احصاء حيوي

الكورس الثاني

(موضوع المحاضرة)

مقايس العلاقة بين العوامل الحياتية

Dr.Safwan Nathem Rashed

معنوية الفرق بين المعدلات
Significance of Differences in Rate

١- مقارنة بين معدلين: Comparison of Two Rates

ان مقارنة معدلين مثل $\left(\frac{a_1}{n_1} , \frac{a_2}{n_2}\right)$ اساساً هي نفس الطريقة المستخدمة لمقارنة الفرق بين متوسطين $\left(\bar{X}_A , \bar{X}_B\right)$ المعروف لدينا سابقاً

ويشترط ان يكون هذان المعدلان تم اخذهم من عينتين مستقلتين.

- **اختبار χ^2 Chi-Square test باستخدام توزيع t.**

اذا كان لدينا عينتين حجم كل منهما n_1,n_2 على التوالي وهما كبيرتين بشكل معقول اي يزيد عن 30 فان الفرضية المناسبة هو التوزيع الطبيعي

الذي يكون مقارب لتوزيع المعاينة، وان العينتين تم سحبهما من نفس التوزيع للمجتمع بوسط حسابي قدره (p) وانحراف معياري قدره $(\sqrt{pq})$ لذا سيكون الخطأ القياسي للفرق بين العينتين هو $\left(\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right)$ ليكون المختبر الاحصائي للنسب بالصيغة الاتية:

$$t_{\infty}=\frac{\left(\frac{a_1}{n_1}-\frac{a_2}{n_2}\right)}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

ويتم ايجاد احتمال الرفض او القبول من جدول t بدرجة حرية (∞) ومستوى معنوية (α) اي $t(\alpha,\infty)$ ، واذا كانت قيمة t_{∞} اكبر من القيمة الجدولية المقابلة لها عندئذٍ ترفض فرضية العدم. وتكون شكل الفرضية كما يلي:

شكل الفرضية كما يلي:

$$H_0 : \frac{a_1}{n_1} = \frac{a_2}{n_2}$$

$$\text{v.s.} \begin{cases} H_1 : \dfrac{a_1}{n_1} \neq \dfrac{a_2}{n_2} \\ H_1 : \dfrac{a_1}{n_1} > \dfrac{a_2}{n_2} \\ H_1 : \dfrac{a_1}{n_1} < \dfrac{a_2}{n_2} \end{cases}$$

ملاحظة (١) : عند مقارنة معدلين فأننا نأخذ أفضل تقدير للاحتمالية عندئذٍ وتحت فرضية العدم سوف نفترض ان كل من عينتان الحالة ليكون التقدير لـ **p** كما يلي:

$$\boxed{p = \frac{a_1 + a_2}{n_1 + n_2}}$$

ملاحظة (٢) : ان اختبار χ^2 واختبار t_∞ لمعدلين هو نفسه لان χ^2 هي احصائية ناتجة عن تربيع التوزيع الطبيعي القياسي ذات وسط حسابي صفر وتباين قدره واحد وان t_∞ هي كذلك قياس عشوائي يمكن كتابته بالشكل الاتي :

الشكل الاتي:

$$X \sim N\left(\mu, \sigma^2\right)$$

يتم تحويل المتغير العشوائي الى قياسي على النحو الاتي:

$$t = \frac{X-\mu}{\sigma} \sim N\left(0,1\right)$$

بتربيع الطرفين نحصل على:

$$t^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_{(1)}$$

$$\chi^2_{(1)} = t^2_{\infty} = \frac{\left(\frac{a_1}{n_1} - \frac{a_2}{n_2}\right)^2}{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

ويمكن تبسيط هذه الصيغة من خلال فرض ان:

$$\left.\begin{aligned} A &= a_1 + a_2 \\ N &= n_1 + n_2 \end{aligned}\right\} \quad p = \frac{a_1 - a_2}{n_1 + n_2} = \frac{A}{N}$$

$$\chi^2_{(1)} = \frac{\left(\dfrac{a_1 n_2 - a_2 n_1}{n_1 n_2}\right)^2}{p(1-p)\left(\dfrac{n_1 + n_2}{n_1 n_2}\right)} = \frac{\dfrac{(a_1 n_2 - a_2 n_1)^2}{(n_1 n_2)^2}}{\dfrac{A}{N}\left(1 - \dfrac{A}{N}\right)\left(\dfrac{N}{n_1 n_2}\right)}$$

$$\chi^2_{(1)} = \frac{\dfrac{(a_1 n_2 - a_2 n_1)^2}{(n_1 n_2)^{\cancel{2}}}}{\dfrac{A}{N}\left(\dfrac{N - A}{\cancel{N}}\right)\left(\dfrac{\cancel{N}}{\cancel{n_1 n_2}}\right)}$$

لنحصل على المختبر الاحصائي لمربع كاي بالصيغة الاتية:

$$\therefore \chi^2_{(1)} = \frac{\left(a_1 n_2 - a_2 n_1\right)^2 \times N}{AB\left(n_1 n_2\right)} \quad ; \; B = N - A$$

ويكون ايجاد الاحتمالية او القيمة الجدولية، اما من الصف الاخير من جدول (**t**) مع أخذ القيمة التربيعية له او من خلال جدول χ^2 بدرجة حوية (واحد) ومستوى معنوية معين فاذا كانت القيمة المحسوبة اكبر منالجدولية نرفض فرضية العدم H_0 ونقبل بالفرضية البديلة H_1. وبما ان خطوات اختبار عدة نسب من توزيع ذي الحدين فيكون الاختبار ذي طرفين او طرف واحد بصيغة لمربع كاي واحدة وهي تختلف عن اختبار عدم التجانس هو:

$$\therefore \chi^2_{(\alpha,v)} = \chi^2_{(\alpha,(r-1)(c-1))}$$

- أ- استخدام اختبار χ^2 Chi-Square test للمقارنة بين معدلين (حسن المطابقة).

لغرض المقارنة بين معدلين عندما N تكون كبيرة يفترض ان:

١. الفرق بين المعدلين يتوزع توزيعاً طبيعياً عندما N تكون كبيرة.

٢. يتم استخدام القانون السابق لـ χ^2 اذا كان التوزيع مستمر وفي حالة كون التوزيع متقطع يمكن اجراء تصحيح لصيغة χ^2 المعروفة سابقاً بالشكل الاتي:

$$\boxed{\therefore \chi^2_{(1)} = \frac{\left(\left|a_1 n_2 - a_2 n_1\right| - \frac{N}{2}\right)^2 \times N}{AB\left(n_1 n_2\right)}}$$

نلاحظ ان معامل التصحيح هم $\frac{1}{2}$ حجم العينة $\frac{N}{2}$

مثال:

الجدول الاتي يحتوي على حالات الوفاة للمصابين بمرض معين خلال فترة زمنية محددة وكما يلي:

النسبة المئوية للوفاة	عدد المتوفين a	عدد المصابين n	فترة
10/40=0.25=25%	10	40	1994-2000
5/61=0.082=8.2%	5	61	2001-2007
	A=15	N=101	

هل ان معدلات الوفيات تختلف معنوياً للفترتين تحت مستوى معنوية $\alpha = 0.05$.

Sol/

$$1 = \begin{cases} H_0 : \dfrac{a_1}{n_1} = \dfrac{a_2}{n_2} \\ H_1 : \dfrac{a_1}{n_1} \neq \dfrac{a_2}{n_2} \end{cases} \quad ; \quad \begin{cases} a_1 = 10 \quad , \quad n_1 = 40 \\ a_2 = 5 \quad , \quad n_2 = 61 \end{cases}$$

$$\left.\begin{aligned} A &= a_1 + a_2 = 15 \\ N &= n_1 + n_2 = 101 \end{aligned}\right\} ; \; B = N - A = 101 - 15 = 86$$

$$2 - \alpha = 0.05 \;\; ; d.f. = (r-1)(c-1) = (2-1)(2-1) = 1 \times 1 = 1$$

$$\therefore \chi^2_{(\alpha,v)} = \chi^2_{(0.05,1)} = 3.841$$

$$3 - \therefore \chi^2_{(1)} = \frac{\left(|a_1 n_2 - a_2 n_1| - \dfrac{N}{2} \right)^2 \times N}{AB(n_1 n_2)}$$

Sol/

$$\chi^2_{(1)} = \frac{\left(\left|10\times 40 - 5\times 61\right| - \frac{101}{2}\right)^2 \times 101}{15\times 86\times 40\times 61} = 4.15$$

4- القرار حول حسن المطابقة : نلاحظ ان $\chi^2_{\text{table}} > \chi^2_{\text{cal}}$ لذلك نرفض فرضية العدم ونقبل الفرضية البديلة على ان الفرق بين المعدلين اكبر ولم تحدث نتيجة الصدفة وانما وجود فرق جوهري وله سبب معين.

واذا استخدمنا القيمة الجدولية لـ t ستكون $t\left(\frac{0.05}{2},\infty\right) = 1.96$.

$$\chi^2 = t^2 = (1.96)^2 = 3.841 \;;\; \sqrt{\chi^2} = \sqrt{3.841} = 1.96$$

ب- اختبار فيشر الدقيق للمقارنة بين معدلين

Fisher's of test for Comparison Two Rates

يمكن استخدام اختبار فيشر الدقيق في حالة كون N صغيرة اي اقل من 30 او في ايجاد التكرار المتوقع لاحد خلايا الجدول اقل من 5 أو اثنين معاً، والجدول الاتي يوضح كيفية الاختبار لتوضيح حسن المطابقة بين معدلين فاكثر.

Groups	Event	No event	Total
Groups1	a_1	b_1	n_1
Groups2	a_2	b_2	n_2
	A	B	N

ويمكن حساب التكرار المتوقع لخلايا الجدول باستخدام مجاميع الصفوف والاعمدة والمجموع العام.

حيث يتم حساب القيمة المتوقع للخلايا الاربعة الموضحة في الجدول السابق على النحو الاتي:

$$E_{ij}\ ;\begin{cases} i = 1, 2, 3, \ldots, r \\ j = 1, 2, 3, \ldots, c \end{cases}$$

حيث ان:

$$E_{11} = \frac{n_1 A}{N} \ ; \ E_{12} = \frac{n_2 A}{N}$$

$$E_{12} = \frac{n_1 B}{N} \ ; \ E_{22} = \frac{n_2 B}{N}$$

عندما تكون N صغيرة يكون افتراض التوزيع الطبيعي غير صحيح وان استخدام اختبار χ^2 يكون ايضاً غير صحيح ولغرض تطبيق اختبار فيشر الدقيق نجد احتمالات الحوادث $(x=0,1,2,\ldots,a)$ على فرض ان اصغر قيمة في الخلية الموجودة في الجدول هي (a) حيث ان الاحتمالات هي:

$$\boxed{p(0), p(1), p(2), p(3), \ldots., p(a)}$$

نبدأ بحساب هذه الاحتمالات ومن ثم تجميعها مع بعضها ليكون المجموع الناتج يمثل الاحتمالية المطلوب والذي سوف يقارن مع مستوى المعنوية α لغرض اتخاذ القرار المناسب وعليه لابد من ايجاد قيم تساعد على البحث عن القيمة الاحتمالية والمتمثلة بـ(L, D, M) وكالاتي:

١. L : تمثل اقل مجموع من مجاميع الصفوف والاعمدة.

٢. D : يتم ايجادها وفق الصيغة الاتية:

$$\boxed{\therefore D = N - M}$$

٣. M : تمثل مجموع الصف او العمود الذي لا يحتوي على (L) والذي له اقل معدل ويمكن توضيحه وفق الحالات الاتية:

- في حالة كون (A) هو الاقل اي L=A فان:

$$\boxed{\therefore \begin{cases} \text{IF } L = A \text{ and } \dfrac{a_1}{n_1} \leq \dfrac{a_2}{n_2} \ ; \ \text{then } M = n_1 \\ \\ \text{IF } L = A \text{ and } \dfrac{a_1}{n_1} \geq \dfrac{a_2}{n_2} \ ; \ \text{then } M = n_2 \end{cases}}$$

• في حالة كون (B) هو الاقل اي L=B فان:

$$\therefore \begin{cases} \text{IF } L = B \text{ and } \dfrac{b_1}{n_1} \leq \dfrac{b_2}{n_2} \;;\; \text{then } M = n_1 \\ \text{IF } L = B \text{ and } \dfrac{b_1}{n_1} \geq \dfrac{b_2}{n_2} \;;\; \text{then } M = n_2 \end{cases}$$

• في حالة كون (n1) هو الاقل اي L=n1 فان:

$$\therefore \begin{cases} \text{IF } L = n_1 \text{ and } \dfrac{a_1}{A} \leq \dfrac{b_1}{B} \;;\; \text{then } M = A \\ \text{IF } L = n_1 \text{ and } \dfrac{a_1}{A} \geq \dfrac{b_1}{B} \;;\; \text{then } M = B \end{cases}$$

- في حالة كون (n2) هو الاقل اي L=n2 فان:

$$\therefore \begin{cases} \text{IF } L = n_2 \text{ and } \dfrac{a_2}{A} \le \dfrac{b_2}{B} \ ; \ \text{then } M = A \\ \text{IF } L = n_2 \text{ and } \dfrac{a_2}{A} \ge \dfrac{b_2}{B} \ ; \ \text{then } M = B \end{cases}$$

- والان يتم ايجاد القيم الاحتمالية من خلال اتباع الصيغة الرياضية الاتية:

$$p(0), p(1), p(2), p(3), \ldots, p(a)$$

$$p(0) = \frac{(D-L+1)(D-L+2)(D-L+3)\cdots D}{(N-L+1)(N-L+2)(N-L+3)\cdots N}$$

ويتم الحصول على باقي القيم الاحتمالية وفق الصيغ الاتي:

$$p(1) = p(0) \times \frac{L \times M}{1 \times (D - L + 1)}$$

$$p(2) = p(1) \times \frac{(L-1) \times (M-1)}{2 \times (D - L + 1)}$$

$$p(3) = p(2) \times \frac{(L-2) \times (M-2)}{3 \times (D - L + 1)}$$

$$\vdots \; \vdots \; \vdots$$

$$p(r+1) = p(r) \times \frac{(L-r) \times (M-r)}{(r+1) \times (D - L + 1)}$$

r: اصغر عنصر موجود في الجدول

وبعد حساب القيم الاحتمالية، **نوجد حاصل جمع هذه الاحتمالات** ونقارنها مع مستوى المعنوية α .

فإذا كان: $\boxed{p(0)+p(1)+p(2)+p(3)+\cdots+p(r+1)\geq\alpha}$

نرفض فرضية العدم H_0 ونقبل بالفرضية البديلة H_1 (اي يوجد فرق معنوي بين المعدلين).

اما اذا: $\boxed{p(0)+p(1)+p(2)+p(3)+\cdots+p(r+1)\geq\alpha}$

نقبل بفرضية العدم H_0 ونرفض الفرضية البديلة H_1 (اي لا يوجد فرق معنوي بين المعدلين).

مثال:

تم اعطاء مغذي معقم وآخر مضاف اليه مادة معينة لـ 17 شخص مصابين بداء الكلب وتم تسجيل حالات الوفيات من كل مجموعة عند مستوى معنوية 5% كما مبين في الجدول الاتي.

Groups	المتوفين Dead	على قيد الحياة Live	Total
المغذي (١)	$a_1 = 1$	$b_1 = 11$	$n_1 = 12$
المغذي (٢)	$a_2 = 3$	$b_2 = 2$	$n_2 = 5$
	$A = 4$	$B = 13$	$N = 17$

هل هناك فرق معنوي بين معدلات الوفيات بتأثير نوع المغذي.

Sol/

1−
$$\begin{cases} H_0 : \dfrac{a_1}{n_1} = \dfrac{a_2}{n_2} \quad ; \quad H_0 : \dfrac{1}{12} = \dfrac{3}{5} \\ H_1 : \dfrac{a_1}{n_1} \neq \dfrac{a_2}{n_2} \quad ; \quad H_0 : \dfrac{1}{12} \neq \dfrac{3}{5} \end{cases}$$

الايجاد (L, D, M) وكما يلي:

٢− L : نلاحظ الصفوف والاعمدة من الجدول والخاص بالمجموع نجد ان اقل مجموع من بين الصفوف او الاعمدة هو (L= 4 =A):

حيث ان:

N < 30 −−− 17 < 30

وان

Eij < 5 −−− 4 < 5

$$E_{11} = \frac{n_1 A}{N} = \frac{12 \times 4}{17} \; ; \; E_{12} = \frac{n_2 A}{N} = \frac{5 \times 4}{17}$$

$$E_{12} = \frac{n_1 B}{N} = \frac{12 \times 13}{17} \; ; \; E_{22} = \frac{n_2 B}{N} = \frac{5 \times 12}{17}$$

Sol/

٣- M : يتم ايجادها وفق الصيغة المناسبة لها التي تكون L=A فان:

$$\therefore \begin{cases} \text{IF } L = A \text{ and } \dfrac{1}{12} = 0.083 \le \dfrac{3}{5} = 0.6 \;;\; \text{then } M = n_1 = 12 \\ \cancel{\text{IF } L = A \text{ and } \dfrac{1}{12} = 0.083 \ge \dfrac{3}{5} = 0.6 \;;\; \text{then } M = n_2 = 5} \end{cases}$$

٤- $\boxed{D = N - M = 17 - 12 = 5}$

٥- نلاحظ ان اقل قيمة في الجدول وخاصةً الخلايا هو واحد (١) اي ان a=1 لتكون القيم الاحتمالية.

$$\boxed{p(0), p(1)}$$

Sol/

$$p(0), p(1)$$

$$p(0) = \frac{(D-L+1)(D-L+2)(D-L+3)\cdots D}{(N-L+1)(N-L+2)(N-L+3)\cdots N}$$

$$= \frac{(5-4+1)(5-4+2)(5-4+3)(5-4+4)}{(17-4+1)(17-4+2)(17-4+3)(17-4+4)}$$

$$= \frac{(2)(3)(4)(5)}{(14)(15)(16)(17)} = \frac{1}{476} = 0.0021$$

$$p(1) = p(0) \times \frac{L \times M}{1 \times (D-L+1)} = \frac{1}{476} \times \frac{4 \times 12}{1 \times (5-4+1)} = \frac{48}{476 \times 2}$$

$$= \frac{24}{476} = 0.0504$$

Sol/

$$p(0)+p(1) \Leftrightarrow \alpha$$

وعليه فان:

$$\frac{1}{476}+\frac{24}{476}=0.053>\alpha=0.05$$

٥- القرار: نلاحظ ان القيم الاحتمالية المحسوبة هي اكبر من قيمة α مما يدل على رفض فرضية العدم H_0 والقبول بالفرضية البديلة H_1 اي وجود فرق معنوي في تأثير نوع المغذي على معدل الوفيات.