كيفية اختيار العينات المناسبة

* تعريف المجتمع والعينة

المجتمع Population هو مجموعة كل الافراد موضوع الاهتمام في دراسة معينة. وهذا يعني ان المجتمع يشمل كل الأعضاء او ا لعناصر او الصفات لمجموعة معينة. والمجتمع قد يكون كبيرا او صغيرا، متل ذلك كل الرجال في العالم، كل الرجال الذين لهم الحق في التصويت في بلد ما، كل التلاميذ بمدرسة معينة خلال العام الدراسي الحالي، جميع الطلاب بالسنوات النهائية بجامعة محمد لمين دباغين خلال عام جامعي معين.

وعندما يكون حجم المجتمع كبيرا جدا، يكون من الصعب بل من المستحيل على الباحث ان يدرس كل افراد المجتمع موضوع الاهتمام. ولذلك ينتقي الباحث مجموعة أصغر من المجتمع وتسمى العينة Sample بشرط ان تكون ممثلة لذلك المجتمع. وتسمى المؤشرات او المقاييس الاحصائية لخصائص المجتمع بالبارامترات Parameters، بينما تسمى المؤشرات او المقاييس الاحصائية لخصائص العينة بالإحصاءات Statistics.

وللتمييز بين المقاييس الاحصائية الوصفية بالمجتمع وتلك المناظرة لها بالعينة، يرمز لبارمترات المجتمع ولإحصائيات العينة بحروف اغريقية وبحروف رومانية على التوالي مثال ذلك:

المقاييس الاحصائية	الرمز المستخدم بالمجتمع	الرمز المستخدم بالعينة
المتوسط	μ	$\bar{x}$
التباين	σ^2	s^2
الانحراف المعياري	σ	s

ملاحظة:

• على الباحث ان يعرف مجتمع الدراسة تعريفا دقيقا.

• على الباحث ان يميز بين مجتمع الافراد والمجتمع الاحصائي ، فمجتمع الافراد يتكون من الافراد الذين يتم دراستهم ، اما المجتمع الاحصائي فهو يتكون من القياسات او الدرجات التي يتم الحصول عليها من هؤلاء الافراد، وبالمثل يتم التمييز بين عينة الافراد والعينة الاحصائية ، حيث تتكون العينة من الافراد الذين تم اختيارهم من مجتمع الدراسة ، اما العينة الاحصائية فتتكون من الدرجات التي يتم الحصول عليها من عينة الافراد. مع التذكير، انه كلما زاد حجم العينة كلما كان احصاء العينة المتمثل في المتوسط الحسابي اكتر دقة في تقدير بارامتر المجتمع U.

* المعاينةSampling :

وهي عملية اختيار العينات والوحدات على سبيل المثال (الاشخاص، المنظمات...) من خلال المجتمعات موضوع الدراسة. والفكرة الأساسية من وراء هذا المفهوم هي اننا نرغب في تقدير خصائص مجموعة كبيرة من الأشخاص (المجتمع) من خلال خصائص مجموعة صغيرة من الأشخاص (العينة)، ومن ثم تطوير القدرة على التعميم.

كما يوجد هناك نهجين عامين عادة ما يستخدمان في بحوث العلوم الاجتماعية وهما:

> العينات العشوائية الاحتمالية، بحيث يكون لدى جميع الأفراد نفس الفرصة ليكونوا ضمن الدراسة انطلاقا من احتمالات رياضية يتم حسابها وتقديرها.

> العينات غير الاحتمالية، بحيث يتم اختيار الأشخاص على اساس توافرهم (على سبيل المثال المتطوعين) أو بسبب حكم الباحث الشخصي من حيث تمثيلية العينة للمجتمع. والنتيجة المتوقعة من هذا الاجراء تكمن في أن هناك جزء غير معروف من الأشخاص يتم ابتعاده من التطوع.

* تقدير العينات:

يتم تقدير العينات في المقام الأول وفقا للإجراءات التي يتم اختيارها بدلا من تكويناتها النهائية من حيث الحجم. (على سبيل المثال يكون من المستحيل على الباحث معرفة ما إذا كانت العينات غير الاحتمالية تمثيلية حتى ولو كانت خصائص العينات (نسب النساء على الرجال، نسب الناجحين على الفاشلين، نسب السود على البيض) متشابهة بشكل موثوق. ومثل هذه الحالة من شأنها أن يؤدي الى التحيز في النتائج المحصل عليها لأن المجتمع موضوع الاهتمام عادة ما يكون غير معروف. وبالمثل قد تكون النتائج المحصل عليها باستخدام العينات العشوائية أو الاحتمالية غير دقيقة لأسباب عدة، أي أن محاولة استخدام اجراءات العينات الاحتمالية تعد امرا ضروريا لكنها غير تكون عير كافية للحصول على النتائج التي يمكننا تعميمها على جميع أشخاص المجتمع، بحيث تتجلى احدى الاهتمامات الرئيسية حول العينة في الاهتمام فيما إذا كانت نسبة الاجابة على أدوات جمع البيانات كبيرة بما فيه الكفاية، والصيغة المتداول عليها في ذلك هي:

نسبة الاجابة=عدد الأفراد الدين اجابوا بشكل فعلي على ادوات جمع البيانات/العدد الكلي للعينة

مع التذكير الى أنه كلما كانت نسبة الاجابة ضعيفة كلما زاد احتمال ارتكاب الخطأ من النوع الثاني أو خطأ بيتا (ß) وهو قبول الفرض الصفري وهو خاطئ في الواقع.

*خطأ المعاينة Sampling Error:

وهو انحراف العينة المختارة عن الصفات، السمات، السلوكيات والأشكال الحقيقية للمجتمع وذلك نتيجة عدم تمثيل العينة للمجتمع، بحيث أنه:

< كلما زاد حجم العينة كلما كان احصاء $\bar{x}$ أكثر دقة في تقدير بارامتر المجتمع U، مما يؤدي الى القول أن هناك علاقة بين حجم العينة وخطأ المعاينة (كلما زاد حجم العينة كلما قل خطأ المعاينة كذلك، وبالتالي تطوير القدرة على التعميم)

< كلما زاد حجم العينة كلما زاد احصاء الانحراف المعياري للعينة S أكثر دقة في تقدير بارامتر المجتمع σ، مما يؤدي بنا الى القول أن هناك علاقة بين الانحراف العياري للعينة والخطأ المعياري (كلما انخفض الانحراف المعياري S كلما قل الاختلاف بين خصائص العينة وخصائص المجتمع).

*هامش الخطأ Margin Error: وهو الحد الأقصى المتوقع للفرق بين بارامتر المجتمع U وتقدير إحصاءه العينة $\bar{x}$ لذلك البارامتر. كما يجب على هامش الخطأ أن يكون مؤهلا من قبل بيان احصائي غالبا ما يعبر عنه بمستوى الثقة Confidence Interval. على سبيل المثال لو يتم تقرير في دراسة معينة أن 51% من الموظفين يتجهون الى تفضيل التدريب على المهارات الابداعية بهامش خطأ ±3 نقاط وبفاصل أو بمستوى ثقة 95%، فهذا يعني لو أنه لو يتم اجراء الدراسة نفسها على عينة أخرى ، فمن المرجح أن يكون تفضيل الموظفين لهذا النوع من التدريب بمستوى ثقة 95% متراوحا ما بين 48% و54%.

*القوة الإحصائية وحجم العينة:

يمكن تعريف القوة الاحصائية باحتمال رفض الفرض الصفري فحين أن الفرضية البديلة صحيحة. والعوامل المؤثرة في القوة الاحصائية تتجلى في كل من حجم العينة، حجم التأثير (الفرق المتواجد بين الفرض الصفري والفرض البديل، ومستوى الدلالة ($\propto$) بما في ذلك توزيع المعلمة المراد تقديرها. وعليه، فاذا كان الباحث على علم أو متأكد من أن الاحصاءات المحصل عليها في الدراسة تتبع التوزيع الطبيعي أو القياسي (Z)، فهناك نوعين من المعلمات تحتاج الى التقدير من قبل الباحث وهما: متوسط المجتمع U وتباين المجتمع (σ^2). وفي معظم الأوقات يكون الباحث متأكد من احدى المعلمين فقط (U أوσ^2) ومن ثم فهو في حاجة الى تقدير المعلم الآخر. فمثلا، إذا كان الباحث يرغب في معرفة متوسط المجتمع بإمكانه تقدير ذلك باستخدام توزيع العينة والذي ينتهي باستخدام توزيع "ت". كما يتم تقدير القوة الاحصائية للاختبار الاحصائي بشكل عام لغرضين هما:

> يمكن تقدير القوة الاحصائية قبل جمع البيانات استنادا الى معلومات من الدراسات السابقة يتم من خلالها تحديد أو تقدير حجم العينة المطلوب في الدراسة الحالية وذلك باستخدام كل من مستوى الدلالة ($\propto$)، القوة الاحصائية التي يكون الباحث في حاجة اليها، وحجم التأثير، وهذا ما يسمى **A priori**

> كما يمكن تقدير القوة الاحصائية بعد تحليل البيانات. ومثل هذه الحالة عادة ما تحدث عندما تكون النتائج المحصل عليها غير دالة احصائيا، مما يؤدي بالباحث الى حساب القوة الاحصائية للتأكد ما إذا كانت هذه النتائج بسبب العينة أو بسبب عدم وجود قوة احصائية بشكل فعلي وهذا ما يسمى ب **Post Hoc.**

ملاحظة: ان الغرض الأول هو الأكثر أهمية عندما يتعلق الأمر بتحديد حجم العينة المناسب، مع التذكير أنه يوجد هناك برنامج احصائي جد فعال في هذا الشأن وهو يستخدم بشكل مكثف من قبل الباحثين تمت تسميته بـG*Power وهو متوفر على مستوى المواقع الإلكترونية.

* تقدير حجم العينة المناسبة للاختبارات الاحصائية:

> تقدير حجم العينة لعينة واحدة:

في الدراسات التي يتجلى فيها الهدف في تقدير متوسط متغير متصل لمجموعة واحدة مثل ذلك:

$$t = \frac{\bar{x} - \Delta}{\frac{s}{\sqrt{n}}}$$

حيث:

$\bar{x}$: متوسط العينة

Δ : متوسط المجتمع U

s : الانحراف المعياري لمتوسط العينة

n : حجم العينة

فان الصيغة المعمول بها في تحديد حجم العينة الأدنى هي كالتالي:

$$n = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

حيث:

n : حجم العينة الأدنى المطلوب

$Z1-(\propto)/2$: القيمة الخرجة لتوزيع Z القياسي وهي تساوي (1.96) عند مستوى الدلالة 0.025=2/0.05

$Z1-ß$: قيمة القوة الاحصائية من التوزيع الطبيعي التي تحمل $1-ß$ تحتها والتي يرغب فيها الباحث لاجتناب الوقوع في خطأ ß

ES : حجم التأثير، هو يساوي الفرق المتواجد بين متوسط العينة ومتوسط المجتمع على الانحراف المعياري. والصيغة المستخدمة في ذلك هي كالتالي:

$$ES = \frac{|\mu_1 - \mu_0|}{\sigma}$$

مثال: اراد باحث في السلوك التنظيمي تقدير متوسط الكفاءة الذاتية لدى العمال ذوي تقدير الذات المرتفع، وذلك باستخدام فاصل ثقة بنسبة 95% وقوة احصائية تقدر ب 0.8. كما تجدر الاشارة، الى أن حجم التأثير غير معروف مما أدى بالباحث الى استخدام حجم تأثير مرتفع قدر ب 0.50 وهذا وفقا لما اوصى به الباحث Cohen في هذا الشأن.

المطلوب: حدد حجم العينة الأدنى والمناسب.

وبناء على ذلك يكمن الحل فيما يلي:

-القيام بكشف قيمة Z المعيارية المناظرة للقوة الاحصائية تساوي 0.80

باستخدام جدول التوريع المعياري وبعد التحري على أقرب قيمة ممكنة الى 0.80 داخل التوزيع تبين أنها تساوي 0.7995، وهي تتقاطع مع Z=0.04 وZ=0.8، ومن ثم فهي تساوي 0.84

ومنه:

$$\boldsymbol{n} = (\mathbf{1.96} + \mathbf{0.84}|\mathbf{0.5})^2$$

$$\boldsymbol{n} = (\mathbf{2.80}|\mathbf{0.5})^2$$

$$\boldsymbol{n} = [\mathbf{5.6}^2]$$

$$\boldsymbol{n} = \mathbf{31.36}$$

وبناء على ذلك، حتى يتمكن الباحث من ضمان فاصل ثقة بنسبة 95%، وتقدير متوسط الكفاءة الذاتية لدى العمال ذوي تقدير الذات المرتفع بالشكل المطلوب، يجب عليه استخدام على الأقل 31 مشاركا وهذا حتى يتمكن من اجتناب الوقوع في الخطأ من النوع الثاني.

> تقدير حجم العينة لعينتين مترابطتين:

في الدراسات التي تسعى خطتها الى تقدير متوسط الفرق المركز على البيانات المترابطة، مثل ذلك:

$$t = \frac{\overline{d}}{S_d / \sqrt{n}}$$

حيث:

$\bar{d}$: متوسط الفرق الملاحظ بين القياس القبلي والقياس البعدي

Sd: الانحراف المعياري للفرق الملاحظ بين القياس القبلي والقياس البعدي

n : حجم أزواج الملاحظات

فان الصيغة المعمول بها في تحديد حجم العينة الأدنى المطلوب هي كالتالي:

$$n = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

حيث:

n : حجم العينة

Z1−(∝)/2 : القيمة الخرجة لتوزيع Z القياسي (1.96) عند مستوى 0.05

Z1− ß : قيمة القوة الاحصائية التي يرغب فيها الباحث لاجتناب الوقوع في خطأ ß

ES : حجم التأثير والصيغة المستخدمة في ذلك هي كالتالي :

$$ES = Ud/\theta d$$

Ud هو متوسط الفرق بين القياس القبلي والقياس البعدي و θd هو الفرق في الانحراف المعياري بين القياس القبلي والقياس البعدي.

مثال: على اعتبار أن حجم التأثير في الدراسة المراد اجراؤها حول مدى تأثير البرنامج التدريبي في الأداء المهني وذلك بقياس قبلي وقياس بعدي يساوي 0.30. ومن هذ المنطلق، فاذا اراد الباحث استخدام فاصل

ثقة بنسبة 95% وقوة احصائية تقدر ب 0.9، احسب حجم العينة الأدنى والمطلوب لاجتناب الخطأ من النوع الثاني.

وبناء على ذلك، من خلال القيام باستخدام جدول التوريع المعياري والتحري على أقرب قيمة ممكنة الى 0.90 داخل التوزيع تبين أنها تساوي 0.8997، وهي تتقاطع مع Z=0.08 وZ=1.2، ومن ثم فهي تساوي 1.28.

ومنه:

$$\boldsymbol{n} = (\mathbf{1.96} + \mathbf{1.28}|\mathbf{0.3})^2$$

$$\boldsymbol{n} = (\mathbf{3.24}|\mathbf{0.3})^2$$

$$\boldsymbol{n} = \lfloor \mathbf{10.80}^2 \rfloor$$

$$\boldsymbol{n} = \mathbf{116.64}$$

وبناء على ذلك، حتى يتمكن الباحث من ضمان فاصل ثقة بنسبة 95% وتقدير بالشكل المطلوب متوسط تأثير البرنامج التدريبي في الأداء المهني يجب عليه استخدام على الأقل 117 مشاركا وهذا حتى يتمكن من اجتناب الوقوع في الخطأ من النوع الثاني.

> تقدير حجم العينة لعينتين مستقلتين:

في الدراسات التي تسعى خطتها الى تقدير الفرق بين متوسطي متغير متصل بمجموعتين مستقلتين مثل ذلك،

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_p^2(\frac{1}{N_1} + \frac{1}{N_2})}}$$

فان الصيغة المعمول بها في تحديد حجم العينة الأدنى هي كالتالي:

$$n_i = 2\left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES}\right)^2$$

ni : حجم العينة في كل مجموعة من المجموعتين

Z1−(∝)/2 : القيمة الحرجة لتوزيع Z القياسي (1.96) عند مستوى 0.05

Z1− ß : قيمة القوة الاحصائية التي يرغب فيها الباحث لاجتناب الوقوع في خطأ ß

ES : حجم التأثير والصيغة المستخدمة في ذلك هي كالتالي :

$$ES = \frac{|\mu_1 - \mu_2|}{\sigma}$$

حيث:

$|U1-U2|$ هو الفرق المطلق بين متوسط المجموعة الأولى ومتوسط المجموعة الثانية

σ هو الانحراف المعياري المجمع، والصيغة المستخدمة في ذلك هي كالتالي:

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

مثال: على اعتبار أن حجم التأثير في الدراسة المراد اجراؤها حول مدى تأثير البرنامج التدريبي في الأداء المهني بالنسبة للذكور والاناث هو 0.26. ومن هذ المنطلق، فاذا اراد الباحث استخدام فاصل ثقة بنسبة 95% وقوة احصائية تقدر ب0.80، احسب حجم العينة الأدنى والمطلوب في كل مجموعة من المجموعتين ومن ثم اجتناب الخطأ من النوع الثاني.

وفي هذا الصدد، ومن خلال استخدام جدول التوزيع المعياري وبعد التحري على أقرب قيمة ممكنة الى 0.80 داخل التوزيع تبين أنها تساوي 0.7995، وهي تتقاطع مع Z=0.04 وZ=0.8، ومن ثم فهي تساوي 0.84

ومنه:

$$ni = 2(1.96 + 0.84|0.26)^2$$

$$(2.8/0.26)^2 ni = 2$$

$$ni = 2(10.769)^2$$

$$ni = 231.95$$

وفي ضوء هذه النتيجة، فلكي يتمكن الباحث من اجتناب الخطاء من النوع الثاني ومن ثم ضمان فاصل ثقة بنسبة 95% من حيث أن الاختبار الاحصائي المستخدم تكون لديه قوة احصائية تساوي 80% لاكتشاف الفرق الحقيقي المتواجد في الواقع بين متوسطي الذكور والاناث في الأداء المهني وفقا للبرنامج التدريبي يجب عليه استخدام على الأقل 232 مشارك في مجموعة الذكور والحجم ذاته في مجموعة الاناث.

جدول التوزيع الطبيعي

Standard Normal Distribution Table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	**.7995**	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	**.8997**	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964

2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998