مجلة الدراسات المالية والمحاسبية والإدارية
ISSN 2352-9962
العدد السابع - جوان 2017
OEBUniv.Publish

التـقـديـر بنوعيـه في الاسـتدلال الإحصائـي مع تطبيق نظرية النهاية المركزية

Estimation with its two types in statical inference theorem With application of the central limit

د. سمية بلقاسمي، أ.د. بوعشة مبارك، جامعة باتنة 1، جامعة قسنطينة2، الجزائر.

تاريخ التسليم:(22/ 2017/01)، تاريخ التقييم:(19/ 2017/03)، تاريخ القبول:(12/ 2017/04)

الملخص

يهدف المقال إلى عرض أحد قسمي الاستدلال الإحصائي المستخدم على نطاق واسع في مختلف الدراسات النظرية والتطبيقية، وهو التقدير بنوعيه النقطي والتقدير بفترة، مع إنشاء فترة ثقة لمتوسط مجتمع يتوزع توزيعا طبيعيا عندما يكون تباينه معلوما، وفي حال كونه مجهولا. بالإضافة إلى عرض وتبيان أهمية بعض النظريات المفيدة في الاستدلال الإحصائي بصفة عامة وفي التقدير بصفة خاصة والمتمثلة في متراجحة تشيبتشف، قانون الأعداد الكبيرة، مع محاولة البرهنة على نظرية النهاية المركزية التي تسمح بتحديد توزيع المعاينة لأهم الإحصاءات، في حالة خاصة كافية من أجل معظم تطبيقات الإحصاء. أما الجانب في التطبيقي من المقال تم التعرف على حجم العينة الكاف لتحقق نظرية النهاية المركزية، مع تطبيقها بتوظيف تقنية المحاكاة في الحزمة الإحصائية ℝ في حالة البيانات المستخرجة من توزيع بواسون والتوزيع الأسي، وذلك بأسلوب خال من التعقيد الرياضي وبذلك يصبح بإمكان غير الرياضيين الاستفادة من الإحصاء واستخدام النظرية بسهولة في مختلف مجالاتهم.

Abstract

This article aims to present one of the two sections of the statical inference used Widely in various theoretical and pratical studies, wich is estimation of both types, point estimation and period estimation, with creationg a confidence interval to the average of society distributed naturally with a known variance, or unknow variance. Moreover it aims to present and indicate the importance of some beneficial theories in statical inference in general and in Estimation in particular, of between it the low of large numbers, with an attempt to prove the central limit theorem in special case sufficient for most applications of statistics. As for the pratical side of the article we tried to identify the sample's size sufficient to achieve the central limit theorem, and applying this theorem using simulation technology in the statical package ℝ In the case of data extracted from the Poisson distribution and exponential distribution, in a manner free from mathematical complexity , thus it get easier for non-statisticians to use it easily in their various fields.

تـمـهـيـد:

عندما ينجز الباحث العلمي تجربته يتوصل إلى نتائج علمية محددة يبينها على أساس المعلومات التي تحصل عليها من التجربة، وتتعدى النتائج المستخلصة حدود المواد التي تحتويها التجربة الخاصة التي قام بها الباحث، أي أنه يقوم بتعميم هذه النتائج إلى كافة التجارب المماثلة. يدعى هذا النوع من التعميم من الخاص إلى العام بالاستقراء[1].

سيتم في هذا المقال التطرق لأحد أشكال الاستقراء وهو الاستدلال الإحصائي الذي يقصد به استنتاج مقاييس إحصائية خاصة بالمجتمع (التي تعتبر مجهولة بالنسبة للباحث) بالاعتماد على بيانات ومقاييس إحصائية خاصة بعينة عشوائية، ويتكون الاستدلال الإحصائي من موضوعين أساسيين : التقدير واختبار الفروض[2].

عند الاستدلال على مقياس من المقاييس الإحصائية الخاصة بالمجتمع عن طريق التقدير يتم استخدام إما التقدير بنقطة أو التقدير بفترة. أما النوع الآخر من الاستدلال أي اختبار الفروض فيتم من خلاله اختبار مدى صحة تخمين يتعلق بقيمة تلخص ملامح المجتمع، أي فيما إذا كانت بيانات العينة تؤيد أو تناقض ادعاء يتعلق بتلك القيمة.

وسنحاول من خلال هذا المقال تسليط الضوء بشكل خاص على الاستدلال الإحصائي عن طريق التقدير، وعليه يمكن طرح الإشكالية التالية:

كيف يتم التقدير، وما هي أساسياته وأنواعه؟ وهل يمكن تحديد توزيع المعاينة لإحصاء ما بصورة تقريبية إذا كان لدينا عينة عشوائية واحدة ؟

للإجابة على الإشكالية المطروحة تم تقسيم المقال لأربعة محاور كالتالي:

المحور الأول: الـتـقـديـر بـنـقـطـة.

المحور الثاني: أهـم الـنظريات الإحـصـائـيـة الـمـفـيـدة فـي الـتـقـديـر.

المحور الثالث: الـتـقديـر بـفـتــرة.

المحور الرابع: تــطـبـيــق نــظــريــة الـنـهـايــة الـمـركـزيــة.

المحور الأول- التقدير بنقطة:

قبل التطرق لتعريف التقدير بنقطة يجب الإجابة عن الأسئلة التالية: ماذا نقصد بمصطلح مجتمع؟ ولماذا لا نتناول المجتمع بأكمله؟ وما هي الصفات التي يجب أن تتوفر في العينة؟.

أولا- أساسيات التقدير بنقطة:

يقصد بالمجتمع الإحصائي مجموعة العناصر التي تشكل هدف الدراسة والتي يرغب في التنبؤ بمعلومات حولها، حيث أن المجتمعات غالبا ما تكون من الضخامة بحيث يكون إخضاع كل عنصر فيها للدراسة نوعا من المستحيل، وحتى عندما يكون ذلك ممكنا من الناحية النظرية على الأقل فإن ما تتطلبه الدراسة من جهود وزمن ونفقات طائلة تجعلها من الناحية الواقعية أمرا غير عمليا البتة.

لهذا يتم الاستعانة بأسلوب المعاينة الذي يؤدي إلى خفض التكاليف ويضمن السرعة كما يمكن أن يحقق دقة عالية بسبب محدودية الجهد بجانب إشراف أكثر فعالية[3]. وعلى هذا الأساس تطور استخدامه تطورا سريعا في شتى الميادين وأصبح يلعب دورا مهما في الكثير من الدراسات النظرية والتجريبية ومن أساليب المعاينة الأكثر شيوعا أسلوب المعاينة العشوائية، والهدف من العشوائية هو عدم التحيز في الاختيار بإتاحة الفرص لجميع عناصر العينة (كل عنصر من عناصر العينة متغير عشوائي) وعندما تكون الدراسة عن طريق عينة عشوائية هنا يبدأ الحديث عن الاستدلال الإحصائي[4].

وعليه فالمعالجة النظرية التي سنراها مستقبلا تستدعي أن تحقق طريقة الانتقاء لعينة إحصائية خاصيتين حتى نتمكن من استخدام الحساب الاحتمالي، الأولى تتمثل في كون عناصر العينة مستقلة عن بعضها البعض أما الثانية فيجب أن يبقى تابع الكثافة نفسه عند كل عنصر من عناصر العينة، وهو نفسه تابع كثافة المجتمع المدروس. وطريقة الانتقاء التي تحقق هذين الأمرين تسمى "**الانتقاء العشوائي لعينة إحصائية**"

وعليه إذا كان للمتحولات العشوائية $x_1, x_2, ..., x_n$ تابع كثافة مشترك من الشكل:

$g(x_1, x_2, ..., x_n) = f(x_1).f(x_2)....f(x_n) = [f(x)]^n$، عندئذ يمكن القول أن $x_1, x_2, ..., x_n$ هي عينة عشوائية حجمها n مسحوبة من مجتمع تابع كثافته $f(x)$[5].

يتم تلخيص صفات العينة كما هو الشأن بالنسبة للمجتمع باستخدام الكميات، ويطلق على الكمية العددية التي تلخص بعض أوجه العينة بإحصاء، أما التي تلخص بعض أوجه المجتمع فتسمى معلم[6]. يقصد بالمقدر النقطي العلاقة الرياضية التي تعطي تقديرا لأحد ملامح المجتمع من خلال العينة[7]. وتمثل كل من $\bar{x}$, S^2, p (الوسط الحسابي والتباين للعينة، النسبة في العينة) على الترتيب أفضل المقدرات التي تستخدم في تقدير المعالم μ, δ^2, π (الوسط الحسابي، التباين، والنسبة في المجتمع) على الترتيب. والمقدر الجيد هو الذي يحقق عدة خواص أكثرها شيوعا عدم التحيز، الكفاءة، الاتساق، والكفاية[8].

وفي حالة التقدير بنقطة تستخدم الإحصاءات في تقدير المعالم مباشرة، فهو **قيمة المقدر بعد التعويض في الصيغة أو الدالة بالبيانات الفعلية للعينة.**

من المعروف أنه لا يمكن القيام بتعميمات تتصف باليقين المؤكد بصفة عامة، وعليه بما أن نتائج التقدير غير صحيحة تماما وهي نتائج محتملة فكيف نصف الخطأ في التقدير؟

ثانيا - الإحصاءات متغيرات عشوائية:

يتذبذب كل إحصاء في قيمته عشوائيا من عينة إلى أخرى بينما تبقى قيمة المعلمات ثابتة، القيمة التي نحسبها لإحصاء من عينة واحدة تعتمد على مفردات العينة، وعليه فكل إحصاء تحددت قيمته من عينة عشوائية هو متغير عشوائي.

يسمى توزيع قيم إحصاء ما لكل العينات العشوائية المتساوية الممكنة بتوزيع المعاينة لذلك الإحصاء وبالتالي له نفس الخصائص الإحصائية لأي توزيع احتمالي (متوسط، انحراف معياري، شكل بياني...). ويطلق على الانحراف المعياري لجميع القيم الممكنة لإحصاء ما في جميع العينات العشوائية المتساوية بالخطأ المعياري أو الخطأ في التقدير[9].

لكن كيف يمكن تحديد توزيع المعاينة لإحصاء ما إذا كان لدينا عينة عشوائية واحدة؟ بالطبع عينة واحدة بمفردها لا تعطي هذه المعلومات ومع ذلك فبدمج المعلومات المستمدة من عينة عشوائية واحدة مع بعض النظريات الإحصائية يمكن على الأقل تحديد توزيع المعاينة بصورة تقريبية[10].

سنحاول فيما تبقى من الجانب النظري من المقال التطرق إلى بعض النظريات الإحصائية الهامة والمفيدة في التقدير إلى حد كبير، مع محاولة البرهان على نظرية النهاية المركزية التي تسمح بتحديد توزيع المعاينة لأحد الإحصاءات الهامة وهو متوسط العينة $\bar{x}$، بالإضافة إلى عرض أساسيات التقدير بفترة.

المحور الثاني: أهم النظريات الإحصائية المفيدة في التقدير

من بين النظريات ذات الدور الكبير في عملية التقدير: متراجحة تشيتشف، قانون الأعداد الكبيرة ونظرية النهاية المركزية.

أولا- متراجحة تشيتشف: تنطبق على أي متحول شريطة أن يكون له توقع و تشتت.

ليكن المتغير العشوائي x توقعه $E(x) = u$, وتشتته $V(x) = \delta^2$، عندئذ إذا كان a عدد موجب فإن:[11]

$$Pr\left\{[x-u] \geq a\right\} \leq \frac{\delta^2}{a^2}$$ أو:

$$Pr\left\{[x-u] < a\right\} \geq 1-\frac{\delta^2}{a^2}$$

كما تكتب متراجحة تشيبتشف على الشكل:

$$Pr\left\{[x-E(x)] \geq K.\delta(x)\right\} \leq \frac{1}{k^2}$$

أو

$$Pr\left\{[x-E(x)] < K.\delta(x)\right\} \geq 1-\frac{1}{k^2}$$

و عليه لا توضح متراجحة تشيتشف فقط أن الانحراف المعياري يعتبر مؤشرا جيدا لمدى انتشار التوزيع الاحتمالي حول متوسطه بل تبين أيضا أن $100\{1-\frac{1}{k^2}\}\%$ من المشاهدات على الأقل يجب أن تقع داخل k من الانحرافات المعيارية بعدا عن المتوسط، و عليه 75% من المشاهدات يجب أن تقع داخل

وحدتي انحراف معياري بعيدا عن المتوسط و89% من المشاهدات على الأقل يجب أن تقع داخل ثلاث وحدات انحراف معياري بعيدا عن المتوسط.

وبإسقاط هذه المتراجحة على متوسط عينة حجمها n، مع العلم أن:

$$\mathbf{E}(\breve{\mathbf{X}}) = E\left(\sum_{i=1}^{n}\frac{x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}E(x_i) = \frac{1}{n}.nE(x) = \boldsymbol{u}$$

$$\mathbf{V}(\breve{\mathbf{x}}) = \mathrm{E}(\breve{x}-\mathrm{E}(\breve{x}))^2 = \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)-\frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}x_i\right)\right]^2$$

$$= \mathrm{E}\left[\frac{1}{n}\left(\sum_{i=1}^{n}x_i-\mathrm{E}\left(\sum_{i=1}^{n}x_i\right)\right)\right]^2$$

$$\mathbf{V}(\breve{\mathbf{x}}) = \frac{1}{n^2}\,\mathrm{V}\left(\sum_{i=1}^{n}x_i\right) = \frac{1}{n^2}\,n\,\mathrm{V}(x) = \frac{\delta^2}{\mathbf{n}}$$

فإننا نجد:

$$\boldsymbol{p_r}\{|\ \breve{\boldsymbol{x}}-\boldsymbol{u}\ |< a\} \geq \mathbf{1}-\frac{\boldsymbol{\delta^2}}{\boldsymbol{na^2}} \text{ أو } \boldsymbol{P_r}\{|\ \breve{\boldsymbol{x}}-\boldsymbol{u}\ |< k\frac{\boldsymbol{\delta}}{\sqrt{\boldsymbol{n}}}\} \geq \mathbf{1}-\frac{\mathbf{1}}{\boldsymbol{k^2}}$$

حيث: $\mathbf{k}\frac{\delta}{\sqrt{n}} = \boldsymbol{a}.$ $\leftarrow$ $\frac{\mathbf{1}}{\boldsymbol{k^2}} = \frac{\boldsymbol{\delta^2}}{\boldsymbol{na^2}}$[12]

ثانيا - قانون الأعداد الكبيرة:

يمكن هذا القانون من الوصول إلى معرفة تتمتع بدرجة عالية من الثقة أو القيام باستقراء حول متوسط المجتمع u ، وذلك بدءا من المعلومات التي توفرها عينة حجمها محدود وليكن n .

مضمون هذه النظرية أنه إذا كان $f(x)$ تابع كثافة بمتوسط u و تشتت δ^2، وكان $\breve{x}$ متوسط عينة عشوائية مأخوذة من $f(x)$ حجمها n ، ليكن c وb أي عددين موجبين، عندئذ يمكن إيجاد n_0 بحيث:

$P_r\{|\ \breve{x}-u\ |< c\} \geq 1-b$ من أجل جميع قيم $n > n_0$.

بمعنى آخر باحتمال قريب من الواحد بالقدر الذي نريده يمكن أن نجعل الكمية $|\ \breve{x}-u\ |$ صغيرة بالقدر الذي نريده وذلك من خلال اختيار حجم العينة[13].

ويمكن الوصول إلى حجم العينة اللازم لتكون $(\breve{x}-u)$ أقل من العدد الموجب c باحتمال قريب من الواحد بالقدر b إنطلاقا من تطبيق متراجحة تشيبتشف كما يلي:

$$P_r\{|\ \breve{x}-u\ |< c\} \geq 1-\frac{\delta^2}{nc^2}$$

يكفي أن نأخذ $b = \frac{\delta^2}{nc^2}$ أي $n_0 = \frac{\delta^2}{bc^2}$ وفقا لقانون الأعداد الكبيرة فإنه باحتمال قريب من الواحد بالقدر الذي نريده (b) يمكن جعل الفرق بين متوسط العينة $\breve{x}$ ومتوسط المجتمع u صغير بالقدر الذي نريده (أقل من c) وذلك باختيار حجم العينة أكبر من المقدار $\frac{\delta^2}{bc^2}$.

ثالثا - نظرية النهاية المركزية:

* إذا كانت $x_1, x_2, \dots x_n$ متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي وإذا كان δ_1^2, u_1 المتوسط والتشتت لكل منها على الترتيب، عندئذ يتقارب توزيع المجموع $\sum_{i=1}^{n} x_i$ من التوزيع الطبيعي $N\sim(nu_1, \sqrt{n}\delta_1)$ عندما تؤول n إلى اللانهاية[14].

* كما أن توزيع $\breve{} = \sum_{i=1}^{n} \frac{x_i}{}$ يتقارب من التوزيع الطبيعي $N\sim(u_1, \delta_1/\sqrt{n})$ مع زيادة حجم العينة[15].

أول من عرض هذه النظرية الأساسية هو لابلاس من عام 1812 وبرهنها ليابونوف عام 1901، وسنحاول البرهنة عليها في حالة خاصة كافية من أجل معظم تطبيقات الإحصاء كما يلي:

ليكن $\boldsymbol{X} = x_1 + x_2 + \cdots x_n$ مجموع متغيرات عشوائية مستقلة لها نفس التوزيع حيث δ_1^2 , u_1 متوسط وتباين كل منها. وليكن u وδ^2 متوسط وتباين المجموع وعليه $\boldsymbol{\delta^2} = n\delta_1^2$, $\boldsymbol{U} = nU_1$ والشكل المعياري للمجموع هو : $Z = \frac{X-nU_1}{\delta_1\sqrt{n}}$ أي: $Z = \frac{1}{\delta_1\sqrt{n}}\sum_{i=1}^{n}(x_i - u_1)$ حيث لـ x_i , u_1 نفس التوزيع.

وليكن $\varphi_1(t)$ التابع المميز لأي منها أي: $\varphi_1(t) = \varphi_{x_i-u_1}(t)$

$$i = 1,2 \dots n$$

ويمكن الحصول على $\varphi_z(t)$ التابع المميز للمجموع المعياري Z باستخدام خواص* التابع المميز كما يلي:

$$\varphi\frac{x_i - u_1}{\delta_1\sqrt{n}} = \varphi_{x_i-u_1}\left(\frac{t}{\delta_1\sqrt{n}}\right) = \varphi_1(\frac{t}{\delta_1\sqrt{n}})$$

وعليه: $^1\varphi_z(t) = [\varphi_1(\frac{t}{\delta_1\sqrt{n}})]^n$

وبما أن العزم الأول والثاني لـ $_i - u_1$ هما 0 و δ^2 على الترتيب نجد:**

$$\varphi_1(t) = 1 + 0\,i\,t - \frac{1}{2}\delta^2\,t^2 + \frac{1}{6}u'_3(it)^3 + \dots..$$

وعليه للحصول على الدالة المميزة للمجموع المعياري Z نستبدل t بـ: $\frac{t}{\delta_1\sqrt{}}$ ونرفع قوى النشر إلى n على الشكل:

---

* يمكن البرهنة أن: $\varphi_{ax+b}(\ \ ) = e^{itb}\varphi_x(at)$

$$\varphi_{x_1+x_2}(t) = \varphi_{x_1}(t).\varphi_{x_2}(t).$$

$$\varphi_{x_{1,}x_2\dots x_n}(t) = [Q_{x1}(t)]^n\dots.$$

حيث لـ $x_1, x_2, x_3 \dots . x_n$ نفس التوزيع الاحتمالي

**عند نشر $\varphi(t)$ بنشر ماكلوران نجد.... $\varphi_x(t) = 1 + \grave{u}_1 it - \frac{1}{2!}\grave{u}_2 t^2 + \grave{u}_3\frac{(it)^3}{3!}$..

$$\boldsymbol{\varphi_z(t) = [1 - \frac{\frac{1}{2}t^2}{n} + \frac{\varepsilon(n.t)}{n}]^n}$$

حيث تنتهي الكمية $\varepsilon(n.t)$ الى الصفر عندما تؤول n إلى اللانهاية، وعليه $\boldsymbol{\varphi(t) \to e^{-1/2t^2}}$ وهي الدالة المميزة للتوزيع الطبيعي المعياري[16]، وعليه فإن توزيع $\sum_{i=1}^{n} x_i$ يتقارب من التوزيع الطبيعي $N\sim(nu_1, \delta_1\sqrt{n})$.

على أساس ما سبق يمكن أن نستنتج أن المتوسط $\breve{} = \sum_{i=1}^{n} \frac{x_i}{}$ يتقارب إلى التوزيع الطبيعي $N\sim(u_1, \frac{\delta_1}{\sqrt{n}})$ كما يلي:

نعرف أن مجموع متغيرات مستقلة $_1, x_2, \dots x_n$ لها نفس التوزيع يتوزع وفق التوزيع الطبيعي $N\sim(nu_1, \delta_1\sqrt{n})$

وعليه بالاعتماد على خواص التابع المميز نجد:*

$$\varphi_{\sum_{i=1}^{n} x_i}(t) = \varphi_{(\delta_1\sqrt{n})Z+nu_1}(t) = e^{itnu_1}\varphi_Z(\sqrt{n}\delta_1 t) = e^{itnu_1}e^{-\frac{1}{2}t^2\delta_1^2 n}$$

وعليه الدالة المميزة للمتوسط $\sum_{i=1}^{n} \frac{x_i}{n}$ تكتب على الشكل:

$$\varphi_{\breve{x}}(t) = \varphi_{\frac{(\sqrt{n}\delta_1)Z + nu_1}{n}}(t) = \varphi_{\Sigma(xi)}\left(\frac{t}{n}\right) = e^{itu_1}e^{-\frac{1}{2}t^2\frac{\delta_1^2}{n}}$$

وهي الدالة المميزة لتوزيع طبيعي بوسط حسابي $_1$ وانحراف معياري $\frac{\delta_1}{\sqrt{n}}$ ، $N\sim\left(u_1, \frac{\delta_1}{\sqrt{n}}\right)$.

- تبدو أهمية نظرية النهاية المركزية من زاويتين، فهي توضح أولا نزوع العديد من المتغيرات العشوائية إلى أن يكون توزيعها بصورة تقريبية هو التوزيع الطبيعي إذ يمكن أن نتصور طول الإنسان حصيلة عدد كبير من المورثات العشوائية كطول الأب، طول الأم، نشاط الغدد ذات علاقة بالطول، البيئة، التغذية. وإذا كانت هذه العوامل تضاف إلى بعضها البعض لتنتج واقعا معينا بالنسبة إلى طول الإنسان فعندئذ يمكن اعتبار الطول كحصيلة لعدد كبير من المتغيرات العشوائية وهكذا تنطبق نظرية النهاية المركزية ويكون توزيع متغير الطول هو التوزيع الطبيعي على وجه التقريب.[17]

كما يمكن استخدام نظرية النهاية المركزية لتقريب التوزيع الثنائي من التوزيع الطبيعي، فإذا اصطلحنا على أن يوافق حالة النجاح العدد1 وحالة الفشل العدد صفر، فعندئذ تكون نتائج التكرارات المستقلة n عبارة عن متتالية من المتغيرات المستقلة $x_1, x_2, x_3 \dots . x_n$ حيث يأخذ x_1 إما القيمة 1 أو 0 في المتتالية الواحدة. ويكون عدد النجاحات X هو عدد مرات ورود العدد 1 أو مجموعها: $X = \sum_{i=1}^{n} x_i$، وبما أن x_i

*نعلم أن: $\sum_{i=1}^{n} \ _i = (\delta_1\sqrt{n})Z + nu_1$

$$\breve{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{(\delta_1\sqrt{n})Z + nu_1}{n}$$

يتوزع وفق توزيع برلوني عندئذ تصبح نتائج التكرارات المستقلة n: $x_1, x_2, x_3 \ldots . x_n$ هي عبارة عن عينة من مجتمع برلوني ويصبح X مجموع العينة يتوزع وفقا لنظرية النهاية المركزية توزيعا طبيعيا بمتوسط np وانحراف معياري $\sqrt{npq}$ حيث $N\sim(np, \sqrt{npq})$ (p احتمال النجاح في التجربة الواحدة، q احتمال الفشل في التجربة الواحدة و يساوي 1-p) [18] عندما تكون n كبيرة كفاية.

أما في حالة توزيع بواسون فإذا أخذنا مجموعة من متغيرات بواسون لها المتوسط نفسه λ و جمعناها فإننا نحصل على توزيع بواسون بمتوسط متزايد ويمكن توضيح ذلك كما يلي:

نعلم أن الدالة المميزة لتوزيع بواسون: * $\varphi_x(t) = e^{-\lambda}e^{\lambda e^{it}}$

وبالاستناد إلى خواص التابع المميز نجد:

$$\varphi_{\Sigma_{xi}}(t) = e^{-n\lambda}e^{n\lambda e^{it}}$$

وهي الدالة المميزة لتوزيع بواسون بمتوسط $n\lambda(n^2p)$ [19]، وبما أن مجموع متغيرات عشوائية مستقلة لها نفس التوزيع يؤول إلى التوزيع الطبيعي مع زيادة حجم العينة فإن توزيع بواسون يؤول إلى التوزيع الطبيعي وفقا لنظرية النهاية المركزية عندما يزداد المتوسط [20].

- أما العطاء الثاني لنظرية النهاية المركزية فيتعلق بالاستقراء الإحصائي، فالعديد من الإحصاءات التي تستخدم للقيام باستقراءات حول معلمات توزيع ما تأخذ شكل مجموع أو متوسط لقياسات العينة وإذا كان الحال كذلك وكانت n كبيرة كفاية يمكن اعتبار التوزيع الطبيعي تقريبا جيدا للتوزيع الاحتمالي لتلك الإحصاءات [21] بصرف النظر عن توزيع المجتمع الذي سحبت منه العينة.

المحور الثالث: أساسيات التقدير بفترة

نعلم أن الخطأ في التقدير هو الانحراف المعياري لقيم إحصاء ما خلال كل العينات العشوائية الممكنة. ولكن إلى أي مدى يمكن أن تبتعد قيمة الإحصاء عن قيمة المعلم كنتيجة للخطأ المعياري؟ وكيف نثق بأن الفترة المحددة هي فعلا تضم قيمة المعلم؟ .

وعليه عندما نكون مهتمين بتقدير معلمة ما يكون من المهم أن يتوفر لدينا : 1- **التقدير بنقطة، 2- حجم الخطأ المحتمل في التقدير بنقطة، أو مدى من القيم من المحتمل أن يضم القيمة الفعلية للمعلم، 3- درجة الثقة التي تصاحب عبارة أن قيمة المعلمة تقع داخل المدى المحدد**[22].

---

*الدالة المميزة لتوزيع بواسون:

$$\varphi_X(t) = E(e^{itx}) = \sum_{x=0}^{n} e^{itx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{n}\frac{(e^{it}\lambda)^x}{x!} = e^{-\lambda}e^{\lambda e^{it}}$$

وكأساس لذلك يهدف التقدير بفترة إلى إعطاء الإحصاء الذي تختلف قيمته من عينة إلى أخرى مدى يتحرك فيه.

ولكن كيف لنا أن نحدد حجم الخطأ ودرجة الثقة؟

يمكن توضيح الصيغة العامة لفترة الثقة لمعلم ما بشكل عام كما يلي:

- إذا كان المعلم المراد تقديره هو θ وكانت T_1, T_2 إحصائيتين بحيث $(T_1 < T_2)$، وكان $P(T_1 < \theta < T_2) = 1-\propto$ حيث $1-\propto$ لا يعتمد علىθ. فإن الفترة (T_1, T_2) تسمى فترة الثقة بدرجة ثقة $100\%(1-\propto)$ للمعلم θ . ويسمى $_1$ حد الثقة الأدنى و T_2 حد الثقة الأعلى والمقدار $(1-\propto)$ معامل الثقة .

- لإنشاء فترة الثقة نحتاج الحصول على الكمية المحورية وهي دالة في المقدر والمعلم وتوزيعها لا يعتمد على المعلم θ. وعليه إذا تم سحب n عينة من المجتمع فإن القيمة الحقيقية للمعلم θ تقع بين (t_1, t_2)* في $100(1-\propto)$% من العينات.

ويمكن توضيح إنشاء فترة ثقة لمعلم ما من خلال المثالين التاليين:

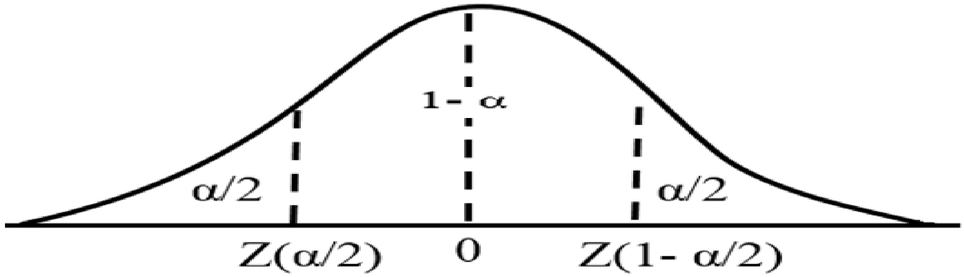
أولا- إنشاء فترة ثقة للوسط الحسابي u في التوزيع الطبيعي عندما تكون δ^2 معلومة:

بما أن الوسط الحسابي للعينة ˘ هو أفضل مقدر للمعلم u فسوف يستخدم لإنشاء فترة الثقة لـ u. نعرف أن الوسط الحسابي لعينة حجمها n من توزيع طبيعي $N\sim(u,\delta)$ يتبع التوزيع الطبيعي $N\sim(u,\frac{\delta}{\sqrt{n}})$ وبالتالي فالكمية المحورية المناسبة هي: $Z = \frac{\breve{X}-u}{\delta/\sqrt{n}}$ وهي دالة في المقدر $\breve{X}$ والمعلم u وتوزيعها هو التوزيع الطبيعي المعياري $N\sim(0,1)$ الذي لا يعتمد على المعلم u[23].

بفرض أن مستوى الثقة المرغوب فيه هو $100\%(1-\propto)$ حيث $\propto$ تمثل نسبة العينات التي قيم متوسطاتها تقع خارج Z خطا معياري، وعليه $(1-\propto)$ تمثل مساحة توزيع المعاينة لـ $\breve{X}$ ، النقط التي تحدد الحدود العليا والدنيا لهذه المساحة المركزية تسمى بالقيم الجزئية التي توضح أقصى عدد من الأخطاء المعيارية التي يمكن أن تنحرف بها $\breve{x}$ عن u باحتمال $(1-\propto)$. وعليه يمكن استنتاج أن مساحة كل طرف من طرفي توزيع المعاينة والتي تناظر قيم $\breve{x}$ وتقع خارج القيم الجزئية تساوي $\propto/2$. ويمكن توضيح ما سبق من خلال الشكل التالي:[24]

شكل 01: المساحة المركزية $(1-\propto)$ والقيم الجزئية المناظرة لها

$(t_1, t_2)^*$ قيمتي (T_1, T_2)



المصدر: جورج كانافوس، دون ميلر، ترجمة سلطان محمد عبد الحميد، "الإحصاء للتجاريين مدخل حديث"، دار المريخ للنشر، الرياض، 2004، ص322.

باستخدام خاصية التناظر للتوزيع الطبيعي المعياري نجد: $Z\left(\propto/_2\right) = -Z(1-\propto/_2)$ وعليه:

$$P_r\left\{-Z_{(1-\propto/_2)} < Z < Z_{(1-\propto/_2)}\right\} = 1{-}\propto$$

و بالتعويض عن Z نحصل على :

$$P_r\left\{-Z_{(1-\propto/_2)} < \frac{\breve{X}-u}{\frac{\delta}{\sqrt{n}}} < Z_{(1-\propto/_2)}\right\} = 1{-}\propto$$

التي تكتب على الشكل :

$$P_r = \left\{\breve{X} - Z_{(1-\propto/_2)}\frac{\delta}{\sqrt{n}} < u < \breve{X} + Z_{(1-\propto/_2)}\frac{\delta}{\sqrt{n}}\right\} = 1{-}\propto$$

وإذا كانت $\breve{}$ هي القيمة المشاهدة لـ$\breve{X}$ فان فترة الثقة لـ u هي $\breve{x} \pm Z(1-\propto/_2)\frac{\delta}{\sqrt{n}}$ ومستوى الثقة $(1{-}\propto)100\%$.[25]

ثانيا -إنشاء فترة ثقة للوسط الحسابي u في التوزيع الطبيعي $N\sim(,\delta^2)$ إذا كانت δ^2 غير معروفة:

نعلم أنه من المفروض أن يتوزع متوسط عينة عشوائية حجمها n مسحوبة من مجتمع طبيعي $N\sim(u,\delta^2)$ (δ^2معلوم) وفق التوزيع الطبيعي $N\sim(u,\frac{\delta^2}{n})$ وعليه المتغير المعياري يساوي $Z = \frac{\breve{X}-u}{\delta\sqrt{n}}$، لكن في حال كون تباين المجتمع δ^2 مجهول فمن المعقول استبدال المعلم δ بالإحصاء S لتصبح الصيغة الجديدة $T = \frac{\breve{X}-u}{s\sqrt{n}}$[26]، و التي يمكن توضيح أنها تتوزع وفق توزيع **Student** بـ(n-1) درجة حرية:

$$T_{(n-1)} = \frac{Z}{\sqrt{x_2/_{n-1}}} = \frac{\breve{X}-u}{\delta/\sqrt{n}} \div \sqrt{\frac{(n-1)S^2}{\delta^2(n-1)}} = \frac{\breve{X}-u}{\delta/\sqrt{n}}.\frac{\delta}{S} = \frac{\breve{X}-u}{S\sqrt{n}}$$

وعليه فالكمية المحورية المناسبة هي: $T = \frac{\breve{X}-u}{s\sqrt{n}}$ وهي دالة في المقدر $\breve{x}$ والمعلم u وتوزيعها يعتمد على توزيع Student بـ(n-1) درجة حرية والذي لا يعتمد على المعلم u (توزيع Student يشبه التوزيع

الطبيعي المعياري في كونه متناظر ومتماثل حول الصفر لكنه أكثر تشتتا لأن إضافة الإحصاء الثاني S للصيغة يزيد من تشتت T من عينة لأخرى).

وعلى أساس ما سبق :

$$P_r\left\{-T_{\left(1-\propto/_2\right)} < T < T_{\left(1-\propto/_2\right)}\right\} = 1-\propto$$

وبالتعويض عن T نجد:

$$P_r = -T_{\left(1-\propto/_2\right)} < \frac{\check{X}-u}{s\sqrt{n}} < T_{\left(1-\propto/_2\right)}\} = 1-\propto$$

والتي يمكن أن تكتب على الشكل:

$$P_r\left\{\check{X} - T_{\left(1-\propto/_2\right)}S\sqrt{n} < u < \check{X} + T_{\left(1-\propto/_2\right)}S\sqrt{n}\right\} = 1-\propto$$

وعليه إذا كانت s ,˘ هما القيمتان المشاهدتان لـ$\check{X}$ وS فإن فترة الثقة بدرجة ثقة $(1-\propto)100\%$ هي:

$$\check{x} \pm T\left(1-\propto/_2\right)s/\sqrt{n}$$ [27]

وبناء على كل ما سبق يمكن كتابة الهيكل العام لفترة الثقة لمعلم ما من الشكل:

التقدير بنقطة ± قيمة جزئية x الخطأ المعياري للتقدير بنقطة.
أو:
التقدير بنقطة ± هامش خطأ المعاينة.

المحور الرابع: تطبيق نظرية النهاية المركزية

- **حالة بيانات مستخرجة من توزيع أسي وتوزيع بواسون:**

سنعطي فيما يلي التعليمات الخاصة باستدراج 5000 عينة حجم كل منها 36 من مجتمع يتوزع توزيع بواسون، كما سنولد عينات عشوائية حجم كل واحدة منها 36 من التوزيع الأسي* (الذي يستخدم في دراسة الكفاءة لتمثيل فترة حياة الدارات الالكترونية) بمعدل $\frac{1}{10}$ أي أنها تابع أسي بوسط حسابي 10 وتباين 100.

ولتحقيق ذلك نستخدم التعليمات أدناه:

---

*تابع الكثافة متغير عشوائي يتوزع توزيعا أسيا :

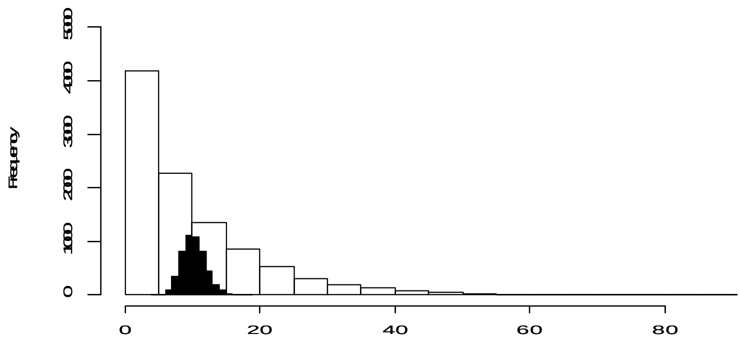
$$f\,(x) = \lambda e^{-\lambda x} \quad E(x) = \frac{1}{\lambda} \qquad var(x) = \frac{1}{\lambda^2}$$

```
N=10000
> # rpois génère une variable de N observations qui a une deviation de Poisson
> #de moyenne 10 et de variance 10
> rendement=rpois(N, lambda=10)
> #on arrondit rendement à zero chiffres après la virgule
>redement=round(rendement)
> #initialisation du vecteur
> M=5000
> moyennes=rep(0,M)
> #on tire 5000 échantillons de 36
> #et le résultat est mis dans le vecteur moyenne
>for( i in 1:M) moyennes[i]=mean(sample(rendement,36))
>hist (rendement ,xlim = c(0,20),ylim = c(0,3000),main="",xlab="")
>par(new=TRUE)
>hist (moyennes ,xlim = c(0,20),ylim = c(0,3000),main="",xlab="", col=1)
>sd(rendement)
[1] 3.241304
>sd(moyennes)
[1] 0.5565438
>sd(rendement)/sqrt(36)
[1] 0.5402173
```

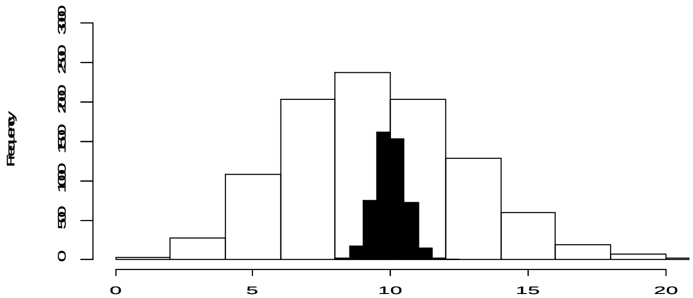
```
> N=10000
> #rexpgenère une variable de N observations qui a une distribution exp
> rendement=rexp(N,rate=1/10)
> # on arrondit rendement a zero 0 chiffres après la virgule
> rendement = round(rendement)
> #initialisation du vecteur
> M=5000
> moyennes = rep(0,M)
> #on tire 5000 échantillon de taille 36 .
> # et le resultat est mis dans le vecteur moyenne
> for ( iin 1:M ) moyennes [i] = mean (sample(rendement ,36))
>hist (rendement ,xlim = c(0,87),ylim = c(0,5000),main="",xlab="")
>par(new=TRUE)
>hist (moyennes ,xlim = c(0,87),ylim = c(0,5000),main="",xlab="", col=1)
>sd(rendement)
[1] 10.19886
>sd(moyennes)
[1] 1.718560
>sd(rendement)/sqrt(26)
[1] 2.000161
```

المصدر: نتائج الحزمة الإحصائية R

الشكل 02: توزيع المعاينة للوسط الحسابي للتوزيع الأسي



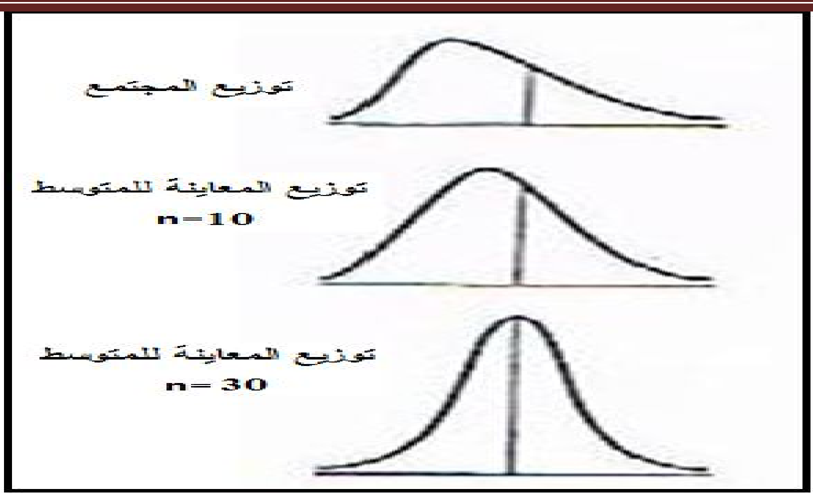
الشكل 03: توزيع المعاينة للوسط الحسابي لتوزيع بواسون



المصدر: نتائج الحزمة الإحصائية R

من الشكلين السابقين (المدرج بالأسود) يتبين لنا أن توزيع المعاينة للمتوسط لكل من توزيع بواسون والتوزيع الأسي هو التوزيع الطبيعي، بانحراف معياري يساوي انحراف المجتمع مقسوما على $\sqrt{36}$ و$\sqrt{26}$ على التوالي، وعليه يتبين لنا أن نظرية النهاية المركزية قد خلقت لدينا إحساسا بتخمين معين، حيث نجد أن توزيع المعاينة لـ $\overline{X}$ قريبا من التماثل وله شكل ربوة عندما تكون n كبيرة ولو كان توزيع المجتمع ليس طبيعيا، ففي العينات العشوائية الكبيرة نحصل بيانات نموذجية تحتوي على عدد كبير من القيم التي هي أعلى وأدنى من متوسط المجتمع وعليه تكون فرصة وقوع $\overline{X}$ أعلى من U مساوية لفرصة وقوعها أدنى من U وعليه إذا كانت العينة ذات حجم كبير فإن توزيع النتائج الممكنة لـ $\overline{X}$ سيكون متماثلا وله قمة وحيدة، ومن الشائع علميا استخدام n=30 كمقياس لتحديد ما إذا كان من الأمان افتراض أن توزيع المعاينة لـ $\overline{X}$هو تقريبا طبيعي أم لا، ويمكن توضيح ذلك من خلال الشكل التالي:

شكل 04: تأثير حجم العينة على شكل توزيع متوسط العينة:

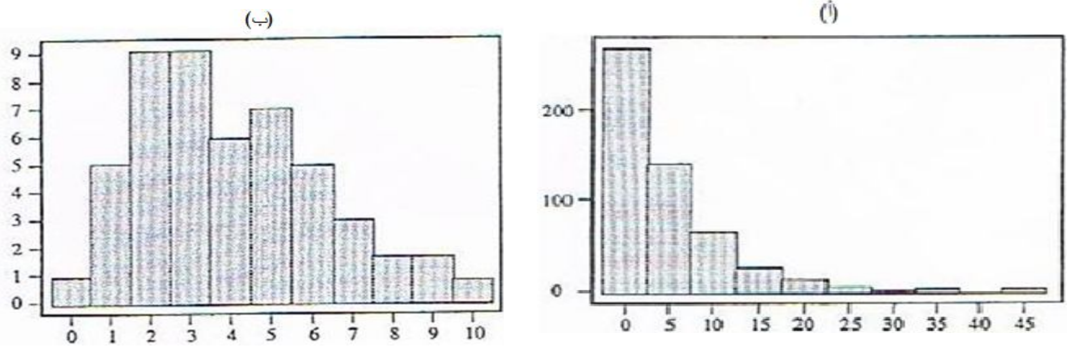


المصدر: من إعداد الباحثة

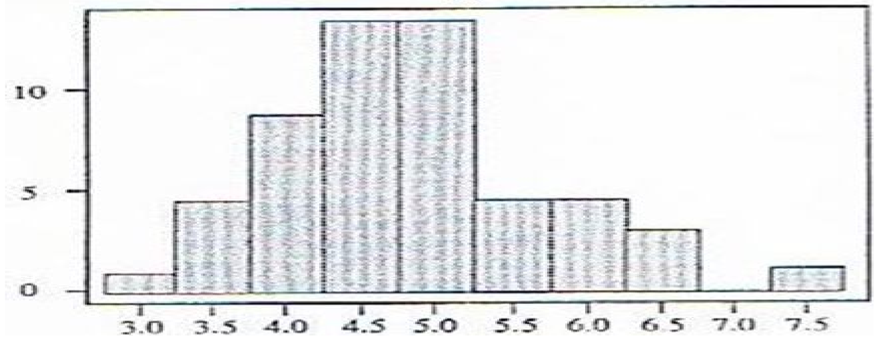
ويمكن توضيح نظرية النهاية المركزية بشكل أكثر تفصيلا كما يلي:[28]

تم محاكاة 50 عينة كل ذات الحجم n=10 من مجتمع توزيعه ملتوي بشدة، ثم أعيد محاكاة 50 عينة كل منها ذو الحجم n=40، ويمكن توضيح النتائج كما يلي:

شكل 05: تأثير حجم العينة على توزيع متوسط العينة عندما يكون n=10 و n=40:



(جـ)



في الجزء (أ) من الشكل تم توضيح توزيع المجتمع حيث يلاحظ الالتواء بشدة، أما في الأجزاء (ب) و(جـ) عرضت توزيعات المعاينة لـ لجميع العينات الخمسين عندما تكون n=10 ثم n=40 على التوالي. ونلاحظ في الجزء (ب) أنه عندما تكون n=10 فإن توزيع المعاينة لـ يبقى ملتويا بعض الشيء لكن في الجزء (جـ) أين n=40 فإن توزيع المعاينة لـ يكون كتماثل وله شكل ربوة.

استنتاجات وتوصيات:

* في حالة التقدير بنقطة تستخدم الإحصاءات في تقدير المعالم مباشرة، وكل إحصاء تحددت قيمته من عينة عشوائية هو متغير عشوائي يسمى توزيع قيمه لكل العينات العشوائية المتساوية الممكنة بتوزيع المعاينة لذلك الإحصاء.

* بدمج المعلومات المستمدة من عينة عشوائية واحدة مع بعض النظريات الإحصائية يمكن على الأقل تحديد توزيع المعاينة لإحصاء ما بصورة تقريبية.

* وفقا لقانون الأعداد الكبيرة فإنه باحتمال قريب من الواحد بالقدر الذي نريده يمكن جعل الفرق بين متوسط العينة ومتوسط المجتمع u صغير بالقدر الذي نريده وذلك باختيار حجم العينة.

* تبين نظرية النهاية المركزية أنه كلما تم تكرير التجربة كلما كان خطأ المعاينة صغيرا وبالتالي كلما كان تقدير المتوسط أكثر دقة، وكلما كان حجم العينة كبير كفاية ($n \geq 30$) يمكن اعتبار التوزيع الطبيعي تقريبا جيدا للتوزيع الاحتمالي للإحصاءات التي شكل مجموع أو متوسط لقياسات العينة بصرف النظر عن توزيع المجتمع.

* عندما نكون مهتمين بتقدير معلمة ما يكون من المهم أن يتوفر لدينا التقدير بنقطة، حجم الخطأ المحتمل في التقدير بنقطة، ودرجة الثقة.

* يمكن لغير الإحصائيين تطبيق النظريات الإحصائية في مختلف مجالاتهم بطريقة مبسطة دون التعمق في التعقيد الرياضي باستخدام البرمجيات الإحصائية.

الـهــوامـــش والإحــالات:

1- ناظم الشمري، "**الإحصاء الرياضي**"، الكتاب الثالث، دمشق، 1978، ص 176.

2-أحمد عودة أحمد عبد المجيد، "**مشكلات استخدام الإحصاء في تحليل البيانات للرسائل العلمية والأطروحات**"، الملتقى الأول حول تجويد الرسائل والأطروحات وتفعيل دورها في التنمية الشاملة، الرياض، 2011، ص ص2-3.

3- Pierre-Gheury, **«Guide pratique de l'analyse de donnée** », De Boeck, Belgique, 2010, P 62.

4- أحمد عودة عبد المجيد، "**مشكلات استخدام الإحصاء في تحليل البيانات للرسائل العلمية والأطروحات**"، مرجع سبق ذكره، ص2.

5- ناظم الشمري، "**الإحصاء الرياضي**"، مرجع سبق ذكره، ص ص179-180.

6- جورج كانافوس، دون ميلر، ترجمة سلطان محمد عبد الحميد، "**الإحصاء للتجاريين مدخل حديث**"، دار المريخ للنشر، الرياض، 2004، ص26.

7- ياسر الموسى، "**الإحصاء التطبيقي باستخدام الحاسب الآلي**"، دار الرضا للنشر، دمشق، 2003، ص210.

8- راجع أحمد عودة عبد المجيد، زين العابدين عبد الرحيم البشير، "**الاستدلال الإحصائي**"، مطابع جامعة الملك سعود، الرياض، 1997، ص ص 58،90.

9- جورج كانافوس، دون ميلر، "**الإحصاء للتجاريين مدخل حديث**"، مرجع سبق ذكره، ص ص 264،268، 269.

10- نفس المرجع السابق، ص270.

11- ناظم الشمري، "**الإحصاء الرياضي**"، مرجع سبق ذكره، ص183.

12- Jean- Jacques Croutsche , « **Statistique et analyse des donnée en marketing et gestion »,** Edition ESKA, Paris, 1997, pp 109-110.

13- ناظم الشمري، "**الإحصاء الرياضي**"، مرجع سبق ذكره، ص ص 186-187.

14- نفس المرجع السابق، ص189.

15- Bernard Verlant, Geneviève Saint- Pierre, « **Statistiques & probabilités** », Berti Edition, Algérie, 2008, p 157.

16- ناظم الشمري، "**الإحصاء الرياضي**"، مرجع سبق ذكره، ص ص 188-189.

17- أنيس إسماعيل كنجو، "**الاحصاء والاحتمال**"، مرجع سبق ذكره، ص ص 356-357.

18- Bernard Verlan, Geneviève saint- pierre, « **Statistiques et probabilités** », op,c.i.t, p112 .

19- ناظم الشمري، "**الإحصاء الرياضي**"، مرجع سبق ذكره، ص163.

20- روبر مارتن بلان، "**المدخل إلى الإحصاء الطبي**"، المركز العربي للتعريب والترجمة التأليف والنشر، دمشق، 2000، ص 136.

21- أنيس إسماعيل كنجو، "**الإحصاء والاحتمال**"، مرجع سبق ذكره، ص357.

22- جورج كانافوس، دون ميلر، "**الإحصاء للتجاريين مدخل حديث**"، مرجع سبق ذكره، ص314.

23- أحمد عودة عبد المجيد، زين العابدين عبد الرحيم البشير، "**الاستدلال الإحصائي**"، مرجع سبق ذكره، ص ص141-143.

24- جورج كانافوس، دون ميلر، "**الإحصاء للتجاريين مدخل حديث**"، مرجع سبق ذكره، ص ص 321-322.

25- أحمد عودة عبد المجيد، زين العابدين عبد الرحيم البشير، "**الاستدلال الإحصائي**"، مرجع سبق ذكره، ص ص 143-144.

26- Murray R- Spiegel, « **Probabilités et statistique** », Série Schaun, Paris, 1984, p 169.

27- أحمد عودة عبد المجيد، زين العابدين عبد الرحيم البشير، "**الاستدلال الإحصائي**"، مرجع سبق ذكره، ص ص 48، 146، 147.

28- جورج كانافوس، دون ميلر، "**الإحصاء للتجاريين مدخل حديث**"، مرجع سبق ذكره، ص384.