نظرية النهاية المركزية و الاستدلال الإحصائي

د. ديلمي لخضر جامعة باتنة

د. سحنون محمد جامعة قسنطينة

الخلاصة:

يهدف المقال إلى عرض نظرية النهاية المركزية بأسلوب خال من التعقيد التقني والرياضي بتوظيف تقنية المحاكاة في الحزمة الإحصائية R. وبهذا يمكن لغير الرياضيين الاستفادة من الإحصاء وتوظيفه في مجالات عملهم. ولقد توصل البحث إلى:

1- أن تطور المعلوماتية قد أحدث تغيرات هامة في علم الإحصاء حيث أصبح بالإمكان فهم وتطبيق الأفكار الإحصائية باستخدام المعلوماتية.

2- بينت نظرية النهاية المركزية أنه كلما كررنا التجربة كلما كان خطأ المعاينة صغيرا وبالتالي كلما كان تقدير متوسط المجتمع أكثر دقة.

مقدمة:

تتمثل المسألة المركزية للاستدلال الإحصائي في استخدام بيانات العينة للتعرف على ملامح المجتمع الذي سحبت منه. كأن نحاول تقدير متوسط المجتمع μ بناءا على متوسط العينة $\bar{X}$. على أن مثل هذا التقدير يتطلب – ضمن أمور أخرى – معرفة توزيع معاينة إحصاء العينة. و يستخدم الإحصائيون بيانات العينة و النظريات الإحصائية لتحديد هذا التوزيع. و من النظريات الإحصائية التي تعتمد لتحديد توزيع معاينة متوسط العينة، نظرية النهاية المركزية. و يقر الإحصائيون أن برهان هاته النظرية معقد نوعا ما بالنسبة لغير الرياضيين(1). و إذا كان معظم العاملين في الإحصاء هم في الحقيقة من الرياضيين، إلا أن البعض منهم ليسو كذلك. فهناك العديد منهم من يمتلك المقدرة على تطبيق الإحصاء في الحقول التي يعملون فيها. و ربما كان من غير المفيد كثيرا دراسة الجانب النظري في الإحصاء دون الاهتمام بتطبيقاته .

و في هذا البحث سنبذل جهدا لنبين أنه بالإمكان فهم و تطبيق نظرية النهاية المركزية دون استخدام أي رياضيات تذكر، من خلال توظيف تقنية المحاكاة في الحزمة الإحصائية $\mathcal{R}$.

إن هدفنا هو تجسير الهوة بين المفاهيم الرياضية و مختلف مستخدمي الإحصاء، وتبيان أنه بالإمكان فهم و تطبيق العلاقات الإحصائية دون الخشية من الغرق في خضم الرياضيات.

و لتحقيق الهدف أعلاه قسم البحث إلى ثلاث أقسام و خاتمة ، تناول القسم الأول بعض المفاهيم الإحصائية المساعدة في فهم نظرية النهاية المركزية ، أما القسم الثاني فا ستعرض تعليمات $\mathcal{R}$ التي يمكن بواسطتها محاكاة بعض التوزيعات الاحتمالية و في القسم الثالث تم التحقق من نظرية النهاية المركزية باستخدام الحزمة $\mathcal{R}$ أما الخلاصة فقد خصصت للاستنتاجات.

1- الإحصاءات هي متغيرات عشوائية:

يستخدم الإحصائيون كلمة مجتمع في سياق المجموعة التي نرغب في الحصول على معلومات عن الصفات (المتغيرات) المدروسة . و تستخدم المعلمات كالوسط الحسابي و الانحراف المعياري و النسبة لتلخيص هاته الصفات. و تعطي المعاينة وسائل جذابة للتعرف على ملامح المجتمع مقارنة مع ما توفره التعدادات الشاملة. فالمعاينة يمكن أن تؤدي إلى تخفيض التكاليف و تضمن السرعة كما يمكن أن تحقق دقة عالية بسبب محدودية الجهد بجانب إشراف أكثر فعالية. ومن الأساليب الأكثر شيوعا أسلوب المعاينة العشوائية البسيطة الذي يكمن في اختيار عينة بحيث يكون لكل مفردة من مفردات المجتمع فرص مستقلة و متساوية لاختيارها ضمن العينة.بمعنى أن مفردات المجتمع لا يفضل بعضها في عملية الاختيار. كما يعطي أسلوب المعاينة العشوائية الوسيلة التي يمكن بواسطتها تقدير الدقة و التحكم في خطأ المعاينة من خلال اختيار حجم العينة[2]. و يتم تلخيص صفات العينة كما هو الشأن بالنسبة للمجتمع باستخدام الكميات كالوسط الحسابي و الانحراف المعياري و النسبة و تسمى في هاته الحالة إحصاءات و تستخدم لتقدير المعلمات و مبدئيا لابد أن نعرف أن كل إحصاء يتذبذب في قيمته من عينة إلى أخرى بينما تبقى المعلمات ثابتة و هي عادة مجهولة. فالقيمة التي نحسبها من عينة واحدة تعتمد على مفردات العينة التي يتم اختيارها لذلك فإن كل إحصاء تحددت قيمته من عينة عشوائية هو متغير عشوائي و بالتالي فهناك تنويعات مختلفة من النتائج للمتوسط الحسابي $\bar{X}$ و الانحراف

المعياري (sd) و النسبة p. إن مفتاح فهم فائدة إحصاء ما يكمن في فهم كيف يتغير هذا الإحصاء من عينة عشوائية إلى أخرى. بمعنى أننا نحتاج إلى فهم توزيع القيم الممكنة لهذا الإحصاء خلال كل العينات العشوائية الممكن سحبها من المجتمع.

توزيع قيم إحصاء ما لكل العينات العشوائية الممكنة يسمى بتوزيع المعاينة . إن توزيع المعاينة لإحصاء ما له نفس الخصائص الإحصائية التي لأي توزيع احتمالي. فمثلا له متوسط وانحراف معياري و شكل بياني معين. لذا فإن توزيع المعاينة هو ببساطة توزيع احتمالي يستخدم بصفة خاصة للنتائج الخاصة لإحصاء ما لكل العينات العشوائية الممكنة المتساوية الحجم والمسحوبة من المجتمع موضع الاهتمام. وكما أشرنا أعلاه فإن الإحصاءات تستخدم كتقديرات لمعلمات المجتمع، و في هذا السياق فإن الانحراف المعياري لتوزيع معاينة ما (أي الانحراف المعياري لإحصاء ما) يقيس دقة هذا الإحصاء، أو بمعنى آخر فالانحراف المعياري لإحصاء ما يدل على الدرجة التي يمكن أن تبتعد بها قيم هذا الإحصاء عن قيم المعلمة الحقيقية من عينة إلى أخرى(3). و عليه يكون من المرغوب فيه أن يكون للإحصاء انحراف معياري صغير بحيث أنه و لأي عينة، يكون من غير المحتمل أن تبعد هذا الإحصاء كثيرا عن القيم المعلمة. يسمى الانحراف المعياري لإحصاء ما بالخطأ المعياري . و عليه فإن حساب الخطأ المعياري الفعلي أو الحقيقي لإحصاء ما يتطلب الأخذ في الاعتبار جميع العينات العشـــوائية الممكنة و التي هي من نفس الحجم، و من الواضح أن هذا أمر مستحيل. وعمليا فإننا نختار عينة واحدة في أي وقت ثم نسجل قيمة واحدة فقط من القيم الممكنة لإحصاء ما.

لنفرض أننا سحبنا عينة عشوائية من مجتمع ما متوسطه μ غير معلوم و لنفرض أكثر أن متوسط العينة المسحوبة كان $\bar{X} = 30$. و تعد هذه القيمة أفضل تقدير لــ μ.

والآن الى أي مدى تبتعد هذه القيمة عن μ ؟ بالطبع نحن لا نعرف ذلك بكل تأكيد لأننا لا نعرف قيمة محددة لــ μ. و مع ذلك فتوزيع المعاينة لهذا الإحصاء $\bar{X}$ يخبرنا الى أي مدى تنحرف قيمة $\bar{X}$ في العينة عن قيمة μ. فمثلا إذا عرفنا أن قيمة $\bar{X}$ تقع في حدود 4 وحدات (زائدا أو ناقصا) من μ و ذلك في 95% من جميع العينات العشوائية المتساوية الحجم عندئذ نكون على ثقة بأنه مع هذه العينة سنجد أن القيمة الفعلية لــــــ μ تقع ما بين 26 و 34. و

يو ضح هذا المثال أن أول خطوة في تحديد منفعة إحـصاء لعمل ا ستنتاج حول معلمة هو التعرف على توزيع المعاينة لهذا الاحصاء.

والآن كيف يمكن تحديد توزيع المعاينة لإحصاء ما إذا كان لدينا عينة واحدة فقط ؟

بالطبع عينة واحدة بمفردها لا تعطي هذه المعلومات، و مع ذلك و بدمج المعلومات المستمدة من عينة عشوائية واحدة مع بعض النظريات الإحصائية، يمكننا على الأقل تحديد توزيع المعاينة بصورة تقريبية [4]. و فيما تبقى من هذا البحث سنستعرض المفاهيم الإحصائية التي تسمح لنا بتحديد توزيع المعاينة لأحد الإحصاءات الهامة و هو متوسط العينة و هذا بالاستناد الى نظرية النهاية المركزية.

وتعرض نظرية النهاية المركزية و تحت شروط عامة جدا أن كلا من المجموع و متوسط عينة عـشوائية مـسحوبة من مجتمع ما، يمتلك عند تكرار هذه العينات، عددا كبيرا من المرات، توزيعا له تقريبا شكل الجرس. و لتوضيح الفكرة الأساسية التي تتضمنها النهاية المركزية نعرضها في العبارة المبسطة التالية: إذ سحبنا عينات عـشوائية حجم كل منها n من مجتمع متو سطهμ وانحرافه المعياري σ محدودان، فإن توزيع متو سط العينة $\bar{X}$، يتطابق مع التوزيع الطبيعي بمتو سط يساوي μ و انحراف معياري $\frac{\sigma}{\sqrt{n}}$و تزداد دقة التقريب كلما ازدادت n [5].

و يمكن اعادة صـــياغة النظرية لتتفق مع $\sum_{i=1}^{n} X_i$ بدلا من $\bar{X}$ فنقول أن توزيع $\sum_{i=1}^{n} X_i$ يسعى الى أن يصبح طبيعيا بمتوسط $n \times \mu$ و انحراف معياري$\sigma\sqrt{n}$.

و تبدو أهمية نظرية النهاية المركزية من زاويتين: فهي توضـــح أولا نزوع العديد من المتغيرات العـشوائية الى أن يكون توزيعها، بـصورة تقريبية هو التوزيع الطبيعي، إذ يمكن مثلا أن نتـصور طول الانـسان حـصيلة عدد كبير من المؤثرات العـشوائية ، مثل طول الأب و طول الأم والمورثات (وعددها كبير) و نـشاط الغدد ذات العلاقة بالطول و البيئة و المحيط بأنواعه والتغذية الخ... وإذا كانت آثار هذه العوامل تـضاف بعـضها إلى بعض، لتنتج واقعا معينا بالنـسبة لطول الإنـسان فعندئذ يمكن اعتبار الطول كحـصيلة لعدد كبير من المتغيرات العـشوائية. وهكذا تنطبق نظرية النهاية المركزية ويكون الطول هو على درجة التقريب التوزيع الطبيعي و ذلك بصـــرف

النظر عن توزيع أي من المتغيرات التي تؤثر في تحديد الطول[6]. وهذا بالطبع محاولة للتعليل ليس أكثر إذ ما يجري في الواقع غير معروف لنا بصـــورة دقيقة، ولكن ما يمكن قوله، على كل حال هو أن نظرية النهاية المركزية توضح سبب وجود العديد من المتغيرات العشوائية التي نصادفها في حياتنا والتي نعتبر أن توزيعها الاحتمالي هو التوزيع الطبيعي.

كما يمكن بالاعتماد على نظرية النهاية المركزية تقريب التوزيع الثنائي الذي نصادفه في مختلف التطبيقات و خ صو صا الطبية منها من التوزيع الطبيعي. فإذا ا صطلحنا على أن يوافق النتيجة (S) أو النجاح العدد 1 و يوافق النتيجة (E) (أو الفشل) العدد صفر 0 . فعندئذ تكون نتائج التكرارات المستقلة لـــ n عبارة عن متتالية من المتغيرات المستقلة: $X_1 \dots X_n$ حيث يأخذ X_i إما القيمة 1 أو القيمة 0 و يكون عدد النجاحات X هو بالضـــبط عدد مرات ورود 1 في تلك المتتالية أو مجموعها :

$$X = \sum_{i=1}^{n} X_i$$

وبما أن X_i يتوزع وفق توزيع برلوني، عندئذ تصبح نتائج التكرارات المستقلة الــــــ n هي $X_1 \dots X_n$ عبارة عن عينة عشوائية من مجتمع برلوني و يصبح X مجموع العينة وفقا النهاية المركزية أعلاه يكون التوزيع التقريبي لــ X في حالة n كبير بكفاية هو التوزيع الطبيعي بمتوسط $n \times p$ (حيث p احتمال النجاح في التجربة الواحدة) و تباين يســـــاوي $n \times p \times q$ وبالتالي يمكن ا ستخدام جدول التوزيع الطبيعي القيا سي لح ساب احتمالات تتعلق بالمتغير X بصورة تقريبية[7].

أما من أجل توزيع بوا سون فالأمر يختلف، فإذا أخذنا مجموعة من متغيرات بوا سون لها المتوسط نفسه و جمعناها معا فنحصل على متغير يمثل عدد الحوادث العشوائية في مجال زمني كبير (مجموع المجالات المرافقة للمتغيرات المختلفة). وهو يحقق توزيع بوا سون بمتو سط متزايد. وبما أن مجموع عدد من المتغيرات الع شوائية الم ستقلة التي لها نفس التوزيع ي سعى الى التوزيع الطبيعي، عندما يزداد المتوسط فإن توزيع بواسون يؤول إلى التوزيع الطبيعي عندما يزداد المتوسط[8].

ومن زاوية أخرى نجد أن العطاء الأكثر أهمية لنظرية النهاية المركزية يتعلق بمسألة الإ ستدلال الإحـ صائي. فالعديد من الإحـ صاءات تـ ستخدم للقيام با ستدلال حول معلمات التوزيعات مثل احتمال النجاح في التوزيع الثنائي (p) و متو سط التوزيع الطبيعي(μ) الخ... هذه الاحصاءات تأخذ شكل مجموع لقياسات العينة أو متوسط هذه القياسات. وإذا كان الحال كذلك و كانت n كبيرة بكفاية فيمكننا اعتبار التوزيع الطبيعي تقريبا جيدا للتوزيع الاحتمالي لذلك الاحصاء. وهو ما تمس الحاجة اليه عند القيام بأي استدلال إحصائي. ورغم هذه الإ ستخدامات المفيدة لنظرية النهاية المركزية فإن برهانها يتسم بالتعقيد بالنسبة بغير الريا ضيين. ولذا سنكتفي هنا بالبرهان على أن:

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} \text{ و } \mu_{\bar{X}} = \mu_X$$

وسنترك التحقق من التوزيع الاحتمالي للقسم الثالث من البحث وهذا بعد تقديم موجز للحزمة الإحصائية $\mathcal{R}$.

و في الواقع لدينا بموجب تعريف التوقع الرياضي ما يلي :

$$\mu_{\bar{X}} = E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu_X = \mu_X$$

$$\sigma_{\bar{X}}^2 = E(\bar{X} - \mu_{\bar{X}})^2 = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu_{\bar{X}}\right)^2$$

$$= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{n}{n}\mu_{\bar{X}}\right)^2$$

$$= \frac{1}{n^2} E\left(\sum_{i=1}^{n} X_i - n\mu_{\bar{X}}\right)^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E(X_i - \mu_X)^2 = \frac{1}{n^2} n\sigma_X^2 = \frac{\sigma_X^2}{n}$$

2- تعليمات الحزمة الإحصائية $\mathcal{R}$ لمحاكات التوزيعات الإحتمالية: .

إن الحزمة الإحصائية هي لغة برمجة و بيئة تطوير رياضي تستخدم لمعالجة البيانات و يمكن استخدام هذه الحزمة للقيام بتحليل إحصائي بسيط أو معقد للبيانات كالطرق الخطية و اللاخطية، إختبارات الفروض، نمذجة السلاسل الزمنية والتصنيف كما تتوفر الحزمة على العديد من الدوال البيانية المفيدة ذات المواصفات الجيدة .

و يمكن إعتبار الحزمة $\mathcal{R}$ كلغة برمجة كاملة و بهذا فهي تختلف عن بقية الحزم الإحصائية . كما تتماشى مع العديد من أنظمة التشغيل كالماكنتوش و الويندوز و لينكس. و لقد تم كتابة الحزمة بلغة البرمجة C و C++ و جافا و هي لغة برمجة بالكائنات. و لقد تم تطوير هذه الحزمة من طرف الأســـتاذين Ross Ihaka و Robert Genthmen من جامعة أوكلند في نيوزلندا الجديدة. و الحزمة الإحصائية $\mathcal{R}$ هي حزمة مجانية يمكن تحميلها مباشرة من الشبكة العنكبوتية وهي تتطور بسرعة حيث تظهر نسخة جديدة كل ستة أشهر.

ويمكن بواســـطة الحزمة الإحصـــائية $\mathcal{R}$ محاكاة الظواهر التي يتم نمذجتها عن طريق التوزيعات الإحتمالية ويعطي الجدول أدناه التوزيعات المتوفرة في الحزمة $\mathcal{R}$ [10].

الجدول –1– أنواع التوزيعات الإحتمالية في $\mathcal{R}$.

الوسيط	اسم التوزيع في الحزمة $\mathcal{R}$	نوع التوزيع
الحجم ، الاحتمال ، np ,	binom	الثنائي
$\frac{1}{mean}$	exp	الأسى
درجة الحرية$df1$ ، درجة الحرية $df2$	f	فيشر
$prob$	geom	الهندسي
df	chisq	كاي مربع
$mean, sd$	norm	الطبيعي
λ	pois	بواسون
df	t	ستودنت
min ,max	unif	المنتظم

يوضح الجدول أنه يتم التعبير عن كل توزيع باسم ووسطاء. فمن أجل التوزيع الطبيعي هناك المتوســـط و الانحراف المعياري. و يمكن القيام بأربع عمليات لكل توزيع من التوزيعات أعلاه:

1– توليد أرقام عشوائية من التوزيع و ذلك بوضع الحرف r قبل اسم التوزيع

2– حساب الاحتمال الموافقة لقيمة من قيم الوسط و ذلك بوضع الحرف p قبل اسم التوزيع .

3– حساب قيم الوسيط الموافقة لاحتمال معطى و ذلك بوضع الحرف q قبل اسم التوزيع .

4– تمثيل تابع كثافة الإحتمال و ذلك بوضع الحرف d قبل اسم التوزيع.

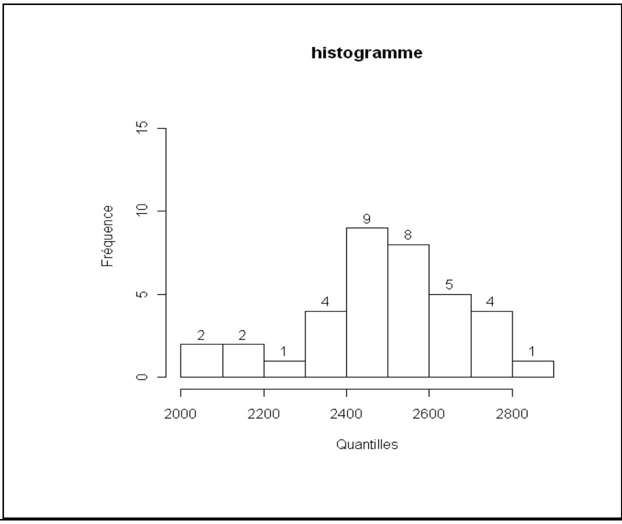
فلتوليد عينة ع شوائية حجمها 36 من مجتمع يتوزع توزيعا طبيعيا بو سط ح سابي 2500 كغ وانحراف معياري قدره 200 كغ و من ثم ح ساب متو سط و الانحراف المعياري للعينة، يمكن استخدام التعليمات التالية:

```
> # On tire un échantillon de 36 individus d’une
> # population de moyenne 2500 kg et d’un ecart-type 200 kg
> échantillon = rnorm ( n=36 , mean = 2500 , sd = 200)
> #Arrondir à 0 chiffres après la virgule
> échantillon = round ( échantillon , 0 )
> #Calculer la moyenne de l’echantillon
> Moyenne = mean (échantillon )
> #Calculer l’ecart type
> EcartType = sd(échantillon )
> # Impriment tout
> échantillon
 [1] 2739 2578 2318 2412 2671 2543 2325 2335 2401 2789 2594
2615 2557 2477 2085
[16] 2408 2665 2417 2030 2818 2429 2472 2552 2181 2167 2395
2662 2799 2510 2785
[31] 2492 2469 2594 2584 2231 2659
> Moyenne
[1] 2493.278
> EcartType
[1] 199.3638
```

كما يمكن ترسيب عناصر العينة المستخرجة و حساب وسيطها وربيعها الأول و الثالث و كذلك انشاء توزيعها التكراري و هذا باستخدام التعليمات التالية :

```
> #ordonner les éléments de l’échantillon
> sort ( Echantillon )
Erreur dans sort(Echantillon) : objet 'Echantillon' introuvable
> #calcul de la médiane et 1e et 30 quartile
> summary (échantillon )
 Min. 1st Qu. Median Mean 3rd Qu. Max.
 2030 2400 2501 2493 2626 2818
> # construction de l’histogramme
> hist(échantillon, labels=TRUE, ylim=c(0,17), xlab="Quantilles",
ylab="Fréquence",main="histogramme")
```

أما لحساب إحتمال أن تكون قيمة المتغير أقل أو تساوي نل 2400 فنستخدم التعليمة




```
> pnorm(q=2400,mean=2500,sd=200)
[1] 0.3085375
```

و لحساب قيمة المتغير الموافقة للاحتمال $p = 0.31$ فنستخدم التعليمة

```
> qnorm(p=0.31,mean=2500,sd=200)
[1] 2400.83
```

بعد أن تعرفنا على التعليمات و الأوامر التي تســـتخدم لمعالجة التوابع الاحتمالية في الحزمة الإحصائية $\mathcal{R}$، نهتم الآن بمحاكاة نظرية النهاية المركزية .

3- التحقق من نظرية النهاية المركزية باستخدام الحزمة الإحصائية $\mathcal{R}$:.

في هذا القسم سنتحقق من نظرية النهاية المركزية باستخدام الحزمة الإحصائية $\mathcal{R}$. ولهذا فقد قمنا باســـتخراج عينات من مجتمع يتوزع توزيعا طبيعيا و آخر له توزيع هندســـي و أخيرا مجتمع يتوزع توزيع بواسون .

A- المجتمع يتوزع توزيعا طبيعيا

لنفترض أن لدينا مجتمعا يتكون من 10000 قطعة أرض ، يبلغ الانتاج المتوسط لهذا المجتمع 2500 كغ للهكتار و يبلغ الإنحراف المعياري 200 . يمكن محاكاة هذا الوضع باستخدام الحزمة $\mathcal{R}$ على النحو التالي:

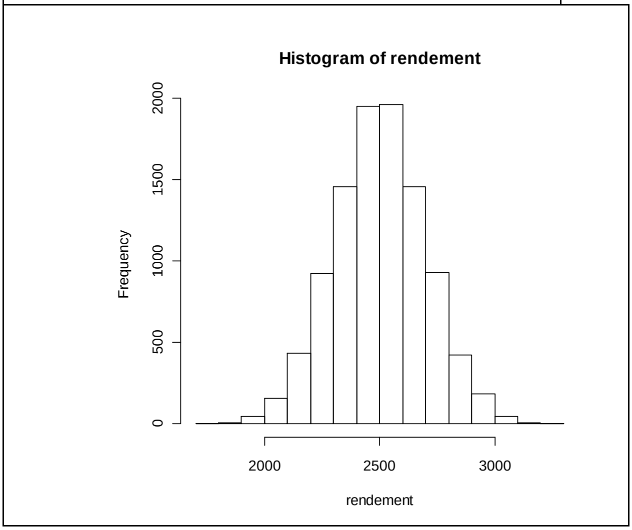
```
N =10000
> # rnorm genere une variable de N observations qui a une deviation
Normale
> #de moyenne 2500 et d'un ecart type 200.
> rendement = rnorm (N,mean =2500,sd = 200 )
> #on arrondit à zero chiffres après la virgule
> rendement = round (rendement )
```

وبهذا نحصل على شعاع يتكون من 10000 مركبة تمثل كل منها انتاج القطع الأرضية و يمكن استخراج بعض هذه المركبات على النحو التالي :

```
> rendement [1:20]
 [1] 2651 2384 2845 2346 2503 2643 2562 2648 2575 2571 2743
2724 2570 2646 2447
[16] 2307 2617 2266 2590 2440
```

و لهذا المجتمع وسط حسابي حقيقي قدره 2500 و انحراف معياري حقيقي قدره 200 و الشكل أدناه يعطي المدرج التكراري لهذا التوزيع

```
> hist (rendement )
```



نفترض أنه لا تتوفر لدينا الإمكانيات الا لســـحب عينة حجمها 36، ولنحاكي هذا السحب ولهذا الغرض نستخدم التعليمات التالية :

```
> hist (rendement )
> # la commande sample() permet d'effectuer le tirage aleatoire de 36 rendements sur 10000
> echantillon=sample (rendement, 36 )
> echantillon
 [1] 2690 2447 2506 2410 2301 2541 2308 2551 2477 2331 2305 2432 2586 2345 2603
[16] 2319 2240 2404 2797 2593 2747 2904 2485 2120 2375 2472 2703 2971 2447 2600
[31] 2649 2072 2541 2748 2754 2242
```

يبلغ متو سط هذه العينة $\overline{\quad} = 2500.444\ \mathrm{kg}$ و من أجل مجتمع به 10000 قطعة يمكننا ســـحب 2.5×10^{102} عينة ممكنة حجم كل واحدة 36 . و لكل عينة من هاته العينات متوسط و انحراف معياري .

```
> choose (10000,36)
[1] 2.523898e+102
```

و اذا كنا لا نستطيع أن نحصل على انتاج مليار قطعة أرض فالإمكان التزول الى الميدان وقياس إنتاج 36 قطعة و الحـ صول على متو سط انتاج هذه القطع ، و الـ سؤال المطروح: هل يمكن الحـ صول على تقدير للإنتاج المتو سط للقطع المؤلفة للمجتمع إذا علمنا الإنتاج المتو سط للعينة ؟

لنشكل شعاعا عدد مركباته 5000 تمثل كل واحدة منها متوسط انتاج عينة حجمها 36 مأخوذة من مجتمع به 10000 مفردة.

```
# initialisation du vecteur
> M=5000
> moyennes = rep(0,M )
> #on tire 5000 Echantillon de 36 elements
> # et le resultat est mis dans le vecteur moyenne
> for ( i in 1 : M ) moyennes[i] = mean(sample(rendement,36))
```

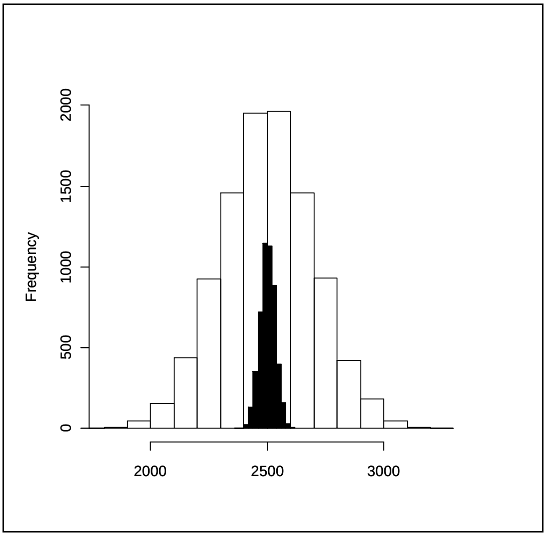
ويمكننا الحصول على مستخرج من هذا الشعاع على النحو التالي :

```
moyennes[1:20]
```

```
 [1] 2517.667 2514.667 2498.750 2498.750 2493.556 2494.528 2518.722
2525.389
 [9] 2484.000 2521.750 2484.083 2476.944 2495.111 2557.194 2448.500
2536.667
[17] 2539.389 2504.111 2515.250 2463.778
```

لنمثل المدرج التكراري الذي يمثل المجتمع و المدرج التكراري الذي يمثل توزيع المعاينة في نفس المعلم :

```
>
hist(rendement,xlim=c(1800,3400),ylim=c(0,2000),main="",xl
ab="")
> par(new=TRUE)
>hist(moyennes,
xlim=c(1800,3400),ylim=c(0,2000),main="",xlab="",col=1)
```



ومن الشكل (المدرج بالأسود) يتبين لنا أن توزيع المعاينة للمتوسط هو التوزيع الطبيعي، بمتوسط يساوي متوسط المجتمع و انحراف معياري يساوي بالتقريب انحراف المجتمع مقسوما على $\sqrt{36}$ و يمكن التأكد من ذلك عن طريق التعليمات التالية :

```
> sd (rendement)
[1] 199.1647
> sd (moyennes)
```

```
[1] 32.90549
> (sd(rendement))/(sqrt (36))
[1] 33.19411
```

B– توزيع معاينة الوسط الحسابي : حالة البيانات المستخرجة من التوزيع الأسي

نعلم أن تابع الكثافة متغير عشوائي يتوزع توزيعا أسيا

$$f(x) = \lambda e^{-\lambda x}$$

ويمكننا التأكد أن :

$$E(X) = \frac{1}{\lambda} \qquad var(X) = \frac{1}{\lambda^2}$$

يســـتخدم هذا التوزيع في دراســـة الكفاءة لتمثيل فترة حياة الدارات الالكترونية[11].

وي سمى الثابت $\frac{1}{\lambda}$ (mean time between failure) و λ هو معدل الإخفاق ذلك أن :

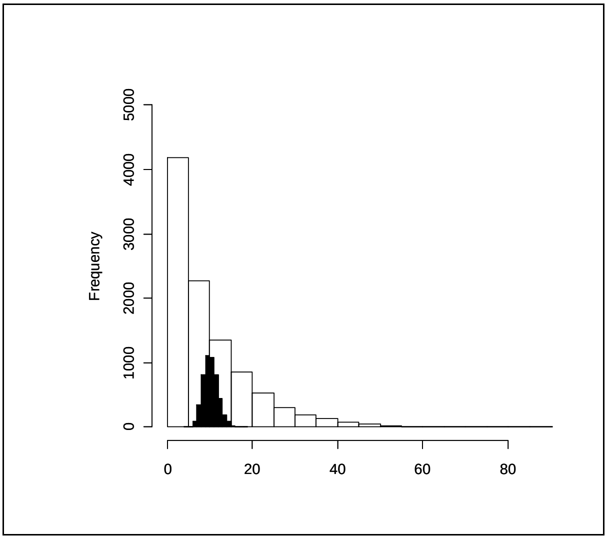
$$h(X) = \frac{f(x)}{1 - F(X)}$$

لتكن لدينا عملية تتبع التوزيع الأ سي بمعدل $\frac{1}{10}$ أي أنها تابع أ سي بو سط حسابي 10 و تباين 100.

لنولد عينات عشـــوائية حجم كل واحدة منها 36 من هذا التوزيع و لتحقيق ذلك نستخدم التعليمات أدناه

```
> N=10000
> #rexp genère 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 ( i in 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
```



C- توزيع المعاينة : حالة البيانات مستخرجة من توزيع بواسون

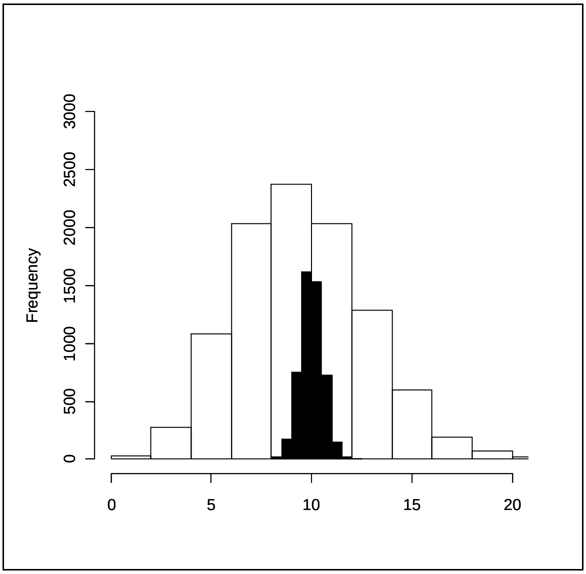
إذا كان X متغير عشوائي يمثل عدد مرات حدوث حوادث مستقلة عن بعضها البعض تقع بمعدل متوسط ثابت عبر الزمن ، الفراغ أو الحجم ، فإن x يكون لها توزيع بواسون له دالة إحتمال

$$p(X = x) = p(n; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

ويلاحظ أن القيم الممكنة للمتغير العشوائي البوا سوني هي الصفر و الأرقام الصحيحة الموجبة، ويلاحظ أيضا أن متوسط توزيع بواسون من أجل عدد من الحوادث في وحدة الزمن هو ببساطة λ كما أن تباين هذا التوزيع يساوي λ أيضا و لهذا التوزيع أهيمة خاصة. إذ أن الوفيات في أمراض كثيرة يمكن النظر اليها على أنها حوادث ع شوائية وم ستقلة في المجتمع[13]، ونعطي فيما يلي التعليمات الخاصة باستدراج 5000 عينة من مجتمع يتوزع توزيع بواسون.

تعليمات $\mathcal{R}$:

```
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
```



من أعلاه يتبين لنا أن نظرية النهاية المركزية قد خلقت لدينا إحـ سا سا بتخمين معين ، فبالرجوع إلى الأشكال السابقة نجد أن توزيع المعاينة لـ $\overline{X}$ قريبا من التماثل و له شكل ربوة عندما تكون n كبيرة ولو كان توزيع المجتمع ليس طبيعيا فما السـبب في ذلك؟ في العينات الكبيرة نكون أكثر قناعة بأننا نحصـل على عينة بيانات نموذجية تحتوي على كل القيم التي هي أعلى و أدنى من متو سط المجتمع. والنتيجة أنه لأي عينة عـ شوائية كبيرة تكون فر صة وقوع $\overline{X}$ أعلى قليلا من μ مساوية لفر صة وقوعها أدنى قليلا من μ و بالتالي اذا كانت العينة ذات حجم كبير بدرجة كافية ($n \geq 30$) فان توزيع النتائج الممكنة لـــ $\overline{X}$ سيكون متماثلا و له قمة وحيدة و نعلم أنه إذا كان لمتوسط العينة توزيع طبيعي فإن $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ يكون له توزيع طبيعي معياري. هذه النتيجة تفترض مـ سبقا أن σ هي ثابت معلوم. ولكن اذا كانت σ غير معلومة ، فإن Z تكون دالة في معلمة غير معلومة و من ثم لا يمكن تحديد قيمة Z لعينة محدودة. ويبدو أن هذا يخلق مشكلة، حيث أنه من الناحية العلمية، نادرا ما تكون قيمة الانحراف المعياري في المجتمع معلومة. وقد نتساءل لماذا لا يتم إستبدال σ بمقدرها أي الانحراف المعياري في العينة؟ ان استبدال σ بالتقدير s في الصيغة أعلاه يؤدي الى الكمية t حيث :

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

إن توزيع المعاينة لهذا الإحصاءt ليس توزيع طبيعي معياري حتى ولو كان توزيع المجتمع هو التوزيع الطبيعي. ولفهم الســـبب في ذلك نقارن الكميتين $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ و $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ فبالنسبة لـ Z هناك متغير عشوائي واحد هو $\bar{X}$ ($\mu\sigma$, ثوابت) أما الإحصاء t فيعتمد على متغيرين ع شوائيين $\bar{X}$ s, . إن إدخال إح صاء إ ضافي يزيد من اختلاف قيمة t من عينة الى أخرى مقارنة بـ Z لذا يجب الا نتوقع أن تكون توزيعات Zt , هما نفس الشيء. في الحقيقة فان توزيع الإح صاء t لجميع العينات العشوائية ذات الحجم n والمسحوبة من مجتمع له توزيع طبيعي يسمى توزيع ستودنت. وهذا التوزيع يشبه التوزيع الطبيعي المعياري من حيث انه متماثل ومركز حول ال صفر و لكنه أكثر ت شتتا و إختلافا و يعتمد هذا الت شتت على حجم العينة، فاذا كانت n كبيرة بدرجة كافية فإن s تصبح تقديرا دقيقا جدا لـــ σ و يكون التشتت في t قليل جدا. واذا كان حجم العينة n صغيرا الى حد بعيد فإن s تكون تقديرا غير دقيق لـــ σ و تظهر t تباينا أكثر. لذا التشتت في التوزيع t يعتمد على حجم العينة n و كلما زادت n فإن التوزيع t يظهر تشتتا أقل و أقل و يصبح متشابها أكثر وأكثر للتوزيع الطبيعي المعياري. في الحقيقة إنهما ي صبحان متطابقين من الناحية النظرية كلما اقتربت n من اللانهاية. وهذا يعني أنه اذا كانت n كبيرة بدرجة كافية، فإن التوزيع الطبيعي المعياري يعد تقريبا جيدا لتوزيع t و يمكن أن يستخدم بدلا منه. والقاعدة المقبولة على نطاق واســـع أن التقريب يعد مقبولا إذا كانت $n \geq 30$. ويلاحظ أن هذه القاعدة الإر شادية تتطابق بـ صورة ملائمة مع القاعدة الإر شادية لتطبيق نظرية النهاية المركزية.

خلاصة و استنتاجات :

* يتسم برهان نظرية النهاية المركزية بالتعقيد بالنسبة لمستخدمي الإح صاء من غير الريا ضيين و هذا ما دفعنا إلى ا ستخدام تقنية المحاكاة لعرض النظرية في صورة مبسطة. و منه يبين أن تطور

الحوا سيب و المعلوماتية قد أحدث تغيرات هامة في علم الإحـصاء حيث أ صبح بالإمكان فهم و تطبيق الأفكار الإحصائية باستخدام المعلوماتية.

* بينت نظرية النهاية المركزية أنه كلما كررنا التجربة كلما كان خطأ المعاينة صـــغيرا و بالتالي كلما كان تقدير المتو سط أكثر دقة. وهي نتيجة هامة بالنسبة لمصممي التجارب حيث تبين أن تكرار التجربة هو شيء مرغوب فيه ما لم تكن هناك قيود تمنع ذلك.

المراجع:

1) Harvey J. Motulsky (2007) : Biostatistique, une approche intuitive. De Boeck, Belgique pp. 58

2) Pierre Ghewy (2010) : guide pratique de l'analyse de données, De Boeck, Belgique p 62

3)- جورج كانافوس، دون ميلر (2004): الاحصـــاء للتجاريين مدخل حديث، دار المريخ للنشـــر، الرياض، المملكة العربية السعودية ص ص 269 ترجمة أ.د سلطان محمد عبد الحميد. أ.د محمد توفيق البلقيني.

4)- جورج كانافوس(2004):مرجع سابق ص 271

5)- اسماعيل كنجو (2000): الاحصاء والاحتمال، مكتبة العبيكان، الرياض المملكة العربية السعودية ص 355

6)- اسماعيل كنجو (2000) مرجع سابق ص 360

7) Bernard Verlaut (2008) : statistique et probabilités. Editions Berti Algérie pp 112

8)- روبر مارتن بلان (2000): المدخل إلى الاحصاء الطبي، المركز العربي للتعريب والترجمة والتأليف والنشر-دمشق- سوريا ص 136

9) Cristian P Robert George Casella (2010) : Méthodes de monte-Carlo avec R. Springer France p 5