الفصل 7 نظرية توزيع المعاينة

مفاهيم إحصائية ـ توزيعات المعاينة للمتوسطات ـ توزيع المعينة للنسبة ـ توزيع المعاينة للفروق والمجاميع ـ توزيع المعاينة للتباين ـ توزيع المعاينة لنسبة تباينين

في مجتمعاتنا المعاصرة لا يمر يوم دون أن يعلن عن نتائج استقصاء أجرته مجلة أو جامعة حول مواضيع سياسية أو اجتماعية متعددة، منها الاستقصاءات المثيرة للجدل حول الآراء السياسية للمواطنين أثناء الحملات الانتخابية. في عالم الأعمال تقوم المؤسسات عن طريق مصالح التسويق ومصالح البحوث والتطوير بإجراء استقصاءات للإطلاع على توجهات المستهلكين، ... فما هي الأسس النظرية الرياضية التي تستند عليها الاستقصاءات المختلفة ؟ أو كيف يمكن الاستدلال من خلال بيانات عينة على خصائص المجتمع الذي أخذت منه؟ الإجابة على هذه الأسئلة وغيرها تتطلب فهم العلاقات الرياضية بين الخصائص المختلفة للمجتمع مثل المتوسط، التباين وغيرها، والخصائص المناظرة لها في العينة وهو ما سنتناوله في هذا الفصل. في الفصول المقبلة سندرس عددا من التطبيقات لهذه العلاقات الرياضية.

7-1 مفاهيم إحصائية

المجتمع والعينة
العينة النفادية والعينة غير النفادية
العينة العشوائية
معالم مجتمع
إحصائية المعاينة

7-1-1 المجتمع والعينة

نشرح مصطلحي المجتمع والعينة (Population et échantillon) و نعرفهما من خلال الأمثلة التالية:

- يرغب باحث زراعي في تقدير متوسط عدد حبات القمح لسنبلة القمح المحلي الجزائري "محمد البشير". يكون الباحث عينة من 100 سنبلة. يعد الحبات في السنابل فيحصل على مجموعة من 100 عدد، يقسم المجموع على 100 فيحصل على العدد المتوسط لحبات القمح في السنبلة في العينة. النتيجة تستخدم لتقدير متوسط عدد حبات القمح من النوع المحلي.

يمكن أن نعبر عن المجتمع المدروس (هو هنا عدد حبات سنبلة محمد البشير) **بمتغيرة عشوائية X** لها متوسط (توقع رياضي) μ وتباين σ^2، وغيرهما من **معالم المجتمع**.

في المقابل كل قيمة محصل عليها في العينة x_i هي تحقق (réalisation) لمتغيرة ع نرمز لها ب X_i. نستنتج أن **العينة يمكن تعريفها رياضيا بتوليفة من المتغيرات أو بمتغيرة متعددة الأبعاد**[1]: **$\{X_1, \dots, X_n\}$**.

من جهة أخرى، كل قيمة محصل عليها في العينة هي أيضا تحقق لمتغيرة المجتمع X، لذلك فإن **التوقع الرياضي لكل متغيرة X_i هو نفسه التوقع الرياضي لمتغيرة المجتمع X، وتباينها هو تباين المجتمع:**

[1] في حالة كون العينة من مفردتين تمثل بثنائية أو ما يسمى أيضا متغيرة ذات بعدين (أنظر فصل المتغيرة ثنائية البعد).

$$V(X_i) = \sigma^2 \qquad , \qquad E(X_i) = \mu$$

يجب التمييز بين خصائص المجتمع، خصائص كل متغيرة من متغيرات العينة وخصائص العينة (نرمز لخصائص العينة بأحرف لاتينية – M للمتوسط المرجح، S للانحراف المعياري وS^2 للتباين – لتمييزها عن خصائص المجتمع μ، σ وσ^2) وتكتب كما يلي:

$$M = \frac{1}{n}\sum X_i \qquad , \qquad S^2 = \frac{1}{n}\sum (X_i - \mu)^2$$

لاحظ. خصائص العينة هي دوال في متغيرات العينة.

مثال2. ترغب هيأة معينة بالبحوث السياسية في تقدير نسبة الناخبين المساندين لمرشح معين في 10 الولايات، فتقوم باستجواب 100 ناخب من كل ولاية. الناخبون في الولايات العشر يمثلون المجتمع بينما ال 1000 ناخب المستجوبون يمثلون العينة.

مثال3. من أجل معرفة مدى دقة صنع قطعة نقدية ترمى القطعة 100 مرة ونحسب عدد مرات الحصول على الصورة والكتابة، حجم العينة هنا هو 100.

مثال4. لتقدير نسبة الكرات داخل صندوق، التي من لون معين، نقوم عدد من المرات بسحب كرة نسجل لونها ثم نعيدها. عدد الكرات المسحوبة يمثل حجم العينة.

لاحظ:

– مصطلح المجتمع يقصد به القياسات أو القيم وليس الأفراد أو الأشياء التي تم قياسها (مجتمع الأوزان، مجتمع آراء الناخبين..).
– المجتمع قد لا يكون محدودا (نتائج رميات قطعة النقد). نرمز لحجم المجتمع ب N، ولحجم العينة ب n.

7-1-2 العينة العشوائية – العينة العشوائية البسيطة – العينة النفادية

عند القيام بالمعاينة نبحث عادة عن عينة ممثله للمجتمع حتى تعكس خصائصه في حدود الإمكانيات المالية والوقت المتاح. العينة العشوائية (Echantillon aléatoire) – وتسمى أيضا العينة الاحتمالية – هي أحد الطرق لتكوين عينة و هي التي يكون لكل مفردة في المجتمع احتمالا معلوما وأكبر من الصفر لكي تظهر في العينة. سنتطرق في دراستنا لنوع محدد من العينات العشوائية هو **العينة العشوائية البسيطة** وهي التي تكون احتمالات جميع المفردات للظهور في العينة **متساوية** وأكبر من الصفر[1]. من طرق تكوين عينة عشوائية بسيطة طريقة القرعة (لاختيار عدد من الطلبة عشوائيا نكتب كل أسماء الطلبة في قصاصات، ثم نسحب منها بالصدفة عدد من القصاصات بمقدار حجم العينة) وطريقة الأعداد العشوائية (نرقم مفردات المجتمع ثم نحدد العينة من خلال مجموعة من الأعداد تؤخذ من الجداول الإحصائية للأعداد العشوائية[2] أو عن طريق برنامج بالحاسوب).

في هذا الفصل سنميز بين حالتين: حالة المعاينة النفادية (Echantillonnage exhaustif) وحالة المعاينة غير النفادية. نقصد بالعينة غير النفادية (non exhaustif) العينة التي يكون فيها السحب من المجتمع **بالإرجاع**، لأنها لا تؤدي إلى نفاد مفردات المجتمع مع تكرار السحب لأنه **يمكن أن تظهر مفردة من المجتمع أكثر من مرة في العينة**[3]

[1] في الملحق نتعرض لأنواع مختلفة من العينات.
[2] أنظر جدول الأعداد العشوائية.
[3] هناك فرضيتان تتكرران في عدد من العلاقات الرياضية التي سنراها لاحقا، هما فرضية أن متغيرات العينة Xi مستقلة والمجتمع لانهائي، يتحقق شرط الاستقلال إذا كانت المعاينة غير نفادية، وفي هذه الحالة يمكن اعتبار المجتمع لانهائي.

وفيها **تكون متغيرات العينة X_i مستقلة ولها نفس التوزيع**. على العكس نسمي المعاينة بدون إرجاع معاينة نفادية حيث لا يمكن أن تظهر مفردة في عينة أكثر من مرة و في هذه الحالة لا تكون نتائج السحب (متغيرات العينة) مستقلة[1] مبدئيا.

3-1-7 معالم المجتمع

معالم المجتمع (Paramètres d'une population) هي خصائصه مثل المتوسط μ، التباين σ^2، نسبة خاصية ما فيه p، ... كما نعد من خصائص المجتمع أيضا طبيعة توزيعه الاحتمالي f(x) كأن يكون طبيعيا أو غيره.

4-1-7 إحصائية المعاينة

لتقدير معالم المجتمع (متوسط المجتمع μ، تباين المجتمع σ^2 النسبة p ...) ننطلق من بيانات العينة، حيث نحتاج إلى حساب معالم مثل متوسط العينة m ، تباين العينة S^2، النسبة في العينة[2] p'. بصفة عامة، نسمي كل قيمة تحسب انطلاقا من بيانات العينة من أجل تقدير قيمة معالم المجتمع إحصائية المعاينة . رياضيا، إحصائية المعاينة (Statistique de l'échantillonnage) هي كل دالة في المتغيرات العشوائية التي تمثل القيم المحصل عليها في العينة.

2-7 توزيع المعاينة للمتوسطات

متوسط توزيع المعاينة للمتوسطات
تباين توزيع المعاينة للمتوسطات
طبيعة توزيع المعاينة للمتوسطات

1-2-7 متوسط توزيع المعاينة للمتوسطات

مسألة: ليكن المجتمع 1، 3، 5، 6، 8. ما هي القيمة المتوقعة لمتوسط عينة (نرمز لمتوسط العينة ب M أو $\overline{X}$ ونرمز للقيمة التي تأخذها هذه المتغيرة بحرف صغير) مسحوبة بالإرجاع، بدون إرجاع مكونة من مفردتين؟

1- في حالة السحب بالإرجاع: نحتاج إلى تحديد جميع الحالات الممكنة للمتوسط m_i حسب كل عينة والتي عددها: $5^2 = 25$

المشاهدة 1 \ المشاهدة 2	1	3	5	6	8
1	1	2	3	3.5	4.5
3	2	3	4	4.5	5.5
5	3	4	4.5	5.5	6.5
6	3.5	4.5	5.5	6.5	7
8	4.5	5.5	6.5	7	8

التوزيع الاحتمالي ل M :

m_i	1	2	3	3.5	4	4.5	5.5	6.5	7	8
p_i	1/25	2/25	3/25	2/25	2/25	5/25	4/25	3/25	2/25	1/25

[1] مبدئيا على الأقل لأنه عندما تكون العينة تمثل نسبة صغيرة من مجتمع كبير يتضاءل تأثير نتيجة سحب على نتائج السحوبات الموالية ومن ثم يمكن اعتبار متغيرات العينة مستقلة.

[2] نرمز لخصائص المجتمع بأحرف يونانية، ما عدا النسبة، ونرمز لخصائص العينة بأحرف لاتينية.

$$E(M) = \sum m_i\,(p_i) = 4.6$$

نلاحظ أن التوقع الرياضي ل M مساوي ل μ متوسط المجتمع: $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

2- في حالة السحب بدون إرجاع.

المشاهدة 2 / المشاهدة 1	1	3	5	6	8
1	–	2	3	3.5	4.5
3	2	–	4	4.5	5.5
5	3	4	–	5.5	6.5
6	3.5	4.5	5.5	–	7
8	4.5	5.5	6.5	7	–

التوزيع الاحتمالي ل M

m_i	2	3	3.5	4	4.5	5.5	6.5	7
p_i	2/20	2/20	2/20	2/20	4/20	4/20	2/20	2/20

$$E(M) = \sum m_i \,.\, p_i = 4.6$$

نلاحظ مرة أخرى أن التوقع الرياضي لمتوسط العينة مساوي لمتوسط المجتمع : $\mu = 4.6$

نظرية 1. إذا كانت X م ع تمثل مجتمع ما وM متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع (بالإرجاع أو بدون إرجاع)، فإن $E(M) = \mu_m = \mu$

البرهان:

$$E(M) = E\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n}\sum_i E(Xi) = \frac{1}{n}\sum_i \mu = \frac{1}{n}n\mu = \mu.$$ (متغيرات العينة X_i)

لاحظ. تنطبق هذه النتيجة على حالتي السحب بالإرجاع أو بدونه، وتعني أننا نتوقع إذا سحبنا عينة من مجتمع أن يكون متوسط العينة مساويا لمتوسط المجتمع، لذلك يستخدم متوسط العينة لتقدير متوسط المجتمع إذا كان هذا الأخير مجهولا، فنكتب:

$$\hat{\mu} = M$$

و نقول إن الإحصائية M هي مقدرة لمعلمة المجتمع μ.

الفرق بين القيمة الحقيقية لمعلمة المجتمع والقيمة التقديرية يسمى **خطأ المعاينة** ويتم قياسه بتباين **المقدر**، في هذه الحالة σ^2_m.

2-2-7 تباين توزيع المعاينة للمتوسطات أو خطأ المعاينة

1-2-2-7 حالة المعاينة بالإرجاع

مثال. أحسب تباين المجتمع في المسألة 1، أحسب التباين (والانحراف المعياري) لتوزيع المعاينة للمتوسطات σ^2_m علما أن العينة مسحوبة بالإرجاع، قارن بين تباين المجتمع وتباين متوسطات العينات الممكنة (توزيع المعاينة للمتوسطات).

من الجدول 1 ، الذي يبين العينات الممكنة ومتوسطاتها، نحسب تباين هذه المتوسطات.

$\sigma^2_m = \sum_i (m_i - \mu_m)^2 p_i = 2.92;$

$\sigma^2 = [\sum_i (x_i - \mu)^2]/5 = 5.84$

نلاحظ أن تباين M أقل من تباين المجتمع.

هذا المثال يمهد للنظرية التالية عن خطأ المعاينة في حالة السحب بالإرجاع:

نظرية 2. إذا كانت X م ع تمثل مجتمع ما وM متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع بالإرجاع، فإن تباين M (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي: $\sigma_m^2 = \frac{\sigma^2}{n}$ حيث n حجم العينة.

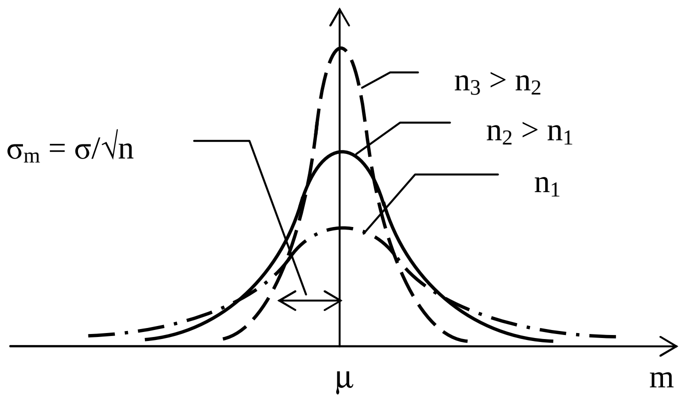
البرهان:

$$\sigma^2{}_m = V(M) = V\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n^2}V\left(\sum_i X_i\right)$$

بما أن السحب بالإرجاع فإن X_i مستقلة ولها نفس التوزيع إذا لها نفس التباين ومنه:

$$\sigma^2{}_m = \frac{1}{n^2}\sum_i V(X_i) = \frac{1}{n^2}nV(X_i) \Rightarrow \sigma^2{}_m = \frac{\sigma^2}{n}$$

لاحظ أن خطأ المعاينة يتأثر طرديا بتباين المجتمع وعكسيا بحجم العينة، وهو أمر متوقع: كلما كانت العينة أكبر والمجتمع أكثر تجانسا (أقل تشتتا) كان التقدير أدق ما يعني خطأ معاينة أقل (أنظر الرسم أدناه).



رسم 7 – 1 خطأ المعاينة يتناقص بزيادة حجم العينة

2-2-2-7 **حالة المعاينة بدون إرجاع.**

في المسألة 1 أحسب تباين المتوسطات الممكنة للعينة $\sigma^2{}_m$ في حالة المعاينة بدون إرجاع، قارن بين تباين المجتمع وتباين المتوسطات الممكنة للعينة.

استنادا إلى العينات الممكنة ومتوسطاتها (الجدول 2) نحسب تباين M : $\sigma^2{}_m = \sum_i (m_i - \mu_m)^2 p_i = 2.19$

تباين المجتمع:

$$\sigma^2 = [\sum_i (x_i - \mu)^2]/N = E(X^2) - E(X)^2 = (1 + 9 + 25 + 36 + 64) / 5 - 4.6^2 = 4.85$$

لاحظ مرة أخرى أن تباين M أقل من تباين متغيرة المجتمع X. لاحظ أيضا أن تباين M أقل منه في حالة السحب بالإرجاع $2.92 > 2.19$؛ مما يعني أن السحب بدون إرجاع يعطي خطأ معاينة أقل. هذا يمهد للنظرية التالية عن خطأ المعاينة في حالة المعاينة بدون إرجاع:

نظرية 3. إذا كانت X م ع تمثل مجتمع ما حجمه N وM متغيرة ع تمثل متوسط عينة حجمها n مسحوبة من ذات المجتمع بدون إرجاع، فإن تباين M (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي:

$$\sigma_m^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

البرهان:

$$\sigma^2{}_m = V\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n^2}V\left(\sum_i X_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} v(x_i) + \sum_{i\neq j=1}^{n}\mathrm{cov}(x_i, x_j)\right]$$

لكن التباين المشترك في حالة متغيرات عينة عشوائية نفادية يساوي $-\sigma^2/(N-1)$ وتوجد A^2_n ترتيبة ممكنة لهذه المتغيرات مثنا مثنا. إذن:

$$\sigma^2{}_m = \frac{1}{n^2}\left[n\sigma^2 + A_n^2\left(\frac{-\sigma^2}{N-1}\right)\right] = \frac{1}{n^2}\left[n\sigma^2 + \frac{n(n-1)(-\sigma^2)}{N-1}\right]$$

$$= \frac{n\sigma^2}{n^2}\left[1 + \frac{-(n-1)}{N-1}\right] = \frac{\sigma^2}{n}\left[\frac{N-1-n+1}{N-1}\right] = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$$

لاحظ:

- يتضاءل تأثير النسبة $\frac{N-n}{N-1}$ (وتسمى معامل الإرجاع) ومن ثم يمكن إهماله إذا كانت العينة صغيرة جدا بالمقارنة مع حجم المجتمع $(n/N < 0.05)$[1] لأنه يقترب من 1.
- قيمة معامل الإرجاع أقل من 1 كلما كان $n>1$، هذا يعني أن خطأ المعاينة في حالة المعاينة بدون إرجاع أقل منه في حالة المعاينة بالإرجاع ما يعني أن المعاينة بدون إرجاع تعطي تقديرا أكثر دقة لمعلمة المجتمع μ. يزيد هذا الفرق بين دقة التقدير في الحالتين كلما زاد حجم العينة n.

3-2-7 طبيعة توزيع الإحصائية *M*

بعد دراسة القيمة المتوقعة ل M وخطأ المعاينة، من المهم دراسة طبيعة توزيع هذه الإحصائية لاستخدام ذلك في تقدير متوسط المجتمع μ بمجال حيث نحتاج إلى تحديد احتمال الصواب والخطأ في هذا التقدير مما يقتضي تحديدي طبيعة توزيع الإحصائية. ندرس طبيعة توزيع متوسط توزيع المعاينة للمتوسطات من خلال النظريتين التاليتين:

نظرية 4. إذا كان المجتمع موزع طبيعيا بمتوسط μ وتباين σ^2 فإن متوسط العينة المسحوبة منه عشوائيا يتبع أيضا التوزيع الطبيعي بمتوسط μ وتباين σ^2/n، ونكتب $M \sim N(\mu, \sigma/\sqrt{n})$ أو $z = \frac{m-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

لاحظ

- في حالة المجتمع محدود والمعاينة نفادية نستبدل العبارة $\sigma/\sqrt{n}$ ب $\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
- هذه النظرية محققة في الحالتين المعاينة بالإرجاع وبدونه، وأيا كان حجم العينة.

[1] تسمي النسبة n/N معدل الاستقصاء، و يرمز لها ب f.

في الحقيقة يمكن استخدام التوزيع الطبيعي حتى إذا كان المجتمع ليس بالضرورة طبيعيا (مجهول التوزيع) بشرط أن لا تكون العينة صغيرة وذلك استنادا إلى نظرية النهاية المركزية[1]:

نظرية 5. إذا كان المجتمع ذا متوسط μ وتباين σ^2 لكن ليس بالضرورة طبيعيا فإن متوسط العينة المسحوبة منه M **يؤول** إلى التوزيع الطبيعي إذا كانت العينة كبيرة ($n \geq 30$)، ونكتب $M \approx N(\mu, \sigma_m)$.

تطبيق: مجتمع حجمه 900 بمتوسط $\mu = 20$ و$\sigma = 12$. نستخرج كل العينات الممكنة ذات حجم n.

أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة: (1) حجم العينة $n = 36$، (2) $n = 64$

في حالة ($n = 36$) أحسب احتمال أن يكون M محصورا بين 18 و22،

أحسب نفس الاحتمال في حالة $n = 64$.

$E(M) = \mu = 20$

(1) $n = 36: \quad n/N = 36/900 = 0.04 < 0.05 \Rightarrow \sigma_m = \sigma/\sqrt{n} = 12/\sqrt{36} = 2$

(2) $n = 64: N = 900 \Rightarrow \dfrac{n}{N} = \dfrac{64}{900} = 0.071 > 0.05$

$$\Rightarrow \sigma_m = \frac{12}{\sqrt{64}}\sqrt{\frac{900-64}{900-1}} = 1.92$$

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}} = \frac{18-20}{12/\sqrt{36}} = -1, \quad Z_2 = 1 \Rightarrow P(18 < m < 22) = P(Z_1 < Z < Z_2) = 0.6827$$

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}\left(\frac{N-n}{N-1}\right)} = \frac{18-20}{1.92} = -1.04, \quad Z_2 = 1.04 \Rightarrow$$

$$P(18 < m < 22) = P(-1.04 < Z < 1.04) = 0.70$$

4-2-7 خلاصة

الجدول التالي يبين أهم خصائص توزيع المعينة للمتوسطات.

جدول 7 - 1. أهم القواعد عن توزيع المعاينة للمتوسطات

المجتمع	المعاينة	الخاصية
مجتمع ما بمتوسط μ وتباين σ^2	سحب بالإرجاع	$\mu_m = \mu, \quad \sigma_m^2 = \dfrac{\sigma^2}{n}$
مجتمع حجمه N بمتوسط μ وتباين σ^2	سحب بدون إرجاع	$\mu_m = \mu, \quad \sigma_m^2 = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$
مجتمع موزع طبيعيا بمتوسط μ وتباين σ^2	سحب بالإرجاع أو بدون إرجاع	$M \sim N(\mu, \sigma_m)$
مجتمع بمتوسط μ وتباين σ^2 لكن ليس بالضرورة طبيعيا	عندما يكون n كبيرا ($n \geq 30$)	$M \approx N(\mu, \sigma_m)$

[1] أنظر نظرية النهاية المركزية في الفصل السابق.

3-7 توزيع المعاينة للنسبة

التوقع الرياضي وتباين توزيع المعاينة للنسبة
طبيعة توزيع المعاينة للنسبة

نقصد بالنسبة في المجتمع الكسر: $p = \frac{N_a}{N}$ ، حيث **N** حجم المجتمع و **N_a** عدد المفردات التي تتحقق فيها صفة ما.

ونقصد بالنسبة في العينة الكسر: $p' = \frac{n_a}{n}$ حيث **n** حجم العينة و **n_a** عدد مفردات العينة التي تحقق نفس الصفة.

ندرس خصائص الإحصائية **p’** و توزيعها الاحتمالي لأنها إحصائية تستخدم لتقدير النسبة في المجتمع **p**.

مثال. سحبنا عينة حجمها 31 من نقاط الطلبة في امتحان الإحصاء الرياضي فكانت كالتالي:

3، 18.5، 10.5، 19.5، 11.5، 14، 11، 16.5، 18، 7.5، 3.5، 2.5، 2.5، 8.5، 10.5، 8.5، 2، 5، 12، 1.5، 5.5، 8.5، 6، 15، 6، 6، 10، 2.5، 1، 2.5، 3.

أحسب **p’** : نسبة الطلبة الحاصلين على المعدل في العينة.

لدينا 12 نقطة أكبر أو تساوي العشرة ($n_a = 12$) إذن $p' = 12/31 = 0.387$

1-3-7 التوقع الرياضي والتباين للإحصائية p'

في حالة المعاينة غير النفادية (سحب بالإرجاع) فإن احتمال تحقيق الصفة ثابت بالنسبة لمفردات العينة وهو ذاته النسبة في المجتمع **p** ، هذا يعني أن عدد مفردات العينة التي تحقق الصفة يتبع التوزيع الثنائي.

$$n_a \sim B(n, p); \quad E(n_a) = np; \quad V(n_a) = npq$$

يمكن إذن استنتاج الخصائص العددية للإحصائية **p’** من خصائص **n_a** :

$$\mu_{p'} = E\left(\frac{n_a}{n}\right) = \frac{1}{n} E(n_a) = \frac{np}{n} \Rightarrow \mu_{p'} = p$$

$$\sigma^2_{p'} = V\left(\frac{n_a}{n}\right) = \frac{1}{n^2} V(n_a) = \frac{1}{n^2} npq \Rightarrow \sigma^2_{p'} = \frac{pq}{n} \Rightarrow \sigma_{p'} = \sqrt{\frac{pq}{n}}$$

لذلك يصلح أن نستخدم **p’** لتقدير **p** و نكتب: $\widehat{p} = p'$

في حالة المعاينة نفادية والمجتمع محدود فإن **n_a** تتبع التوزيع الهندسي الزائد ونكتب $n_a \sim H(N, n, p)$

في هذه الحالة فإن:

$$E(n_a) = np; \quad V(n_a) = npq\frac{N-n}{N-1}$$

ومنه فإن التوقع الرياضي ل **p’** والتباين يكونان :

$$\mu_{p'} = p; \qquad \sigma^2_{p'} = V\left(\frac{n_a}{n}\right) = \frac{1}{n^2} npq\frac{N-n}{N-1} = \frac{pq}{n}\left(\frac{N-n}{N-1}\right) \Rightarrow \sigma_{p'} = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

يمكن إهمال معامل الإرجاع إذا كانت $n < 0.05N$ لأن قيمة المعامل تؤول إلى 1.

مثال. انطلاقا من بيانات عينة المثال السابق، قدر نسبة الطلبة الحاصلين على العلامة عشرة أو أكثر من مجموع الطلبة وعددهم 300. أحسب خطأ المعاينة للمقدر المستخدم.

لدينا $E(p') = p$ إذن القيمة التي نقدر بها p : نسبة الحاصلين على العشرة في مجموع الطلبة هي النسبة في العينة أي 0.387. ونكتب : $\widehat{p} = p' = 0.387$

لحساب خطأ المعاينة نحتاج على استبدال p ب 'p لأن هذه الأخيرة تبقى مجهولة.

$$\sigma^2_{p'} = \frac{pq}{n}\left(\frac{N-n}{N-1}\right) \cong \frac{p'q'}{n}\left(\frac{N-n}{N-1}\right) = \frac{0.387(1-0.387)}{31}\left(\frac{300-31}{300-1}\right) = 0.0069$$

2-3-7 توزيع المعاينة ل p'

نعلم من نظرية موافر (الفصل السابق) أنه في حالة **n** كبيرة بما فيه الكفاية فإن التوزيع الطبيعي يعطي نتائج مقاربة للتوزيع الثنائي ل$\mathbf{n_a}$ و في هذه الحالة فإن المتغيرة المرتبطة بها خطيا **'p** تؤول هي الأخرى إلى التوزيع الطبيعي.

نظرية 6: إذا كانت X م ع تمثل مجتمعا ما أين p نسبة المفردات ذات صفة معينة، و 'p متغيرة ع تمثل نسبة المفردات ذات نفس الصفة في عينة عشوائية مسحوبة من ذات المجتمع، فإن:

$p' \approx N(p, \sigma_{p'})$: $n \geq 30$ عندما

ملاحظات:

- استخدام توزيع مستمر (التوزيع الطبيعي) لحساب احتمال متعلق بمتغيرة متقطعة يقتضي القيام بتصحيح معين. بالنسبة ل **X** التصحيح يكون بزيادة أو طرح 0.5 لأن كل قيمة $\mathbf{n_a}$ تعتبر مجالا حدوده $\mathbf{n_a - 0.5}$ و$\mathbf{n_a + 0.5}$. بالنسبة ل **'p** – متغيرة متقطعة لأنها مجرد تابع لمتغيرة متقطعة $\mathbf{n_a}$ – التصحيح يكون بإضافة أو طرح **(1/(2n** ويسمى هذا **معامل التصحيح**.
- إضافة معامل التصحيح تكون في حالتي حساب $\mathbf{P(p' < \ldots)}$ أو $\mathbf{P(p' \leq \ldots)}$ و الطرح يكون في حالتي حساب: $\mathbf{P(p' < \ldots)}$ أو $\mathbf{P(p' \geq \ldots)}$.
- يؤول معامل التصحيح إلى الصفر – ومن ثم يمكن إهماله – في حالة **n** كبيرة.

مثال. نلقي قطعة نقدية 20 مرة. ليكن X عدد مرات الحصول على صورة.

- أحسب $P(X = 8)$ ثم أدرس إمكانية استخدام قانون موافر– لابلاس لحساب نفس الاحتمال.

$X \sim B(20, 0.5) \Rightarrow P(X = 8) = C^8_{20}\,(0.5)^8\,(0.5)^{12} = 0.1201$,

لدينا $np = 10 > 5$ وكذلك $nq = 10 > 5$ ، وإذا شئنا استخدام القاعدة الثانية فإننا نجد أيضا أن:

$n = 20$، $np = 10$، $nq = 10$، يمكن إذا اعتبار $\frac{x-\mu}{\sigma_x} = \frac{x-np}{\sqrt{npq}} = \frac{x-10}{\sqrt{20(0.5)(0.5)}} \approx N(0, 1)$

ومن ثم استخدام التوزيع الطبيعي المعياري لحساب احتمال المجال المعبر عن القيمة 8 وهو [7.5, 8.5]

$$P(7.5 \leq X^* \leq 8.5) = P\left(\frac{7.5-10}{2.24} \leq Z \leq \frac{8.5-10}{2.24}\right) = P(-1.12 \leq Z \leq -6.67) = 0.12$$

نسمي عملية استبدال القيمتين 7.5 و8.5 ب القيمة 8 عملية تصحيح.

مثال2. أحسب احتمال أن تكون نسبة مرات الحصول على الصورة عند إلقاء قطعة نقدية أقل تماما من 50% في 30 رمية.

$P(p' < 0.5) = P(X/30 < 0.5)$. $X \approx N(\mu, \sigma) \Rightarrow p' = X/30 \approx (p, \sigma_{p'})$

$\sigma_{p'} = \sqrt{pq/n} = 0.091 \Rightarrow p' \approx (0.5, 0.091)$

$P(p' < 0.5) = P(p' \leq 0.5 - \underline{1/(2n)}) = P(p' \leq 0.48)$

لاحظ: نستخدم معامل التصحيح لأن $P(p' < 0.5)$ هي في الأصل: $P(X < 15)$ و لأن X متغيرة متقطعة (عدد الصور) وجب استبدال بالاحتمال السابق الاحتمال: $P(X \leq 14.5)$ ما يقابل: $P(p' \leq 0.5 - 1/(2n))$.

$P(p' \leq 0.48) = P(Z \leq (0.48 - 0.5)/0.091) = P(Z \leq 0.18) = 0.571$

أحسب احتمال أن تكون نسبة مرات الحصول على الصورة أقل أو تساوي 50 % في 30 رمية.

$P(p' \leq 0.5) = P(p' \leq 0.5 + \mathbf{1/(2n)}) = P(p' \leq 0.517)$

أحسب احتمال أن تكون نسبة مرات الحصول على الصورة أكبر أو تساوي 50 % في 30 رمية.

$P(p' \geq 0.5) = P(p' \geq 0.5 - \underline{1/(2n)}) = P(p' \geq 0.48) \cong 1 - P(p' \leq 0.48)$

أحسب احتمال أن تكون نسبة مرات الحصول على الصورة أكبر تماما من 50 % في 30 رمية.

$P(p' > 0.5) = P(p' > 0.5 + \mathbf{1/(2n)}) = P(p' > 0.517) \cong 1 - P(p' \leq 0.517)$

4-7 توزيع المعاينة للفروق والمجاميع

متوسط وتباين توزيع المعاينة للفروق والمجاميع

طبيعة توزيع المعاينة للفروق والمجاميع

1-4-7 المتوسط والتباين

نظرية 7. ليكن لدينا مجتمعان نسحب من كل مجتمع عينة عشوائية، نحسب في العينة المسحوبة من المجتمع الأولى إحصائية ما S_1 (متوسط العينة، التباين، ...) ونحسب نفس الإحصائية في عينة المجتمع الثاني ونسميها S_2.

$$\mu_{S1 \pm S2} = \mu_{S1} \pm \mu_{S2} \; ; \qquad \sigma^2_{S1 \pm S2} = \sigma^2_{S1} + \sigma^2_{S2}$$

مثال 1. إذا كانت الإحصائية هي **المتوسط** فإن:

$$\mu_{m1-m2} = \mu_{m1} - \mu_{m2} = \mu_1 - \mu_2 \qquad \sigma^2_{m1-m2} = \sigma^2_{m1} + \sigma^2_{m2} = \sigma^2/n_1 + \sigma^2/n_2$$

مثال 2. إذا كانت الاحصائية هي **النسبة** فإن:

$$\mu_{p1-p2} = \mu_{p1} - \mu_{p2} = p_1 - p_2 \qquad \sigma^2_{p1-p2} = \sigma^2_{p1} + \sigma^2_{p2} = p_1q_1/n_1 + p_2q_2/n_2$$

مثال: ليكن المجتمع U_1 : 3، 7، 8. والمجتمع U_2 : 2، 4. تحقق من أن:

$$\mu_{U1-U2} = \mu_{U1} - \mu_{U2} \; ; \qquad \sigma^2_{U1-U2} = \sigma^2_{U1} + \sigma^2_{U2} .$$

$\mu_{U1} = (3 + 7 + 8)/3 = 6 \; ; \; \mu_{U2} = (2 + 4)/2 = 3 \Rightarrow$

$\mu_{U1} - \mu_{U2} = 6 - 3 = 3$

$\mu_{U1-U2} = (1 + 5 + 6 - 1 + 3 + 4)/6 = 3$

$\sigma^2_{U1} = (3^2 + 7^2 + 8^2)/3 - 6^2 = 14/3 \; ;$

$\sigma^2_{U2} = (2^2 + 4^2)/2 - 3^2 = 1 \Rightarrow \sigma^2_{U1} + \sigma^2_{U2} = 17/3$

$\sigma^2_{U1-U2} = (1^2 + 5^2 + 6^2 + 1^2 + 3^2 + 4^2) / 6 - 3^2$

$= (1 + 25 + 36 + 1 + 9 + 16) / 6 - 9 = 17/3$

		U_1		
	$U_1 - U_2$	3	7	8
U_2	2	1	5	6
	4	-1	3	4

2-4-7 طبيعة توزيع المعاينة للفرق بين متوسطين

نظرية 8. في حالة $n_1 \geq 30$ وn_2 ، يقترب توزيع المتغيرة المعيارية للفرق بين متوسطين من التوزيع الطبيعي المعياري، ونكتب:

$$\mu_{m1-m2} \approx N(0, 1)$$