دراسة اشتقاق التوزيع الاحتمالي لمعامل الارتباط

أ.م.د. محمد حبيب الشاروط

قسم الإحصاء/ جامعة القادسية

E mail: malsharood@yahoo.com

الخلاصـــــة

تعرف احصاءة معامل الارتباط المعتمدة على n من أزواج المشاهدات لظاهرتين من عينة عشوائية والممثلة بالمتغيرين العشوائيين $(X_t, Y_t)(t = 1,2,...,n)$ والذي يرمز لها بالرمز R بالصيغة الآتية:-

$$R = \frac{\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y})}{[\sum_{t=1}^{n}(X_t - \bar{X})^2 \sum_{t=1}^{n} (Y_t - \bar{Y})^2]^{\frac{1}{2}}}$$

حيث أن $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$ و $\bar{Y} = n^{-1}\sum_{t=1}^{n} Y_t$

يهدف البحث إلى دراسة اشتقاق التوزيع الاحتمالي لاحصاءة معامل الارتباط R بافتراض تحقق الاستقلالية بين المتغيرين العشوائيين X_t, Y_t وان المتغيرين يتوزعان توزيعاً طبيعياً ثنائي المتغيرات.

Abstract

We define the statistic known as the sample correlation coefficient based on n pairs of observed values of two characters which is represented by random variables $(X_t, Y_t)(t = 1,2,...,n)$ **as:**

$$R = \frac{\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y})}{[\sum_{t=1}^{n}(X_t - \overline{X})^2 \sum_{t=1}^{n} (Y_t - \overline{Y})^2]^{\frac{1}{2}}}$$

The aim of this paper is to find the dist. of R **when :**

(i) (X_i, Y_i) **are mutually independent r.vs if** $i \neq j$ **.**

(ii) X_t, Y_t **has abivariate normal dist.**

دراسة اشتقاق التوزيع الاحتمالي لمعامل الارتباط

المقدمـــــــة

يرمـز للاحصـاءة المعروفـة بمعامـل ارتبـاط العينـة المعتمـدة علـى **n** مـن أزواج المشـاهدات للظاهرتين الممثلتين بالمتغيرين العشوائيين $(X_t, Y_t)(t=1,2,..,n)$ بالرمز R وتحسب كالآتي:-

$$R = \frac{\sum_{t=1}^{n}(X_t-\bar{X})(Y_t-\bar{Y})}{[\sum_{t=1}^{n}(X_t-\bar{X})^2\sum_{t=1}^{n}(Y_t-\bar{Y})^2]^{\frac{1}{2}}}.................................(1)$$

حيث أن $\bar{X}=n^{-1}\sum_{t=1}^{n}X_t$ و $\bar{Y}=n^{-1}\sum_{t=1}^{n}Y_t$

يهدف هذا البحث إلى اشتقاق التوزيع الاحتمـالي لاحصـاءة معامل الارتبـاط R بـافتراض تحقق الآتي:-

(a) تحقق الاستقلالية بين أي زوجين مـن المشـاهدات للمتغيـرين العشـوائيين X, Y أي ان $(X_j, Y_j), (X_i, Y_i)$ مستقلان عن بعضهما **matually indp.** لكل $i \neq j$.

(b) ان التوزيع الاحتمالي المشترك للمتغيرين Y_t, X_t هو :

$$P_{X_t,Y_t}(x,y)=\frac{1}{2\pi\sigma x\sigma y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{X-\xi}{\sigma x}\right)^2-2\rho\left(\frac{X-\xi}{\sigma x}\right)\left(\frac{Y-\eta}{\sigma y}\right)+\left(\frac{Y-\eta}{\sigma y}\right)^2\right\}\right]....(2)$$

$$t=1,2,...,n$$
$$\sigma x>0, \sigma y>0$$
$$-1<\rho<1$$

تمثل الصيغة (2) أعلاه التوزيع الطبيعي ثنائي المتغيرات.

سوف نـدرس فـي هذا المبحث توزيـع الاحصـاءة R باعتبـار ان R متغير عشوائي احـادي وان الصيغة (2) ضمن توزيعات متعدد المتغيرات.

ولتحقيق الهدف من الدراسة سوف نستخدم بعض خصـائص التوزيـع فـي الصيغة (2) مـن اجل تحليل توزيـع R أولهـا ان ρ يمثـل معامـل ارتبـاط المجتمـع والـذي يحسـب بالصيغة الآتية:-

$$P_{x,y}=\frac{E(X_t-E(X_t))(Y_t-E(Y_t))}{\sqrt{Var(X_t)Var(Y_t)}}...........(3)$$

وثانيـاً ان المتغيـرين العشـوائيان X_t و Y_t كـل منهمـا يتـوزع توزيعـاً طبيعيـاً بمتوسـطات $E(X_t)=\xi$ $E(Y_t)=\eta$ وتباينـات علـى التـوالي $Var(X_t)=\sigma_X^2$ و $Var(Y_T)=\sigma_Y^2$ وفـي نفس الوقت سوف نحتاج بعض الافتراضات الإضافية سوف تذكر في حينها .

اشتقاق توزيع الاحصاءة R

بما ان الارتباط بين المتغيرين المعياريين $\frac{X_t-\xi}{\sigma_X}$ و $\frac{Y_t-\eta}{\sigma_Y}$ هو نفسه بين المتغيرين X_t و Y_t لذلك لاتوجد خسارة اذا تم افتراض $\xi=\eta=0$ و $\sigma_X=\sigma_Y=1$ سنركز الآن على دراسة التوزيع الشرطي إلى R للقيم الثابتة $X_1, X_2, ..., X_n$ بما ان التوزيع الشرطي إلى

اذا علمت Y_t يتوزع طبيعياً بمتوسط ρX_t وتباين $(1-\rho^2)$ على افتراض

$\xi=\eta=0$ و $\sigma_X=\sigma_Y=1$ فان: $R(1-R^2)^{-\frac{1}{2}}$ سيتوزع توزيع t اللامركزي مضروباً

بـ $(n-2)^{-\frac{1}{2}}$ بدرجة حرية $(n-2)$ بالمعلمة اللامركزية:- $\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\,\frac{\rho}{\sqrt{(1-\rho^2)}}$

وللحصول على التوزيـع اللاشرطي إلى $R(1-R^2)$ يجب احتسـاب القيمـة المتوقعـة لدالـة التوزيع الناتجة على توزيع المتغيرات $X_1, X_2,..., X_n$.

بما ان توزيع الدالة يعتمد فقط على الـ $X's$ من خلال الدالـة $\sum_{t=1}^{n}(X_t-\bar{X})^2$ فسوف نحتـاج فقط الخاصية التي تنص على ان هذه الدالة لها توزيع مربع كاي χ^2 بدرجة حرية $(n-1)$.

فلو افترضنا ان $V=R(1-R^2)^{-\frac{1}{2}}$ فان التوزيع الاحتمالي الشرطي إلى V هو

$$P_V(v\mid s)=\frac{\exp\left[\frac{-\rho^2 S}{2(1-\rho^2)}\right]}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}n-1\right)}(1+v^2)^{-\frac{1}{2}(n-1)}X\sum_{j=0}^{\infty}\frac{\Gamma\left(\frac{n-1+j}{2}\right)}{j!}\left(\frac{2\rho^2v^2S}{(1-\rho^2)(1+v^2)}\right)^{\frac{1}{2}j}\text{....(4)}$$

حيث ان $S=\sum_{t=1}^{n}(X_t-\bar{X})^2$ وبما ان التوزيع الاحتمالي إلى S هو

$$P_S(s)=\left[2^{\frac{1}{2}(n-1)}\Gamma\left(\frac{n-1}{2}\right)\right]^{-1}s^{\frac{1}{2}(n-3)}e^{-\frac{1}{2}S}\qquad(S>0)$$

وان
$$\int_0^{\infty}s^{\frac{1}{2}j}\exp\left[-\frac{1}{2}\rho^2 s(1-\rho^2)^{-1}\right]PS(s)ds=\frac{2^{\frac{1}{2}j}\,\Gamma\left(\frac{n-1+j}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}(1-\rho^2)^{\frac{1}{2}(n+j-1)}$$

فان التوزيع الاحتمالي إلى V : "2"

$$P_V(v)=\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}n-1\right)}(1+v^2)^{-\frac{1}{2}(n-1)}X\sum_{J=0}^{\infty}\frac{(2\rho)^J\left[\Gamma\left(\frac{n-1+j}{2}\right)\right]^2}{j!}\left(\frac{v^2}{1+v^2}\right)^{\frac{1}{2}j}\text{.....(5)}$$

وبإجراء عملية التحويل للمتغير $V=R(1-R^2)^{-\frac{1}{2}}$ نحصل على توزيع r كما مبين في التوزيع الاحتمالي بالصيغة الآتية :

$$P_R(r)=\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}n-1\right)}\sum_{j=0}^{\infty}\frac{\left[\Gamma\left(\frac{n-1+j}{2}\right)\right]^2}{j!}(2\rho r)^j,(-1\le r\le 1)\text{.......(6)}$$

ويمكن إيجاد صيغة أخرى للتوزيع من خلال استخدام العلاقة الآتية:

$$\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}n-1\right)=2^{-(n-3)}\pi\Gamma(n-2)$$

ومن الصيغة (6) وباستخدام بعض التبسيط الرياضي يمكن إيجاد عدد من الأشكال للتوزيع

الاحتمالي كما يأتي:-

$$P_R(r)=\frac{(n-2)(1-\rho^{2})^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\pi}\int_0^\infty \frac{dw}{(\cosh w-\rho r)^{n-1}}.....(6-1)$$

$$P_R(r)=\frac{(n-2)(1-\rho^{2})^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\pi}\int_1^\infty \frac{dw}{(w-\rho r)^{n-1}(w^2-1)^{\frac{1}{2}}}.....(6-2)$$

$$P_R(r)=\frac{(1-\rho^{2})^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\pi\Gamma(n-2)}\frac{d^{n-2}}{d(r\rho)^{n-2}}\left[\frac{\cos^{-1}(-\rho r)}{(1-\rho^2r^2)^{\frac{1}{2}}}\right].....(6-3)$$

$$P_R(r)=\frac{(n-2)(1-\rho^{2})^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\sqrt{2}(n-1)B\left(\frac{1}{2},n-\frac{1}{2}\right)(1-\rho r)^{n-\frac{3}{2}}}F\left(\frac{1}{2},\frac{1}{2},n-\frac{1}{2},\frac{1}{2}(1+\rho r)\right)......(6-4)$$

$$P_R(r)=\frac{(1-\rho^{2})^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\pi\Gamma(n-2)}\left(\frac{\partial}{\sin\theta\partial\vartheta}\right)^{n-2}\frac{\theta}{\sin\theta}.........(6-5)$$

حيث ان $\theta=\cos^{-1}(\rho r)$

ونبين هنا انه في جميع الصيغ المذكورة أعلاه فان $(-1\le r\le 1)$ **أما F(...) فتمثل دالة الـ (hypergeometric) .**

اما بالنسبة للصيغتين (6-1) و (6-2) فيمكن اشتقاقهما الواحدة من الأخرى باستخدام تحويل المتغير في التكامل أما الصيغة (6-4) هي متتابعة من الصيغة (6) ولحجم (n) متوسط نلاحظ ان متسلسلة (hypergeometric) تتقارب تدريجياً "2" ,"3" ,"4".

ولقد أمكن إيجاد دالة التوزيع الاحتمالي التجميعي لبعض القيم الصغيرة من n كما مبينة في أدناه فإذا افترضنا ان

$$F_{n,R}(r)=P_r[R\le r]$$

$n=3$

$$F_{3,R}(r)=\pi^{-1}\left[\cos^{-1}(-r)-\sqrt{\frac{(1-r^2)}{(1-\rho^2r^2)\cos^{-1}(-\rho r)}}\right].......(7-1)$$

$n=4$

$$F_{4,R}(r)=\rho^{-1}\sqrt{(1-\rho^2)(1-r^2)}F_{3,R}(r)-\pi^{-1}\left[\rho^{-1}\sqrt{1-\rho^2}-\cos^{-1}\rho\right].....(7-2)$$

$n=5$

$$F_{5,R}(r)=\frac{1}{2}\rho^{-1}\sqrt{(1-\rho^2)(1-r^2)}F_{4,R}(r)-\frac{1}{2}r(1-r^{2)}F_{3,R}(r)-\pi^{-1}\left[\frac{1}{2}\rho^{-1}(1+\rho^2)\sqrt{\frac{1-r^2}{1-\rho^2r^2}}\cos^{-1}(-\rho r)-\cos^{-1}(-r)\right]....(7-3)$$

$n=6$

$$F_{6,R}(r)=\frac{1}{3}\rho^{-1}\sqrt{(1-\rho^2)(1-r^2)}F_{5,R}(r)+\frac{1}{3}\rho^{-2}(1-\rho^2)rF_{4,R}(r)-\frac{1}{3}\rho^{-3}(1-\rho^2)^{\frac{3}{2}}\sqrt{1-r^2}F_{3,R}(r)+\pi^{-1}\left[\frac{1}{3}\rho^{-3}(1-4\rho^2)\sqrt{1-\rho^2}-\cos^{-1}\rho\right]....(7-4)$$

استطاع العالم **Garwood "3"** من إيجاد صيغة عامة للتوزيع الاحتمالي التجميعي عندما **n=2S+3** هي كالآتي:

$$F_R(r)=\pi^{-1}\cos^{-1}-(1-r^2)^{\frac{1}{2}}\left[(2S!)\right]^{-1}\pi^{-1}(1-\rho^2)^{S-1}\left[\left\{\Delta^S\rho^{2S}-\ldots-\binom{S}{1}\frac{\partial^2}{\partial\rho^2}\Delta^{S-1}\rho^{2S-2}+\binom{S}{2}\frac{\partial^4}{\partial\rho^4}\Delta^{S-2}\rho^{2S-4}+\ldots+(-1)\frac{\partial^{2S}}{\partial\rho^{2S}}\right\}\frac{\rho}{1-\rho^2}\frac{\cos^{-1}(-\rho r)}{(1-\rho^2r^2)^{\frac{1}{2}}}\right]\ldots\ldots\ldots(7-5)$$

والذي نحصل منها على التوزيع الاحتمالي التجميعي لحجم عينة (عدد فردي) ونلاحظ عند زيادة حجم العينة فان الصيغ أعلاه تصبح أكثر تعقيداً ويمكن تمثيل الدالة الاحتمالية بشكل منحني بسيط ضمن الفترة $-1\le R\le 1$ وبمنوال احادي (**antimode if n<4**) كما نلاحظ ايضاً انه بالإمكان احتساب قيمة

$F_R(0)=P_R[R\le 0]$ بشكل بسيط بما انه $(R\le 0)$ تكافئ إلى $\sum_{t=1}^{n}(X_t-\bar{X})Y_t\le 0$ سوف نحتاج إلى احتساب احتمال هذه الحادثة الأخيرة فإذا أعطينا $X_1,X_2,\ldots,X_n$ فسيصبح الاحتمال كالآتي:-

$$\Phi\left(\frac{-\rho}{\sqrt{1-\rho^2}}\sqrt{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right)=P_r\left[\frac{u}{\sqrt{\sum_{j=1}^{n}(X_j-\bar{X})^2}}\le\frac{-\rho}{\sqrt{1-\rho^2}}\right]$$

باخذ المعدل حول توزيع $\sum_{j=1}^{n}(X_j-\bar{X})^2$ الذي يتوزيع توزيع χ^2 بدرجة حرية (**n-1**) نلاحظ ان

$$P_r[R\le 0]=P_r\left[\frac{t_{n-1}}{\sqrt{n-1}}\le-\frac{\rho}{\sqrt{1-\rho^2}}\right]=P_r\left[t_{n-1}\le\frac{-\rho\sqrt{n-1}}{\sqrt{1-\rho^2}}\right]$$

ان هذه النتيجة لوحظت من قبل **Armsen** و **Ruben "2"**.

ويمكن إيجاد عزوم توزيع **R** من خلال حدود الدوال الهندسية الزائدية **hypergeometric functions**

$$\mu_1'=C_n\rho F\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}(n+1),\rho^2\right)$$

$$\mu_2'=1-\frac{(n-2)(1-\rho^2)}{n-1}F\left(1,1,\frac{1}{2}(n+1),\rho^2\right)$$

$$\mu_3'=C_n\left[\rho F\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}(n+1),\rho^2\right)-\rho^{-1}(n-1)(n-2)\right]$$

$$X\left\{F\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}(n-1),\rho^2\right)-F\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}(n+1),\rho^2\right)\right\}$$

$$\mu_4'=1+\frac{(n-2)(n-4)(1-\rho^2)}{2(n-1)}\left[F\left(1,1,\frac{1}{2}(n+1),\rho^2\right)\right]-$$

$$\frac{n(n-2)(1-\rho^2)}{4\rho^2}\left[F\left(1,1,\frac{1}{2}(n+1),\rho^2\right)-1\right]$$

حيث ان $C_n = \frac{2}{n-1}\left[\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right]^2$

قام Ghosh "3" بالحصول على مفكوك μ_1' وكذلك العزوم المركزية الثلاثة الأولى من خلال القوى المعكوسة إلى m-(n+6) كالآتي:-

$$\mu_1 = \rho - \frac{1}{2}\rho(1-\rho^2)m^{-1}\left[1+\frac{9}{4}(3+\rho^2)m^{-1}+\frac{3}{8}(121+70\rho^2+25\rho^4)m^{-2}\right]+0(m^{-4})$$

$$\mu_2 = \frac{(1-\rho^2)^2}{m}\left[1+\frac{1}{2}(14+11\rho^2)m^{-1}+\frac{1}{2}(98+130\rho^2+75\rho^4)m^{-2}\right]+0(m^{-4})$$

$$\mu_3 = -\frac{\rho(1-\rho^2)^3}{m^2}\left[6+(69+88\rho^2)m^{-1}+\frac{3}{4}(797+1691\rho^2+1560\rho^4)m^{-2}\right]+0(m^{-5})$$

$$\mu_4 = \frac{3(1-\rho^2)^4}{m^2}\left[1+(12+35\rho^2)m^{-1}+\frac{1}{4}(436+2028\rho^2+3025\rho^4)m^{-2}\right]+0(m^{-5})$$

يمكن احتساب التحيز للمقدر R كتقدير للمعلمة ρ بشكل تقريبي كالآتي:

$-\frac{1}{2}\rho(1-\rho^2)n^{-1}$ **وتباين المقدر R يحسب كالآتي "4":**

$$Var(R) = (1-\rho^2)^2 n^{-1}\left[1+\frac{11\rho^2}{2n}+\ldots\right]$$

وبشكل تقريبي يحسب كالآتي "1" , "4":

$$Var(R) = (1-\rho^2)^2 n^{-1}, as \quad n \to \infty$$

References

1- Garthwaite, P.H. and Grawford ,J.R.(2004),"The distribution of the difference between t-var." , Biometrika ,91:987-994.
2- Johnson,N.L. and Kotz,S.(1970),"Distribution in statistics continuous univariate distributions-2.
3- Kenney,J.F. and Keeping ,E.S. (1961), "Mathematics of statistics", Pt.2,2nd ed.5th Printing Hc.Good. ISBN.
4- Konishi,S.(1978),"An approximation to the distribution of the sample correlation coefficient ",Biometrika, 65:654-656.