الدلالة الإحصائية للنتائج يركز على مدى الثقة التي نوليها لنتائج الفروق أو العلاقات بصرف النظر عن حجم هذه الفروق أو تلك العلاقات، فإذا كان الفرق دالاً إحصائيًا عند مستوى (٠.٠٥) بين مجموعتين، فذلك يشير إلى أن الفرق المجموعتين حقيقي وأن مجتمع المجموعة الأولي يختلف عن مجتمع المجموعة الثانية، وأننا نثق في هذا الحكم بنسبة ٩٥%. وهنا نكون في حاجة إلى ما يسمي **بالدلالة العملية** للتعبير عن مقدار هذا الفرق من حيث كونه صغيرًا أو متوسطًا أو كبيرًا، لذا فالدلالة العملية هي الجانب المكمل لمفهوم الدلالة الإحصائية، حيث تشير الدلالة العملية إلى الفرق أو حجم الارتباط بصرف النظر عن مدى الثقة التي نضعها في النتائج، أي أن كلاً من الدلالة الإحصائية والدلالة العملية يكمل منهما الآخر ويعوض النقص الكامن فيه، فهما كوجهي عملة واحدة يؤدي استخدامهما معًا إلى إثراء نتائج البحوث النفسية والتربوية والاجتماعية وكذلك البحوث في مجالات المعرفة الأخرى.

وباختصار فمستويات الدلالة الإحصائية للاختبارات الإحصائية تشير إلى مقدار الثقة في النتائج التي يتم التوصل إليها بغض النظر عن حجم الفروق أو قوة الارتباطات بهذه النتائج وأنها تتأثر بطريقة مباشرة بحجم العينة، أما مقاييس حجم التأثير (أو الدلالة العملية) فلا تتأثر بحجم العينة وتشير إلى حجم الفروق أو قوة الارتباطات بين المتغيرات بغض النظر عن مقدار الثقة.

إن وجود الدلالة الإحصائية لا يعني بالضرورة وجود دلالة عملية، فقد تكون الدلالة الإحصائية نتيجة الحجم الكبير للعينة، فالدلالة الإحصائية تستغل المؤشرات الإحصائية لإثبات وجود فرق دال إحصائيًا بين المجموعات موضع المقارنة، حيث تشير الدلالة الإحصائية أو قيمة الاحتمالية (p-value) إلى احتمالية ملاحظة قيمة الإحصاءة Statistic التي تم اختبارها متطرفة عما لوحظت عليه، على الأقل عندما يكون الفرض الصفري صحيحًا.

أما **الدلالة العملية** Practical Significance فإنها تحاول إثبات وجود تحسن فعلي لدى مجموعة الدراسة (التجريبية مثلاً)، أي أنها تهتم بمدى كون نتائج البحث مفيدة

لعالم الواقع الحقيقي، ولذلك القياسات الإحصائية للدلالة العملية تقرِّر إلى أي مدى يمكن تفسير التغيّر الذي يحدث في المتغير التابع بواسطة المتغير أو المتغيرات المستقلة.

والمثال على ذلك أنه إذا تم تجريب أو دراسة فعالية برنامج معين لخفض الاكتئاب لدى مجموعة من مرضى الاكتئاب (مجموعة تجريبية) في مقابل مجموعة ضابطة لم تتعرض لهذا البرنامج، وبعد الانتهاء من البرنامج وجدنا أن هناك فرقاً دالاً إحصائياً (عند مستوى أقل من ٠.٠٥) بين المجموعتين في درجة الاكتئاب، أي أن البرنامج أدى إلى خفض درجة الاكتئاب بدلالة إحصائية لدى المجموعة التجريبية بالمقارنة بالمجموعة الضابطة، ولكننا في المقابل وجدنا أن مرضى الاكتئاب بالمجموعة التجريبية لم تنخفض لديهم درجة الاكتئاب إلى المعدل الطبيعي أو الدرجة التي عندها يتم تصنيفهم إلى غير مكتئبين، وما زالت درجة الاكتئاب مرتفعة لديهم، بالرغم من انخفاضها انخفاضًا دالاً إحصائيًا عن نظيرتها لدى المجموعة الضابطة، في هذه الحالة الدلالة الإحصائية لا قيمة لها لأنه لا وجود للدلالة العملية.

ومقاييس الدلالة العملية كثيرة وهي ما تُعرف بمقاييس حجم التأثير، حيث تساعد الدلالة العملية (أو حجم التأثير) صنّاع السياسة والمربين على اتخاذ قرارات تتعلق بما إذا كان الفرق الدال إحصائيًا بين البرامج يمكن أن يترجم إلى فرق كاف لتبرير تبني برنامج معين دون الآخر http://www.ecs.org/html/educationIssues/Research/primer/understandingtutorial.asp#sub5).

ماذا تفعل اختبارات الدلالة الإحصائية؟ وما الذي لا تفعله تلك الاختبارات؟

تستخدم اختبارات الدلالة الإحصائية للاستدلال على خصائص المجتمع من خلال العينة المأخوذة منه، وتسمح تلك الاختبارات للفرد بحساب احتمالية الحصول على نتائج العينة بافتراض أن الفرض الصفري صحيحًا (أي افتراض أن المجتمع يتصف بما هو مفترض في الفرض الصفري) (King, 2002, 1).

فاختبارات الدلالة الإحصائية تقدر قيمة احتمالية ($P_{Calculated}$) انحراف نتائج العيّنة عن المحدّد بالفرض الصفري للمجتمع، فهي لا تقيّم الاحتمال بأن نتائج العيّنة تصف المجتمع، وإنما تفترض هذه الاختبارات بأنّ الصفرية تصف المجتمع بالضبط ومن ثم تختبر احتمال

العيّنة، وتؤكد الدراسات الإمبريقية بأن العديد من مستخدمي اختبارات الدلالة الإحصائية في الحقيقة لا يفهمون ما تفعله هذه الاختبارات بالفعل، فبعض مستخدميها يقلصون دورها إلى مجرّد قول (عند مستوى سطحي)، فهذه الاختبارات تقيّم ما إذا كانت النتائج ناشئة عن الصدفة أم لا (Thompson, 2002, 65).

وأكد كيرك (1996) Kirk على أن اختبارات الدلالة الإحصائية تقيّم فقط "علاقات ترتيبية" ordinal relationships (مثل: إذا ما كان متوسطي مجموعتين مختلفان، أو أن أحدهما أكبر من الآخر)، وما نريد أن نعرفه هو حجم الفرق بين (أ)، (ب) والخطأ المرتبط بتقديرنا (وهو ما يسمى بحجم التأثير كمقياس للدلالة العملية)، وتعتبر معرفتنا بأن (أ) أكبر من (ب) ليست كافية. وحجم التأثير مهم جدًا لأن الاختبارات الإحصائية تتأثّر بشدّة بحجم العيّنة، الأمر الذي أدى إلى وجو ب استخدامه جنبًا إلى جنب مع الاختبارات الإحصائية التقليدية. وقد طرح Thompson (1993) المثال التالي ليوضح الفرق بين الدلالة الإحصائية والدلالة العملية: إذا قرر باحث ما أن يقارن بين متوسطي ذكاء مجموعتين كبيرتين بياناتهما كالتالي: (م١ = ١٠٠.١٥ ، ع١ = ١٥، ن١ = ١٢٠٠٠)، (م٢ = ٩٩.٨٥ ، ع٢ = ١٥، ن٢ = ١٨٨٠٠٠) فسيجد أن الفرق بين متوسطي المجموعتين دال إحصائيًا، ويقرر أن المجموعة الأولى أعلى ذكاءً من المجموعة الثانية، بالرغم من أن الفرق الذي يقل عن درجة ذكاء واحدة ليس مهم عمليا أو تربويًا. إنّ النتيجة بديهية جدًا بالنسبة للخطأ المعياري للقياس بمعظم اختبارات الذكاء (الخطأ المعياري للقياس بمعظم اختبارات الذكاء = ٤.٢ أو أكثر)، ومن هنا فنحن عادة نتوقّع أن نحصل على درجات أو متوسطات أضعاف الخطأ المعياري للقياس إذا أردنا أن نجزم بأن الفروق ليست نتيجة أخطاء القياس (Thompson, 2002, 65-66).

ما الذي لا تفعله اختبارات الدلالة الإحصائية؟

اختبارات الدلالة الإحصائية لا تخبرنا بما نريد أن نعرفه عن قيم المجتمع وإمكانية إعادة النتيجة للعينات التي يتم سحبها من نفس المجتمع في المستقبل. فقد استنتج كوهين (١٩٩٤) **بأنّ اختبارات الدلالة الإحصائية "لا تخبرنا بما نريد أن نعرفه، ونحن في حاجة ماسة لمعرفة ما نريد أن نعرف، وعلى الرغم من ذلك لا نيأس ونعتقد بأنّها تفعل ذلك"**. كما أنها لا تُقيِّم ما إذا كانت النتائج مهمة أم لا، فقيم الاحتمالية لا تستخدم لاستنتاج قيمة أو أهمية

نتائج البحث، فبعض الأحداث المحتملة مهمة جدًا حتى ولو كانت غير عادية أو شاذة، وبالرغم من ذلك قد يساوي بعض الباحثين بين احتمالية النتيجة وأهمية النتيجة (Thompson, 2002, 65).

وباختصار فاختبارات الدلالة الإحصائية لا تخبر الباحثين بأهمية أو قيمة نتائجهم، وإنما تخبرهم فقط باحتمالية النتيجة، وما إذا كانت نتائج عيناتهم ناشئة عن الصدفة أو خطأ المعاينة. فنحن نفترض أولاً أن الفرض الصفري صحيح في المجتمع، وأن أي عينة يتم سحبها من هذا المجتمع سوف تعطي نتائج مماثلة تقريبًا للفرض الصفري، ثم نحدد احتمال الحصول على نتائج العينة لحجم العينة المُعطي، وكلما ابتعدت نتائج العينة عن المحدد بالفرض الصفري انخفض احتمال حصولنا على النتائج عن طريق الصدفة (أي انخفضت قيم p المحسوبة). ولسوء الحظ أن اختبارات الدلالة الإحصائية لا تخبرنا بما إذا كانت العينة تصف المجتمع (وهو ما نريد أن نعرفه) وإنما تخبرنا فقط باحتمالية نتائج العينة إذا افترضنا أن الفرض الصفري صحيحًا. فدلالة أو عدم دلالة النتيجة لا يشير إلى أهميتها عمليًا، فقد تكون النتيجة دالة إحصائية نتيجة حجم العينة الكبير إلاَّ أنها غير مهمة عمليًا (مثل الفرق الضعيف جدًا الدال إحصائيًا بين متوسطي عينتين حجمهما كبير جدًا)، وعلى العكس من ذلك قد تكون النتيجة غير دالة إحصائية نتيجة حجم العينة الصغير إلاَّ أنها مهمة عمليًا (مثل الفرق الكبير غير الدال إحصائيًا بين متوسطي عينتين حجمهما صغير)، ومن هنا فاختبارات الدلالة الإحصائية غير كافية لتحديد أهمية نتائج البحث، ولذا نحتاج إلى أحجام التأثير لتقييم الدلالة العملية للنتائج، حيث تساعد مؤشرات حجم التأثير القرَّاء على التمييز بين النتائج الدالة عمليًا practically significant وتلك النتائج الدالة إحصائيًا (Armstrong & Henson, 2004, 13-15).

وقد سبق الإشارة إلى أن **الدلالة الإحصائية** قد تشير إلى وجود فرق دال إحصائيًا بين مجموعتي المقارنة (التجريبية والضابطة)، أو بين القياسين القبلي والبعدي، ولكن هذا **الفرق الدال لا يشير إلى تحسن حقيقي** لدى المجموعة التجريبية في الظاهرة موضع القياس. والأمثلة على ذلك كثيرة منها ما يلي:

- **في تجربة لفحص فعالية استخدام طريقة تدريس جديدة لتحسين مستوى التحصيل الدراسي في مادة الرياضيات** لدى ٨٠ تلميذًا من التلاميذ المتأخرين دراسيًا في مادة

الرياضيات بالصف الثالث الإعدادي، تم تقسيمهم بالتساوي إلى مجموعتين، مجموعة ضابطة (درست محتوى الرياضيات بالطريقة العادية)، ومجموعة تجريبية (درست محتوى الرياضيات بطريقة التدريس الجديدة)، وكان المجموعتان متكافئتان في القياس القبلي للتحصيل الدراسي في الرياضيات (حيث متوسط التجريبية = ٧.٣٥، والضابطة = ٧.٦٣ وذلك من النهاية العظمي لاختبار الرياضيات وهو ٦٠ درجة)، وبعد استخدام طريقة التدريس الجديدة تم حساب الفرق بين متوسطي درجات المجموعتين، فوُجِد أن الفرق بين المتوسطين دال إحصائيًا لصالح متوسط درجات المجموعة التجريبية عند مستوى ٠.٠٣ أي ($P < 0.05$) كما بالجدول التالي:

التطبيق	المجموعة	العدد	المتوسط	الانحراف المعياري	قيمة (ت)	مستوى الدلالة	مربع إيتا
القبلي	التجريبية	٤٠	٧.٣٥	٤.٧٢	٠.٢٨	٠.٧٨	
	الضابطة	٤٠	٧.٦٣	٤.١٩			
البعدي	التجريبية	٤٠	١٠.٤٨	٢.٧٨	٢.١٦	٠.٠٣	٠.٠٥٦
	الضابطة	٤٠	٨.٧٥	٤.٢١			

ولكن رغم أن **الدلالة الإحصائية** تشير إلى وجود تحسن دال إحصائيًا في مستوى التحصيل الدراسي في الرياضيات لدى تلاميذ المجموعة التجريبية بالمقارنة بتلاميذ المجموعة الضابطة إلاَّ أن متوسط التحصيل الدراسي لدى المجموعة التجريبية الذي يساوي ١٠.٤٨ درجة في القياس البعدي يشير إلى أن تلاميذ المجموعة التجريبية مازالوا يعانون من التأخر الدراسي في مادة الرياضيات حيث لم يصل متوسط تحصيلهم بعد استخدام طريقة التدريس الجديدة إلى ٣٠ درجة وهي منتصف النهاية العظمي لاختبار الرياضيات أو المستوى المقبول في التحصيل الدراسي في الرياضيات، كما تشير قيمة مربع إيتا Eta-Square وهي أحد مقاييس الدلالة العملية التي تساوي (٠.٠٤٩) إلى أن طريقة التدريس الجديدة لها تأثير ضعيف في تنمية التحصيل الدراسي في الرياضيات لدى تلاميذ المجموعة التجريبية لأن قيمة مربع إيتا **لم تصل إلى القيمة (٠.٠٥٩) التي تشير وجود تأثير متوسط** لطريقة التدريس كما أشار إليها الجدول الإحصائي لحجم التأثير (عزت حسن، ٢٠١١، ٢٨٤).

- **وفي تجربة أخرى لتجربة عقار معين لعلاج ضغط الدم** لدى عينة من مرضى ضغط الدم قوامها ٣٠ مريضًا، إذا كان متوسط ضغط الدم قبل تجريب العلاج = ١٩٠ وبعد العلاج بالعقار فترة زمنية معينة تم قياس ضغط الدم لدى نفس العينة فكان = ١٧٠ وباستخدام اختبار (ت) لحساب دلالة الفرق بين درجات المجموعة التجريبية قبل وبعد العلاج كانت البيانات كما بالجدول التالي:

المجموعة التجريبية	العدد	المتوسط	الانحراف المعياري	قيمة (ت)	مستوى الدلالة
قبل العلاج	٣٠	١٩٠	٦.٤٨	٦٩.٥٢	٠.٠٠٠٠٠٠١
بعد العلاج	٣٠	١٧٠	٧.١٢		

يتضح من بيانات الجدول السابق أن قيمة (ت) دالة إحصائيًا عند مستوى منخفض جدًا من الشك وهو ($P < 0.000001$), أي أن الدلالة الإحصائية تشير إلى وجود انخفاض دال إحصائيًا في مستوى ضغط الدم لدى عينة المرضى بعد تجربة العلاج، ولكن في الواقع ما زال ضغط الدم مرتفعًا لدى عينة المرضى حيث متوسط ضغط الدم بعد العلاج = ١٧٠ ويعني أن المرضي لا يزال يعانون من ضغط الدم ولم ينخفض ضغط الدم لديهم إلى الحد الطبيعي لضغط الدم.

ويلخص (King, 2002, 2) ما لا تفعله اختبارات الدلالة الإحصائية في النقاط التالية:

١. **احتمالية أن يكون الفرض الصفري صحيحًا:** رغم أن هذا مرغوب جدًا، إلا أنه لا يمكن حساب مثل هذا الاحتمال.

٢. **الاحتمال بأنّ هذه النتائج تم الحصول عليها بالصّدفة:** فهذا الاحتمال مجهول سواء كانت العيّنة تمثل المجتمع تمثيلاً تامًا، أو كانت النتائج شاذّة.

٣. **احتمال وجود هذه النتائج بالدراسات المستقبلية.**

٤. **احتمال وجود تأثير حقيقي في المجتمع:** فاختبارات الدلالة الإحصائية لا تقيس حجم التأثير بطريقة مباشرة (مثل: مقدار الفروق المشاهدة)، وبدلاً من ذلك تكون قيمة p

موجودة وتدمج حجم التأثير مع حجم العينة، ولهذا تظهر النتائج على أنها دالة إحصائيًا نتيجة حجم التأثير الكبير أو حجم العينة الكبير أو الاثنان معًا. ولذلك، قد تكون النتائج دالة إحصائيًا بسبب حجم العيّنة الكبير، إلاَّ أنها ليس لها دلالة عملية بسبب حجم التأثير الصغير، والعكس صحيح.

وقد أوضح جونسون (Johnson, 1999, 768) الدلالة الإحصائية في مقابل الدلالة العملية الخاصة بردود أفعال الباحث المتعلقة باختبار الدلالة الإحصائية، وكيفية تفسير حجم العينة المرتبط بنتائج اختبار الدلالة الإحصائية، وذلك كما يلي:

الدلالة أو الأهمية العملية للفرق المشاهد		الدلالة الإحصائية	
		غير دالة	دالة
ردود أفعال الباحث المتعلقة باختبار الدلالة الإحصائية	غير مهمة	سعيد Happy	منزعج Annoyed
	مهمة	حزين جدًا Very Sad	مبتهج Elated
تفسير حجم العينة المرتبط بنتائج اختبار الدلالة الإحصائية	غير مهمة	حجم العينة مضبوط n Okay	حجم العينة كبير جدًا n too big
	مهمة	حجم العينة صغير جدًا n too small	حجم العينة مضبوط n Okay

أي أن الجدول السابق يوضح ما يلي:

- أن الباحث يكون سعيدًا عندما تشير النتائج إلى عدم وجود دلالة إحصائية لفرق غير مهم عمليًا، وكذلك يشعر بالابتهاج عندما تشير النتائج إلى وجود دلالة إحصائية لفرق مهم عمليًا. إلاَّ أنه ينزعج عندما تشير النتائج إلى وجود دلالة إحصائية لفرق غير مهم عمليًا، ويحزن جدًا عندما تشير النتائج إلى عدم وجود دلالة إحصائية لفرق مهم عمليًا.
- أن الباحث يفسر حجم العينة على أنه مضبوط ومعقول عندما تشير النتائج إلى عدم وجود دلالة إحصائية لفرق غير مهم عمليًا، وكذلك عندما تشير النتائج إلى وجود دلالة إحصائية لفرق مهم عمليًا. إلاَّ أنه يفسر حجم العينة على أنه كبير جدًا عندما تشير النتائج إلى وجود

دلالة إحصائية لفرق غير مهم عمليًا، ويفسر حجم العينة على أنه صغير جدًا عندما تشير النتائج إلى عدم وجود دلالة إحصائية لفرق مهم عمليًا.

ويلخص (Svyantek & Ekeberg, 1995, 363) العلاقة بين الدلالة الإحصائية والدلالة العملية في أربع الاحتمالات على النحو التالي:

		الدلالة الإحصائية	
		دالة	غير دالة
الدلالة العملية	دالة (مهمة)	**إيجابيات صائبة** true positives يوجد توازن بين الاستدلال الإحصائي واتخاذ القرار، وهنا يكون للنتائج معني وأهمية بالنسبة لكل من المُقيّم وصانع القرار	**سلبيات خاطئة** false negatives تكون الفروق أو العلاقات غير دالة إحصائيًا ولكنها دالة عمليًا بسبب حجم العينة الصغير
الدلالة العملية	غير دالة (غير مهمة)	**إيجابيات خاطئة** False positives تكون الفروق أو العلاقات دالة إحصائيًا ولكنها غير دالة عمليًا بسبب حجم العينة الكبير	**سلبيات صائبة** true negatives عندما تكون الفروق ليس لها معنى أو دلالة من وجهة نظر كل من المُقيّم وصانع القرار

الدلالة الإكلينيكية Clinical Significance

يشير (Durlak, 2009, 8) إلى أنه يتم استخدام المصطلحين: الدلالة العملية Practical significance والدلالة الإكلينيكية Clinical significance بشكل متبادل للإشارة إلى درجة التغيير ذات المعني في حياة المشاركين.

إلاَّ أن موسوعة Wikipedia تشير إلى أن استخدام المصطلحين: الدلالة العملية والدلالة الإكلينيكية بشكل متبادل يعتبر خطأ، فالدلالة العملية تعطي معلومات محددة عن أهمية النتيجة، وحجم التأثير أحد أنواع أو مؤشرات الدلالة العملية الذي يحدد المدى الذي تبتعد فيه العينة عن التوقعات، وتجيب الدلالة العملية على أسئلة مثل: ما فعالية المعالجة أو ما مقدار التغيير الذي تسببه المعالجة؟ فحجم التأثير يمكن أن يزوّدنا بمعلومات مهمة حول نتائج الدراسة، إلاَّ أن أحجام التأثير لها مصادرها الخاصة للتحيّز، وأنها عرضة للتغيير المستند على تغيّر المتغير التابع في المجتمع، وتميل أحجام التأثير إلى التركيز على تأثيرات المجموعة وليس

التغييرات الفردية. وعلى النقيض من ذلك، تعطي الدلالة الإكلينيكية معلومات تتعلق بما إذا كانت المعالجة كافية لتغيير علامة المريض التشخيصية وتجيب الدلالة الإكلينيكية على أسئلة مثل: هل المعالجة فعّالة بما فيه الكفاية لجعل المريض طبيعيًا؟ فمثلا: قد تؤدي المعالجة إلى تغييرات دالة في أعراض الاكتئاب، فيمكن أن يكون التغيير انخفاض كبير في أعراض الاكتئاب (أي: الدلالة العملية أو حجم التأثير)، وأن ٤٠% من المرضى لا تنطبق عليهم المعايير التشخيصية للاكتئاب (أي: الدلالة الإكلينيكية)، وأنه من المحتمل جدًا أن تؤدي المعالجة إلى فرق دال وحجم تأثير متوسط أو كبير، إلا أن المريض لا يتحرك من المدى المرضي إلى المدى الطبيعي (http://en.wikipedia. org/wiki/Clinical_significance).

ففي العيادات يجب أن يتّخذ الممارسون قرارات مطلقة في أغلب الأحيان، فمثلاً: يجب أن يقرّر الطبيب النفسي إذا ما كان المريض معين يحتاج أو لا يحتاج إلى دواء للاكتئاب. أو أن يقرر المرشد بوجوب دخول أو عدم دخول مريض مكتئب المستشفي. هذه القرارات قد تواجّه بالمعايير التشخيصية أو درجات القطع، فمثلاً: قد يضع طبيب معين قاعدة بأن كولوستيرول الدمّ الأكثر من٢٠٠ ملليجرام لكلّ ديسيلتر deciliter يتطلّب دواءً. ومؤشرات حجم التأثير للدلالة العملية قد تكون ذات علاقة جزئية فقط بتطبيقات دليل البحث لهذه الأنواع من المواقف الإكلينيكية، فمتوسطات المجموعة لا تشير إلى نسبة المشاركين الذين تحسّنوا نتيجة المعالجة، ويدرس معيار الدلالة الإكلينيكية السؤال "هل الأفراد الذين تم معالجتهم يتعذر تميزهم عن الأفراد الطبيعيين فيما يتعلق بالشكاوى الأساسية بعد المعالجة؟ فمثلاً: قد تشير نتائج دراستين إلى انخفاض درجات الاكتئاب بمقدار ١٠ درجات، إلا أنه في الدراسة الأولى قد يحتاج معظم المرضى إلى علاج بالمستشفي بعد التدخّل، بينما في الدراسة الثانية التي بنفس حجم التأثير قد لا يحتاج كل المرضى إلى علاج بالمستشفي. وهنا لا يمكن أن نقول بأن التدخّل الأول غير مهم، فمن المنظور الإكلينيكي الدراستان بالمؤشرات المماثلة للتأثير "العملي" ما زالتا تختلفان بشكل واضح فيما يتعلق بنتائج التدخّل للمرضى في مصحّة نفسية. وتُعرّف الدلالة الإكلينيكية على أنها "القيمة العملية أو التطبيقية أو أهمية تأثير التدخّل الذي يؤدي إلى فرق حقيقي في حياة العملاء اليومية أو الآخرين الذين يتفاعل معهم العميل". وأن بعض التدخلات التي ليس لها تأثيرات دالة قد تكون دالة إكلينيكيًا، فمثلاً: التدخل لعلاج الاكتئاب قد لا يكون له

تأثير دال إحصائيًا على مرضى الاكتئاب عند مقارنتهم بالعاديين إلاَّ أن هذا التدخل قد يساعد هؤلاء المرضي على مواجهة أعراضهم ويحسنوا من نوعية حياتهم (Thompson, 2002, 66).

وتقاس الدلالة الإكلينيكية بخمس طرق شائعة، حيث تُنسب كل طريقة من هذه الطرق إلى مؤلفيها، وهذه الطرق هي:

Jacobson-Truax method, Gulliksen-Lord-Novick method, Edwards-) Nunnally method, Hageman-Arrindell method, hierarchical linear modeling)، فمثلاً في الطريقة الأولى (طريقة Jacobson-Truax) يتم حساب مؤشر ثبات التغيير Reliability Change Index (RCI) الذي يساوي الفرق بين درجات الاختبارين القبلي والبعدي لدى المريض مقسومًا على الخطأ المعياري للفرق. وبعد ذلك تُحدد درجات القطع التي تحدد مكان المريض أو المشارك في إحدى المجموعات الأربع: (تعافي، تحسّن، بدون تغيير، أو تدهور) (http://en.wikipedia. org/wiki/Clinical_significance).

والسؤال: كم نوع من أنواع الدلالة نحتاج إليها؟

الدلالة الإحصائية لا تكفي بمفردها كمعيار لتقييم أهمية نتائج البحث، فالدراسات غير الدالة إحصائيًا قد تظهر تأثيرات مهمة وقابلة للإعادة، فمثلا: قد تسفر نتائج دراسات علاج ضد ورم جديد عن وجود تأثيرات مهمة إلا أن قيمة الاحتمالية المحسوبة = ٠.٠٦ (= $P_{calculated}$ 0.06)، وهنا طالما أن لهذا العلاج تأثيرًا مهمًا من الناحية الإكلينيكية نأخذ به بغض النظر عن الدلالة الإحصائية، وخاصة أنها قريبة من مستوى الدلالة الإحصائية (٠.٠٥)، ومن هنا فبدلاً أن يكون سؤال المعالج: هل العلاج (أ) أفضل من العلاج (ب)؟ وهو ما تهتم به اختبارات الدلالة الإحصائية، يكون سؤاله: إلى أي مدى العلاج (أ) أفضل؟ أي يأخذ مدى حجم التأثير في الحسبان، وهو ما تهتم به الدلالة العملية (Thompson, 2002, 67).

الخلاصة

في ضوء ما سبق وفي ضوء عدم كفاية الدلالة الإحصائية بمفردها كمعيار لتقييم أهمية نتائج البحث، نستنتج أن البحوث النفسية والتربوية والاجتماعية تحتاج إلى حساب كل من

الدلالة الإحصائية والدلالة العملية، وإذا كان البحث يتعلق بالأمراض فمن الضروري حساب الدلالة الإكلينيكية بالإضافة إلى الدلالة الإحصائية.

المراجع

١) رشدي فام منصور (١٩٩٧). حجم التأثير الوجه المكمل للدلالة الإحصائية. *المجلة المصرية للدراسات النفسية، ١* (١٦)، ٥٧-٧٥.

٢) عزت عبد الحميد محمد (٢٠١١). *الإحصاء النفسي والتربوي: تطبيقات باستخدام برنامج SPSS18* . القاهرة: دار الفكر العربي

3) Durlak, J. A. (2009). How to Select, Calculate, and Interpret Effect Sizes. Journal of Pediatric Psychology,1–12. Advance Access published February 16, 2009, doi:10.1093/jpepsy/jsp004.

4) Armstrong, S. A. & Henson, R. K. (2004). Statistical and practical significance in the IJPT: A research review from 1993-2003. *International Journal of Play Therapy, 13*(2), 9-30.

5) King, J. E. (2002). Statistical Significance Versus Practical Significance: Implications for Analyzing CME Data. *Poster presented at the annual meeting of the Alliance for Continuing Medical Education, Orlando, FL, February, 2002.*

6) Johnson, D. H. (1999). The insignificance of statistical significance testing. *Journal of Wildlife Management.* 63(3), 763-772.

7) Svyantek, D. J. & Ekeberg, S. E. (1995). Strong hypothesis testing in organization: Alternative appoarches to evaluation of practical significance. *The International Journal of Organizational Analysis*. 3(4), 361–374.

8) 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.

9) http://en.wikipedia.org/wiki/Clinical_significance

10) http://www.ecs.org/html/educationIssues/Research/primer/ understandingtutorial .asp#sub5.