

الإجراءات الإحصائية المرتبطة بالمقاسات والاختبارات الفسيولوجية

المصدر:

الهزاع، هزاع محمد. فسيولوجيا الجهد البدني: الأسس النظرية والإجراءات العملية للمقاسات الفسيولوجية. الفصل الرابع. كتاب مقدم للنشر.

اختبارات الصدق والثبات

عرفنا في الفصل الأول من هذا الكتاب أن أهم ما يميز الاختبارات والقياسات الفسيولوجية الجيدة هو مقدار كل من صدقها (Validity) وثباتها (Reliability). في الفقرات التالية نستعرض باختصار تعريف كل من الصدق والثبات، وما هي أهم الأنواع الشائعة للصدق، وكيفية حساب كل منها.

الصدق (Validity):

يعد الصدق أهم عنصر من العناصر التي ينبغي أخذها بالحسبان عند انتقاء الاختبار المستخدم وتقويمه. ويعني الصدق مدى الدقة التي يتصف بها الاختبار عند قياس الصفة المراد قياسها، أو بمعنى آخر، مدى القرب من قياس الصفة التي يرمي إلى قياسها. أي أن الاختبار الصادق هو الذي يقيس الصفة بشكل تام أو قريب جداً من ذلك. وهناك أنواع متعددة من الصدق، تبعاً لطبيعة التطبيق المستخدم، لكننا سنركز هنا على أنواع الصدق الأكثر استخداماً في الاختبارات والقياسات الفسيولوجية (٢٤، ٢٧، ٢٨، ٣٠).

١ - الصدق الظاهري (Face validity):

ويسمى أيضاً الصدق المنطقي (Logical validity)، ويعني مدى مناسبة الاختبار المقصود لقياس العناصر التي يهدف إلى قياسها. ويعتمد تحديد ذلك الصدق على المنطق والمعرفة بالجوانب المحيطة بالصفة المراد قياسها، فمثلاً: عند قياس المرونة، لا بد للاختبار المستخدم أن يقيس فعلاً المدى الحركي للمفصل، وألا فإنه لا يصبح صادقاً في قياس المرونة من الناحية الظاهرية (أو المنطقية). كما أنه من غير المعقول أن نستخدم اختباراً مدته بضع ثوان لقياس صفة التحمل الدوري التنفسي، لأن سيكون حينئذ اختباراً غير صادق.

٢ - صدق المحتوى (Content validity):

يرتكز هذا النوع من الصدق على حكم الخبراء والمختصين في المجال المعني، من أجل تحديد فقرات الاختبار (أو المقياس) المراد استخدامه، ومدى تمثيلها لقياس الصفة المراد قياسها. وغالباً ما يستخدم هذا النوع من الصدق في بداية تطوير الاختبار أو المقياس، فمثلاً لو أردنا تطوير اختبار (استبيان) لقياس مستوى النشاط البدني لدى الأطفال، لكان علينا أن نأخذ رأي المختصين في طبيعة الأسئلة المستخدمة في الاستبيان ومدى صلاحيتها لقياس النشاط البدني عموماً، وعن ملاءمتها لفئة الأطفال على وجه الخصوص. وكما هو واضح، فإن هذا النوع من الصدق لا يستخدم فيه إجراء إحصائي

محدد، بل يعتمد على رأي المختصين في تحديد فقرات الاختبار ومدى ملاءمتها، أو إضافة فقرات أخرى أو حذف بعض منها.

٣- الصدق التلازمي (Concurrent validity):

وهو أحد أنواع صدق المحك (Criterion validity)، ويتطلب هذا النوع من الصدق إجراء اختبار إحصائي لتحديد مقداره، من خلال العلاقة الارتباطية بين الاختبار المستخدم ومقياس آخر محكي، أو مقارنة نتائج هذا الاختبار المستخدم بنتائج المقياس المحكي (المعروف أصلاً صدقه). وعلى سبيل المثال، فلو أردنا مقارنة نسبة الشحوم باستخدام مقياس سمك طية الجلد (قياس غير مباشر) مع تحديد نسبة الشحوم بواسطة الوزن تحت الماء (مقياس محكي)، فإننا نتأكد من صدق قياس سمك طية الجلد في الدلالة على نسبة الشحوم من خلال مقارنة العلاقة الارتباطية بين نتيجة الطريقتين، كما نحسب مدى الاتفاق بينهما في تحديد متوسط نسبة الشحوم، ومقدار خطأ التقدير المعياري (Standard error of estimate). وكمثال آخر على هذا النوع من الصدق، يمكن لنا مقارنة نوع جديد من أجهزة قياس مستوى النشاط البدني بطريقة أخرى معروف صدقها في قياس النشاط البدني (مقياس محكي) مثل استخدام الماء غير المشع (Doubly-labeled water)، بحيث يتم القياس بالطريقتين معاً ثم نقارن نتائجهما ومدى توافق الطريقة الجديدة مع المحك.

وهناك صدق آخر يسمى الصدق التقاربي (Convergent validity) شبيه بالصدق التلازمي، يتم فيه مقارنة اختبار ما بأخر معروف (ليس بالضرورة محكي) ثم حساب العلاقة الارتباطية بينهما ومدى الاتفاق بين نتائجهما، فمثلاً يمكن لنا حساب الصدق التقاربي لجهاز قياس الخطى (Pedometer) بناء على مقارنته مع نتائج جهاز قياس الحركة (Accelerometer)، الذي يعد أكثر دقة في قياس مستوى النشاط البدني للفرد.

٤- الصدق التنبؤي (Predictive validity):

وهو نوع آخر من أنواع صدق المحك (Criterion validity)، ويتم فيه على سبيل المثال التنبؤ بمستوى النشاط البدني في الكبر من خلال مستوى النشاط البدني في الصغر. كما يمكن استخدام هذا النوع من الصدق في التنبؤ بالاستهلاك الأقصى للأكسجين من خلال نتائج اختبار ١٦٠٠ متر جري، أو التنبؤ بالأداء البدني للرياضي من خلال اختبارات أخرى معملية أو ميدانية، أو من مواصفات جسمية للرياضي. ويستخدم في هذا النوع من الصدق تحليل الإنحدار (Regression analysis) للحصول على المعادلة التنبؤية ومعاملات الصدق.

وفي حالة استخدام عدة متغيرات مستقلة للتنبؤ بالمتغير التابع، فيلزم عادة استعمال عينة كافية العدد (١٥ شخص لكل متغير مستقل)، وإلا فالملاحظ عند تطبيق المعادلة التنبؤية المستخرجة من عينة الصدق على عينة أخرى جديدة إنكماش (تضاؤل) معامل الصدق في العينة الجديدة، وهذا الإنكماش يسمى (Shrinkage) (٢٨).

٥- الصدق البنائي (Construct validity):

ويعني مدى ملائمة اختبار ما في قياس صفة أو قدرة لا يمكن قياسها مباشرة (أو غير متوفر قياس مباشر لها). فمثلاً عند تصميم اختبار لقياس اللياقة القلبية التنفسية، وفي ظل عدم وجود إمكانية لمقارنة هذا الاختبار بنتائج قياس مباشرة للاستهلاك الأقصى للأكسجين، فيمكن إجراء الاختبار على مجموعتين من الأفراد، أحدها من الرياضيين ذوي اللياقة البدنية العالية، والأخرى من العاديين الذين لا يمارسون أي نشاط بدني على الإطلاق، فإذا كان الاختبار المستخدم يتميز بالصدق البنائي فينبغي عليه أن يميز بين أداء الفئتين من الأفراد في درجاته، وبالتالي يحصل الرياضيون على درجات أعلى من غير الرياضيين في اختبار اللياقة القلبية التنفسية، وإلا فإن الاختبار يصبح غير صادق من الناحية البنائية. وكمثال آخر على الصدق البنائي، لنفترض أننا طورنا اختباراً لقياس مهارة من مهارات كرة السلة، ومن ثم طلبنا من مجموعتين من لاعبي كرة السلة المهرة وغير مهرة القيام بأداء ذلك الاختبار، فإذا كان الاختبار صادقاً من الناحية البنائية، فإنه سيميز بين لاعبي المهرة وغير المهرة وإلا فيعد ذلك الاختبار غير صادق.

ولحساب المعاملات الإحصائية للصدق بين القياس المباشر (المحكّي) وغير المباشر، فإننا نستخدم معامل ارتباط بيرسون (أو معامل ارتباط سبيرمان رو (Rho) للبيانات الرتبية). كما يتم حساب خطأ التقدير المعياري (Standard error of estimate) على النحو التالي:

$$= \text{الانحراف المعياري للتقدير} \sqrt{1 - \text{مربع معامل بيرسون بين القياس والتقدير}}$$

ويمكن كذلك حساب الخطأ الكلي الذي يجمع بين خطأ التقدير المعياري (SEE) ومتوسط الفروق بين التقدير والقياس، وذلك على النحو التالي (٢٣):

$$= \sqrt{\text{مجموع (التقدير - القياس)}^2 \div \text{عدد العينة}}$$

الثبات (Reliability):

تأتي صفة الثبات ثانياً بعد الصدق كأساس للحكم على مدى سلامة الاختبار وملاءمته للاستخدام بكل اعتمادية. والثبات يعني مقدار الموثوقية في الاختبار ليعطي دوماً النتائج نفسها، والاختبار الثابت يعطي النتائج نفسها عند تكرار إجراءه، بدون تذبذب أو تفاوت من وقت إلى آخر. ويُعد التحقق من الثبات أسهل عموماً من التحقق من الصدق. وعادة ما يكون الحد الأدنى المطلوب في معامل الثبات (**Reliability coefficient**) أعلى من الحد الأدنى المطلوب تحقيقه في معامل الصدق، حيث من المتوقع أن يصل معامل الثبات المقبول إلى ٠,٨٠ أو أكثر^(٢٦). والجدير بالتنويه أنه لا بد من تدوين مقدار معامل الثبات عند عرض النتيجة، وعدم الاكتفاء بالقول أن معامل الثبات ذا دلالة معنوية (إحصائية)، حيث من المعروف أن عدد أفراد العينة يؤثر على مستوى الدلالة الإحصائية، وبالتالي فمن الممكن الحصول على معامل ذي درجة منخفضة لكنه ذا دلالة إحصائية عند مستوى ٠,٠٥، وذلك بسبب أن حجم العينة كبير جداً.

ومن أجل التأكد من ثبات الاختبار نستخدم العديد من الأدوات الإحصائية، والتي من أهمها إجراء الاختبار وإعادة الاختبار، ثم نحسب معامل الارتباط الأحادي (**Intraclass correlation coefficient**) مستخدمين تحليل التباين الأحادي. والجدير بالإشارة هنا أن البعض يستخدم معامل ارتباط بيرسون (أو ما يشابهه) لتحديد معامل الثبات في حالة الاختبار وإعادة الاختبار، وهذا قطعاً إجراء غير صحيح للأسباب التالية:

أولاً: يعد معامل ارتباط بيرسون معاملاً ثنائياً (**Bi-variate**) يستعمل في حالات فحص العلاقة بين متغيرين منفصلين، بينما في حالة الاختبار وإعادة الاختبار فإن المتغير واحد فقط تم اختباره مرتين، والأولى في هذه الحالة استخدام معامل ارتباط أحادي (**Uni-variate**).

ثانياً: لا يمكن لمعامل بيرسون أن يفحص التغيرات الحاصلة في المتوسطات والانحرافات المعيارية بين نتائج الاختبار الأول والثاني، بينما يمكن القيام بذلك في تحليل التباين الأحادي (**ICC**).

ثالثاً: في حالة استخدام معامل ارتباط بيرسون، فإنه يمكن أن نحصل على قيمة ارتباط عالٍ بين الاختبار الأول والثاني إذا تغيرت كل الحالات بالزيادة أو النقصان بنفس المستوى، على الرغم من عدم ثباتها^(١٠، ١٧).

وحتى نتمكن من تفسير معامل الثبات بشكل أكثر فائدة، ينبغي لنا أن نحسب مقدار الخطأ المعياري لعملية القياس (Standard error of measurement)، وذلك على النحو التالي:

$$\text{خطأ القياس المعياري (SEM)} = \sqrt{\text{متوسط الانحراف المعياري}^2 - \text{معامل الثبات الأحادي (ICC)}}$$

وكلما كان خطأ القياس المعياري صغيراً تصبح الثقة في الاختبار عالية. والمعتاد أنه إذا كان خطأ القياس المعياري (SEM) يساوي ربع الانحراف المعياري (SD) أو أقل، فإن الاختبار يعد ذو ثبات مقبول^(١٧). ويتم تفسير نتائج الخطأ المعياري كما هو الحال في تفسير الانحراف المعياري، ولنفترض أننا أجرينا اختباراً للقفز العمودي وكانت النتيجة ٤٠ سم مع خطأ معياري يساوي ٢ سم، يمكننا القول في هذه الحالة أن النتائج ستتراوح بين ٣٦ سم و ٤٤ سم في ٩٥% من الحالات، وبين ٣٨ سم و ٤٢ سم في حوالي ٦٨% من الحالات. كما يمكن استخدام اختبار بلاند و ألتمان (Bland & Altman) وحساب حدود الاتفاق عند معدل ثقة ٩٥%، من خلال رسم الفروق بين القياسين الأول والثاني مقابل متوسط البيانات^(١٣)، وسيأتي مثال على استخدام حدود الثقة لبلاند و ألتمان لاحقاً عند الحديث عن الخطأ الفني للقياس.

تحليل القدرة الإحصائية (Statistical power analysis)

في العديد من البحوث، خاصة التجريبية منها، حيث يصعب الحصول على عينه كبيرة أو تكون زيادة عدد الأفراد مكلفة جداً، يصبح من الضروري البدء في تحليل القدرة الإحصائية، من أجل تحديد عدد أفراد العينة الذي يعطي قدرة ملائمة (غالباً ٠,٨٠ - ٠,٩٠) تساعد الباحث على رفض الفرضية الصفرية بكل ثقة عندما تكون تلك الفرضية غير صحيحة. ولتحليل القدرة الإحصائية ينبغي أن نلم بالعناصر المرتبطة بذلك وهي:

١. عدد أفراد العينة (sample size).
٢. مستوى الدلالة الإحصائية (ألفا) (a-level).
٣. حجم التأثير (Effect size).
٤. مقدار القدرة الإحصائية (Statistical power) التي تساوي $1 - \beta$.

وغالباً ما يتم استخدام رقماً لمعامل بيتا يعادل ٤ أضعاف حجم الدلالة الإحصائية (أربعة أضعاف قيمة ألفا)، وبالتالي فعلى افتراض أن ألفا = ٠,٠٥ فسيكون مقدار بيتا = $0,05 \times 4 = 0,20$ وعليه فتصبح القدرة الإحصائية المطلوبة = $1 - 0,20 = 0,80$ (في بعض الحالات يمكن استخدام قدرة أعلى من ذلك مثل ٠,٩٠ أو أدنى من ذلك مثل ٠,٧٠)، غير أن زيادة

القدرة الإحصائية تتطلب منا زيادة عدد أفراد العينة المطلوبين في البحث، كما أن زيادة حجم التأثير المطلوب (زيادة مقدار ألفا) يؤدي إلى زيادة حجم العينة المطلوب.

والمعروف أن حجم التأثير (Effect size) يعني الفرق المعياري بين متوسطي المجموعتين التجريبية والضابطة بغض النظر عن المقياس المستخدم (Measurement scale). فإذا كان الفرق بين المجموعتين التجريبية والضابطة في الاستهلاك الأقصى للأوكسجين يساوي ٤ ملي لتر في الدقيقة (٤٠ ملي لتر في الدقيقة مقابل ٤٤ ملي لتر في الدقيقة) والانحراف المعياري للمجموعة التجريبية كان ٨ ملي لتر في الدقيقة، فإن حجم التأثير هنا يساوي ٠,٥ وهو رقم يقع في خانة الحجم المتوسط كما سنرى ذلك لاحقاً (١٠، ١٥، ٢٧، ٢٨).

ومن المعلوم أن لحجم التأثير أهمية تطبيقية، فالمطلوب في البحث هو أن نجد فرقاً بين المجموعتين التجريبية والضابطة يعد ذو أهمية عملية (تطبيقية) وليس ذو أهمية معنوية (ذو دلالة إحصائية) فقط، حيث يمكن أن نزيد من عدد أفراد العينة بشكل كبير مع تثبيت القدرة الإحصائية وبالتالي نحصل على فرقاً دالاً إحصائياً بين المجموعتين عند مستوى ٠,٠٥، لكن هل هذا الفرق سيُعد ذو أهمية عملية؟ أي أن هذا الفرق قد يكون من الصغر بمكان، الأمر الذي لا يجعل لهذا الفرق أي قيمة تطبيقية (مثلاً: حدوث انخفاض في ضغط الدم الانقباضي مقداره ٢ ملي متر زئبقي بعد برنامج تدريبي، وهو انخفاض ضئيل يدخل ضمن خطأ القياس الحاصل عند قياس ضغط الدم الشرياني، وليس له دلالة إكلينيكية. لذا فمعظم البحوث تحدد قيمة حجم التأثير المقبول منذ بداية تصميم التجربة، لكي يتم على ضوءه وعلى ضوء العوامل الأخرى تحديد حجم عينة البحث المطلوبة، بحيث لا يصبح حجم العينة صغيراً جداً بحيث يقود إلى عدم ظهور أي فرق بين المجموعتين التجريبية والضابطة على الرغم من وجود حجم تأثير ملموس (فرق معياري بين المجموعتين)، أو يصبح حجم العينة كبيراً جداً بحيث يمكن إيجاد فرق إحصائي دال بين المجموعتين وأن كان هذا الفرق (حجم التأثير) صغيراً جداً وديم الأهمية التطبيقية، ناهيك عن ارتفاع تكليف انجاز البحث بشكل غير مبرر من جراء زيادة عدد عينة البحث.

كيفية حساب حجم التأثير (Effect size)

يمكن حساب حجم التأثير (ES) بعد الحصول على نتائج البحث، بناءً على بعض المعلومات الأساسية كما سنستعرض ذلك في الفقرات التالية، غير أنه يمكن لنا أيضاً أن نحدد حجم التأثير مسبقاً قبل البدء بإجراء البحث، مما يساعدنا في تحديد حجم العينة المطلوبة التي تعطي أفضل قدرة إحصائية.

١. في حالة استخدام اختبار - ت للمجموعتين المستقلتين (Independent t-test):

$$\text{Effect size} = (M1 - M2) / SD$$

حيث:

M1 = متوسط المجموعة التجريبية.

M2 = متوسط المجموعة الضابطة.

SD = الانحراف المعياري للمجموعة الضابطة.

وفي حالة عدم وجود مجموعة ضابطة، يصبح المقام الفرق بين متوسط المجموعة الأولى والثانية، والانحراف المعياري هو متوسط انحرافيهما معاً على النحو التالي:

$$S_p = \{S_1^2 (n1-1) + S_2^2 (n2-1)\} / (n1 + n2) - 2$$

حيث:

S_1^2 = مقدار التباين في المجموعة الأولى (مربع الانحراف المعياري للمجموعة ١).

S_2^2 = مقدار التباين في المجموعة الثانية (مربع الانحراف المعياري للمجموعة ٢).

n1 = حجم العينة في المجموعة رقم ١.

n2 = حجم العينة في المجموعة رقم ٢.

٢. في حالة استخدام اختبار - ت للمجموعة الواحدة (Dependent sample- t-test):

$$\text{Effect size} = (M \text{ post} - M \text{ pre}) / SD \text{ pre}$$

حيث:

M post = متوسط المجموعة بعد المعالجة (الاختبار البعدي).

M pre = متوسط المجموعة قبل المعالجة (الاختبار القبلي).

SD pre = الانحراف المعياري للمجموعة قبل المعالجة (انحراف الاختبار القبلي).

ويمكن استخدام المعايير التالية للدلالة على مقدار حجم التأثير، بناءً على اقتراح من كوهين (Cohen) (١٥):

أقل من ٠,٤١ = حجم تأثير صغير

من ٠,٤١ - ٠,٧٠ = حجم تأثير متوسط

أكبر من ٠,٧٠ = حجم تأثير كبير

٣. في حالة استخدام اختبارات تحليل التباين أو تحليل الانحدار:

يمكن استخدام نسبة التباين في المتغير التابع الذي يمكن تفسيره بالمتغير المستقل

(The proportion of variance of the dependent variable accounted for by the independent variable (s))

وذلك بالاعتماد على العوامل التالية:

مربع معامل الارتباط (r^2)

مربع معامل الارتباط المتعدد (R^2)

مربع إيتا (h^2) η^2

مربع أوميغا (w^2) Ω^2

مربع إبسون (e^2) ϵ^2

ويمكن استخدام المعايير التالية للدلالة على مقدار حجم التأثير:

حجم تأثير صغير = ٠,٠١

حجم تأثير متوسط = ٠,٠٦

حجم تأثير كبير = ٠,١٥

كما يمكن تحويل نسبة التباين المشترك (Proportion of variance) إلى حجم تأثير مباشرة باستخدام المعادلة التالية (٢٧):

$$\text{حجم التأثير (ES)} = \sqrt{\epsilon \times (PV - 1) \div (PV)}$$

حيث يمثل PV نسبة التباين المشترك.

كما يمكن تحويل حجم التأثير إلى نسبة التباين المشترك، كالتالي:

$$\text{نسبة التباين المشترك (PV)} = \text{مربع حجم التأثير} \div (\text{مربع حجم التأثير} + \epsilon)$$

الطرق التي يمكن من خلالها زيادة القدرة الإحصائية (Statistical power)

١- زيادة التباين بين المجموعات (Between groups)، أي زيادة الفروق بين متوسطي المجموعتين التجريبية والضابطة (أو بين المجموعة القبلية والمجموعة البعدية)، وذلك من خلال تطبيق معالجة قوية، مثلاً في بحث للقدرة الهوائية: تطبيق البرنامج التدريبي لمدة ٣ شهور بدلاً من شهر واحد، أو زيادة شدة التدريب من ٤٠% من القدرة الهوائية القصوى إلى ٧٠%، من أجل الحصول على تأثير أكبر وبالتالي فرق أكبر بين متوسطي المجموعتين التجريبية والضابطة.

٢- التقليل من التباين الحاصل داخل المجموعات (Within group)، وذلك من خلال:

- ضبط المعالجة بشكل جيد، وتطبيقها على جميع أفراد المجموعة التجريبية بالكامل، ومنع تغيب البعض عن أيام المعالجة (التغيب عن التدريب مثلاً)، كذلك اختيار مشاركين متجانسين داخل المجموعة الواحدة.
- استخدام تصميم بحثي يعتمد على القياسات المتكررة (Repeated measure design)، حيث يتم في هذا النوع من التصميم استبعاد التفاوت الحاصل بين الأفراد داخل المجموعة (Within group) من الخطأ (Residual)، مما يزيد من القدرة الإحصائية (Power). ويعد هذا النوع من التصميم أكثر اقتصادية في حجم العينة، أي أنه يحتاج إلى عدد أقل من الأفراد في كل مجموعة مقارنة بغيره من التصاميم البحثية (٢٨).

٣- زيادة حجم العينة، على أن هذا الإجراء ليس فقط مكلف مادياً بل وصعب في كثير من الأحيان، خاصة في البحوث التجريبية.

٤- رفع مستوى الدلالة الإحصائية (ألفا) من ٠,٠١ إلى ٠,٠٥ (لاحظ أن ذلك يقود إلى زيادة احتمالات رفض الفرضية الصفرية (Null hypothesis) إذا كانت صحيحة (أي زيادة احتمالات حدوث خطأ من النوع الأول، الذي يعني أن هناك فرق بينما في الواقع لا يوجد ذلك فرق)).

العلاقة الارتباطية ومفهوم السببية

يعتقد بعض الباحثين المبتدئين وطلاب الدراسات العليا للوهلة الأولى أن مجرد وجود علاقة قوية بين متغيرين لهو دليل كافٍ على وجود علاقة سببية (Causality) بينهما، أي أن أحدهما يعد سبباً لحدوث الآخر. لكن سرعان ما يكتشفوا فيما بعد أن العلاقة الارتباطية ليست دليلاً كافٍ وحدها على وجود السببية، فقد يكون المسبب عامل ثالث له علاقة قوية مع كلا المتغيرين محل الدراسة. ومن المؤكد أنه يلزم في البدء أن يكون هناك علاقة ارتباطية قوية بين متغيرين ليتمكن لنا فيما بعد أن نحاول إثبات السببية فيما بينهما.

ولعلنا نذكر للقارئ أنه في بدايات بحوث الذكاء في القرن الماضي، وجد العلماء أن هناك علاقة طردية قوية بين مستوى الذكاء لدى الطفل (بناء على اختبارات الذكاء) ومقاس حذاه، مما جعلهم يستنتجون في البداية أن الطفل صاحب القدم الكبيرة هو الأذكى، والعكس صحيح بالنسبة للطفل صاحب القدم الصغيرة، لكن سرعان ما اكتشفوا فيما بعد أن العلاقة الحقيقية هي بين مستوى الذكاء وعمر الطفل، وأن مقاس الحذاء يزداد مع زيادة العمر، أي أن

هناك علاقة طردية بين عمر الطفل ومقاس حذائه، وأنهم لو أبعثوا تأثير العمر على العلاقة بين مقاس الحذاء ومستوى الذكاء لاختفت العلاقة بينهما.

ولا شك أنه في البحث التجريبي فقط، وليس الارتباطي، يمكن لنا محاولة إثبات السببية، وذلك من خلال التحكم بالمتغير المستقل (المسبب للتأثير Cause) والنظر في تأثير ذلك على المتغير الثابت (الخاضع للتأثير Effect). والجدير بالذكر أن استخدام الأسلوب الإحصائي وحده غير كافٍ للجزم بالعلاقة السببية بين متغيرين، بل ينبغي أن يكون هناك أساس منطقي متوافق مع النتيجة الإحصائية. وعليه، فلكي يمكن الجزم بأن عامل الخطورة (س) هو سبب حدوث المرض (ص)، أو أن تحسن الأداء البدني للرياضي كان بفعل تناوله المشروب الرياضي (ع)، ينبغي أن تتوفر مجموعة من الشروط. هذه الشروط الواجب توفرها في العلاقة السببية هي على النحو التالي:

١. قوة العلاقة بين المؤثر والمتأثر (Strength):

أي أن العلاقة بين المتغير المستقل (المؤثر أو المسبب) والمتغير التابع (المتأثر) قوية ومن غير الممكن حدوثها بطريقة الصدفة.

٢. اتساق العلاقة وثباتها (Consistency):

أي أن هناك اتساقاً وانسجاماً بين هذين المتغيرين، فوجود أحدهما يعني دائماً وجود الآخر (أو حدوثه). وبمعنى آخر، متى ما وجد السبب وجدت النتيجة، ومتى ما حصلت النتيجة كان السبب موجوداً.

٣. التزامن (Temporality):

أي أن العلاقة الزمنية بين السبب والنتيجة دائماً موجودة. وبمعنى آخر أن السبب يسبق دائماً النتيجة، وليس العكس.

٤. المنوعة (Specificity):

أي أن عدم وجود السبب، ينفي وجود النتيجة، فعدم وجود عامل الخطورة يقضي بعدم إمكانية حدوث المرض.

٥. القبول المنطقي (Plausibility):

أن العلاقة السببية هي علاقة منطقية، ويوجد لها أساس بيولوجي، كما يمكن توقع حدوثها نظرياً.

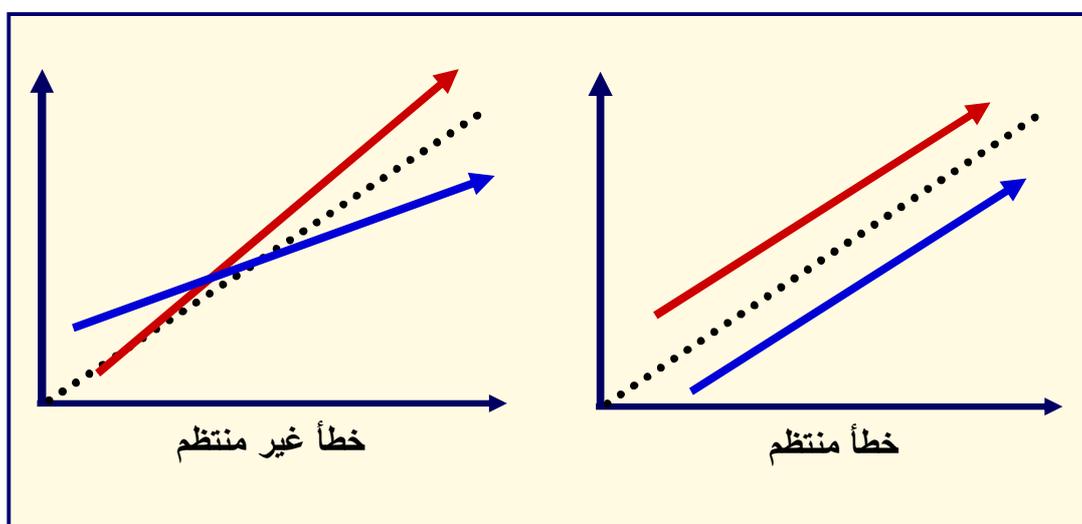
خطأ القياس (Measurement error)

تتميز كثير من القياسات المعملية المستخدمة في فسيولوجيا الجهد البدني بقدر كبيراً من الدقة والثبات، على أن معظم تلك القياسات يتطلب أيضاً حد أدنى من التدريب على إجرائها حتى يمكن إجادتها، وبالتالي التقليل من فرصة حدوث خطأ في القياس. ومع ذلك نلاحظ حدوث بعض التفاوت بين قياسات متكررة للفاحص الواحد، أو قياسات متعددة لفاحصين أو أكثر، خاصة في حالة عدم توفر الخبرة الكافية والتدريب اللازم للأشخاص القائمين على عملية القياس. ولا شك أننا لا يمكن في جميع الأحوال أن نتخلص كلية من التفاوت الحاصل بين قياس وآخر، لكن يمكن حصره في نطاق ضيق مقبول. ومن المعلوم أنه يمكن أن يحصل تفاوت في القياس نتيجة لعوامل بيولوجية (داخلية) وأخرى خارجية تتعلق بالجهاز المستعمل أو بالشخص الذي يقوم بالقياس أو نتيجة لعوامل خارجية أخرى واقعة على المفحوص نفسه. ويعد التفاوت البيولوجي في معظم المتغيرات الفسيولوجية صغيراً، على عكس المتغيرات النفسية والسلوكية على سبيل المقارنة. فضربات القلب في الراحة، خاصة إذا تم قياسها في الصباح بعد الاستيقاظ من النوم مباشرة، لا يتعدى التفاوت البيولوجي الحاصل في معدلها بين يوم وآخر لدى شخص ما بضع ضربات في الدقيقة، وينطبق القول في الحقيقة على معظم المتغيرات الفسيولوجية الأخرى لدى الإنسان، كاستهلاك الأكسجين في الراحة أو معدل التنفس، أو درجة حرارة الجسم أو عدا ذلك من متغيرات.

ويمكن تقسيم خطأ القياس إلى خطأ منتظم (Systematic) وآخر غير منتظم (Random). ويحدث الخطأ المنتظم في اتجاه واحد، إما في الاتجاه الموجب أو في الاتجاه السالب (يعني ذلك أن هناك زيادة مستمرة أو نقص مستمر بين القياسين الحقيقي والمُشاهد). ويوضح الشكل البياني رقم (١) صور متعددة لهيئة كل من الخطأ المنتظم والخطأ غير المنتظم، علماً بأن الخط المتقطع (...) يمثل خط السواء. ونلاحظ من الشكل انتظام الخطأ المنتظم، لكنه عشوائي في حالة الخطأ غير المنتظم. كما نلاحظ أيضاً أن الخطأ غير المنتظم يمكن أن يكون محدوداً عند المقادير الصغيرة لكنه يزداد كلما ازدادت المقادير (أنظر الخط الأحمر)، أو يكون محدوداً جداً في الوسط لكنه يصبح كبيراً كلما اتجهنا إلى الأرقام المتطرفة، أي الصغيرة أو الكبيرة (أنظر الخط الأزرق).

وعندما تزداد قيمة الخطأ العشوائي مع زيادة قيمة الشيء الخاضع للقياس (أنظر الشكل رقم (٢) الذي يوضح اختبار بلاند وألتمان)، فإن ذلك يسمى (Heteroscedasticity)، أي أن الشخص الذي حقق أعلى قيمة في اختبار ما أظهر أعلى خطأ قياس (وهذا يؤثر على قدرة الاختبار الإحصائي المستخدم مما يزيد من احتمالية ارتكاب خطأ من نوع ٢). وعندما تكون

البيانات بهذا الشكل فينبغي أن يتم قياسها على هيئة مقياس نسبي (Ratio scale)، أو تحويل البيانات إلى لوغاريتمية. وعندما لا يكون هناك علاقة بين مقدار خطأ القياس وقيمة المتغير قيد القياس، فيقال عن البيانات أنها (Homoscedasticity)، وعليه يتم التعبير عنها بنفس وحدة القياس التي هي عليه. ومن أمثلة الأسباب المؤدية لحدوث الخطأ المنتظم في الاتجاه الموجب حدوث التعلم أو زيادة الدافعية لدى الشخص، حيث نلاحظ أن أداء المفحوص في القياس الثاني يكون أفضل من الأول نتيجة للتعلم أو التعود على القياس، أو بسبب زيادة الدافعية لدى الشخص. كما يمكن لعوامل مثل حدوث التعب أو انخفاض الدافعية أو عدم كفاية فترة الاسترداد أن تقود إلى حدوث خطأ منتظم في الاتجاه السالب.

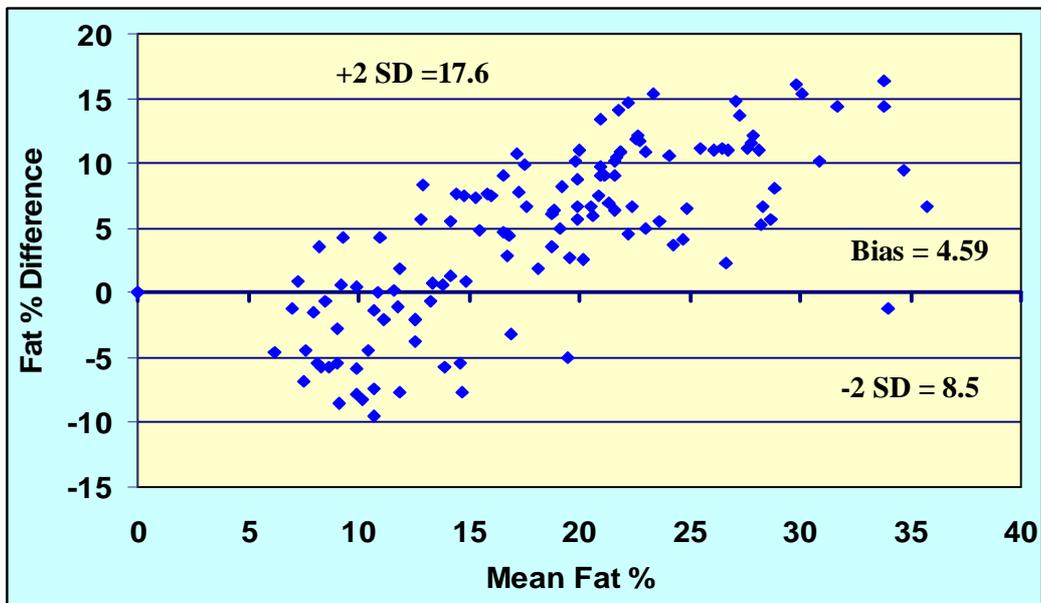


شكل رقم (١): صور متعددة لهيئة كل من الخطأ المنتظم وغير المنتظم.

أما الخطأ غير المنتظم (أو العشوائي) فيحدث نتيجة لمتغيرات بيولوجية (في الإنسان نفسه) أو لأسباب ميكانيكية (في الجهاز) أو من جراء اختلاف البروتوكول المستخدم في القياس (في الإجراءات). والخطأ غير المنتظم هو خطأ عشوائياً كما يدل على ذلك الاسم، وهو يسبب مشكلة كبيرة يصعب في أحوال عديدة اكتشافه، لذا يتطلب الأمر الحرص والدقة في ضبط إجراءات القياسات ومواعيدها والتقليل من حدوث العوامل البيولوجية المؤثرة على القياس، مثل أن يكون الرياضي مريضاً أو مرهقاً أو لم يحصل على كفايته من النوم، مع التأكد من دقة أجهزة القياس.

ويمكن معرفة اتجاه خطأ القياس، وما إذا كان منتظماً أو عشوائياً، وذلك باللجوء لعدة اختبارات إحصائية تستخدم لهذا الغرض، من أهمها اختبار مدى الاتفاق لبلاند وألتمان (Bland

Altman & (13) ، ويوضح الشكل البياني رقم (٢) رسماً لمدى الاتفاق بين طريقتين من طرق قياس نسبة الشحوم على مجموعة من الطلاب الجامعيين السعوديين تم إجراءها في مختبر فسيولوجيا الجهد البدني بجامعة الملك سعود، إحدى هاتين الطرق كانت الطريقة المحكية التي تمت بواسطتها مقارنتها بالطريقة الأخرى الأقل دقة. ويظهر من الرسم عدد المحاولات (كان عدد المشاركين في البحث ١٣٢ شخصاً) ومقدار إنحرافين معياريين في كل من الاتجاه الموجب (يساوي ١٧,٦% شحوم) والاتجاه السالب (يساوي ٨,٥% شحوم)، كما يتضح من الرسم أن متوسط خطأ القياس في نسبة الشحوم بين الطريقتين يكون سالباً عندما تكون نسب الشحوم منخفضة (أقل من ١٥%)، وموجباً عندما تكون نسبة الشحوم مرتفعة (أكثر من ١٥%). وهذا يشير إلى وجود خطأ قياس عشوائي، علماً بأن مقدار التحيز أو الانحراف (bias) عن خط السواء قد بلغ ٤,٥٩%.



شكل رقم (٢): رسم بياني يوضح كيفية حساب خطأ القياس باستعمال اختبار بلاند وألتمان، مستخدمين بيانات لنسب الشحوم لطلاب جامعيين تم قياسهم بطريقتين (المصدر: الهزاع، وآخرون: . تحديد نسبة الشحوم في الجسم لدى الطلاب الجامعيين من خلال قياس كل من حجم الجسم وكثافته بواسطة إزاحة الهواء والمقاومة الكهروحيوية وسمك طية الجلد. الرياض: مركز البحوث، كلية التربية، جامعة الملك سعود، ١٤٢٣هـ).

حساب الخطأ الفني للقياس (TEM)

تتأثر دقة القياس بمقدار الخبرة والتدريب الذي تلقاه الفاحص، فالفاحص غير المتدرب تشطح قياساته عن تلك التي للفاحص الذي يمتلك الخبرة، كما تتفاوت قياسات الفاحص غير المتدرب بين بعضها البعض عند إجراء محاولات متكررة للمتغير ذاته، وجرت العادة أن يتم تدريب الفاحص تدريباً كافياً حتى يجيد عملية القياس، ويصبح التفاوت (أو التباين) بين محاولاته المتكررة (أو بينه والفاحص المتدرب) ضئيلة جداً. وكقاعدة عامة، يُوصي دائماً بإجراء قياسات متكررة للمفحوص ذاته ثم مقارنة الفرق بينهما، فإذا كان الفرق يصل إلى ١٠% أو أكثر بالزيادة أو النقصان، فلا بد للفاحص من إجراء مزيداً من التدريب. وللتأكد من أن الفاحص قد أمّتك الخبرة الكافية التي تجعل قياساته ثابتة ومعتمدة، نقوم بقياس مقدار التباين بين محاولتين يجريهما الفاحص لعدد من المفحوصين، ثم نقوم بحساب الانحراف المعياري للفرق بين المحاولتين، وذلك ما يسمى "بالخطأ الفني للقياس" (Technical error of measurements) (٢٧، ١٩)، حيث يمكن إجراءه بين محاولات الفاحص الواحد، أو بين محاولات فاحصين اثنين، على النحو التالي:

$$S = [(\sum d^2 / (2 n))^{0.5}]$$

ويمكن إعادة صياغة المعادلة بالشكل التالي:

$$= \sqrt{\text{مجموع مربعات الفروق} \div (2 \times \text{عدد المفحوصين})}$$

وفي الواقع، فإن قياس مقدار التباين بين المحاولات المتكررة للقياس يعد أيضاً مقياساً للاتساق (Consistency)، ويمكن أيضاً اعتماده كمقياس لثبات القياس (Reproducibility). ويمكن أن يتم حسابه بين المحاولات لدى نفس الفاحص (within) أو بين محاولات فاحصين (Between). وإذا كان التباين المراد فحصه لأكثر من قراءتين، أو لأكثر من فاحصين، فنستعوض عن المعادلة السابقة بإجراء اختبار الثبات مستخدمين معامل الارتباط أحادي التباين المسمى (Intraclass correlation coefficient)، أو اختصاراً (ICC)، وهو غير معامل الارتباط الثنائي، كمعامل بيرسون على سبيل المثال.

مثال لكيفية حساب مقدار الخطأ الفني للقياس:

لنفترض أن شخصاً تدرب على قياس سمك طية الجلد، وأردنا أن نتعرف على مقدار دقته في إجراء القياس، وهل التفاوت الحاصل لديه بين قياسين لموقع واحد من مواقع قياس

سمك طية الجلد مقبولاً أم لا؟ أي ما مقدار اعتمادنا على دقة قياسات هذا الشخص؟ لذا قمنا بالطلب منه أن يقوم بقياس سمك طية الجلد مرتين لخمس مفاصين، وقارنا بين القراءتين الأولى والثانية لكل مفوص. وكانت البيانات التي حصلنا عليها كما هو موضحاً في الجدول رقم (١) أدناه. وعليه يمكننا حساب الخطأ الفني للقياس وذلك بتطبيق المعادلة السابقة، على النحو التالي:

$$\text{الخطأ الفني للقياس} = \text{مجموع مربعات الفروق} \div (٢ \times \text{عدد المفوصين})$$

$$= \sqrt{(٥ \times ٢) \div ١٤} = ١,١٨ \text{ ملم}$$

جدول رقم (١): بيانات سمك طية الجلد لـ٥ الفاصصين الذي قام بقياس خمسة مفوصين مرتين.

رقم المفوص	القراءة الأولى	القراءة الثانية	الفرق بينهما	مربع الفرق
١	١٢	١٠	٢	٤
٢	١٤	١٣	١	١
٣	١٨	١٩	١-	١
٤	٨	١٠	٢-	٤
٥	١٦	١٤	٢	٤
	متوسط = ١٣,٦	متوسط = ١٣,٤	٢+	المجموع = ١٤

لكن ماذا يعني لنا هذا الناتج؟ هل هو مقبول؟ أم غير مقبول؟ وهل لو كان بهذا المقدار في حالة قياس الوزن أو الطول أو نسبة الشحوم في الجسم، هل يمكن اعتباره خطأ مقبولاً؟ أم غير مقبول؟ في الواقع لا بد من إجراء خطوة إضافية أخرى، حتى يكون تفسير هذا الناتج ذا معنى. هذه الخطوة هي في حساب الخطأ الفني النسبي للقياس (Relative error)، وذلك بنسبته إلى قيمة المتوسط من القراءتين، وذلك على النحو التالي:

$$\text{الخطأ الفني النسبي} = (١,١٨) \div ((١٣,٤ + ١٣,٦) \div ٢) \times ١٠٠ = ٨,٧ \%$$

الآن، هل يعد مقدار الخطأ النسبي هذا مقبولاً أم لا؟ يعتمد الأمر على الحالة التي نحن بصدد قياس سمك طية الجلد لها، فإذا كان الأمر يتعلق بتدريب أحد الأشخاص على القياس، وبالتالي التأكد من إجادته للقياس، فإننا قد نقبل خطأً فنياً في قياس سمك طية الجلد مقداره ٥%، أما في حالة معرفة الفرق بين سمك طية الجلد لأحد الرياضيين النخبة قبل التدريب وبعده، أو بين حالتين لسمك طية الجلد في برنامج خفض الوزن، فإننا قد لا نقبل أكثر من ٢% أو بحد أقصى ٣%. وتوصي الجمعية الدولية لتطوير القياسات الجسمية (ISAK) بأن لا يتجاوز الخطأ الفني لقياس سمك طية الجلد لمن يمنح شهادة المستوى الأول في إجادة القياسات الجسمية مقدار ١٠%، ولمن يمنح شهادة المستوى الثاني أو الثالث مقدار ٧,٥%. كما توصي الجمعية بأن لا تتجاوز نسبة الخطأ الفني للقياس لدى الأشخاص العاملين في مختبرات فسيولوجيا الجهد البدني المعتمدة لأجراء الاختبارات الفسيولوجية على الرياضيين الأستراليين (١٩) ما يلي:

- > ٣,٥ % للقياسات المتكررة لسمك طية الجلد في اليوم نفسه.
- > ٥,٠ % للقياسات المتكررة لسمك طية الجلد في أيام مختلفة.
- > ١,٥ % للقياسات المتكررة لسمك طية الجلد في سبع مناطق من الجسم في اليوم نفسه.
- > ٣,٠ % للقياسات المتكررة لسمك طية الجلد في سبع مناطق من الجسم في أيام مختلفة.

ويستعرض الجدول رقم (٢) بعض الحدود العليا المعتمدة لخطأ القياس الفني المسموح فيها عند إجراء القياسات الفسيولوجية على الرياضيين الأستراليين، حيث نلاحظ أن معظم تلك المتغيرات الفسيولوجية التي يتم قياسها يتراوح خطأ القياس المسموح به من ٢% إلى ٥%، ما عدا قياس حمض اللبنيك الذي يكتنفه أصلاً العديد من العوامل المؤثرة على دقة قياسه، الأمر الذي يجعل من قيمة خطأ القياس المسموح به أكبر من بقية المتغيرات الأخرى.

جدول رقم (٢): مقدار الخطأ الفني للقياس المستهدف طبقاً لمعايير القياسات الفسيولوجية على الرياضيين الأستراليين.

الخطأ الفني المستهدف	وحدة القياس	المتغير
> ٢%	ملم	٧-٨ مناطق لطية الجلد
> ٥%	سم	القفز العمودي
> ٣%	شمعة	القدرة اللاهوائية (درجة الجهد لمدة ١٠ ث)
> ٣%	لتر/ دقيقة	الاستهلاك الأقصى للأكسجين
> ٥%	لتر/ دقيقة	التهوية الرئوية القصوى
> ٢%	ضربة/ دقيقة	ضربات القلب القصوى
> ١٥%	ملي مول/ لتر	حمض اللبنيك

المصدر: Gore C. (Ed): Physiological Testing for Elite Athletes, 2000, p.

معامل الثبات (Coefficient of reliability)

بالإضافة إلى الخطأ الفني للقياس، يمكن أيضاً حساب معامل الثبات الذي يدل على مقدار الاعتمادية والوثوق في الاختبار المستخدم، وهو إجراء لا يتطلب إلا وجود قرائنتي الخطأ الفني للقياس والانحراف المعياري، وذلك على النحو التالي:

$$\text{معامل الثبات} = 1 - ((\text{الخطأ الفني للقياس})^2 \div (\text{الانحراف المعياري})^2)$$

ويتراوح معامل الثبات من صفر (أي أن القياس لا يعتمد عليه) إلى واحد صحيح (ثبات تام)، ويقترح بعض المختصين أن معامل الثبات الموصى به ينبغي أن لا يقل عن ٠,٩٥ (أي ٩٥%) (٢٠, ٢٩).

معامل التفاوت (Coefficient of variation)

لا شك أن معرفة مقدار أي من التباين أو الانحراف المعياري يعطي مؤشراً جيداً على تباين القياسات، لكن ننظر إلى البيانات الموضحة في الجدول رقم (٤-٣). سيئين لنا أن مقادير الانحراف المعياري لا تختلف كثيراً بين الطول والوزن ونسبة الشحوم، وهي أعلى

بكثير من الانحراف المعياري لقوة القبضة. وهذا في الواقع صحيح، لكن لا يمكن استخدام هذه المعلومات لمقارنة التباين في القياسات بين المتغيرات الأربعة وبالتالي مقارنة درجة ثبات تلك القياسات، بل نستخدم ما يسمى بمعامل التفاوت، والذي يساوي:

$$\text{معامل التفاوت} = (\text{الانحراف المعياري} \times 100) \div \text{المتوسط الحسابي}$$

لقد تم حساب معامل التفاوت للبيانات نفسها، كما هو موضحاً في الجدول رقم (٣). ويتضح من الجدول أن مقدار التفاوت كان الأعلى في نسبة الشحوم ووزن الجسم، ثم في قوة القبضة، بينما كان الأدنى في طول الجسم. لا بد من الإشارة أيضاً أن معامل التفاوت يُحسب فقط في حالة البيانات ذات المقياس النسبي (Ratio scale)، أي التي لها صفراً حقيقياً والفرق بين رقم وآخر يعني الشيء نفسه (أي الفرق بين ١٢ كجم و ١٣ كجم يساوي الفرق نفسه بين ٢٥ كجم و ٢٦ كجم في مقياس الوزن)، وليس عندما تكون البيانات ذات مقياس فترتي (Interval)، مثل درجات الحرارة، أو استخدام درجات z لحساب الطول تبعاً للعمر (Height for age) أو الوزن تبعاً للطول (Weight for height) كما في دراسات النمو (٢٠٠١).

جدول رقم (٣): بيانات الطول والوزن ونسبة الشحوم والقوة العضلية لدى عينة من الأطفال السعوديين البدناء.

المتغير	المتوسط الحسابي	الانحراف المعياري	معامل التفاوت
الطول (سم)	١٣٩,١	٨,٢	٥,٩ %
الوزن (كجم)	٤٢,٣	١٠,٧	٢٥,٣ %
نسبة الشحوم (%)	٢٩,٠	٧,٧	٢٦,٦ %
قوة القبضة (كجم)	١٦,٣	٢,٩	١٧,٨ %

مصدر البيانات: الهزاع، هزاع، ١٤١٧هـ.

المقاييس المعيارية والمقاييس المحكية

جرت العادة عند إجراء قياسات على عدد كبير من الأفراد أن يتم تدوينها على أسس معيارية أو ما يسمى بالمقاييس المعيارية (Normative-referenced standards)، وتسمى أختصاراً بالمعايير (Norms)، حيث تقسم على هيئة درجات مئوية (Percentiles)، كما هو

متبع على سبيل المثال في قياسات الوزن والطول أو اختبارات اللياقة البدنية، ويوضح الجدول رقم (٤) الدرجات المئينية لقوة القبضة بالكيلوجرام لدى الأطفال السعوديين البنين من مدينة الرياض^(٤). وتمثل الدرجات العليا من المعايير (المئين ٨٠ فما فوق) المستويات التي حققها الأفراد المتميزون، أي الذين يمثلون العشرون بالمائة الأعلى، بينما يمثل المئين ٥٠ متوسط الأفراد (أي أن نصف المجتمع أدنى من هذا المستوى بينما النصف الآخر أعلى من هذا المستوى)، وتمثل الدرجات الدنيا من المعايير (المئين ٢٠ فما دون) المستويات التي حققها غالبية الأفراد (أي ٨٠% من الأفراد).

جدول رقم (٤): الدرجات المئينية لقوة القبضة بالكيلوجرام لدى الأطفال السعوديين البنين من مدينة الرياض (عدد العينة = حوالي ١٢٠٠ طفلاً).

العمر بالسنوات								الدرجة المئينية
١٣	١٢	١١	١٠	٩	٨	٧	٦	
٢٧,٣	٢٢	١٩	١٧	١٥	١٣	١٠	١٠	٩٥
٢٥	٢٠,٤	١٧,٥	١٥	١٣,٥	١١,٥	٩,٩	٨	٩٠
٢٢	١٨	١٦	١٣,٥	١٢	١٠,٦	٨	٧	٨٠
٢٠	١٦,٥	١٥	١٢	١٠,٥	١٠	٧	٦	٧٠
١٨,٥	١٦	١٤,٥	١١,٥	١٠	٩	٦,٥	٦	٦٠
١٦,٧	١٥	١٣,٥	١٠,٥	٩,٥	٨	٦	٥	٥٠
١٥	١٤	١٣	١٠	٩	٧	٥,٥	٥	٤٠
١٤	١٣	١٢	٩	٧,٥	٦,٥	٥	٥	٣٠
١٣	١٢	١١	٨	٧	٥,٦	٤,٥	٣	٢٠
١٢	١٠	١٠	٧	٦	٥	٣,٥	٣	١٠
١١,٨	٨	٧	٦	٥	٤	٣	٢,٥	٥

المصدر: الهزاع، هزاع. كتاب فسيولوجيا الجهد البدني لدى الأطفال والناشئين، ١٩٩٧م.

وتعد المقاييس المعيارية الطريقة المثلى للتعرف على واقع الصفة محل البحث في المجتمع، فمثلاً لو قمنا بإجراء قياس لنسبة الشحوم في مجتمع الطلاب الجامعيين وأتضح أن المئين ٥٠ لنسبة الشحوم في هذا المجتمع يساوي ٣٢%، فهذا يعني أن نصف الطلاب

الجامعيين لديهم نسبة شحوم أكثر من ٣٢% والنصف الآخر لديه نسبة شحوم أقل من ٣٢%، لكن لو قمنا بقياس نسبة الشحوم فيما بعد لأحد الطلاب وكانت ٢٦%، فهذا لا يعني أن هذا الطالب لديه نسبة شحوم طبيعية نظراً لأنها أقل من المئين ٥٠ للطلاب الجامعيين، فالمقاييس المعيارية لا تعني أنها المقاييس الطبيعية (أو المثالية) بل هي المقاييس الفعلية أو الواقعية، حيث يمكن أن تكون البدانة على سبيل المثال متفشية في هذا المجتمع. من هذا المنطلق جاء إتجاه آخر ينادي باستخدام المقاييس المحكية (Criterion-Referenced standards)، والتي تعني المقاييس المطلوبة أو المرغوبة، فمثلاً ينبغي أن تكون نسبة الشحوم المرغوبة لذلك الشاب الجامعي هي ٢٠% فما دون، بينما مقارنته مع المقاييس المعيارية قد لا تظهر أنه بدين.

يتضح إذاً أن المقاييس المحكية هي مستويات مرغوبة (مطلوبة) ينبغي تحقيقها من قبل الشخص. والمقاييس المحكية تكون موضوعاً مسبقاً، وذلك بناءً على أسس نظرية أو تطبيقية، كأن تكون بناءً على المستوى المرتبط بالصحة والخلو من المرض، كما في حدود السمنة لمؤشر كتلة الجسم لدى الكبار (٣٠ كجم/ متر ٢ أو أكثر)، أو تلك المرتبطة بالإنجاز البدني، وما شابه ذلك، وليست بناءً على المستويات الفعلية لأفراد المجتمع محل الدراسة. على أنه يتم أحياناً تصنيف المقاييس المعيارية بناءً على نسب مئوية كأن يقال من المئين ٨١ - ١٠٠ يعادل ممتاز، ومن المئين ٦١ - ٨٠ يساوي فوق المتوسط، ومن المئين ٤١ - ٦٠ يعني المتوسط، ومن المئين ٢١ - ٤٠ يساوي دون المتوسط، ومن المئين ١ - ٢٠ يساوي ضعيف، أو ما شابه ذلك من تصنيفات.

وعلى الرغم من المميزات الجيدة والمبررات القوية لاستخدام المقاييس المحكية في اختبارات الأداء البدني واللياقة البدنية المرتبطة بالصحة، إلا أن البعض يرى أن استخدامها غير ملائم في بعض الأحيان، حيث يتم وضع تلك المستويات بشكل عشوائي وغير مبني على أسس منطقية، بل أنها توضع في بعض الأحيان بشكل تعجيزي، مما يجعل تحقيقها من قبل الأفراد يصير صعب المنال، الأمر الذي يولد الإحباط لمن لا يتمكن من الوصول إلى هذه المستويات، خاصة بعد محاولات عدة. وكمثال على المقاييس المحكية نورد مستويات الأداء البدني لجري مسافة ميل واحد والذي هو ضمن اختبار اللياقة البدنية المسمى (Fitnessgram)، الصادر من معهد البحوث الهوائية في مدينة دالاس بولاية تكساس الأمريكية، كما هو موضحاً في الجدول رقم (٥)، ويتبين من الجدول أنه يلزم الطفل الذي في عمر ٥ سنوات أن يقطع مسافة الميل جرياً أو مشياً في مدة لا تزيد عن ١٦ دقيقة للبنين أو ١٧ دقيقة للبنات، أما الطفل الذي في عمر عشر سنوات فعليه أن يقطع مسافة الميل في ١٠ دقائق على الأكثر، ويقابل

ذلك لدى الفتيات ١٢ دقيقة، وهكذا لبقية الفئات العمرية الأخرى طبقاً للمستوى المبين في الجدول المذكور (١٦).

جدول رقم (٥): المستويات المحكية في اختبار جري ميل واحد للأطفال ضمن اختبار معهد البحوث الهوائية في ولاية تكساس الأمريكية.

المستويات المحكية (بالدقائق والثواني)		العمر بالسنوات
بنات	بنين	
١٧	١٦	٥
١٦	١٥	٦
١٥	١٤	٧
١٤	١٣	٨
١٣	١٢	٩
١٢	١١	١٠
١٢	١١	١١
١٢	١٠	١٢
١١,٣٠	٩,٣٠	١٣
١٠,٣٠	٨,٣٠	١٤
١٠,٣٠	٨,٣٠	١٥
١٠,٣٠	٨,٣٠	١٦
١٠,٣٠	٨,٣٠	١٧

المصدر: Cureton & Warren, RQES, 1990

” لا شيء يكتنف قياسه الخطأ

كما جسم الإنسان“

“Nothing is measured with greater error than the human body”

Beneke, 1878

المراجع

المراجع العربية:

١. الأحمدى، محمد، وهزاع الهزاع. مصداقية أستبانة قياس مستوى النشاط البدني لدى الشباب من ١٥ - ٢٥ سنة باستخدام قياس حركة الجسم ورصد ضربات القلب. *الدورية السعودية للطب الرياضي*، ١٤٢٤هـ، ٧: ٢-١٤.
٢. الجباري، علي، وخالد المزيني. مصداقية اختبار جري ٢٠م متعدد المراحل واختبار جري/مشي ٦٠٠م للتنبؤ بالاستهلاك الأقصى للأكسجين لدى الشباب من ١٨-٢٥ سنة. *الدورية السعودية للطب الرياضي*، ١٤٢٠هـ، ٤ (١): ٩٤ - ٢٤٤.
٣. الهزاع، هزاع محمد. تجارب معملية في وظائف أعضاء الجهد البدني. الرياض: جامعة الملك سعود، ١٤١٣هـ.
٤. الهزاع، هزاع محمد. فسيولوجيا الجهد البدني لدى الأطفال والناشئين. الرياض: الإتحاد السعودي للطب الرياضي، ١٤١٧هـ.
٥. الهزاع، هزاع محمد. معادلات التنبؤ بنسبة الشحوم في الجسم: تطبيقات على الشباب الجامعي. *مجلة جامعة الملك سعود (العلوم التربوية)*، ١٤١٥هـ، ٧ (١): ١٧-٣٢.
٦. الهزاع، هزاع محمد. العلاقة بين الاستهلاك الأقصى للأكسجين والأداء البدني أثناء جري كل من ٦٠٠ متر و ١٠٠٠ متر لدى الأطفال. *مجلة جامعة الملك سعود (العلوم التربوية)*، ١٤١٥هـ، ٦ (٢): ٢٤٧ - ٢٥٩.
٧. الهزاع، هزاع محمد. الدليل الإرشادي لاختبار اللياقة البدنية المرتبطة بالصحة لدول مجلس التعاون، للفئات العمرية من ٧-١٨ سنة، الرياض: الأمانة العامة لدول مجلس التعاون لدول الخليج العربية، ١٤٢٢هـ/ ٢٠٠١م.
٨. الهزاع، هزاع محمد، محمد الأحمدى. أستبانة قياس مستوى النشاط البدني لدى الشباب من ١٥-٢٥ سنة: تطويرها ومعاملات صدقها وثباتها. *المجلة العربية للغذاء والتغذية*، ٢٠٣، ٤ (٨): ٢٧٩-٢٩١.
٩. الهزاع، هزاع محمد، وخالد المزيني، وسعيد الرفاعي. تحديد نسبة الشحوم في الجسم لدى الطلاب الجامعيين من خلال قياس كل من حجم الجسم وكثافته بواسطة إزاحة الهواء والمقاومة الكهروحيوية وسمك طية الجلد. الرياض: مركز البحوث، كلية التربية، جامعة الملك سعود، ١٤٢٣هـ.

المراجع الإنجليزية:

10. Atkinson G, Nevill A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; 26: 217-238.
11. Bassett D, Ainsworth B, Swartz A, Strath S, O'Brien W, King G. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc* 2000; 32: S 471- S 480.
12. Bedogni G, Malavolti M, Severi S, Poli M, Mussi C, Fantuzzi A, Battistini N,. Accuracy of an eight-polar bioelectrical impedance method in the assessment of total body water. *Eur J Clin Nutr* 2002; 56: 1143-1148.

13. Bland J, Altman D. Measuring agreement in method comparison studies. *Statist Methods Med Res* 1999; 8: 135-160.
14. Burke E. Validity of selected laboratory and field tests of physical working capacity. *Res Q Sport Exerc* 1976; 47: 95-104.
15. Cohen J. A power primer. *Psychol Bulletin* 1992; 112: 155-159.
16. Cureton K, Warren G. Criterion-referenced standards for youth health-related fitness tests: A tutorial. *Res Quart Exerc Sports* 1990; 61: 7-19.
17. Docherty D. (Ed.). *Measurement in Pediatric Exercise Science*. Champaign, IL: Human Kinetics Publishers, 1996.
18. Durant R, Baranowski T, Davis H, Rhodes T, Thompson W, Greaves K, Puhl J. Reliability and variability of indicators of heart rate monitoring in children. *Med Sci Sports Exerc* 1992; 24: 389-395.
19. Gore C. (Ed.): *Physiological Tests for Elite Athletes*. Australian Sports Commission. Champaign, IL: Human Kinetics, 2000.
20. Goto R, Mascie-Taylor C. Precision of measurement as a component of human variation. *J Physiol Anthropol* 2007; 26: 253-256.
21. Latin R, Berg K, Smith P, Tolle R, Woodby-Brown S. Validation of cycle ergometry equation for predicting steady-rate VO₂. *Med Sci Sports Exerc* 1993; 25: 970-974.
22. Leger L, Thivierge M. Heart rate monitors: validity, stability, and functionality. *Phys Sportsmed* 1988; 16 (5): 143-151.
23. Lohman, T. Skinfold and body density and their relation to body fatness: a review. *Human Biology* 1981; 53: 181-225.
24. Mathews D. *Measurement in Physical Education*. Philadelphia: W. B. Saunders Company, 1978.
25. MacDougall J, Wenger H., Green H. (Eds.) *Physiological Testing of the High – Performance Athlete*. Champaign, IL: Human Kinetics Publishers. 1990.
26. Morrow J, Jackson A. How significant is your reliability? *Res Q Exerc Sport* 1993; 64: 352-355.
27. Shultz B, Sands W. Understanding measurement concept and statistical procedures. In: *Physiological Assessment of Human Fitness*, Maud P, Foster C. (Eds.). Champaign, IL: Human Kinetics Publishers, 1995: 257-287.
28. Thomas J, Nelson J. *Research Methods in Physical Activity*. Campaign, IL: Human Kinetics, 1996.
29. Ulijaszek S, Kerr D. Anthropometric measurement error and the assessment of nutritional status. *Br J Nutr* 1999; 82: 165-177.
30. Vincent WJ. *Statistics in Kinesiology*. Champaign, IL: Human Kinetics, 1999.
31. Winter E, Eston R, Lamb K. Statistical analysis in the physiology of exercise and kinanthropometry. *J Sports Sci* 2001; 19: 761-775.