المحاضرة السادسة في الإحصاء الاستدلالي: المعامل الزادي(Z-test)

يعد اختبار z أحد أنواع الاختبارات الإحصائية المعلمية ويتم استخدامه في الحالات التالية:

1) اختبار Z للنسب:

- **اختبار Z لعينة واحدة للنسب:** يهدف هذا الاختبار إلى تحديد ما إذاكان هناك فرق ذو دلالة إحصائية بين نسبة العينة والقيمة المفترضة لنسبة المجتمع. أي انه يستخدم لتقييم الفرق بين نسبة عينة واحدة (نسبة معينة وجدت في هينة من البيانات) وقيمة مفترضة لنسبة المجتمع (محددة مسبقا نفترض انها تمثل النسبة الحقيقية للمجتمع)) $\hat{p} - P_0$. ولتحقيق ذلك يشترط ان تكون:

- البيانات المستخدمة لحساب $\hat{p}$ عشوائية وغير متحيزة أي أن كل فرد في العينة لديه نفس فرصة الاختيار، وأن كل مشاهدة مستقلة عن الأخرى
- ان يكون حجم المجتمع N كبيرا جدا مقارنة بحجم العينة n، وهذا يضمن أن سحب عينة لن يؤثر بشكل كبير على نسبة المجتمع. لذلك يمكن افتراض ان قيمة نسبة المجتمع Pثابتة

يتم حساب Zوفق المعادلة:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

حيث: $\hat{p}$ تمثل النسبة المحسوبة من العينة (وتحسب بقسمة عدد افراد العينة حاملي الصفة على حجم العينة)

P_0 تمثل النسبة المفترضة في المجتمع

n حجم العينة

$\sqrt{P_0(1-P_0)/n}$ يمثل الانحراف المعياري لتوزيع نسب العينة

وتكون صياغة الفرضيات كالتالي:

- One-tailed test
  - $H_0: \hat{P} = P_0$
  - $H_1: \hat{P} < P_0$ or $H_1: \hat{P} > P_0$
- Two-tailed test
  - $H_0: \hat{P} = P_0$
  - $H_1: \hat{P} \neq P_0$

مثال :لنفترض أننا أجرينا استطلاعًا لـ 100 شخص وسألناهم عما إذاكانوا يفضلون الذهاب الى الصالة الرياضية صباحا او مساءا. يمكننا استخدام هذا الاختبار لمقارنة النسبة التي حصلنا عليها في عينتنا (على سبيل المثال، 60% يفضلون ممارسة الرياضة صباحا) مقارنة بالنسبة التي كنا نتوقعها في المجتمع بأكمله (على سبيل المثال، 50%).

الحل:

- تحديد المشكلة:

هل النسبة التي تحصلنا عليها من عينة من 100 شخص يفضلون ممارسة الرياضة صباحًا (60%) تختلف بشكل كبير عن النسبة التي نتوقعها في المجتمع بشكل عام (50%) ؟ (نختار مستوى دلالة $\alpha = 0.05$)

- صياغة الفرضيات:
  - فرضية العدم :(H_0) النسبة الحقيقية للأشخاص الذين يفضلون ممارسة الرياضة صباحًا في المجتمع تساوي 50% . ($p = 0.5$)
  - فرضية البديل :(H_1) النسبة الحقيقية للأشخاص الذين يفضلون ممارسة الرياضة صباحًا في المجتمع لا تساوي 50 % . ($p \neq 0.5$)

- حساب Z لدينا: $\hat{P}$ النسبة المقاسة في العينة = 0.6

P_0 النسبة المتوقعة في المجتمع = 0.5

n حجم العينة = 100

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

بالتعويض بالقيم:

$Z = (0.6 - 0.5) / \sqrt{(0.5 * (1-0.5) / 100)} = 2$

تذكير بالقيم الجدولية لاختبار Z

مستوى الدلالة	اختبار احادي الطرف	اختبار ثنائي الطرف
$\boldsymbol{\alpha = 0.05}$	1.65	1.96
$\boldsymbol{\alpha = 0.01}$	2.33	2.58

- تحديد القيمة الحرجة: بما ان الاختبار ثنائي الطرف، باستخدام جدول التوزيع الطبيعي القياسي نجد أن القيمة الحرجة لـ α = 0.05 هي ±1.96.
- مقارنة القيمة المحسوبة بالقيمة الحرجة :قيمة Z المحسوبة (2) أكبر من القيمة الحرجة (1.96)
- **اتخاذ القرار**: بما أن قيمة Z المحسوبة تقع في منطقة الرفض، فإننا **نرفض فرضية العدم**. وبالتالي يمكننا القول أن هناك فرقًا ذو دلالة إحصائية بين النسبة التي حصلنا عليها في العينة (60%) والنسبة المتوقعة في المجتمع (50%)، أي أن النسبة الحقيقية للأشخاص الذين يفضلون ممارسة الرياضة صباحًا في المجتمع تختلف عن 50%

اختبار Z لعينتين للنسب: يستخدم هذا الاختبار لمقارنة نسبتين من عينتين مستقلتين حجمهما n_1 وn_2 (أي لتقييم الفرق بين نسبتي عينتين $(\hat{p}_1 - \hat{p}_2)$ كدليل لاختبار الفرضية الصفرية التي تنص على عدم وجود فرق بين نسبتي المجتمعين$(P_1 - P_2)$ .. الهدف هو معرفة ما إذاكان هناك فرق ذو دلالة إحصائية بين هاتين النسبتين، أي ما إذاكان الفرق الذي نلاحظه بين العينتين ناتج عن الصدفة أم أنه يعكس فرقًا حقيقيًا بين المجتمعين اللذين تم أخذ العينتين منهما. ويشترط ان يكون:

- الحصول على البيانات بطريقة عشوائية وغير متحيزة
- قيم n_1 و n_2 ثابتة .
- أحجام المجتمعين N_1 و N_2 أكبر من 100 ضعف حجمي العينتين n_1 و n_2 على التوالي. وبالتالي، يمكن افتراض أن قيم P_1 و P_2 ثابتة .
- أحجام العينة كبيرة بما يكفي لضمان صحة التقريب الطبيعي للتوزيع العيني.

يتم حساب Z كالتالي:

$$z_{cal} = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{(n_1 + n_2)}$$

حيث:

$\hat{p}_1$ النسبة التقديرية في العينة الأولى .

$\hat{p}_2$ النسبة التقديرية في العينة الثانية.

$\bar{P}$ النسبة المجمّعة في كلا العينتين .

n_1 حجم العينة الأولى .

n_2 حجم العينة الثانية

وتكون صياغة الفرضيات كالتالي:

- **اختبار أحادي الطرف** :
  - $H_0: (P_1 - P_2) = 0$
  - $H_1: (P_1 - P_2) < 0$ أو $H_1: (P_1 - P_2) > 0$
- **اختبار ثنائي الطرف** :
  - $H_0: (P_1 - P_2) = 0$
  - $H_1: (P_1 - P_2) \neq 0$

مثال: لنفترض أننا نريد المقارنة بين نظام تدريب جديد يركز على التقوية والمرونة ونظام تدريب تقليدي من حيث نسبة الاصابات، عند مستوى دلالة (0.05) ولدينا المعطيات التالية:

المجموعة	نظام التدريب	عدد اللاعبين	عدد الإصابات	نسبة الإصابات $(\hat{p})$
الأولى	جديد	100	10	0.1
الثانية	تقليدي	100	15	0.15

الفرضية الصفرية :(H_0) لا يوجد فرق في نسبة الإصابات بين اللاعبين الذين يتبعون نظام التدريب الجديد واللاعبين الذين يتبعون النظام التقليدي.

الفرضية البديلة :(H_1) تختلف نسبة الإصابات بين المجموعتين، أي أن نسبة الإصابات تختلف بشكل كبير بين المجموعتين.

اختيار نوع الاختبار: اختبار z للنسب لعينتين هو الأنسب لمقارنة نسبتين في عينتين مستقلتين

$\hat{p}1= 10/100 = 0.1$ $\qquad$ $\bar{P} = (10+15)/(100+100) = 0.125$

$\hat{p}2= 15/100 = 0.15$

نقوم بتعويض القيم في معادلة Z:

$$z = (0.1 - 0.15) / \sqrt{(0.125 * 0.875 * (1/100 + 1/100))}$$

$$\mathbf{Z = -1.069}$$

القيمة الحرجة عند المستوى دلالة 0.05 في اختبار ذو طرفين هي 1.96 ، ونلاحظ ان القيمة المحسوبة ل Zاصغر من القيمة الجدولية وبالتالي نقبل الفرضية الصفرية أي ان نوع التدريب المتبع لا يؤثر على نسبة الإصابات لدى اللاعبين في هذه الدراسة.

2) اختبار Z للوسط الحسابي:

- في حالة حجم العينة كبير $n \geq 30$ والتباين δ^2 معلوم: يستخدم معامل Z في هذه الحالة لاختبار دلالة الفروق بين المتوسط الحسابي للعينة والمتوسط الحسابي للمجتمع الذي سحبت منه العينة.

ويتم حسابه وفق المعادلة:

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

حيث: $\overline{x}$: المتوسط الحسابي للعينة.

μ : المتوسط الحسابي للمجتمع. أو المتوسط الافتراضي للمجتمع.

σ : الانحراف المعياري للمجتمع.

n : حجم العينة

مثال:

وجد باحث في دليل أحد الاختبارات البدنية للطلبة أن متوسط الأداء العام هو 85 بانحراف معياري يساوي 8، فإذا أراد هذا الباحث أن يعرف ما إذاكان أداء الطلبة في مدينته يختلف عن متوسط الأداء العام، فاختار عينة عشوائية مؤلفة من 30 طالب، واستخرج متوسط أدائهم في هذا الاختبار فوجد أنه يساوي 84.

المطلوب: قم باختبار الفرضية الصفرية مقابل البديلة بدلا منه مستخدما مستوى دلالة $(\alpha = 0.05)$

الحل: لدينا المعطيات التالية

متوسط الأداء العام (المجتمع): $\mu = 85$

الانحراف المعياري العام: $\sigma = 8$

حجم العينة: $n = 30$

متوسط أداء العينة: $\bar{X} = 84$

صياغة الفروض الإحصائية:

لا توجد فروق بين المتوسط الحسابي لأداء العينة والمتوسط الحسابي للأداء العام $H_0: \mu = \mu_0$

يوجد فروق بين المتوسط الحسابي لأداء العينة والمتوسط الحسابي للأداء العام. $H_1: \mu \neq \mu_0$

حساب Z

$$Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$$
$$= (84 - 85) / (8 / \sqrt{30}) \approx -0.68$$

بما أن -0.68 محصورة بين -1.96 و 1.96، فإنها تقع ضمن منطقة القبول، أي نقبل الفرضية الصفرية التي تقر بعدم وجود فروق ذات دلالة إحصائية بين أداء عينة الباحث وبين الأداء العام عند مستوى دلالة 0.05.

ملاحظة: يمكن استخدام اختبار Z في بعض الحالات أين يكون حجم العينة أقل من 30 بشرط ان يكون التوزيع طبيعي والانحراف المعياري للمجتمع معلوم، أما في نفس الحالة (حجم العينة أقل من 30 والتوزيع يكون طبيعي) لكن الانحراف المعياري للمجتمع غير معلوم فنستخدم اختبار T

تمرين: ترغب إدارة مدرسة ثانوية في مقارنة نسبة الطلاب الذين يمارسون الرياضة بانتظام بين الطلاب الذكور والاناث، فتم اختيار عينة عشوائية من 120 طالب من الذكور ووجد ان 90 طالب يمارسون الرياضة بانتظام، وتم اختيار عينة عشوائية من 100 طالب، ووجد ان 70 طالبة يمارسن الرياضة بانتظام.

المطلوب: قم باختبار ما اذا وجدت المدرسة أن نسبة الطلاب الذكور الذين يمارسون الرياضة بانتظام **اكبر** من نسبة الطالبات؟

الحل:

صياغة الفرضيات:

- $H_0: (P_1 - P_2) = 0$ (لا يوجد فرق ذو دلالة إحصائية بين نسبة الطلاب ونسبة الطالبات الذين يمارسون الرياضة بانتظام)
- $H_1: (P_1 - P_2) > 0$ (نسبة الطلاب الذكور الذين يمارسون الرياضة بانتظام اكبر من نسبة الطالبات)

الاختبار المناسب هو Z للنسب لعينتين مستقلتين ونوع الاختبار هو اختبار ذو طرف واحد (الفرضية البديلة تتضمن إشارة $>$)

- حساب النسب:

$\hat{p_1} = 90/120 = 0.75$ (نسبة الطلاب الذكور الذين يمارسون الرياضة بانتظام)

$\hat{p_2} = 70/100 = 0.7$ (نسبة الطالبات اللاتي يمارسن الرياضة بانتظام)

$$\bar{P} = (90+70)/(120+100) = 0.72 \text{ (النسبة المجمعة)}$$

$$z = (0.75 - 0.7) / \sqrt{0.72\,(1-0.72) * (1/120 + 1/100)}$$

$$z = 0.82$$

نختار مستوى دلالة $\alpha = 0.01$، نجد ان القيمة الحرجة او القيمة الجدولية ل Z هي: 2.33، بما ان Z المحسوبة اقل من القيمة الجدولية وبالتالي نقبل الفرضية الصفرية. أي انه لا يوجد فرق بين نسبة الطلاب في الجنسين فيما يخص ممارسة الرياضة بانتظام.