المحاضرة الثامنة

5.6.3 مبرهنات النهايات الحدية:

وجدنا من خلال المبرهنة (2) أنّه إذا كانت لدينا عينة عشوائية $X_1, X_2, \dots, X_n$ من المجتمع الطبيعي $N(\mu, \sigma^2)$ ، فإنّ المتغير العشوائي (مجموع عناصر هذه

العينة) له التوزيع الطبيعي $N(n\mu, n\sigma^2)$ ومِنْ ثَمّ يكون للمتغير العشوائي $Z = \frac{\sum_{i=1}^{n} \mathrm{X_i} - n\mu}{\sqrt{n}\sigma}$ التوزيع الطبيعي المعياري $N(0,1)$.

وكذلك فإنّه يكون للمتغير العشوائي $\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} \mathrm{X_i}}{\mathrm{n}}$ (المتوسط الحسابي لعناصر العينة) التوزيع الطبيعي $N(\mu, \frac{\sigma^2}{n})$ ، ومِنْ ثَمّ يكون للمتغير العشوائي $Z = \frac{\overline{\mathrm{X}} - \mu}{\frac{\sigma}{\sqrt{n}}}$ التوزيع الطبيعي المعياري $N(0,1)$ وذلك بصرف النظر عن قيمة حجم العينة n .

ولحسن الحظ إنّ ما سبق ذكره يبقى صحيحاً من أجل أي عينة عشوائية $X_1, X_2, \dots, X_\mathrm{n}$ مأخوذة من أي مجتمع ، بشرط أن يكون حجم العينة n كبيراً بالقدر الكافي، وهذه ما تفيد به مبرهنة النهاية المركزية التي سنقبلها دون برهان.

<u>مبرهنة (4):مبرهنة النهاية المركزية:</u>

إذا كانت $\mathrm{X_1, X_2, \dots, X_n}$ عينة من مجتمع $\mathrm{F(t)}$ متوسطه μ وتباينه σ^2 $(\sigma^2 \neq 0)$، فإنّ توزيع المتغير العشوائي $\mathrm{Y} = \sum_{i=1}^{n} \mathrm{X_i}$ يتقارب من التوزيع الطبيعي $N(n\mu, n\sigma^2)$.

كذلك الأمر فإنّ توزيع $\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} \mathrm{X_i}}{\mathrm{n}}$ يتقرب من التوزيع الطبيعي $N(\mu, \frac{\sigma^2}{n})$ ونتيجة لهذه المبرهنة ، فإنّ توزيع :

- $\frac{\sum_{i=1}^{n} \mathrm{X_i} - n\mu}{\sqrt{n}\sigma}$ يتقارب من التوزيع الطبيعي المعياري $N(0,1)$.
- $\frac{\overline{\mathrm{X}} - \mu}{\frac{\sigma}{\sqrt{n}}}$ يتقارب من التوزيع الطبيعي المعياري $N(0,1)$.

وذلك من أجل قيم n الكبيرة.

تبيّن هذه المبرهنة مدى نزوع مجموع عناصر عينة إلى التوزيع الطبيعي. والسؤال الذي يطرح نفسه هو : كم يجب أن يكون حجم العينة n حتى يصبح التقريب الناشئ عن تطبيق هذه المبرهنة تقريباً جدياً من وجهة النظر العملية.

لسوء الحظ لا يوجد جواب عام ومحدد تماماً لهذه السؤال، ويتعلق الأمر بالمجتمع المأخوذ منه العينة. فكلما كانت درجة التناظر كبيرة في توزيع المجتمع المأخوذ منه العينة كان التقريب جيداً أكثر. ويُلحظ أنّه من أجل $n \geq 30$ ، فإنّ التقريب يكون جيداً، وذلك مهما يكن المجتمع المأخوذ منه العينة.

مثال(48.3):

إذا كان $X_1, X_2, ..., X_n, ...$ متتالية من المتغيرات العشوائية المستقلة التي لكل منها التوزيع البواسوني بوسيط $\lambda = 2$

فإذا كان : $Y_{100} = X_1 + X_2 + \cdots + X_{100}$

فأوجد: $P(190 < Y_{100} < 210)$

الحل:

من المبرهنة 5 نعلم أن :

$$E(X_i) = V(X_i) = \lambda = 2 \qquad (i = 1,2, ...)$$

وحسب مبرهنة النهاية المركزية فإن المتغير العشوائي Y_{100} تقريباً التوزيع: $N(100\mu, 100\sigma^2)$ أي التوزيع $N(200,200)$

ويكون : $P(190 < Y_{100} < 210)$

$$= P\left(\frac{190-200}{10\sqrt{2}} < \frac{190-200}{10\sqrt{2}} < \frac{200-210}{10\sqrt{2}}\right)$$

$$= P(-0.707 < Z < 0.707) = 2\phi(0.707) - 1 \cong 0.52$$

6.6.3 تقريب التوزيع الثنائي بالتوزيع الطبيعي:

قمنا فيما سبق بعدة تطبيقات للتوزيع الثنائي اقتضت جميعها حساب احتمال أن يأخذ X وهو عدد النجاحات من بين n تكراراً ، قيمة معينة أو أن يقع ضمن مجال معين ، وقد اقتصرنا التطبيقات على أمثلة تكون n صغيرة ، وذلك بسبب مشقة الحسابات عندما n كبيرة وتقدم ، مبرهنة النهاية المركزية حلاً لهذه المشكلة، ذلك لأنه يمكن النظر إلى عدد النجاحات X عند تكرار التجربة البرنولية n تكراراً مستقلاً على أنه مجموع لـ n متغيراً مستقلا ولكل منها التوزيع البرنولي بوسيط p.

وهكذا نلحظ أنه إذا كان X متغيراً عشوائياً له التوزيع الثنائي بوسيطين n و p فإن : $X = \sum_{i=1}^{n} X_i$

حيث $X_1, X_2, \dots, X_n$ متغيرات عشوائية برنولية لكل منها الوسيط p ووفقاً لنظرية النهاية المركزية يكون التوزيع التقريبي لـ x في حالة n كبيرة كفاية التوزيع الطبيعي بمتوسط np وتباين npq . وهكذا يمكننا من جديد استخدام جدول التوزيع الطبيعي المعياري لحساب احتمالات تتعلق بالمتغير العشوائي الحداني، ولكن بصورة تقريبية والقاعدة العملية لهذا التقريب تقتضي بأن يتحقق الشرطان الآتيان:

$$nq \geq 5 \quad , \qquad n\mathrm{p} \geq 5$$

وعندها يكون :

$$P(x_1 \leq X \leq x_2) = P(x_1 - \tfrac{1}{2} \leq Y \leq x_2 + \tfrac{1}{2})$$

حيث : $Y \sim N(n\mathrm{p}, \mathrm{npq})$

وقد قمنا بتعديل طرفي مجال تحولات Y لأننا نقوم بتقريب توزيع منفصل بتوزيع مستمر والتعديل الذي أجريناه تبين أنه يحسن التقريب كثيراً وتسمى إضافة أو طرح $\frac{1}{2}$ عملية تصحيح من أجل الاستمرار.

مثال(49.3):

تم اختبار لقاحٍ جديدٍ ضد الزكام ، وقد أعطي اللقاح لمئة شخص ، وتم مراقبتهم من جهة إصابتهم بالزكام لمدة عام ، ولقد نجا 68 منهم من الإصابة بالزكام، ولنفترض أننا نعلم من معلومات سابقة أن احتمال عدم الإصابة بالزكام هو بصورة طبيعية وبدون استخدام اللقاح 0.50 والمطلوب:

أيّ نتائج يمكن استخلاصها من هذه التجربة حول فعالية اللقاح؟

الحل:

لنحسب احتمال نجاة 68 أو أكثر من الإصابة بالزكام تحت الفرض أن $P = 0.5$

فإذا كان X المتغير الدال على عدد الذين نجوا من الإصابة بالزكام خلال العام عندئذ:

$$X \sim b\left(n = 100; p = \frac{1}{2}\right) \approx N(\mu = np; \sigma^2 = npq)$$

حيث: $\mu = np = (100)\left(\frac{1}{2}\right) = 50$

$$\sigma^2 = npq = (100)\left(\frac{1}{2}\right).\left(\frac{1}{2}\right) = 25$$

ومنه : $X \sim N(\mu = 50; \sigma^2 = 25)$

$$P[X \geq 68] = P[X \geq 68 - 0.5] = P[X \geq 67.5]$$

$$P\left[\frac{X-\mu}{\sigma} \geq \frac{67.5-\mu}{\sigma}\right] = P[Z \geq \frac{67.5-50}{5}]$$

$$= P[Z \geq 3.5] = 1 - P[Z \leq 3.5] = 1 - 0.9998 = 0.0002$$

النتيجة تدل على الاعتراف بفعالية اللقاح في الوقاية من الزكام.

مثال(50.3): تدعي شركة لصنع الأدوية بأن أحد الأدوية التي تنتجها تؤدي إلى شفاء %80 من المرضى الذين يعالجون به ، ولاختيار هذا الادعاء أخذت عينة من 100 مريض ، وعولجوا بهذا الدواء ، واتخذ القرار بقبول هذا الادعاء إذا شفي منهم 75 مريضاً أو أكثر.

والمطلوب:

1. ما احتمال رفض ادعاء الشركة إذا كان احتمال الشفاء باستخدام هذا الدواء هو 0.80 فعلاً؟
2. ما احتمال قبول ادعاء الشركة إذا كان احتمال الشفاء لا يتجاوز 0.70 ؟

الحل:

النموذج المدروس يتبع التوزيع الثنائي كون التجربة ثنائية (شفاء أو عدم شفاء) ومكررة تكراراً مستقلاً n=100 مرة.

وإذا كان X عدد المرضى الذين تم شفاؤهم نتيجة استخدام هذا الدواء فإن:

الطلب (1):

$$X \sim b(n = 100; P = 0.80) \approx N(\mu = np; \sigma^2 = npq)$$

$$\mu = np = (100)(0.80) = 80$$

$$\sigma^2 = npq = (100)(0.80).(0.2) = 16$$

ومنه : $X \approx N(\mu = 80; \sigma^2 = 16)$

حيث تم اعتماد التقريب الطبيعي للتوزيع الثنائي لأن n كبيرة.

$$P[X < 75] = P[X < 74.5] = P\left[\frac{X-\mu}{\sigma} < \frac{74.5-80}{4}\right]$$

$$= P[Z < -1.38] = 0.0853$$

وهو احتمال رفض ادعاء الشركة.

الطلب (2): من أجل قبول ادعاء الشركة يجب أن نحسب:

$$\mu = np = (100)(0.7) = 70$$

$$\sigma^2 = npq = (100)(0.70).(0.30) = 21$$

$$\sigma = \sqrt{21} = 4.58$$

$$P[X \ \geq 75] = P[X \geq 74.5] = P\left[\frac{X-\mu}{\sigma} \geq \frac{74.5-70}{4.58}\right]$$

$$= P[Z \geq -0.982] = 1 - P[Z \leq 0.982]$$

$$= 1 - 0.8365 = 0.1635$$

وهو احتمال قبول ادعاء الشركة.

7.3 تمارين غير محلولة (للقسم العملي):

1. تدل الإحصائيات الطبية على أنه 40% من المدخنين يصابون بسرطان الرئة. فما احتمال أن يصاب 4 من المرضى من 10 أشخاص ؟ وما احتمال إصابة 3 على الأقل؟

2. منطقة ريفية يعتقد بأنّ 60% من منازلها مؤمن ضد الحريق ، اختير أربعة مالكي منازل ريفية من المنطقة عشوائياً وقد أمن X منهم ضد الحريق ، اكتب جدول التوزيع الاحتمالي لـ X.

وما احتمال أن يكون ثلاثة منهم على الأقل قد أمنوا بيوتهم ضد الحريق؟

3. لنفترض أن المحركات الأربعة لطائرة تجارية مصممة بحيث تعمل مستقلة عن بعضها عن بعض ، واحتمال عطل أي منها والطائرة في الجو هو 0.01 . والمطلوب: ما احتمال

1. ألا يقع أي عطل ؟ 2. ألا يقع أكثر من عطل؟

4. في مصنع للأدوية ينتج أكياس مصل معين، إذا كان 20% من إنتاجه من هذه الأكياس معيب الصنع ، فاحسب احتمال أن يكون هناك أربعة أكياس معيبة الصنع في عينة من 100 كيس . ثم احسب احتمال أن يكون هناك كيس على الأقل معيب الصنع ضمن العينة المدروسة.

5. تقع حوادث اصطدام الطرق في منطقة معينة بمعدل حادث واحد لكل يومين. والمطلوب:

أ- احسب الاحتمالات الموافقة لـ 0 ، 1، 2 ، 3، 4، 5، 6. حوادث اصطدام في الأسبوع في تلك المنطقة.

ب- ما عدد الاصطدامات الأسبوعية الأكثر احتمالاً؟

ت-كم يوماً في الأسبوع نتوقع أن يمر دون اصطدامات.

؟$\left(\mu = 7 \times \frac{1}{2} = 3.5\right)$

6. إذا كانت نسبة الأشخاص الذين يتوفون بسبب نوع معين من المرض هو 0.002 وكان عدد الذين أمنوا على حياتهم ضد هذا النوع من المرض هو 1000 شخص. وبفرض X يمثل عدد الأشخاص الذين يتوفون بسبب هذا المرض. والمطلوب:

أ- أوجد احتمال ألا تدفع شركة التأمين لأي منهم.

ب-أوجد احتمال أن تدفع الشركة لعدد يتراوح بين شخص واحد وثلاثة أشخاص.

ت-ما العدد المتوقع الذي ستدفعه شركة التأمين لهؤلاء الأشخاص؟

7. إذا كان معدل عدد حوادث المرور في مدينة معينة خلال يوم ماطر هو 3 حوادث. والمطلوب:

أ- ما احتمال وقوع حادث واحد فقط في هذه المدينة خلال يوم ماطر؟

ب- ما احتمال وقوع ثلاثة حوادث على الأكثر في هذه المدينة خلال يوم ماطر؟

8. إذا كان طول الشخص عبارة عن متغير عشوائي X له التوزيع الطبيعي $N(\mu = 175, \sigma^2 = (7.5)^2)$ فكيف يحدد مهندس ارتفاع أبواب الغرف في منزل يقوم بتصميمه بحيث لا يضطر أكثر من 5% من الأشخاص إلى تخفيض رؤوسهم عند الدخول أو الخروج.

9. إذا كان احتمال أن يصاب مريض القلب بأزمة قلبية أثناء المعالجة هو 0.2 فإذا كان لدينا 10 مرضى يعانون الظروف الصحية نفسها فالمطلوب:

أ- ما احتمال إصابة 6 مرضى بأزمة قلبية أثناء العلاج؟

ب- ما العدد المتوقع من هؤلاء المرضى أن يصاب بأزمة قلبية أثناء العلاج؟

10. إذا كانت أعمار أحدث المصابيح تتوزع وفق التوزيع الطبيعي بمتوسط يساوي 5 سنوات وانحراف معياري يساوي سنة واحدة.

احسب احتمال أن يقع متوسط عينة حجمها n=9 من هذه المصابيح في المجال [4.4, 5.2]

11. بافتراض أن احتمال الحصول على طفل أزرق العينين هو $\frac{1}{3}$ ، فإذا كان في الأسرة 6 أطفال . المطلوب:

أ- ما احتمال عدم الحصول على طفل أزرق العينين؟

ب- ما احتمال أن نصف الأطفال ذوو عيون زرقاء؟

ت- ما احتمال الحصول على طفل واحد على الأقل ذي عيون زرقاء؟

ث- ما العدد المتوقع للحصول على أطفال ذوي عيون زرقاء؟

12. إذا كان احتمال وجود شخص يستخدم يده اليسرى في الكتابة في مجتمع ما هو 0.02 تم اختيار عينة عشوائية من الحجم 500 شخصٍ من ذلك المجتمع فالمطلوب:

أ- احسب احتمال وجود ثلاثة أشخاص على الأقل من هؤلاء الأشخاص في العينة المسحوبة يستخدمون اليد اليسرى في الكتابة.

ب- ما العدد المتوقع والتباين من هؤلاء الأشخاص في العينة المسحوبة للذين يستخدمون اليد اليسرى في الكتابة؟

13. إن معدل أطوال طلاب جامعة دمشق هو 165 $c.m$ وبانحراف معياري قدره 8 $c.m$ والمطلوب: ما نسبة الطلاب الذين تتجاوز أطوالهم 175 $c.m$؟

14. إن درجات 300 طالب نجحوا في امتحان مقرر الإحصاء الحيوي كانت تتوزع وفق التوزيع الطبيعي بمتوسط 65 وبانحراف معياري 10 والمطلوب:

أ- ما عدد الطلاب الذين وقعت درجاتهم بين 70 و 80 درجة.

ب- ما أقل علامة حصل عليها طالب من الـ 12% الأوائل ؟ وما أعلى علامة حصل عليها طالب من المجموعة الباقية؟

15. في عيادة أحد الأطباء 15 مريضاً، وقد بدأ باستقبالهم في الرابعة مساءً ، ففي أي ساعة سيكون واثقاً 95% بأنه سينتهي عمله إذا كان يعلم من خبرته السابقة أن أزمنة مقابلة المرضى تتوزع توزعاً طبيعياً بتوقع $\mu = 12$ دقيقة وبانحراف معياري 4 دقائق.

المحاضرة التاسعة

الفصل الرابع

الاستدلال الإحصائي و اختبار الفرضيات

Statistical Inference and Hypothesis Testing

1.4) عزوم العينة و دوالها :

1.1.4) تمهيد : إن هدف الإحصاء كعلم هو القيام باستقراء حول خصائص الجمهرة (المجتمع) اعتماداً على عينة مأخوذة من هذا المجتمع، وكل مسألة إحصائية تبدأ بعينة من القياسات أو الملاحظات ، و لا يخفى أن لطريقة اختيار العينة أثراً حاسماً في الدراسة ، و عندما لا يكون هناك أرجحية لانتقاء عينة دون أخرى ، أي عندما يكون لجميع العينات ذات الحجم نفسه الإمكانات نفسها في السحب ، فمن شأن ذلك أن يدفعنا إلى عد القيم المميزة للعينة هي القيمة المميزة للمجتمع (للجمهرة) الذي أخذت منه العينة ، و يجب الانتباه إلى إن ما يوضع في الحسبان ليس عينة واحدة ، سحبناها و عيّنا خصائصها ، و لكن مجمل العَّينات التي يمكن أن نحصل عليها من المجتمع المدروس إلا أننا لا نعتمد على تلك المعلومات ككيان معزول قائم بذاته ، و إنما نعتمد عليها في سياق سلسلة متكاملة تتضمن العينة المدروسة وغيرها من العينات الممكنة . من أجل ذلك علينا أن نصيغ مفهوم العينة بشكل آخر و يتناسب مع هذا التصور .

لقد لاحظنا أن المتغيرات العشوائية و قوانين توزيعها ما هي إلا نماذج رياضية لدراسة المجتمعات الإحصائية ، و لذلك فإن دراسة المجتمع إحصائياً تعود لدراسة المتغير العشوائي X الذي يصف هذا المجتمع ، وإذا كان لـ X التوزيع $F(x)$ فيمكن القول إن للمجتمع التوزيع $F(x)$ ، و هكذا يمكن التعبير عن القيم المميزة للمجتمع مثل متوسطه μ و تباينه σ^2 بدلالة القيم المميزة للمتغير العشوائي X الذي يمثل هذا المجتمع كما يأتي :

$$\sigma^2 = V(X) \quad , \quad \mu = E(X)$$

و بملاحظة إنَّ عناصر العينة تتغير من تكرار إلى آخر، فيمكن القول إنها قيم لمتغيرات عشوائية مستقلة تمثل المجتمع المدروس .