اختبار كولوموجروف سيمرنوف

يستخدم اختبار كولوموجراف سيمرنوف بدلا من اختبار مربع كاي في بعض الأحيان وعندما لا تتحقق الشروط اللازمة لإجراء اختبار مربع كاي، اختبار كولوموجروف سيمرنوف لا بديل عنه عندما يكون مجموع التكرارات أقل من 30 أو أن التكرارت المتوقعة لكل خلية أقل من 5 وعملية ضم الخلايا التى اشرنا اليها في اختبار مربع كا ^{2} ستؤدي الي فقد كثير من درجات الحرية الأمر الذي يتعذر معه اجراء الاختبار، وخطوات إجراء الاختبار كالتالي:

(1) نحسب الجدول التكراري للتكرارت المشاهدات والتكرارات المتوقعة كما سبق في اختبار مربع كا^2.

(2) نحسب التوزيع الاحتمالي للتكرارات المشاهدة والتكرارات المتوقعة (التكرار النسبي الذي يساوي تكرار المشاهدة مقسوما على مجموع التكرارات).

(3) نحصل على التوزيع الاحتمالي المتجمع للتكرارات المشاهدة $F(x)$ والتكرارات المتوقعة $S(x)$.

(4) نحسب أقصى فرق بين الاحتمال المتجمع المشاهد والمتوقع ونرمز له بالرمز D_0 وهو إحصاءة الاختبار، $D_0 = \max|F(x) - S(x)|$. ثم نستخدم جدول كولوموجراف سيمرنوف عند درجة حرية تساوي مجموع التكرارات N وبمستوى معنوية α للحصول على قيمة $D_{\alpha,N}$.

(5) نرفض فرض العدم H_0 إذا كانت القيمة المحسوبة $D_0 = \max|F(x) - S(x)|$ أكبر من القيمة الجدولة $D_{\alpha,N}$، أي نرفض H_0 عندما $D_0 > D_{\alpha,N}$.

مثال(1): (في مثال اختبار كا ^{2} السابق) يراد اختبار فيما اذا كانت البيانات التالية تتبع التوزيع الطبيعي بمتوسط 30 وتباين 100.

الحل:

$H_0: N(30,100)$ العينة لمتغير يتبع التوزيع الطبيعي بمتوسط 30 وتباين 100

H_1: العينة لمتغير لا يتبع التوزيع الطبيعي بمتوسط 30 وتباين 100

الجدول التكراري للمشاهدات والتوقع

المجموعة	1	2	3	4	المجموع
المشاهدة O_i	8	4	3	5	20
المتوقعة E_i	5	5	5	5	20

جدول التوزيع الاحتمالي للقيم المشاهدة والقيم المتوقعة

المشاهدة O_i	0.4	0.2	0.15	0.25
المتوقعة E_i	0.25	0.25	0.25	0.25

جدول التوزيع التراكمي للقيم المشاهدة والقيم المتوقعة

المشاهدة $O_i = F(x)$	0.4	0.6	0.75	1.00
المتوقعة $E_i = S(x)$	0.25	0. 5	0.75	1.00
$\|F(x) - S(x)\|$	0.15	0.1	0	0

قيمة الإحصاءة $D_0 = \max|F(x) - S(x)| = 0.15$،

عند مستوى معنوية $\alpha = 0.05$ ودرجة حرية 20 نجد أن القيمة الجدولية $D_{0.05,20} = 0,294$.

وحيث أن $D_0 = 0.15 < D_{0.05,20} = 0.294$ ، لا نستطيع رفض H_0 . أي أن البيانات تتبع التوزيع الطبيعي $N(30, 100)$

مثال(2): نفرض أن لدينا البيانات التالية:

Variable	1	2	3	4	5	Total
المشاهدة O_i	9	11	8	4	3	35

اختبر الفرض القائل أن نسبة المشاهدات لكل متغير هي 20%.

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 0.2$$

$$H_0 : p_1 \neq p_2 \neq p_3 \neq p_4 \neq p_5 \neq 0.2$$

جدول التوزيع الاحتمالي للقيم المشاهدة والقيم المتوقعة

المشاهدة O_i	0.257	0.314	0.229	0.114	0.086
المتوقعة E_i	0.2	0.2	0.2	0.2	0.2

جدول التوزيع التراكمي للقيم المشاهدة والقيم المتوقعة

المشاهدة $O_i = F(x)$	0.257	0.371	0.8	0.914	1.00
المتوقعة $E_i = S(x)$	0.2	0. 4	0.6	0.8	1.00
$\lvert F(x) - S(x) \rvert$	0.057	0.029	0.2	0.114	0

قيمة الإحصاءة $D_0 = \max \lvert F(x) - S(x) \rvert = 0.2$

عند مستوى معنوية $\alpha = 0.05$ نجد ان القيمة الجدولية $D_{0.05,35} = 0.224$. وحيث أن

$D_0 = 0.2 < D_{0.05,20} = 0.224$ ، لا نستطيع رفض H_0 . أي أن نسبة المشاهدات لكل متغير هي 20%.

اختبار الوسيط

يتم أخذ عينة عشوائية حجمها n_i من المجتمع $i=1,2,\dots,i$ ثم يحدد وسيط إجمالي العينات $N=\sum n_i$ ، ثم نحدد من كل عينة i عدد الوحدات التي تتجاوز الوسيط O_{1i} (عدد الوحدات اكبر من الوسيط)، وعدد الوحدات التي لا تتجاوز الوسيط O_{2i} (عدد الوحدات أصغر من أو تساوي الوسيط) ثم توضع في جدول $2\times c$ كما يلي

العينة Sample	1	2	…	c	
أكبر من الوسيط	O_{11}	O_{12}	…	O_{1c}	***a***
ليست أكبر من الوسيط	O_{21}	O_{22}	…	O_{2c}	***b***
	n_1	n_2	…	n_c	***N***

حيث a إجمالي عدد الوحدات من جميع العينات التي تتجاوز الوسيط ، b إجمالي عدد الوسيط من جميع العينات التي لا تتجاوز الوسيط.

الافتراضات

(1) عينات عشوائية كل مفردة فيها مستقلة عن البقية.

(2) عينات مستقلة عن بعضها البعض.

(3) طريقة القياس ترتيبية على الأقل.

(4) اذا كان لجميع العينات نفس الوسيط كان لها نفس الاحتمال p ان عينة تتجاوز الوسيط.

الاختبار

H_0: جميع العينات لها نفس الوسيط

H_1: مجتمعين على الأقل لهما وسيطين مختلفين.

تحسب القيمة T على النحو التالي:

$$T=\frac{N^2}{ab}\sum_{i=1}^{c}\left[O_{1i}-\frac{a\,n_i}{N}\right]^2 \Big/ n_i$$

T تتبع تقريبا توزيع مربع كا^2 بدرجة حرية c-1 . ونرفض فرض العدم H_0 إذا كانت $T>\chi^2_{1-\alpha,\,c-1}$.

مثال(1): يراد اختبار الفرق بين أربع طرق في زراعة الذرة، جربت كل طريقة بشكل عشوائي على عدد من قطع الأرض المتشابهة فنتج المحصول لكل "ايكر" لكل طريقة كما يلي

الطريقة	المحصول
1	83, 91, 94, 89, 89, 96, 91, 92, 90
2	91, 90, 81, 83, 84, 83, 88, 91, 89, 84
3	101, 100, 191, 93, 96, 95, 94
4	78, 82, 81, 77, 79, 81, 80, 81

اختبر بمستوى معنوية $\alpha=0.01$ تساوي وسيطات هذه المجتمعات. (اختبر تساوي وسيط المحصول الناتج من استخدام كل من الطرق الأربعة).

الحل:

H_0: جميع الطرق تعطي نفس الوسيط

H_1: مجتمعين على الأقل لهما وسيطين مختلفين.

عدد الوحدات الإجمالي N هو 34 فيكون الوسيط هو الوسط الحسابي للمفردتين السابعة عشر والثامنة عشر (بعد ترتيب جميع القيم في الطرق الأربع) فنحصل على القيمة 89 وبالتالي في كل عينة نحدد عدد الوحدات التي تتجاوز 89 ويكون O_{1i} وعدد الوحدات التي لا تتجاوز 89 وهو O_{2i} كما يلي

الطريقة	1	2	3	4	المجموع
أكبر من 89	6	3	7	0	16
ليست أكبر من 89	3	7	0	8	18
المجموع	9	10	7	8	

$$T=\frac{N^2}{ab}\sum_{i=1}^{c}\left[O_{1i}-\frac{a\,n_i}{N}\right]^2\Big/ n_i$$

$$=\frac{34^2}{16\times18}\left[\frac{\left(6-\frac{16\times9}{34}\right)^2}{9}+\frac{\left(3-\frac{16\times10}{34}\right)^2}{10}+\frac{\left(7-\frac{16\times7}{34}\right)^2}{7}+\frac{\left(0-\frac{16\times8}{34}\right)^2}{8}\right]$$

$$=4.01(0.34+0.29+1.97+1.78)=17.6$$

وحيث ان $\chi^2_{0.99,\,3}=11.3449$ ، وعليه $T=17.6>\chi^2_{0.99,\,3}=11.3449$.

نرفض فرض العدم H_0. أي ان هناك فرق بين وسيطات هذه الطرق.

ولتحديد $\hat{\alpha}$ أقل مستوى للرفض

$$1-\hat{\alpha}=p(X>T)=P(X>17.6)\qquad\Rightarrow\qquad 1-\hat{\alpha}>0.999\qquad\Rightarrow\qquad \hat{\alpha}<0.001$$

<u>تمرين(1)</u>: اختبر ان العينات التالية أخذت من مجتمعات لها نفس الوسيط

البيانات										العينة
21	29	17	29	31	15	30	42	42	35	الأولى
40	16	33	35	28	42	17	26	38	34	الثانية
30	38	23	31	30	41	36	30	29	17	الثالثة
25	29	36	24	33	42	27	22	34	39	الرابعة

اختبار كوكران "للمشاهدات المرتبطة"

عدد c من المعالجات تفرض عشوائيا وبشكل مستقل على عدد r من القطاعات ونتيجة كل معالجة تسجل 1 أو 0 (نجاح او فشل). وتوضع في جدول c من المعالجات و r من القطاعات، R_i تمثل مجموع الصف i و c_j مجموع العمود j .

فروض الاختبار

(1) نختار القطاعات عشوائيا من بين مجتمع جميع القطاعات الممكنة.

(2) تكون جميع المعالجات موحدة في النتيجة اما 1 او 0.

الاختبار

H_0 : تأثير المعالجات متساوي

H_1 : توجد فروق في تأثير المعالجات (معالجتين على الأقل مختلفتين)

ثم نحسب القيمة T كما يلي

$$T = c(c-1)\frac{\sum_{j=1}^{c}\left(c_j - \frac{N}{c}\right)^2}{\sum_{i=1}^{r} R_i\,(c-R_i)} = c(c-1)\frac{\sum c_j^2 - (c-1)N^2}{c\,N - \sum R_i^2} = (c-1)\frac{\sum R_j^2 - \frac{\left(\sum R_i\right)^2}{c}}{\sum R_i - \frac{\sum c_i^2}{c}}$$

حيث: c عدد المعالجات، R_i تمثل مجموع الصف i ، c_j مجموع العمود j ، $N = \sum R_i = \sum c_j$.

T تتبع تقريبا توزيع كاي تربيع بدرجة حرية $c-1$. نرفض فرض العدم H_0 اذا كانت $T > \chi^2_{1-\alpha,\, c-1}$.

مثال(1): اختيرت 12 مباراة عشوائيا من بن مباريات الدوري العام وسئل ثلاثة من الخبراء حول توقعاتهم نتيجة كل منها وتم تسجيلها. ثم بعد بعد لعب المباريات وتسجيل النتيجة الحقيقية قورنت مع توقعات كل خبير وفي حالى تطابق التوقع مع النتيجة الحقيقية تسجل رقم 1 وفي حالة الاختلاف يسجل الرقم 0 وكانت النتيجة كما في الجدول التالي:

الخبراء	1	2	3	4	5	6	7	8	9	10	11	12	المجموع
1	1	1	0	1	0	1	1	1	0	0	1	1	8
2	1	1	1	1	0	1	1	1	0	1	1	1	10
3	1	1	0	0	0	1	1	0	1	0	1	1	7
المجموع	3	3	1	2	0	3	3	2	1	1	3	3	25

اختبر عند مستوى معنوية $\alpha = 0.1$ الاختلاف في تقرير الخبراء.

الحل:

H_0 : توقعات الخبراء متطابقة (لايوجد اختلاف في تقرير الخبراء)

H_1 : هناك اختلاف في توقعات الخبراء (اثنين من الخبراء على الأقل مختلفين)

بحساب نحسب القيمة T نجد أن

$$c = 1, \qquad \sum c_j^2 = 3^2 + 3^2 + 1^2 + 2^2 + 0^2 + 3^2 + 3^2 + 2^2 + 1^2 + 1^2 + 3^2 + 3^2 = 65$$

$$N = \sum R_i = 10 + 8 + 7 = 25, \qquad \sum R_i^2 = 10^2 + 8^2 + 7^2 = 213$$

$$T=(c-1)\frac{\sum R_i^2-\frac{\left(\sum R_i\right)^2}{c}}{\sum R_i-\frac{\sum c_j^2}{c}}=(3-1)\frac{213-\frac{(25)^2}{3}}{25-\frac{65}{3}}=2.8$$

وحيث أن $\chi^2_{0.90,2}=4.6052$. فنجد أن $T=2.8<\chi^2_{0.90,2}=4.6052$

لا نستطيع رفض H_0 . أي ان هناك تطابق بين توقعات الخبراء والنتيجة الحقيقية للمباراة.

أقل مستوى للرفض $\hat{\alpha}$ يكون

$$1-\hat{\alpha}=p(X>T)=P(X>2.8)\Rightarrow\hat{\alpha}<1-0.75=0.25$$

<u>تمرين(1)</u>: اختيرت وعان من طرق المبيعات على 12 من النساء ، كل امرأة تعرضت للنوعين من المبيعات وطلب منها شراء منتج معين واستخدام نفس المنتج مع جميع النساء بالطريقتين ، وفي نهاية المر وضعت الدرجة 1 اذا شعرت المراة بالموافقة على شراء المنتج والدرجة 0 اذا لم توافق وكانت النتيجة كما في الجدول التالي:

طريقة المبيعات	1	2	3	4	5	6	7	8	9	10	11	12	المجموع
1	1	1	1	1	1	0	0	0	1	1	0	1	8
2	0	1	1	0	0	0	0	0	1	0	0	1	4
المجموع	1	2	2	1	1	0	0	0	2	1	0	2	12

اختبر عند مستوى معنوية $\alpha=0.001$ وجود اختلاف بين طريقتي المبيعات. ثم أوجد $\hat{\alpha}$.

<u>تمرين(2)</u>: استخدم اختبار كوكران للمشاهدات المرتبطة لاختبار هل هناك فرق معنوي بين المعالجات التالية عند مستوى معنوية $\alpha=0.01$.

المعالجات	1	2	3	4	5	6	7	8	9	10	المجموع
A	0	0	1	0	1	0	0	1	0	0	3
B	1	0	1	0	1	0	0	1	1	1	6
C	0	1	1	0	1	1	0	1	1	0	6
D	1	1	1	1	1	0	1	1	1	0	8
E	1	1	1	1	1	0	1	1	1	1	9
المجموع	3	3	5	2	5	1	2	5	4	2	32

ثم اوجد $\hat{\alpha}$.

اختبار مان هوتني لمجتمعين مستقليين

اذا كانت $F_X(x)$ دالة التوزيع الاحتمالي للمجتمع الأول X و $G_Y(x)$ دالة التوزيع الاحتمالي للمجتمع الثاني Y. وكان $X_1, X_2, \dots, X_n$ عينة من المجتمع الأول و $Y_1, Y_2, \dots, Y_m$ عينة من المجتمع الثاني. نجمع العينات في عينة واحدة حجمها $n+m$ وترتب من الأصغر الى الأكبر ثم نأخذ $R(X_i)$ هو ترتيب المفردات X_i ، $R(Y_j)$ هو ترتيب المفردات Y_j في العينة المشتركة لكل من i, j.

فروض الاختبار

(1) عينة عشوائية من كل مجتمع.

(2) العينتان مستقلتان.

(3) مقياس البيانات ترتيبي على الأقل.

$H_0: F_X(x) = G_Y(x)$ المجتمعين متساويين

$H_1: F_X(x) \neq G_Y(x)$ المجتمعين مختلفين

(A) اختبار من طرفين $H_0: P(X<Y) = 1/2$ $H_1: P(X<Y) \neq 1/2$

أي أن احتمال مفردات العينة الأولى X تقل عن مفردات العينة الثانية Y يساوي النصف أي تساوي كونها أقل مع كونها اكبر.

(B) اختبار من طرف أيمن $H_0: P(X<Y) \leq 1/2$ $H_1: P(X<Y) > 1/2$

أي أن فرض العدم يقول بأنه احتمال أن تقل X عن Y لا يتجاوز النصف.

(C) اختبار من طرف أيسر $H_0: P(X<Y) \geq 1/2$ $H_1: P(X<Y) < 1/2$

أي أن فرض العدم يقول بأنه احتمال ان تقل X عن Y لا يقل عن النصف.

ويكون الاختبار بحساب إحصاءة الاختبار T كما يلي

$$T = S - \frac{n\,(n+1)}{2}$$

حيث $S = \sum R(X_i)$ مجموع ترتيبات المجتمع X. و n حجم العينة X.

ويتم رفض فرض العدم H_0 عند مستوى معنوية α

(A) إذا كانت $T > w_{1-\alpha/2}$ أو $T < w_{\alpha/2} = -w_{1-\alpha/2}$. جدول A_7

(B) إذا كانت $T > w_{1-\alpha}$.

(C) إذا كانت $T < w_{\alpha} = -w_{1-\alpha}$.

مثال(1): أخذت عينة حجمها 48 طالب من إحدى المدارس لمعرفة اذا كان البنيان الجسدي مختلف بين طلاب الريف وطلاب المدينة وكان في العينة 12 منهم من الريف و 36 من المدينة وكانت بياناتهم كالتالي

	1	2	3	4	5	6	7	8	9	10	11	12
X_i	4.8	7.3	6.5	6.3	9	4.2	10.6	12.5	12.9	16.1	11.4	2.7
Y_j	12.7	14.2	12.6	2.1	17.7	11.8	16.9	7.9	16	10.6	5.6	5.6
	7.6	11.3	8.3	6.7	3.6	1	2.4	6.4	9.1	6.7	18.6	3.2
	6.2	6.1	15.3	10.6	1.8	5.9	9.9	10.6	14.8	5	2.6	4

استخدم اختبار مان هويتني بمستوى معنوية $\alpha = 0.05$ لاختبار هل هناك فرق في البنيان الجسدي لأبناء الريف وأبناء المدينة.

الحل:

بعد ترتيب البيانات يكون $S = \sum R(X_i) = 321$. و $n = 12$

$H_0 : P(X < Y) = 1/2$ البنيان الجسدي لأبناء الريف والمدينة متساوي

$H_1 : P(X < Y) \neq 1/2$ البنيان الجسدي لأبناء الريف والمدينة غير متساوي

$$T = S - \frac{n\,(n+1)}{2} = 321 + \frac{12(13)}{2} = 243$$

نبحث في جدول A_7 عند $n = 12$ و $m = 36$ وعندما لم توجد بالجدول نستخدم القانون

$$w_{\alpha/2} = \frac{nm}{2} + x_{\alpha/2}\frac{\sqrt{nm(n+m+1)}}{\sqrt{12}} = \frac{12(36)}{2} + 1.96\frac{\sqrt{12(39)(49)}}{\sqrt{12}} = 133.68$$

$$w_{1-\alpha/2} = \frac{nm}{2} + x_{\alpha/2}\frac{\sqrt{nm(n+m+1)}}{\sqrt{12}} = \frac{12(36)}{2} - 1.96\frac{\sqrt{12(39)(49)}}{\sqrt{12}} = 298.32$$

حيث $x_\alpha \approx N(0,1)$. حيث أن $T = 243 > w_{\alpha/2} = 133.68$ ، $T = 243 < w_{1-\alpha/2} = 298.32$

لا نستطيع رفض H_0 . أي البنيان الجسدي لأبناء الريف والمدينة متساوي.

اختبار ويلكوكسون لعدة مجتمعات مستقلة

اذا كانت $F_X(x)$ دالة التوزيع الاحتمالي للمجتمع الاول X و $G_Y(x)$ دالة التوزيع الاحتمالي للمجتمع الثاني Y. وكان $X_1, X_2, \ldots, X_n$ عينة من المجتمع الأول و $Y_1, Y_2, \ldots, Y_m$ عينة من المجتمع الثاني. نجمع العينات في عينة واحدة حجمها $n+m$ وترتب من الأصغر الى الأكبر ثم نأخذ $R(X_i)$ هو ترتيب المفردات X_i ، $R(Y_j)$ هو ترتيب المفردات Y_j في العينة المشتركة لكل من i, j.