استخدام الأشكال البيانية في عملية النمذجة الإحصائية

هاسميك أنترانيك وارطان

الكلية التقنية الإدارية الموصل / هيئة التعليم التقني

الاستلام 27 / 09 / 2011

القبول 03 / 04 / 2012

Abstract

Through studying numerical summaries and graph of data movement, to achieveproper and suitable type of data distribution, in simple, clear and minimizing errors, in the same time the graphs and summaries illustrate the possibility of unavailing of available distribution that we have, and the possibility of create data distribution in the case of disadvantages of identification to available data distribution.

المستلخص

من خلال دراسة الملخصات العددية والرسومات التوضيحية لحركة البيانات نحاول التوصل إلى نوع التوزيع المناسب للبيانات بشكل واضح وسهل وأقل احتمالاً للخطأ. وقد أوضحت لنا الرسومات والملخصات أنه من الممكن أن لا تفيد مجموعة التوزيعات المتوفرة لدينا وانه من الممكن ابتكار توزيع للبيانات في حالة عدم الانتفاع بمطابقة البيانات للتوزيعات المعروفة.

المقدمة

تعد النمذجة اللبنة الاساسية لفهم البيانات، فعن طريقها يمكن الوصول الى النموذج الصحيح الممثل للبيانات،والباقي هو تطبيق هذا النموذج للوصول الى ما يبغيه الباحث من التنبؤ أو أختبار فرضية معينة.

وهنالك العديد من النماذج الاحصائية التي بنيت على مدار السنين يستطيع الباحث أستخدامها وهي التوزيعات الاحتمالية، ولكن يجب التأكد من البيانات قيد المعالجة إذا كانت تتبع هذا النموذج أو ذاك بنسبة عالية وبخلافه فأن إستخدام النماذج الجاهزة مباشرة قد تؤدي الى

إغفال عنصر مهم في تحديد مساره، بالاضافة الى أن هنالك بعض الحالات من البيانات التي لا تتبع نموذج معين بالذات،فإيجاد النموذج الملائم لها مسألة مهمة جداً.

النماذج التقليدية الجاهزة الموجودة في العديد من المصادر تتعامل مع العناصر المستقرة من البيانات وتبني عليها النموذج بتحديد علاقة معينة. ولكن دراسة الدالة الكمية للبيانات تعطينا منظور جديد وفي بعض الاحيان أوضح للبيانات وتساعدنا في تحديد النموذج الملائم أخذين بنظر الاعتبار العنصرين (المستقر والعشوائي) وتحديد منظور جديد لبناء النموذج.

تعد الاختبارات الإحصائية إحدى الوسائل الأساسية والمهمة في دراسة البيانات وفهم طبيعتها والتوزيع المناسب لها. لا يوجد إختبار أمثل فلكل إختبار نسبة من الخطأ الذي من الممكن أن يقع أثناء إتخاذنا للقرار، والذي نسميه بالتعبير الإحصائي الخطأ من النوع الأول والثاني. وكذلك لكل إختبار ميزاته وعيوبه التي تميزه عن الاخر. علماً أن تطبيق عدة إختبارات لغرض معين مثل إختبار إتباع البيانات للتوزيع الطبيعي، لا يعني أنها ستؤدي الى نفس النتيجة، لكن في نفس الوقت عندما تكون نتائج الاختبارات متطابقة فهذا سوف يعزز النتيجة التي تم التوصل إليها.

تعاريف ومصطلحات

للتعرف على الرسومات التي سوف تحدد طبيعة البيانات سوف نستخدم بعض الدوال، وهي:

❖ **الدالة التراكمية Cumulative Distribution Function (CDF)**[1]

لأي متغير عشوائيX، فأننا نعرف الدالة $F(x)$

$$\mathbf{F(x) = P(X \le x)}$$

حيث أن x هو عدد حقيقي، فأن الدالة $F(x)$ تسمى الدالة التراكمية (أو دالة التوزيع)، ويرمز بألحرف الكبير للدلالة على المتغير العشوائي والحرف الصغير لقيمة محددة، وبالتالي فأن الدالة هي إحتمال أن يكون المتغير X أقل أو يساوي قيمة معطاة x. ومن أهم خصائصها أنها متزايدة و $0 \le F(x) \le 1$.

❖ **دالة الكثافة الاحتمالية probability Density Function (PDF)**[2]

ويرمز لها عادة $F(x)$، وهي تزودنا بوسيلة أخرى لتعريف التوزيع وتحدد بالصيغة

$$\mathbf{f(x)dx = P(x \le r.v.x \le x + dx)}$$

حيث أنه تم دراسة x كمتغير مستمر لشيوع الدوال الاحتمالية للمتغيرات المستمرة في الواقع. وأن dx مدى صغير بشكل متناهي لـ x للمساحة تحت المنحنى، حيث أن:

$$\begin{aligned} f(x)dx &= \Pr obability(observation\, in\, dx) \\ &= F(x + dx - F(x) \end{aligned}$$

$$f(x) = \frac{df(x)}{d(x)}$$

❖ **الدالة الكمية Quantile Function (QF)**[2]

ويرمز لها بـ Q(P)، حيث أنه تمثل القيم الكمية للمتغير العشوائي المقابلة للاحتمال P

$$Q(P) = \text{the value of } x \text{ for which probability } (x \le x_p)$$
$$= x_p$$

أي أن الدالة الكمية هي الدالة العكسية للدالة التراكمية. حيث

$$P = F(X) \quad \rightarrow \quad X = Q(P)$$
$$Q(P)=F^{-1}(X)$$

حيث أن $0 < P < 1$ وأن P تمثل الاحتمال المتراكم (قيمة الدالة التراكمية) وغالباً ما يتم ترتيب قيمه تصاعدياً تبدأ من الصفر لتنتهي بالواحد. فالوسيط هو عبارة عن Q(1/2) وكذلك الربيعين Q(1/4) ، Q(3/4). وتعتبر جداول التوزيع الطبيعي أبسط مثال على هذه الدالة حيث نستطيع الحصول على قيمة المتغير لقيمة إحتمالية متراكمة معين.

ولتمثيل الدالة عن الطريق الرسم نحتاج الى الاحتمال المتراكم P (كمدخلات على المحور السيني) مقابل قيم x (كمدخلات على المحور الصادي).

❖ دالة الكثافة الكمية (3)Quantile Density Function(QDF)

بنفس الطريقة التي نشتق فيها دالة الكثافة الاحتمالية من الدالة التراكمية فأننا من مشتقة الدالة الكمية نحصل على دالة الكثافة الكمية (QDF).

$$q(p) = \frac{d\,Q(P)}{d(P)} = \frac{d\,x}{d\,p} = \frac{\Delta x}{\Delta p}$$

وعليه لتمثيل هذه البيانات، يتم رسم الاحتمال المتراكم P على المحور السيني مقابل $\frac{\Delta x}{\Delta p}$ المحور الصادي.

وبما أن $Q(P)$ هي دالة غير متناقصة فهذا يتبعه أن الميل $q(p)$ يكون أيضاً غير متناقص على الاقل للفترة $0 \le P \le 1$.

❖ الدالة الكمية للكثافة (3)Density Quantile Function (P – PDF)

وهي مقلوب دالة الكثافة الكمية وتمثل احتمالية تشبه دالة الكثافة الاحتمالية $f(x)$، إلا أن مدخلاتها هي الاحتمال المتراكم P. وبما أن

$$x = Q(P)$$
$$P = F(x)$$

لأي زوج من قيم (x , p)، ومن تعريف الاشتقاق فان

$$\left(\frac{dx}{dp}\right)\left(\frac{dp}{dx}\right) = 1$$

$$\therefore \frac{d}{dp}Q(P) \cdot \frac{d}{dx}F(X) = 1$$

$$q(p) \cdot \frac{d}{dx}F(x) = 1$$

$$\frac{d}{dx}F(X) = \frac{1}{q(p)}$$

$$f(X) = \frac{1}{q(p)}$$

المدخلات لـ p

إذن تكتب الدالة $F(x)$ بدلالة p ، أي يعاد ترميزها للرمز $f_p(P)$ وبذلك يمكن القول بأن:-

$$f_p(P) = \frac{1}{q(p)}$$

شكل المنحنى

نستطيع تقسيم أشكال المنحنيات الى:

1) منحنيات متماثلة، وهي المنحنيات التي تنطبق أو تشابه الدالة $y = x$ أي أنها تكون ممتدة على خط 45°.
2) المنحنيات الملتوية، أي غير المتماثلة. وهي عندما تتراكم البيانات في أحد أطراف المنحنى وذيل في الطرف الآخر. فإذا كان الذيل جهة اليمين تسمى موجبة الالتواء، أو الذيل نحو اليسار وفي هذه الحالة سالبة الالتواء.
3) المنحنيات على شكل U. وفي هذه الحالة تكون التكرارات عالية في أطراف مدى المتغير العليا والدنيا.
4) المنحنيات على شكل J. وممكن أن تكون J positive (أي على شكل ) أو Jnegative (أي على الشكل ).

وعليه لتقصي شكل البيانات نحتاج الى معرفة ألالتواء (Skewness) وكذلك مدى إمتداد الذيل في التوزيع والذي له علاقة بما يسمى التفلطح (Kurtosis).

مقاييس الشكل

1. الالتواء Skewness

كما أشرنا سابقاً فأن تراكم التكرارات حول الطرف الأيسر او الأيمن للبيانات يدل على الالتواء، ولتقصي ذلك نحتاج الى بعض المقاييس الاحصائية. وهي الوسيط (Median,m) والربيع الأول (Lowerquartile,lq) والربيع الثالث (Upperquartile,uq) والمعروفة طريقة حسابها، ومنه ستظهر بعض مقاييس الالتواء. وهي[3]

Upper quartile deviation(uqd) = uq − m
Lower quartile deviation(lqd) = m − lq

والفرق بينهما هو:

Quartile difference (qd) = uqd – lqd
= (uq – m) – (m – lq) = lq + uq – 2m

فعندما يكون التوزيع متماثلاً فأن (qd) تساوي صفر، وعندما تكون قيمته كبيرة وموجبة فهذا يدل على ذيل طويل نحو اليمين (ألتواء موجب) والعكس صحيح.

بالاضافة الى مقياس بسيط أخر يدل على مدى أنتشار البيانات هو

$$\textbf{interquartile range}\ (iqr) = uq - lq$$

وللحصول على مقياس خالي من وحدات القياس والذي يسمى Galton'sskewnesscoefficient (g)

$$g = \frac{dq}{iqr} = (uq - lqd) \Big/ (uqd + lqd)$$

حيث ان معامل كالتون تقع بين (1,-1). وتمثل القيم طرفين أيسر وأيمن يحسبان بالشكل التالي[3]:

$$\textbf{Lower p} - \textbf{deviation},\ ld(p) = m - \hat{Q}(p)$$

وإذا رمزنا لـ (1- p) بـ q، فيكون لدينا

$$\text{Upper p deviantion},\ ud(p) = \hat{Q}(q) - m$$

وللأغراض التطبيقية نستخدم

$$\hat{Q}(P_r) = X_{(r)}$$
$$\hat{Q}(q_r) = X_{(n+1-r)}$$

وبما أنه في الدراسات العملية نتعامل مع X وليس $Q(P)$ فانه من الممكن الاستعاضة عن $\hat{Q}(P)$ بـ $X_{(r)}$ و $\hat{Q}(q)$ بـ $X_{(n+1-r)}$، حيث $X_{(r)}$ تمثل المشاهدة ذات التسلسل r.

عند رسم ud(p) مقابل lq(p) والذي نسميه Deviation سوف نحصل على خط مستقيم 45° في الشكل البياني المتماثل تماماً. وكذلك يمكن الاستفادة منها لتعريف inter p-range (وتسمى أيضاً .quasi-range H, spread fun) و p- difference حيث

$$\textbf{inter p} - \textbf{range},\ ipr(p) = \hat{Q}(q) - \hat{Q}(p) = ld(p) + ud(p)$$

$$\textbf{p} - \textbf{difference},\ pd(p) = ud(p) - ld(p) = \hat{Q}(q) + \hat{Q}(p) + 2m$$

تكون هذه القيمة صفر عندما يكون التوزيع متماثل.

وللحصول على مقياس للالتواء نحول pd(p) الى الحالة القياسية بقسمتها على (iqr) أو ipr(p) للتخلص من المقاييس، وهذا ينتج لنا

$$\textbf{Galtonp} - \textbf{skewness},\ g(p) = \frac{pd(p)}{iqr}$$

$$\textbf{p} - \textbf{skewness index},\ g^{*}(p) = pd(p) \big/ ipr(p)$$

$$= \left(x_{(r)} + X_{(n+1-r)} + 2m\right) \Big/ \left(X_{(n+1-r)} - x_{(r)}\right)$$

كذلك تتراوح قيمتها مابين $(1,-1)$

2. التفلطح Kurtosis

يوجد مقياس أخر يحدد شكل النموذج وهو، التفلطح وهو متزامن مع مدى وجود قمة او تسطح مركز التوزيع، حيث يتزامن وجود القمة أو تفلطحها في مركز التوزيع مع الذيل.فوجود قمة عالية يتزامن معها ذيل قصير، والقمة المفلطحة مع ذيل طويل. ويمكن قياس تحدب أو تفلطح التوزيع بعدة طرق، أحداها هي [inter p-range[ipr(p)والتي تم تحويلها الى الحالة القياسية بأعتماد (interquartile range(irq، ويكون لدينا ما يعرف بـ(shape index,t(p

$$t(p) = \frac{ipr(p)}{iqr}$$

حيث ترسم هذه الدالة عند $0 \le p \le 0.5$ نحصل عن طريقها على قيمة إمتداد التوزيع (أو ما يسمى الذيل) للقيم الصغيرة من p ومدى تمركزها في منتصف الاحتمال القريب من (0.5)، فعندما تكون p قريبة من (0.5) فالتوزيعات التي لديها قمة عالية تميل أن يكون لديها (t(p أصغر من التوزيعات التي قمتها أقل. وبعكسه للقيم الصغيرة من p أي القريبة من الذيل.فالتوزيعات ذات الذيول الطويلة سيكون لدينا (ipr(p كبيرة وإعتماداً عليها تكون قيم (t(p كبيرة،حيث أنه يمكن أعتبار هذا المقياس مقياساً نسبياً للمقارنة بين أكثر من توزيع.

يستخدم المقياس للتوزيعات المتماثلة والتوزيعات غير المتماثلة من المفيد أن نقيس الذيل على حدى وذلك باستخدام.

$$ut(p) = \frac{ud(p)}{uqd}$$

$$lt(p) = \frac{ld(p)}{lqd}$$

الشكل Q-Q[7]:

تهدف معظم الدراسات الإحصائية إلى تحويل البيانات من مجرد أرقام عشوائية إلى نموذج إحصائي تتمثل بدالة يتم من خلالها فهم توزيع البيانات وهذا ما يسمى بعملية النمذجة الإحصائية. ومن أهم الطرق المستخدمة في النمذجة هي الشكل Q-Q. ففي هذا الشكل يتم رسم الدالة الكمية للعينة مقابل الدالة الكمية للمجتمع أو الدالة الكمية لمجموعة من البيانات مقابل الدالة الكمية لمجموعة ثانية.من أهم خواص هذا الرسم أنه إذا كان الرسم على شكل خط مستقيم بمعامل إنحدار يساوي (1) ويمر خلال نقطة الأصل، تكون العينة منطبقة على المجتمع. ولكن إذا كان الشكل خط مستقيم بأنحدار (1) ولكن لايمر من نقطة الاصل فأن التوزيعين هي نفس التوزيعات ولكن بتغير بثابت معين. أما إذا كان الشكل البياني خط بأنحدار لا يساوي الواحد الصحيح ولا يمر خلال نقطة الأصل، فان التوزيعين متطابقين إلى حد ما. وعليه فأن إنحراف الشكل البياني عن الخط المستقيم يعتبر مقياس بياني يبين إنحراف المتغير عن المتغير الاخر.

فأن الخط غير المستقيم يعطينا مدلول أن العينة والمجتمع قد يكون متطابقين في مكان ما، فعليه عندما يكون الشكل مستقيم مختلف عندما يكون لدينا خط منحني. فعندما نرسم الدالة الكمية للبيانات مقابل الدالة الكمية للتوزيع، فأن نقاط التطابق تعني أن التوزيع متماثل في تلك النقطة وأبتعاده يبين إختلافه وخاصة إذا كانت في الإطراف فأنها تبين إختلافها في الذيول. فمثلاً إذا كانت الدالة الكمية للعينة متجه نحو الأسفل معناها أن توزيع المجتمع له ذيل أعلى من العينة.

الجانب العملي

تعد النمذجة الإحصائية اللبنة الأساس للدراسات الإحصائية أذ أننا مبدئياً ومنذ ألقاء النظرة الأولى على البيانات يجب علينا أن نجري عليها الدراسة،يجب تفهم هذه البيانات وطبيعتها من ناحية تواجد البيانات ضمن مديات معينة بشكل كثيف في القمم أو الذيول ولفهم هذه النقاط فأننا نحتاج إلى إيجاد ملخصات للقيم تم ذكره سابقاً ورسومات للدوال الكمية والاحتمالية.

بداية تم توليد بيانات باستخدام دالة RUND في برنامج أكسل تتوزع توزيع منتظم قياسي (a=0 , b=1) ترتيبها متكونة من 500 مشاهدة، يمكن أعتبار هذه المشاهدات أساس للتحويل الى العديد من التوزيعات التي تمتلك دالة كمية معروفة. كما أن هذه المشاهدات في حال ترتيب القيم تصاعدياً نلاحظ أن P تبدأ بالتزايد من الصفر الى الواحد وهي بهذه الصفة تشابه الدالة التراكمية ولذلك تم التعريف P=F(X). تم تطبيق الدوال والرسومات المطروحة في هذا البحث عن نموذج محاكاة وليس نموذج بيانات من الواقع لأن البحث بحاجة إلى أكثر من نموذج بيانات لاستنتاج الفرق بين التوزيعات ولحل المشاكل التي يوجها الباحث أثناء بناء النموذج.

يتم في الجانب التطبيقي محاكاة البيانات المولدة حسب التوزيع المنتظم القياسي لتحويلها الى متغيرات عشوائية تتوزع أسي بمعلمة $\lambda = 1$ عن طريق الدالة الكمية:

$$Q(P) = -\ln(1-u)$$

حيث أنه $u \sim U_n(0,1)$.

كما تم تحويل بيانات التوزيع المنتظم القياسي إلى توزيع ويبل وفق الدالة الكمية التالية[3]:

$$\mathbf{Q(P)} = \left(-\ln(\mathbf{1}-\mathbf{u})\right)^{2}$$

حيث تم التعويض عن قيمة λ بـ 1 و $\beta = 2$ في الدالة الكمية المشار إليها أنفاً.

تم تصنيف البيانات الأصلية حسب مقاديرها ورتبت بشكل تصاعدي في عمود أُطلق عليه اسم x. هذه الخطوة الأولى والأساسية لكل التحليلات. بعد ذلك تم تخصيص عمود آخر يمثل احتمال كل قيمة من قيم المتغير x. لنفرض أن r تمثل تسلسل المشاهدة.

ولحساب قيمة $\mathbf{P}_r$ المقابل للقيمة x_r نفرض كقدير أولي بأن $\mathbf{P_r} = r/n$، حيث أن $r=1,2,\ldots,n$ وبذلك تكون قيمة $0 \leq p \leq 1$ ولكن المشكلة أن القيم الاحتمالية تكون غير متماثلة ولتحويلها إلى التماثل سنستخدم الصيغة:

$$\mathbf{P}_r = \frac{(r-0.5)}{n}$$

حيث تمثل r تسلسل المشاهدة $=1,\ldots\ldots,500$ r ، و n العدد الكلي للمشاهدات n=500 وبذلك تكون القيم المحسوبة لـ P متماثلة ومتناسبة مع الترتيب التصاعدي لقيم x أي أصبح لدينا زوج من البيانات متمثل بـ (x_r,P_r)، حيث يمثل P الاحتمال ويمثل x كمية الاحتمال P للعينة (sample p-quantile). وبذلك تكون هذه الـ500 مشاهدة تمثل مجتمع التوزيع المنتظم القياسي، ولتحويل التوزيع الى توزيع أسي وتوزيع ويبل للمجتمع تم استخدام الصيغتين التاليتين:

$$Q(P) = -\ln(1-P)$$

$$\mathbf{Q(P)} = \left(-\ln(\mathbf{1}-\mathbf{P})\right)^2$$

وبذلك أصبح لدينا 4 أعمدة نهائية كل عامود يتكون من 500 مشاهدة. تم ترميز العمود الأول على أنه X

$$\mathbf{X} = -\ln(\mathbf{1}-\mathbf{u})$$

المتمثل بتوزيع أسي لعينة حجمها 500 مشاهدة.

و I المتمثل بتوزيع أسي للمجتمع (على فرض 500 مشاهدة)

$$\mathbf{I} = -\ln(\mathbf{1}-\mathbf{P})$$

و Y المتمثل بتوزيع ويبل للعينة حجمها 500 مشاهدة

$$\mathbf{Y} = (-\ln(\mathbf{1}-\mathbf{u}))^2$$

و H المتمثل بتوزيع ويبل للمجتمع (على فرض 500 مشاهدة)

$$\mathbf{H} = (-\ln(\mathbf{1}-\mathbf{P}))^2$$

كما اننا حاكينا بيانات للتوزيع اللوجستي الناتج من جمع التوزيع الاسي مع التوزيع الاسي العكسي كما في الصيغة:

$$C = Q(P) = \ln(u) - \ln(1-u)$$

والدالة الكمية للتوزيع للمجتمع هي:

$$E = Q(P) = \ln(P) - \ln(1-P)$$

وقبل رسم الدوال الخمسة التي تم عرضها في الجانب النظري حاولنا تطبيق أختبار مطابقة البيانات للتوزيع الطبيعي حسب أختبار أندرسون، أختبار كولموجروف سميروف ومربع كاي وتم عرض بعض النتائج في الأشكال 3,2,1.

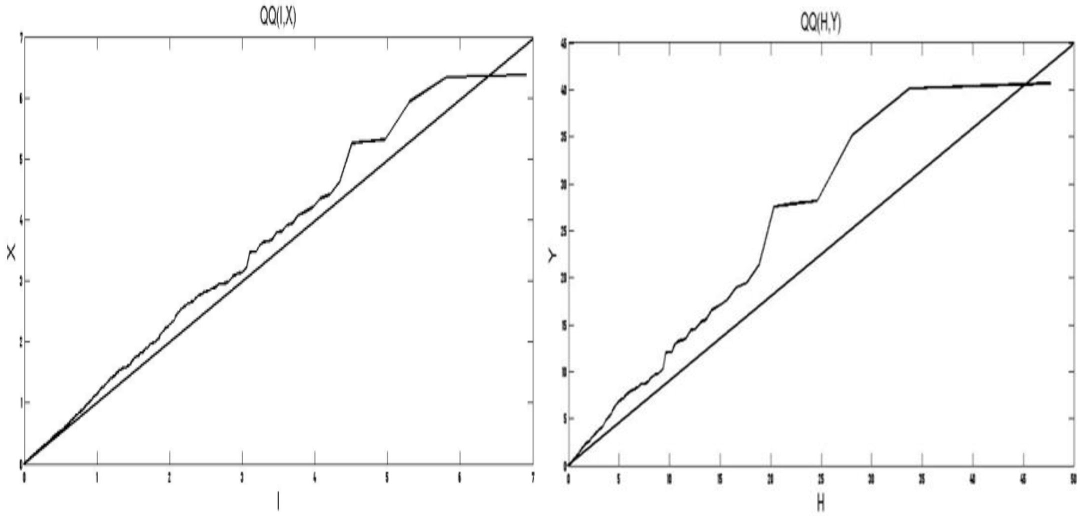
نلاحظ من الشكل (2) المتعلق بمطابقة البيانات التوزيع اللوجستي مع بيانات المجتمع الطبيعي أن التوزيع اللوجستي أنطبق على التوزيع الطبيعي. وهذا خلل كبير لذلك وجب العودة الى الرسومات والملخصات لتقدير نوع التوزيع المناسب قبل التسرع والحكم عليه عن طريق Normality test.

كما أنه تم أختبار المطابقة بـ χ^2 لكل توزيع مع عينة وكانت النتائج كالآتي:-

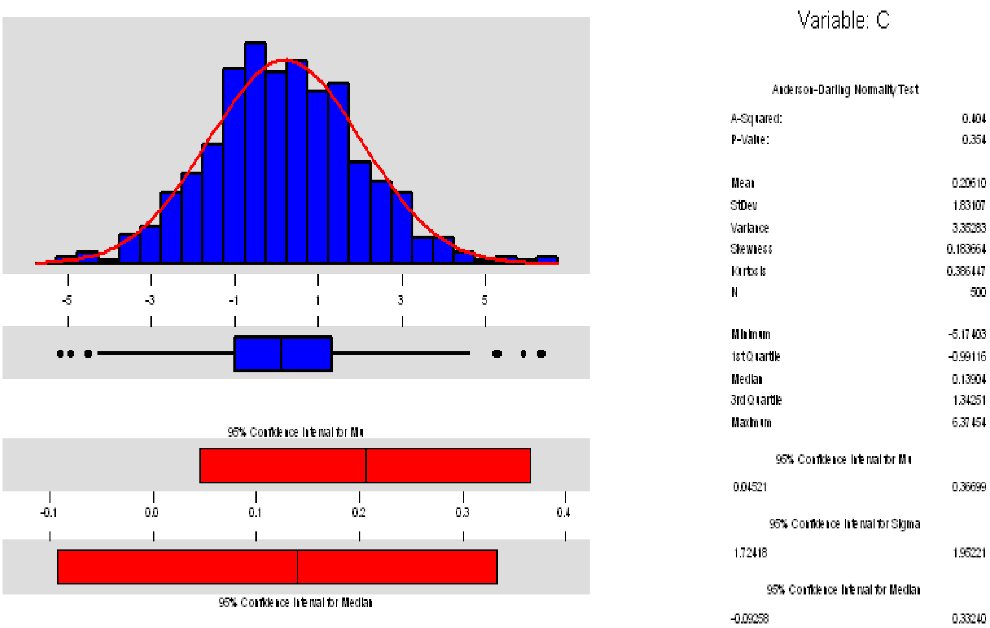
القيمة المحسوبة لـ χ^2	توزيع (مجتمع)	عينة
7600.912	أسي $\lambda = 1$	أسي $\lambda = 1$
61.62	ويبل $\lambda = 1, \beta = 2$	ويبل $\lambda = 1, \beta = 2$

نلاحظ أن قيم χ^2 كبيرة أي أنها تعبر عن وجود فروق معنوية بين القيمة المتوقعة وقيم المجتمع الحقيقية. وبذلك فأنه من الممكن أن يبعدنا هذا الاختبار عن التوزيع الحقيقي الذي تتوزع فيه البيانات حيث أنه أعتبر العينة المأخوذة من المجتمع الاسي لا تنتمي إليه علماً أننا حاكينا أسياً كذلك بالنسبة لتوزيع ويبل.

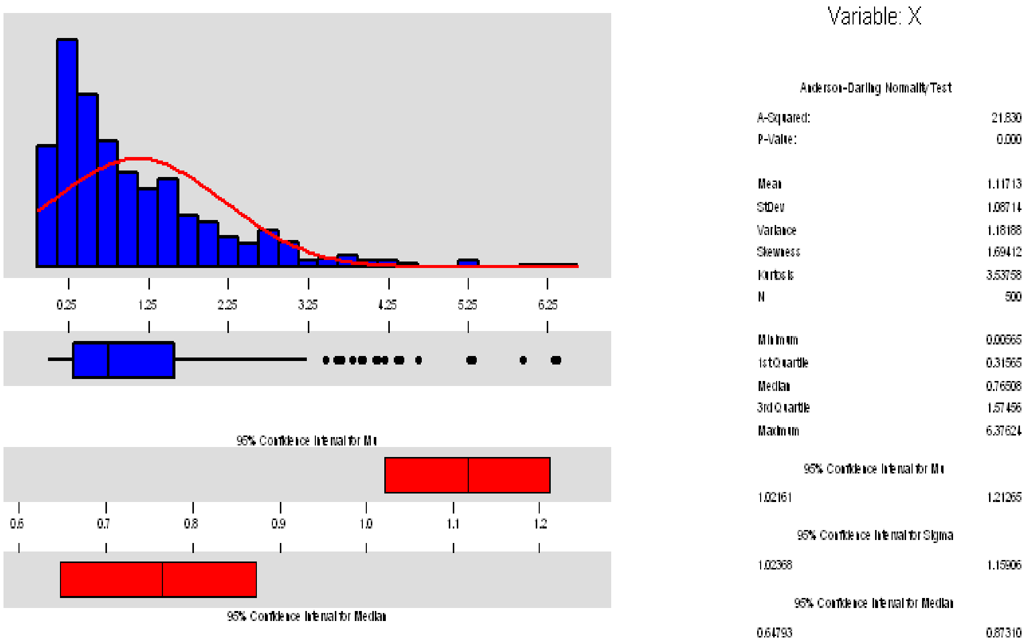
شكل (1): يبين شكلي Q-Q للجدول السابق



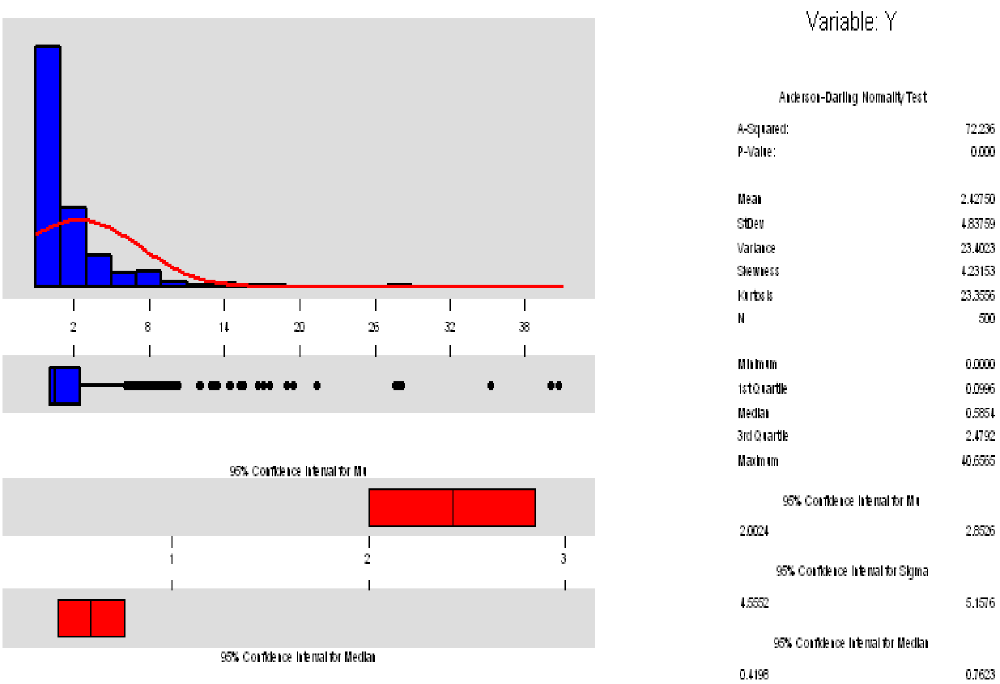
شكل (2): المدرج والمنحنى التكراري إضافة إختبار أندرسن وبعض المقاييس الإحصائية للمتغير *C*



شكل (3): المدرج والمنحنى التكراري إضافة إختبار أندرسن وبعض المقاييس الإحصائية للمتغير X



شكل (4): المدرج والمنحنى التكراري إضافة إختبار أندرسن وبعض المقاييس الإحصائية للمتغير *Y*



مما سبق نلاحظ أن تدقيق في حركة البيانات وأشكالها الهندسية يجب أن يكون قبل إجراء الاختبارات الإحصائية. وقد تم تطبيق لاختبار الطبيعي وأختبار χ^2 على أنواع مختلفة من البيانات منها أحادية الذيل الأيسر (كما في الاسي العكسي) واحادية الذيل الايمن (كما في

التوزيع الاسي وتوزيع ويبل) ومنتهية (كما في التوزيع المنتظم (0,1)) وذوات الذليلين كما في التوزيع الطبيعي واللوجستي المتماثل واللوجستي الملتوي. تم رسم 15 شكل للتوزيعات ذوات الذيل الأيمن للاسي ويبل ولكون الرسومات كثيرة تم انتقاء بعض الأشكال المدرجة أدناه:–

الأشكال البيانية

- **General shape**

Fig	Sample function	(Y)	(X)	Remarks
5	Quantile , $\hat{Q}$ (P)	$\mathbf{X}_{(r)}$	$\mathbf{P}_{(r)}$	دالة غير متناقصة
6	CDF, $\hat{F}(X)$	$\mathbf{P}_{(r)}$	$\mathbf{X}_{(r)}$	دالة غير متناقصة
7	Quantile density, $\hat{q}$ $(\mathbf{P})$	$\mathbf{D_x}/\mathbf{D_p}$	Mid-$\mathbf{P}_{(r)}$	
	p-PDF, $\hat{f}_P(\mathbf{P})$	$\mathbf{D_P}/\mathbf{D_x}$	Mid-$\mathbf{P}_{(r)}$	
	PDF, $\hat{f}(\mathbf{X})$	$\mathbf{D_P}/\mathbf{D_x}$	Mid-$\mathbf{X}_{(r)}$	

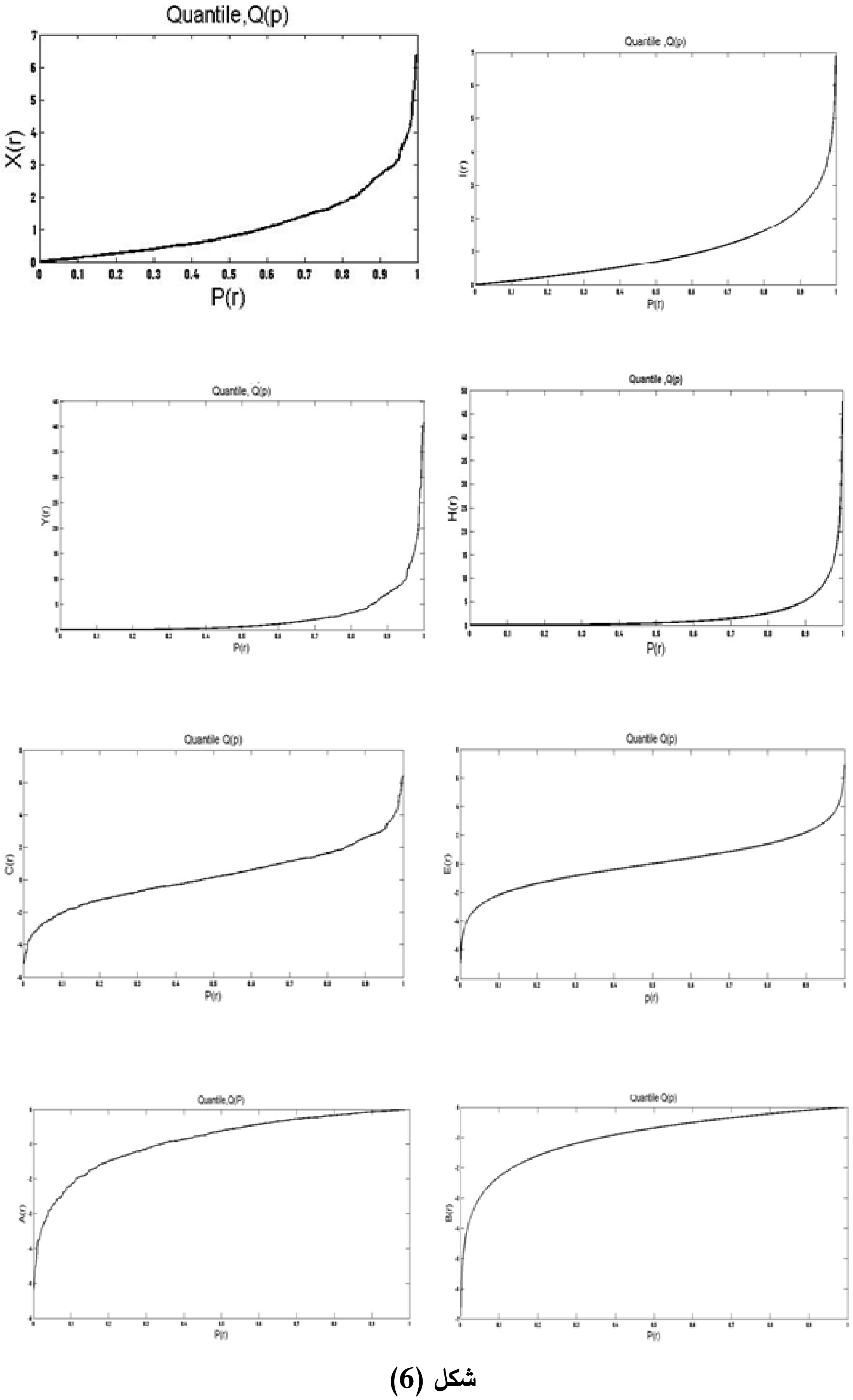
- **Skewnes**

Fig	Plot	(Y)	(X)	Remarks
8	Deviations	$\mathbf{X}_{(n+1-r)} - \mathbf{m}$	m-X(r)	إذا كان التوزيع متماثلاً (ثنائي الذيل) فأن الشكل يكون على شكل خط متناظر يمر بنقطة الاصل بزاوية 45°، بينما التوزيعات أحادية الذيل تكون أشكال متناظرة ولكنها غير متماثلة حول الخط.
	g*(P)	$[X_{(n+1-r)} + X_{(r)} - 2m]$ / $[X_{(n+1-r)} - X_{(r)}]$	$\mathbf{P}_{(r)}$	يحصل تذبذب لبيانات العينة في الوسط بسبب زيادة الفروقات في المنتصف أي عندما يقترب (r) من (n+1-r). أما رسومات المجتمع فلا تحصل هذه الحالة لأن الزيادة في قيم (الاحتمال) المستخدم في توليد X) تكون زيادة (أي بوتيرة متماثلة)، (يراجع رسم اللوجستي مجتمع)

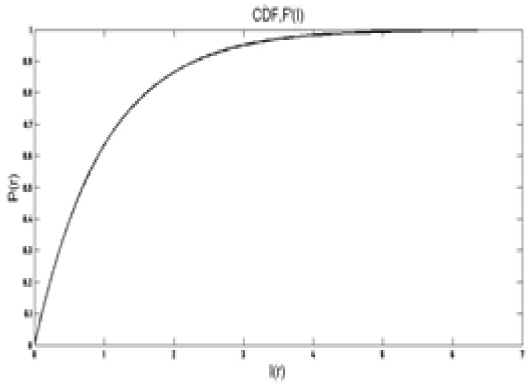
- **Tail shape**

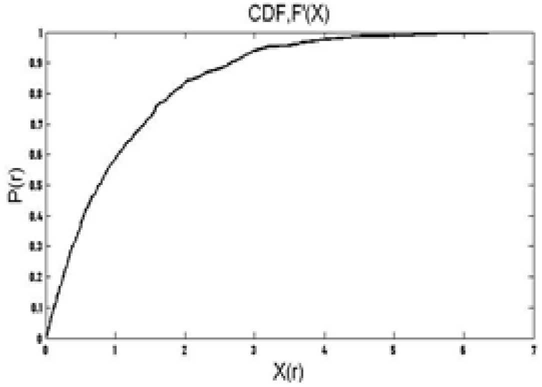
Fig.	*Plot*	*(Y)*	*(X)*	*Remarks*
	Upper shape index, ut(P) $0 < p_{(r)} < 0.5$	$[X_{(n+1-r)} - m]/[uq - m]$	$1 - \mathbf{P}_{(r)}$	يعتبر الشكلان مقعران قريبان للتناظر (في حالة كون التوزيع له ذيل أيسر يكون الشكل محدب)
	Lower shape index, lt(P) $0 < p_{(r)} < 0.5$	$[m - X_{(r)}]/[m - lq]$	$\mathbf{P}_{(r)}$	

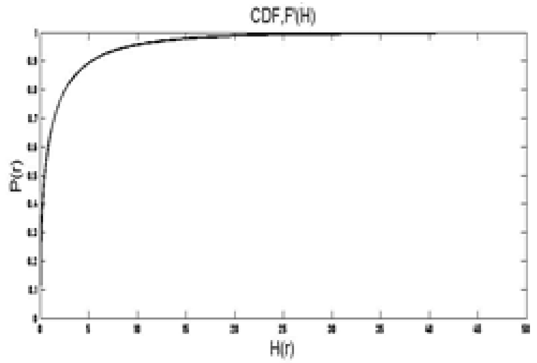
شكل (5)

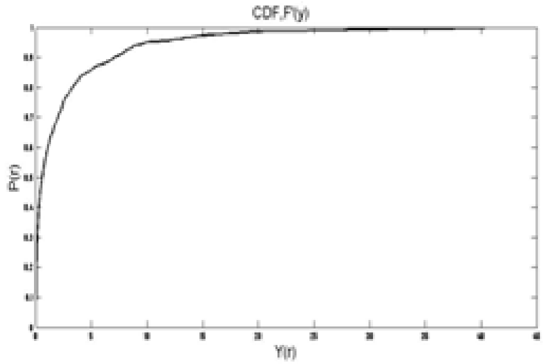

Quantile,Q(p)
X(r)
P(r)
Quantile ,Q(p)
I(r)
P(r)
Quantile, Q(p)
Y(r)
P(r)
Quantile ,Q(p)
H(r)
P(r)
Quantile Q(p)
C(r)
P(r)
Quantile Q(p)
E(r)
p(r)
Quantile,Q(P)
A(r)
P(r)
Quantile Q(p)
B(r)
P(r)

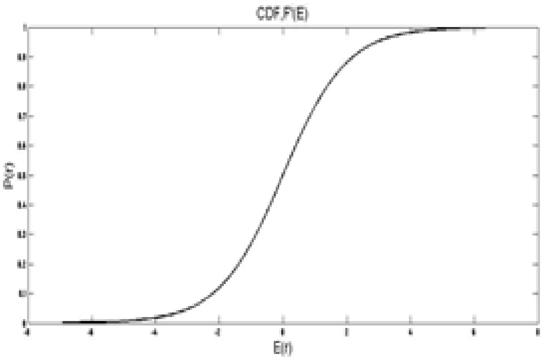
شكل (6)

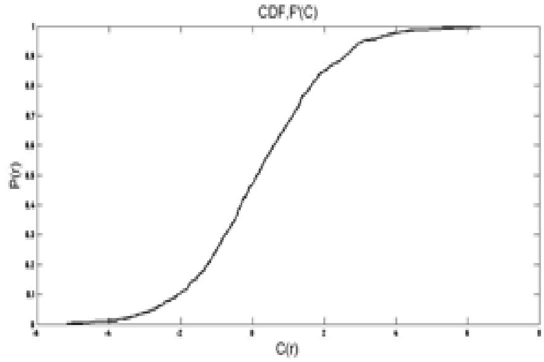

CDF,F'(I)
P'(r)
I(r)

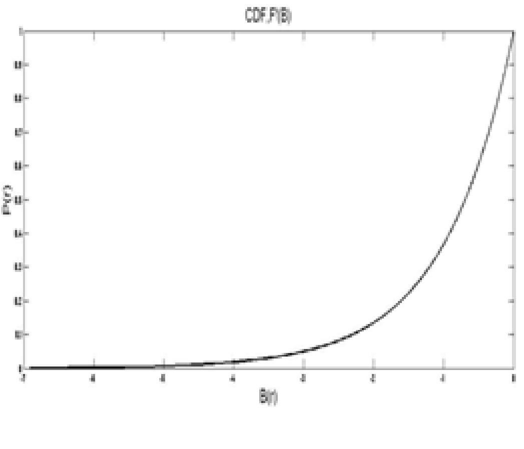

CDF,F'(X)
P'(r)
X(r)

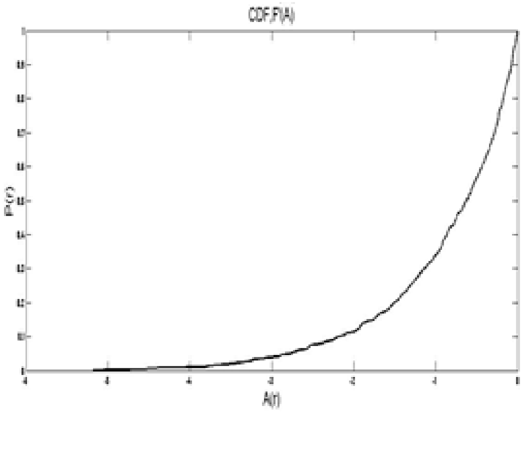

CDF,F'(H)
P'(r)
H(r)


CDF,F'(y)
P'(r)
Y(r)

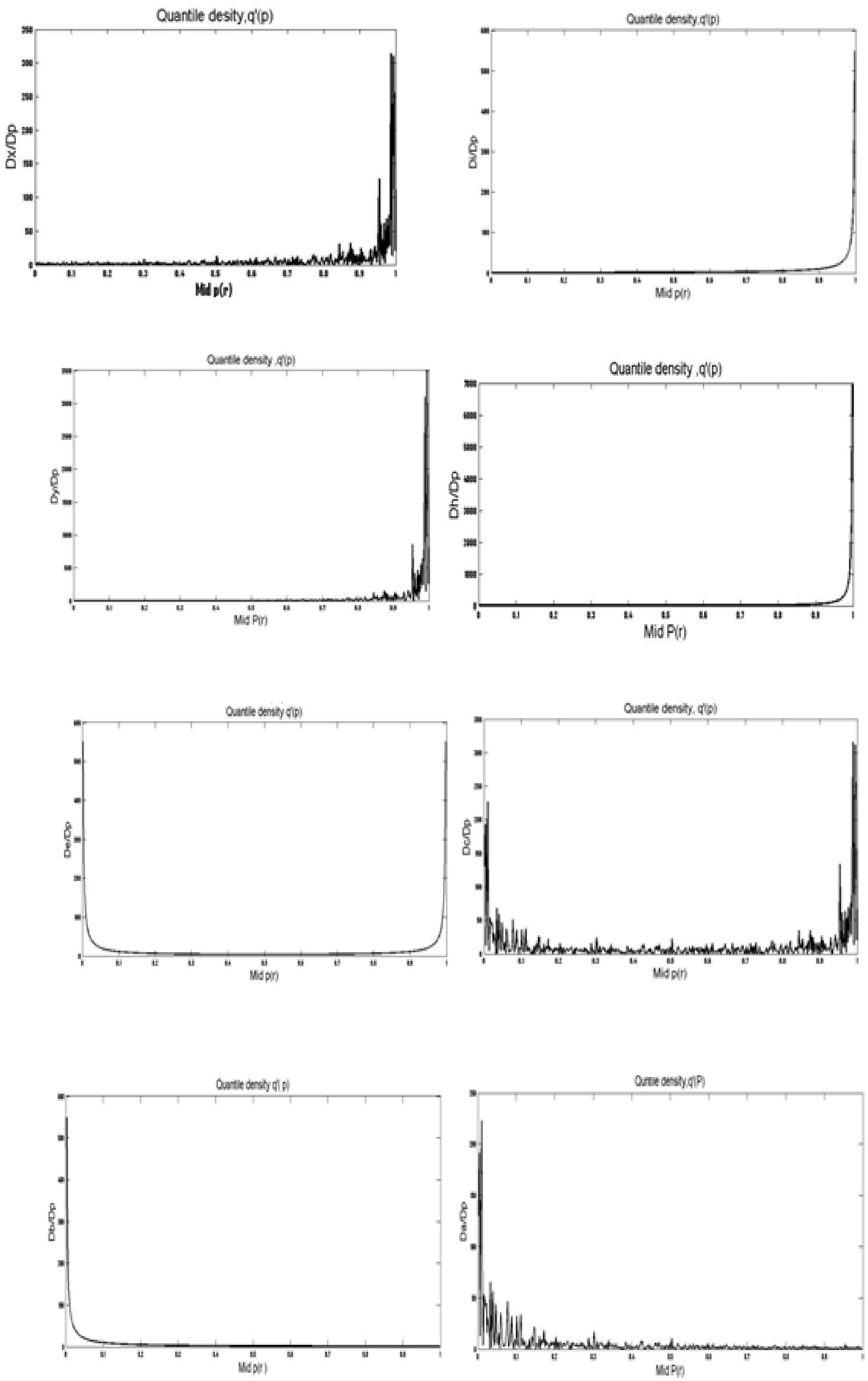

CDF,F'(E)
P'(r)
E(r)


CDF,F'(C)
P'(r)
C(r)

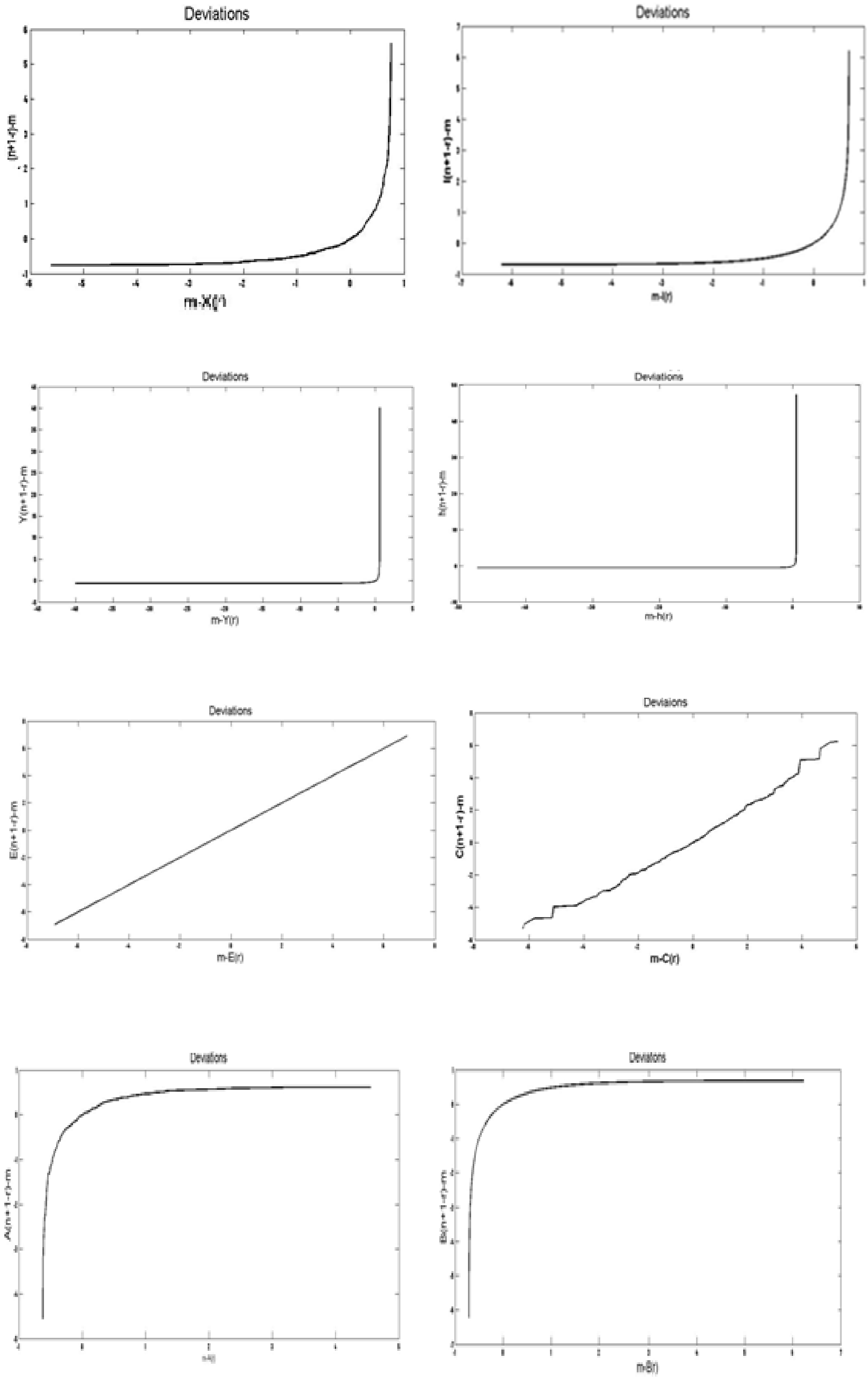

CDF,F'(B)
P'(r)
B(r)


CDF,F'(A)
P'(r)
A(r)

شكل (7)


Quantile desity,q'(p)
Dx/Dp
Mid p(r)
Quantile density,q'(p)
Di/Dp
Mid p(r)
Quantile density ,q'(p)
Dy/Dp
Mid P(r)
Quantile density ,q'(p)
Dh/Dp
Mid P(r)
Quantile density q'(p)
De/Dp
Mid p(r)
Quantile density, q'(p)
Dc/Dp
Mid p(r)
Quantile density q'(p)
Db/Dp
Mid p(r)
Quntile density,q'(P)
Da/Dp
Mid P(r)

شكل (8)


Deviations
Deviations
Deviations
Deviations
Y(n+1-r)-m
m-Y(r)
h(n+1-r)-m
m-h(r)
Deviations
E(n+1-r)-m
m-E(r)
Deviaions
C(n+1-r)-m
m-C(r)
Deviations
A(n+1-r)-m
Deviations
B(n+1-r)-m
m-B(r)

الاستنتاجات والتوصيات

1) إن إلاختبارات اللامعلمية من ضمنها أختبار مربع كاي لا تكفي للتعبير عن التوزيع المتوقع للبيانات.

2) يجب تمحيص البيانات بدراسة مداها وملخصاتها ورسم الدوال لفهم ميزات البيانات من تحدب أو تقعر ونقاطهم وأللتواء قبل الحكم عليها أو قبل ترشيح التوزيع اللازم.

3) قد لا تتناسب البيانات مع توزيع محدد وبذلك وفي حالة فهم الرسومات والملخصات فأنه من الممكن صياغة توزيع مناسب لها.

4) كما ذكرنا فأن هذا الموضوع تم تطبيقه على العديد من العينات المختلفة التوزيع لفهم واقع الرسومات وتحديد ألية أنتقاء التوزيع المناسب لها. وياحبذا لو توضع رسومات قياسية للتوزيعات المشهورة كالجداول الإحصائية حتى يتسنى للباحث رسم بيانات عينته ومقارنة الرسومات مع رسومات المجتمع.

المصادر

1) هرمز، أمير حنا، "الإحصاء الرياضي"، كلية الإدارة والاقتصاد، جامعة الموصل (1990).

2) Gibbons, Jean Dickins on & Chakraborti, Subhabrata, "Nonparametric Statistical Inference", 4th edition, Marcel Dekker Inc., USA (2003).

3) Gilchrist, Warren G., "Statistical modeling with quantilefunctions", Chapman& Hall/CRC, USA (2000).

4) Roussas, G.George, "A course in mathematical statistics", 2nd edition, Academic press, USA,1997.

5) Lilliefors, Hubert W., "On the kolomogrov-smirnov test for normality with mean and variance unknown", American Statistical Association Journal, 399-402., (June 1967).

6) Soong, T.T., "Fundamentals of probability and statistics for engineers", John Wiely& Sons Ltd, London (2004).

7) National Institute of Standards and Technology (NIST) websiteebook, "Engineering Statistics Handbook", 2006

8) WWW.en.Wikipedia.org/wiki/Q-Q_plot.