# بعض المعادلات الإحصائية المستخدمة
# في تحديد عينة البحث

المصدر:

**الهزاع، هزاع محمد. فسيولوجيا الجهد البدني: الأسس النظرية والإجراءات المعملية للقياسات الفسيولوجية. الفصل الرابع. كتاب مقدم للنشر.**

## بعض المعادلات الإحصائية المستخدمة في تحديد عينة البحث

عرفنا فيما سبق أهمية أن يخطط الباحث جيداً قبل بدءه إجراء البحث، من أجل أن يتمكن من دحض الفرضية الصفرية (أي أن يجد الباحث على سبيل المثال فرقاً معنوياً ذو قيمة عملية بين المجموعة التجريبية والضابطة). ولكي يحصل ذلك فمن الضروري حساب عدد أفراد عينة البحث المطلوبين بشكل دقيق، بحيث لا تكون تلك العينة صغيرة (أقل مما ينبغي) فيحصل خطأ من نوع بيتا (نوع ٢)، أي لا يجد الباحث فرقاً ذا دلالة إحصائية بين المجموعتين التجريبية والضابطة بسبب عدم كفاية العينة، وأن لا تكون العينة كبيرة جداً، مما يجعل الباحث يكتشف فرقاً ذا دلالة إحصائية لكنه فرقاً ضئيلاً من الناحية العملية، ناهيك عن زيادة كلفة البحث، مع العلم أنه من الصعوبة بمكان الحصول على عينات كبيرة الحجم في بعض البحوث التجريبية.

ولتحديد عينة البحث المطلوبة (**n**)، يلزمنا أن نعرف المتغيرات التالية:

- التباين (**Variance**) - أو مربع الانحراف المعياري ($S^2$) - للمتغير قيد الدراسة، وعادة يمكن الحصول عليه من دراسات سابقة، أو ينبغي إجراء دراسة استطلاعية لتحديده.
- مقدار درجة زي ألفا ($Z_\alpha$) وهي عند ألفا = ٠,٠٥ تساوي ١,٩٦ (عند مستوى ثقة = ٩٥% في الاختبار ذي الذيلين)، وعند ألفا مقدارها ٠,٠١ فهي تساوي ٢,٥٨.
- مقدار درجة زي بيتا ($Z_\beta$) وعادة ما يستخدم مقدار لبيتا = ٢٠% وتساوي ٠,٨٤ (عند قدرة = ٨٠% في الاختبار ذي الذيل الواحد).
- مقدار الفرق المطلوب الحصول عليه ($d$) في المتغير قيد الدراسة (أي مقدار الفرق المحسوس عملياً).

وعادة ما نستخدم المعادلات التالية تبعاً للتصميم الإحصائي المستعمل:

١. في الدراسات التي تستخدم اختبار - ت للعينات المزاوجة (**Paired t-test**)، (مثل: إجراء اختبار قبلي وبعدي لدى المجموعة نفسها) مع الأخذ بالحسبان قيمة خطأ ألفا فقط:

$$\mathbf{n} = \frac{(Z_\alpha)^2 * (S)^2}{(d)^2}$$

علماً بأن (*) هي علامة ضرب.

**مثال**: **دراسة تأثير برنامج نشاط بدني على نسبة الشحوم في الجسم لدى عينة من الأشخاص البدناء، علماً بأن الانحراف المعياري لنسبة الشحوم من دراسات سابقة على البدناء يساوي ١٠%، ومقدار الفرق الذي نريد أن نكتشفه هو ٥% فأكثر. بتطبيق المعادلة السابقة سيكون الجواب كالتالي:**

$$\mathbf{n} = \frac{(1.96)^2 * (10)^2}{(5)^2}$$

$$\mathbf{n} = \frac{(3.84) * (100)}{(25)} = 15.36$$

**أي أن العدد المطلوب هو ١٦ شخص. لكن لاحظ أن التصميم البحث الأمثل في مثل هذه الحالة ينبغي أن يتضمن مجموعة تجريبية ومجموعة ضابطة، وبالتالي فإن المعادلتين رقم (٣) ورقم (٤) هي الأكثر ملائمة في هذا الصدد.**

٢. في الدراسات التي تستخدم اختبار - ت للعينات المزاوجة (**Paired t-test**) مع الأخذ بالحسبان قيمة خطأ ألفا وقيمة خطأ بيتا:

$$\mathbf{n} = \frac{(Z_\alpha + Z_\beta)^2 * (S)^2}{(d)^2}$$

**مثال**: **لو أخذنا المثال السابق، فإننا بتطبيق المعادلة رقم (٢) سنحصل على التالي:**

$$\mathbf{n} = \frac{(1.96 + 0.84)^2 * (10)^2}{(5)^2}$$

$$\mathbf{n} = \frac{(7.84) * (100)}{(25)} = 31.36$$

**أي أن العدد المطلوب هو ٣٢ شخص، أي ضعف العدد الذي حصلنا عليه في المعادلة السابقة، وهذا العدد يجعلنا مطمئنين أكثر لأننا أخذنا في الاعتبار احتمالات حدوث خطأ من نوع بيتا.**

٣. في الدراسات التي تستخدم اختبار - ت للعينات المستقلة (**Independent t-test**) مع الأخذ بالحسبان قيمة خطأ ألفا فقط:

$$\mathbf{n} = \frac{(Z_\alpha)^2 * 2 * (S)^2}{(d)^2}$$

**والملاحظ أننا نحتاج في هذا النوع من المعادلات ضعف العدد الـذي نحـصل عليـه فـي المعادلة رقم (١)، نظراً لأننا نضرب الأرقام في العدد ٢ كما هو موضحاً في المعادلة، وذلك لأننا نتعامل مع عينيتين مستقلتين.**

٤. في الدراسات التي تستخدم اختبار - ت للعينات المـستقلة (**Independent t-test**) مـع الأخذ بالحسبان قيمة خطأ ألفا وقيمة خطأ بيتا:

$$\mathbf{n} = \frac{(Z_\alpha + Z_\beta)^2 * 2 * (S)^2}{(d)^2}$$

٥. في الدراسات التي تستخدم اختبار الفروق في النسبة المئوية (**Proportion**) مع الأخـذ بالحسبان قيمة خطأ ألفا وقيمة خطأ بيتا:

$$\mathbf{n} = \frac{(Z_\alpha + Z_\beta)^2 * 2 * p\,(1 - P)}{(d)^2}$$

**حيث يمثل الرمز (*P*) النسبة المئوية. ونستخدم هذه المعادلة في الدراسات التي يتم فيهـا إجراء تدخل (Intervention) لتغيير سلوك أو مخاطر صحية (مثل حث النـاس علـى ممارسـة النشاط البدني)، حيث تتم مقارنة احتمالات النجاح والفشل (أي نجاح التغيير من عدمه)، فعنـد احتساب النسبة المئوية المتوقعة للنجاح على أساس ٦٠% ، فإن ذلك يعني أن النسبة الأخـرى المتبقية هي ٤٠% (١ - ٠,٦٠ = ٠,٤٠)، وبالتالي فإن الفرق (*d*) هـو ٢٠% (أي ٦٠ % - ٤٠ %)، علماً بأن النسبة ٥٠% تعني أنه لا يوجد فرق. وكلما كان الفرق المتوقع (*d*) أصـغر (مثلاً: ٥٥% - ٤٥%، أي ١٠%) كان العدد المطلوب من العينة أكبر.**

٦. في الدراسات المسحية التي نريد من خلالها معرفة بيانات وصفية، كمـا فـي حالـة اختبارات اللياقة البدنيـة علـى طـلاب المـدارس أو إجـراء القياسـات الجـسمية (الأنثروبومترية)، يمكن تحديد حجم العينة باستعمال المعادلة التالية:

$$\mathbf{n} = \frac{((Z * \sigma)^2 / (e)^2)}{1 + ((1/N) * ((Z * \sigma)^2 / (e)^2)}$$

**n = حجم العينة المطلوبة**

**N = Population size حجم المجتمع**

**Z = Confidence level, at 95% = 1.960 (مستوى الثقة عند ٩٥%)**

**σ = Standard deviation of the population الانحراف المعياري للمتغير قيد البحث على مستوى المجتمع**
**ويتم أحياناً تقديره من متوسط الانحراف المعياري المتوقع للمتغيرات قيد الدراسة في حالة وجود أكثر من متغير**

e = **the maximum acceptable error as fraction of the standard deviation** **مقدار الخطأ المقبول كجزء من الانحراف المعياري (مثلاً نصف انحراف معياري)**

**علامة قسمة = / ; علامة ضرب = ***

## حساب حدود الثقة (Confidence Limits)

في أحيان كثيرة، يكون من المستحسن - بالإضافة إلى حساب المتوسط والانحراف المعياري للبيانات (**Point estimate**) - القيام بحساب الحدود الدنيا والعليا للثقة (**U & L limits of Confidence**) عند مستوى ٩٥% أو ٩٩%، وذلك على النحو التالي:

- الحدود العليا للثقة عند مستوى ٩٥% (**UL 95% CI**):

**UL 95% CI = M + Zc * SEM**

- الحدود الدنيا للثقة عند مستوى ٩٥% (**LL 95% CI**):

**LL 95% CI = M - Zc * SEM**

حيث: **M = المتوسط الحسابي**

**Zc = درجة زي عند مستوى ٩٥% من الثقة وتساوي ١,٩٦ (وهي تساوي ٢,٥٨ عند مستوى ٩٩% من الثقة 99% CI) .**

**SEM = الانحراف المعياري للمتوسط، ويساوي الانحراف المعياري مقسوماً على الجذر التربيعي لعدد أفراد العينة.**

**مثال:** إذا كان المتوسط الحسابي = ١٠٠ ، والانحراف المعياري = ١٥ ، وعدد أفراد العينة - ٢٥ شخص (يكون الانحراف المعياري للمتوسط = ١٥ ÷ الجذر التربيعي لـ ٢٥ = ٣)، فتكون كل من الحدود العليا والدنيا للثقة عند مستوى ٩٥% كالتالي:

- الحدود العليا للثقة عند مستوى ٩٥% (**UL 95% CI**):

**UL 95% CI = M + Zc * SEM**

**UL 95% CI = 100 + 1.96 * 3 = 105.9**

- الحدود الدنيا للثقة عند مستوى ٩٥% (**LL 95% CI**):

**LL 95% CI = M - Zc * SEM**

**LL 95% CI = 100 - 1.96 * 3 = 94.1.**

أي أن متوسط المتغير قيد الدراسة سيتراوح لدى عينة المجتمع من ٩٤,١ إلى ١٠٥,٩ في ٩٥% من الحالات. أو أننا واثقين بنسبة ٩٥% أن متوسط المجتمع في المتغير قيد الدراسة سيتراوح ما بين ٩٤,١ و ١٠٥,٩.

**ملحوظة**: في حالة زيادة عدد أفراد العينة إلى ١٠٠ شخص (بدلاً من ٢٥ شخصاً)، فإن مدى الثقة (**Confidence range**) سيصبح أصغر، كما يلي:

سيصبح الانحراف المعياري للمتوسط (**SEM**) = ١٥ ÷ الجذر التربيعي لـــ ١٠٠ = ١,٥

الحدود العليا للثقة = ١٠٢,٩ ، الحدود الدنيا للثقة = ٩٧,١

أي أن حدود الثقة أصبحت ضيقة مقارنة بالمثال السابق.