احصاء حيوي

الكورس الثاني

(موضوع المحاضرة)

اختبارات تتعلق بالنسب

Dr.Safwan Nathem Rashed

ملاحظة:

استناداً الى الخطوات الاساسية المتبعة في اختبار اي فرضية والتي تم التعرف عليها سابقاً سوف يتم اتباعها في اختبار النسب.

اختبارات تتعلق بالنسب والتي تقسم الى:

١. اختبار النسبة لمجتمع واحد.

٢. اختبار الفرق بين نسبتين.

ويتم استخدام اختبارات النسب للمقارنة واتخاذ القرار عندما يكون المجتمع يتبع توزيع ثنائي الحدين فضلاً عن (n) كبيرة جداً.

١–اختبار النسبة لمجتمع واحد

نفرض ان (X) متغير عشوائي يشير الى عدد حالات النجاح لـ (n) من المحاولات المستقلة تكون فيها n عدد كبير باحتمال نجاح للمحاولة قدره (p) اما الفشل فيكون (q=1−p) فان الدالة الاحتمالية للتوزيع سيكون:

$$\boxed{p(x) = C_x^n p^x (1-p)^{n-x} \quad ; \; x = 0, 1, 2, \ldots, n}$$

وعليه فان الوسط الحسابي لهذا التوزيع والتباين له هو:

$$\boxed{\mu = n\,p \quad ; \quad \sigma^2 = n\,p\,q}$$

حيث توجد علاقة بين توزيع ثنائي الحدين والتوزيع الطبيعي حيث ان الدرجة المعيارية لـ Z في توزيع ثنائي الحدين يقترب او يتقارب توزيعها من التوزيع الطبيعي القياسي (($N(0,1)$) عندما يكون عدد المحاولات كبير اي n كبيرة وان احتمال النجاح او الفشل الذي يمثل p صغير اي يقترب من الصفر:

$$n \to \infty \qquad ; \qquad p \to 0$$

اي ان:

$$X \sim Ber(n, p) \to X \sim N(n\,p, n\,p\,q)$$

والذي يكون في معيار الاختبار لـ Z:

$$Z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2}} = \frac{\bar{x} - n\,p}{\sqrt{n\,p\,q}}$$

$$\lim_{n \to \infty} Z = \lim_{n \to \infty} \frac{\bar{x} - n\,p}{\sqrt{n\,p\,q}} \sim N(0,1)$$

وهذا الاستنتاج يسمح لنا باستخدام القيم الحرجة الخاصة بالتوزيع الطبيعي القياسي لغرض اختبار فرضية العدم H0 :

$$H_0 : p = p_0$$

p_0 : نسبة معطاة ضد اي فرضية بديلة اخرى إن كانت ذات جانب واحد او جانبين تحت المنحنى.

ولأجل التوصل الى معيار الاختبار الاحصائي المناسب لاختبار نسبة واحدة عند توفر عدد المحاولات مع احتمال النجاح او الفشل فان معيار الاختبار سيكون هو:

$$\boxed{Z = \frac{n\hat{p} - np}{\sqrt{n\hat{p}\hat{q}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \quad ; \quad \therefore \hat{p} = \frac{X}{n} \sim N(p, \frac{pq}{n})}$$

مثال:

توصل فريق طبي الى ان صنع دواء معين بإمكانه شفاء مرضى مصابين بنوع من الطفح الجلدي بنسبة 80% ويهدف تقييم هذا الدواء قامت احدى شركات تصنيع الادوية باختبار هذا الدواء على عينة من المرضى قوامها اي حجمها (120) مريض قبل البدء بعملية التصنيع تجارياً فتبين ان عدد الذين شفوا فعلاً من هذا المرض عددهم (86) مريضاً. هل يمكن القول ان ادعاء الفريق الطبي صحيح عند مستوى معنوية 1% ،علماً ان البيانات تتبع توزيع ذي الحدين.

Sol/

p: تمثل نسبة الشفاء.

1- $$\boxed{\begin{array}{l} H_0 : p = 0.80 \\ H_1 : p \neq 0.80 \end{array}}$$

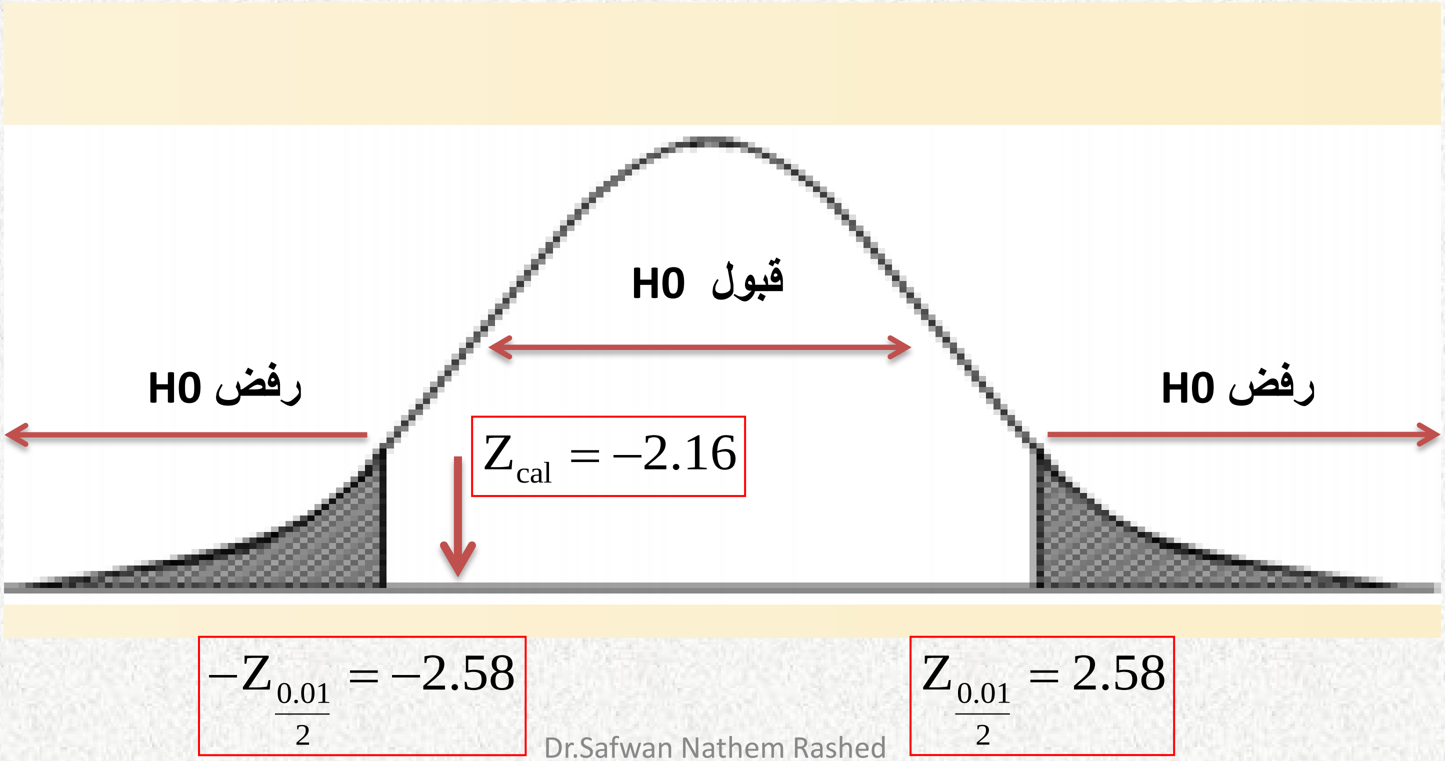
2- $$\boxed{\alpha = 0.01 \rightarrow Z_{\frac{0.01}{2}} = Z_{0.005} = 2.58}$$

3- $$\boxed{\begin{array}{l} \hat{p} = \dfrac{X}{n} = \dfrac{86}{120} = \dfrac{\text{عدد المشاهدات}}{\text{العدد الكلي}} \\ \hat{q} = 1 - \hat{p} = 1 - 0.72 = 0.28 \end{array}}$$

4-
$$Z_{cal} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\,\hat{q}}{n}}} = \frac{0.72 - 0.80}{\sqrt{\frac{(0.72)(0.28)}{120}}} = \frac{-0.08}{0.037}$$

$$Z_{cal} = -2.16$$

5-



6-

القرار: نلاحظ ان القيمة المحسوبة لـ Z_{cal} اقل من القيمة الجدولية لـ Z_{table} اي ان ($Z_{table} > Z_{cal}$) فضلاً عن ان القيمة المحسوبة تقع في منطقة القبول لـ H_0 تحت المنحنى وهذا يدل على رفض الفرضية البديلة والتمسك بفرضية العدم ، ان ادعاء الفريق الطبي صحيح وان الدواء له اهمية في علاج المرضى ويمكن البدء بالتصنيع التجاري لهذا الدواء الجديد.

٢-اختبار الفرق بين نسبتين

افرض ان (X1) متغير عشوائي يشير الى عدد حالات النجاح لـ (n1) من المحاولات المستقلة باحتمال نجاح للمحاولة قدره (p1) حول صيغة معينة، ونفرض ان (X2) متغير عشوائي يشير الى عدد حالات النجاح لـ (n2) من المحاولات المستقلة باحتمال نجاح للمحاولة قدره (p2) حول نفس الصيغة تكون فيها (n1,n2) على التوالي كبيرين عندئذ فان الوسط الحسابي والتباين للحالتين سيكون هو:

$$\boxed{\mu_1 = n_1 p_1 \,;\, \sigma_1^2 = n_1 p_1 q_1 \;|||\; \mu_2 = n_2 p_2 \,;\, \sigma_2^2 = n_2 p_2 q_2}$$

وبفرض ان العينتين مستقلتين عن بعضهما فان:

$$\boxed{\hat{p}_1 = \frac{X_1}{n_1} \quad ; \quad \hat{p}_2 = \frac{X_2}{n_2}}$$

وعليه سوف تكون فرضية العدم H_0 والفرضية البديلة H_1 للفرق بين نسبتين كالاتي:

$$H_0 : p_1 - p_2 = 0 \quad \text{or} \quad H_0 : p_1 = p_2$$

$$\text{v.s.} \begin{cases} H_1 : p_1 - p_2 \neq 0 \quad \text{or} \quad H_0 : p_1 \neq p_2 \\ H_1 : p_1 - p_2 > 0 \quad \text{or} \quad H_0 : p_1 > p_2 \\ H_1 : p_1 - p_2 < 0 \quad \text{or} \quad H_0 : p_1 < p_2 \end{cases}$$

وبناءً على ما تم ذكره سابقاً فانه عندما تكون حجم العينة كبيراً يكون:

$$\hat{p}_1 = \frac{X_1}{n_1} \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right) ; \hat{p}_2 = \frac{X_2}{n_2} \sim N\left(p_2, \frac{p_2 q_2}{n_2}\right)$$

لذلك فان:

$$(\hat{p}_1 - \hat{p}_2) \sim N\left((p_1 - p_2), \left(\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)\right)$$

وبالتالي فان المعيار الاختباري للفرق بين نسبتين الملائم سكيون بالصيغة الرياضية الاتية:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1)$$

وان المشكلة هنا في تحديد قيمة **(p,q)** لأغراض التطبيق

العملي وحيث اننا لانمتلك اية معلومات عن **(p,q)** عندئذٍ يجب

تقديرهما، وان افضل تقدير لـ **(p)** هو:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} \quad ; \quad \hat{q} = 1 - \hat{p}$$

ليكون معيار الاختبار للفرق بين نسبتين اكثر الملائم بالصيغة الرياضية الاتية:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\,\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

مثال:

لغرض دراسة مرض قرحة الاثني عشري تم اختيار عينتين مستقلتين: الاولى من الرجال قوامها (200) رجل حيث وجد ان (19) منهم يشكون من هذا المرض، والمجموعة الثانية من النساء قوامها (100) امرآة وجد بان (5) منهم فقط يشكين من هذا المرض. فهل هناك اختلاف بين الرجال والنساء من حيث الاصابة بهذا المرض عند مستوى معنوية 5%.

Sol/

1−

$$H_0 : p_1 - p_2 = 0 \quad \text{or} \quad H_0 : p_1 = p_2$$

$$H_1 : p_1 - p_2 \neq 0 \quad \text{or} \quad H_0 : p_1 \neq p_2$$

2−

$$\alpha = 0.05 \rightarrow Z_{\frac{0.05}{2}} = Z_{0.025} = 1.96$$

3−

$$\hat{p}_1 = \frac{X_1}{n_1} = \frac{19}{200} = 0.095 \;\; ; \;\; \hat{p}_2 = \frac{X_2}{n_2} = \frac{5}{100} = 0.05$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{19+5}{200+100} = \frac{24}{300} = 0.08$$

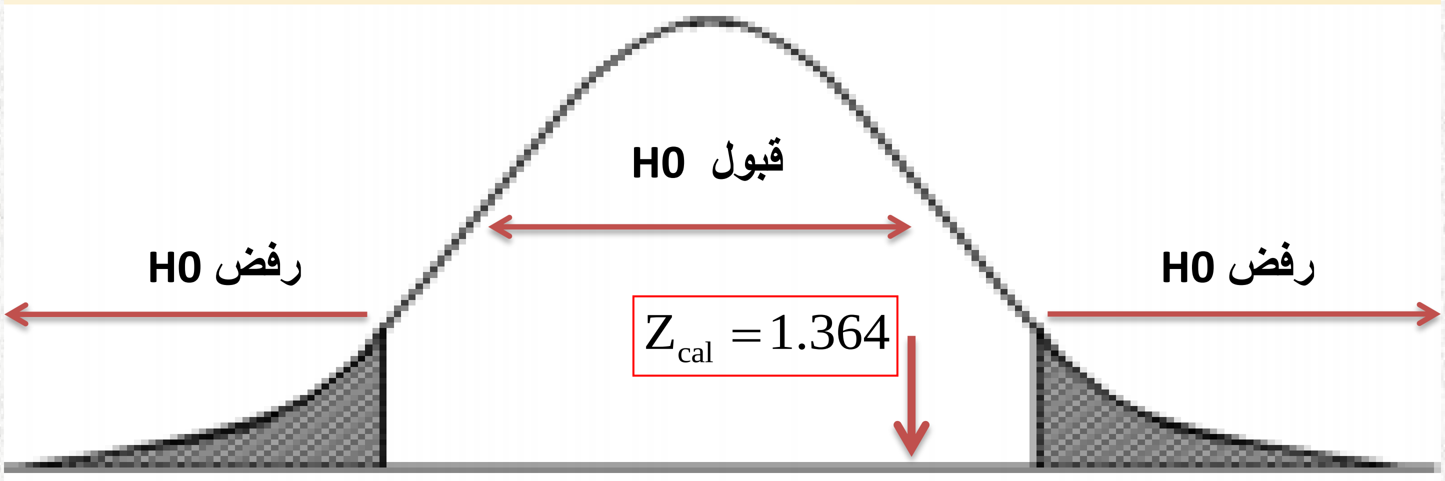
$$\hat{q} = 1 - \hat{p} = 1 - 0.08 = 0.92$$

4−

$$Z_{cal} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.095 - 0.05)}{\sqrt{(0.08)(0.92)\left(\frac{1}{200} + \frac{1}{100}\right)}}$$

$$Z_{cal} = \frac{0.045}{\sqrt{0.001}} = \frac{0.045}{0.033} = 1.364$$

5−



$-Z_{\frac{0.05}{2}} = -1.96$ $Z_{\frac{0.05}{2}} = 1.96$

6-

القرار: نلاحظ ان القيمة المحسوبة لـ Z_{cal} اقل من القيمة الجدولية لـ Z_{table} اي ان $(Z_{table} > Z_{cal})$ فضلاً عن ان القيمة المحسوبة تقع في منطقة القبول لـ H_0 تحت المنحنى وهذا يدل على رفض الفرضية البديلة والتمسك بفرضية العدم ، لا يوجد فرق بين الرجال والنساء بإصابتهم بمرض قرحة الاثني عشري.

مثال:

تم تقسيم (200) مريض يعانون من مرض معين وبصورة عشوائية الى مجموعتين متساويتين. تضم المجموعة الاولى التي عولجت بالطريقة القياسية تم شفاء (78) منهم خلال ثلاث ايام، ومن ضمن المجموعة الثانية التي عولجت بطريقة جديدة تم شفاء (90) منهم خلال ثلاث ايام ايضاً، ويفرض ان العينتين مستقلتين عن بعضهما، فهل يكون القول ان العلاج بالطريقة الجديدة اكثر فاعلية من الطريقة القياسية تحت مستوى معنوية 5% ، علماً ان العينتين يتبعان التوزيع ذي الحدين.

Sol/

1- $H_0 : p_1 - p_2 = 0 \quad \text{or} \quad H_0 : p_1 = p_2$

$H_1 : p_1 - p_2 < 0 \quad \text{or} \quad H_0 : p_1 < p_2$

2- $\alpha = 0.05 \rightarrow Z_{0.05} = 1.65$

$n_1 = n_2 = 100 \quad ; \quad X_1 = 78 \quad ; \quad X_2 = 90$

3- $\hat{p}_1 = \dfrac{X_1}{n_1} = \dfrac{78}{100} = 0.78 \quad ; \quad \hat{p}_2 = \dfrac{X_2}{n_2} = \dfrac{90}{100} = 0.90$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{78 + 90}{100 + 100} = \frac{168}{200} = 0.84$$

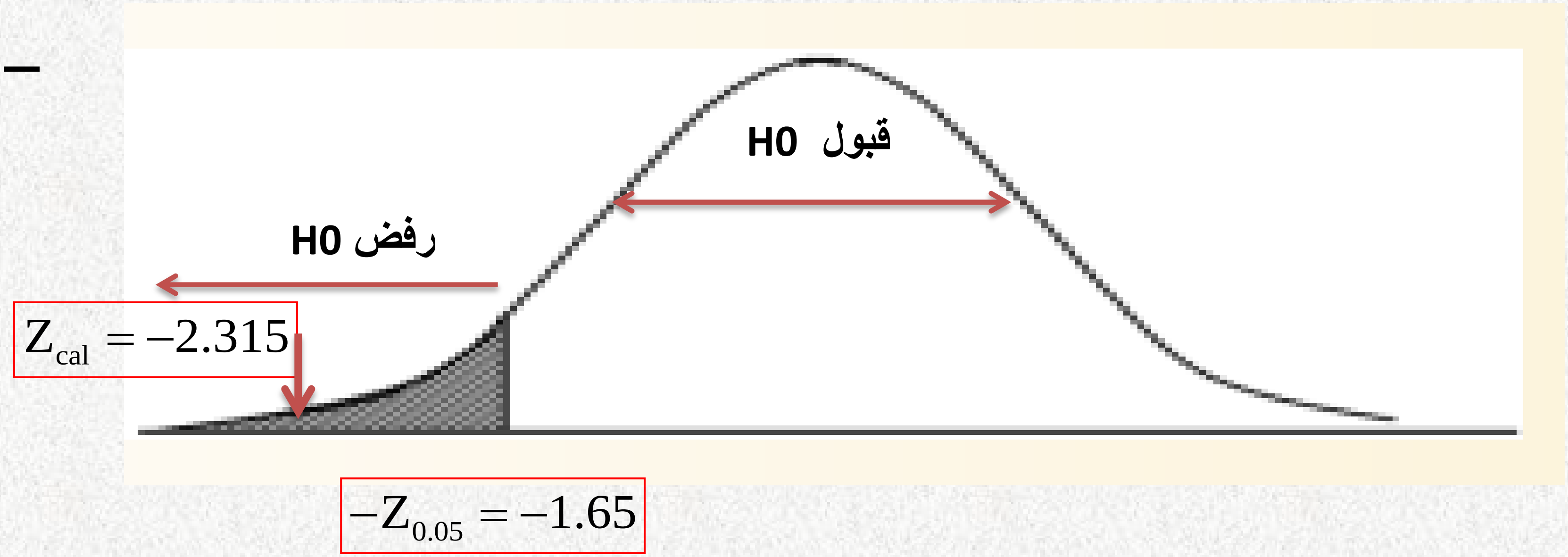
$$\hat{q} = 1 - \hat{p} = 1 - 0.84 = 0.16$$

4−

$$Z_{cal} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\,\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.78 - 0.90)}{\sqrt{(0.84)(0.16)\left(\frac{1}{100} + \frac{1}{100}\right)}}$$

$$Z_{cal} = \frac{-0.12}{0.052} = -2.315$$

5−



6-

القرار: نلاحظ ان القيمة المحسوبة لـ Z_{cal} اكبر من القيمة الجدولية لـ Z_{table} اي ان $(Z_{table} < Z_{cal})$ فضلاً عن ان القيمة المحسوبة تقع في منطقة الرفض لـ H_0 تحت المنحنى وهذا يدل على رفض الفرضية العدم والتمسك بفرضية البديلة ، اي ان هناك اختلاف واضح في العلاج باستخدام الطريقة الجديدة افضل من الطريقة القياسية.