المحاضرة الخامسة: توزيعات المعاينة (يتبع)

3-4. توزيع المعاينة لنسبة $\frac{S_1^2}{\sigma_1^2}$ إلى $\frac{S_2^2}{\sigma_2^2}$: توزيع F

من أجل المقارنة بين تباين مجتمعين فإننا نحتاج لنسبة بين تباين عينتين مأخوذتين من هذين المجتمعين. وسنتطرق إلى توزيع هذه النسبة في حالة المعاينة من مجتمعين طبيعيين مستقلين.

نظرية(09): إذا كانت S_1^2 ، S_2^2 هما تبايني عينتين مستقلتين حجمهما n_1 و n_2 مسحوبتين من مجتمعين لهما توزيعين طبيعيين ذو التباينين σ_1^2، σ_2^2 على الترتيب. فإن المتغير:

$$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (30)$$

يكون له توزيع **فيشر** (F) بدرجات حرية (υ_1,υ_2) أي (n_1-1,n_2-1) .

وما يجب الإشارة إليه أيضا هو أن توزيع F هو دالة في درجات الحرية، حيث يكون لتوزيع F نوعين من درجات الحرية:

- درجات حرية مقترنة بتباين العينة S_1^2 في **البسط**، ويُرمز لها بـ:

$$\upsilon_1=n_1-1$$

- درجات حرية مقترنة بتباين العينة S_2^2 في **المقام**، ويُرمز لها بـ:

$$\upsilon_2=n_2-1$$

أي أن توزيع F يتحدد تماما بدلالة درجات الحرية، ولا يتوقف على أي معالم أخرى. حيث يتمركز حول القيمة واحد، ويرجع ذلك إلى أن تباين المجتمعين يتم تقديرهما بتبايني العينتين، ومنه فمن المتوقع أن يكون كل من: S_1^2/σ_1^2 ، S_2^2/σ_2^2 قريبا من القيمة واحد، لذلك فإن النسبة:

$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ تقترب أيضا من الواحد الصحيح.

توزيع F هو توزيع ملتوي إلى اليمين ومداه نظريا يكون من الصفر إلى ما لا نهاية، أي أن قيم المتغير F لا تكون سالبة، كذلك نجد أن توزيع F يقترب من التوزيع الطبيعي كلما زادت درجتا الحرية.

إن توزيع F مثل توزيع Z و توزيع T و توزيع χ^2 ، نجده متواجدا في معظم البرامج الإحصائية الجاهزة، ولكي نستخدمه يجب أن نحدد أولا الاحتمال المرغوب فيه (محدد بالمساحة المظلة التي تقع على يسار القيمة الجزئية المطلوبة). بعد ذلك نبحث عن:

- درجات الحرية الخاصة بالبسط وهي: ($\upsilon_1 = n_1 - 1$)
- درجات الحرية الخاصة بالمقـــام وهي: ($\upsilon_2 = n_2 - 1$)

ومنه قيمة F التيهقتج من تقاطع درجتي حرية البسط والمقام هي القيمة الجزئية المطلوبة، حيث أن المساحة التي تقع على يسار تلك القيمة تمثل الاحتمال المطلوب.

مثال(11): إذا كان: $n_1 = 16$، $n_2 = 20$، ونرغب في إيجاد إحتمال أن يأخذ المتغير F قيمة لا تزيد عن: 1/ 0.36 ، 2/ 2.23

الحل:

لدينا:

$$\upsilon_1 = n_1 - 1 = 16 - 1 = 15\ ,\ \upsilon_2 = n_2 - 1 = 20 - 1 = 19$$

1- باستخدام جدول توزيع F يكون:

$$P\left(F_{(15,19)} \leq 0.36\right) = 0.025$$

وبالمقابل يكون:

$$P\left(F_{(15,19)} > 0.36\right) = 0.975$$

2 - بنفس الطريقة نجد:

$$P\left(F_{(15,19)} \leq 2.23\right) = 0.95$$

وبالمقابل يكون:

$$P\left(F_{(15,19)} > 2.23\right) = 0.05$$

5-3. توزيع المعاينة لنسبة العينة

يحتاج الباحث في أغلب الدراسات لمعرفة نسبة ظاهرة معينة في المجتمع محل الدراسة،كنسبة المدخنين في ولاية بسكرة، نسبة الذكور في جامعة محمد خيضر ببسكرة، نسبة الوحدات التالفة في إنتاج مصنع معين، نسبة الأيام التي تزيد فيها الحرارة عن 40 درجة مئوية خلال فصل الصيف في منطقة معينة،...،الخ، ففي كل حالة من هذه الحالات نجد أن المجتمع محل الدراسة منقسم إلى قسمين؛

قسم تتوافر فيه الظاهرة محل الدراسة (الخاصية المدروسة)، والقسم الثاني لا تتوافر فيه هذه الظاهرة. ومجتمعات من هذا النوع يكون فيها المتغير وصفيا أي نوعيا لا نستطيع قياسه كميا، وبالتالي تعاد صياغته وتحويله إلى متغير عشوائي نرمز له بالرمز x، ونتعامل في هذا النوع من المجتمعات مع نسبة الظاهرة محل الدراسة في المجتمع، ويرمز لها بالرمز **P** ، ويطلق عليها **نسبة المجتمع** ، وتحسب بالعلاقة التالية:

$$P = \text{(عدد مفردات المجتمع التي تتحقق فيها الظاهرة المدروسة)} / \text{(العدد الكلي لمفردات المجتمع)}$$

وبالتالي فإن P تمثل إحتمال ظهور هذه الظاهرة في المجتمع، ويُرمز لإحتمال عدم ظهور هذه الظاهرة في المجتمع بالرمز q، حيث أن حدث ظهور الظاهرة، وحدث عدم ظهورها، هما حدثان مكملان لبعضهما البعض، إذن:

$$q = 1 - P$$

تعتبرنسبة P من أهم معالم المجتمع التي يرغب الباحث في معرفتها لكي يستطيع وصف المجتمع محل الدراسة وصفا جيدا، ولكن في الكثير من الأحيان لا نستطيع تحديد نسبة المجتمع لعدم توافر بيانات عن كل مفردات المجتمع، ولذلك نقوم بالاستدلال عليها، أي استنتاجها باستخدام نسبة الظاهرة محل الدراسة في العينة العشوائية المسحوبة من هذا المجتمع ، ويرمز **لنسبة العينة** بالرمز $\bar{p}$ وتحسب بالعلاقة التالية:

$$\bar{P} = \text{(عدد مفردات العينة التي تتوفر فيها الظاهرة المدروسة)} / \text{(العدد الكلي لمفردات العينة)}$$

نسبة العينة ($\bar{p}$)، كأي إحصائية تتغير قيمتها من عينة لأخرى، وبالتالي فهي متغير عشوائي له توزيع احتمالي يطلق عليه **توزيع المعاينة لنسبة العينة**.

وسنجد أنه توجد علاقة بين الوسط الحسابي وتباين توزيع المعاينة لنسبة العينة واللذين نرمز لهما على التوالي، بالرمز $\bar{p}$ والرمز-$\sigma_{\bar{p}}^2$، وبين نسبة المجتمع، فنجد أن:

$$\mu_{\bar{p}} = P \text{.............} (31)$$

وعندما يكون المجتمع لا نهائيا أو مجتمعا غير محدودا أو كانت عملية السحب تتم مع الإرجاع (أي $n \leq 0.05N$) فإن:

$$\sigma_{\bar{p}}^2 = \frac{\mathbf{Pq}}{\mathbf{n}} \text{..........} (32)$$

أما إذا كان المجتمع محدود أو عملية السحب تتم دون إعادة (أي $n>0.05N$) فإن:

$$\sigma_{\bar{p}}^2 = \frac{\mathbf{Pq}}{\mathbf{n}} \left(\frac{\mathbf{N-n}}{\mathbf{N-1}}\right) \text{............} (33)$$

لقد عرفنا العلاقة بين الوسط الحسابي وتباين توزيع المعاينة لنسبة العينة، وبين نسبة المجتمع P ، ولكن لاستنتاج قيمة المعلمة P باستخدام نسبة العينة $\bar{p}$ ، يجب معرفة طبيعة توزيع المعاينة لنسبة $\bar{p}$ ، وبما أن التغير الذي يحصل في قيمة P سببه تغير عدد مفردات العينة التي تتوافر فيها الظاهرة محل الدراسة من عينة إلى أخرى فقط، لأن كل العينات التي نستنتج منها توزيع المعاينة حجمها ثابت ويساوي n .

وبما أن توزيع عدد المفردات التي تتوافر فيها الظاهرة محل الدراسة في العينة (عدد المحاولات الناجحة في العينة) تتبع توزيع ذي الحدين (Binomial Distribution) بمعلمتين n و P في حالة سحب مفردات العينة مع الإرجاع.

ونعلم أنه وفقا **لنظرية النهاية المركزية**، نستطيع استخدام التوزيع الطبيعي كتقريب جيد لتوزيع ذات الحدين عندما يكون حجم العينة كبيرا.

وبالتالي عندما تكون n كبيرة، نستطيع استخدام التوزيع الطبيعي كتقريب جيد لتوزيع المعاينة لعدد مفردات العينة التي تتوافر فيها الظاهرة محل الدراسة، والذي هو نفسه توزيع المعاينةلنسبة العينة، وذلك كما هو واضح في النظرية التالية:

نظرية (10): وفقا لنظرية النهاية المركزية، توزيع المعاينةلنسبة $\bar{p}$ يقترب من التوزيع الطبيعي عندما يكون حجم العينة كبيرا بدرجة كافية، ويتحقق ذلك عندما يكون كل من np و nq على الأقل 5؛

أي إذا كان : $np \geq 5$, $nq \geq 5$ ، فإن المتغير العشوائي Z ، حيث:

$$Z = \frac{\bar{p} - \mu_{\bar{p}}}{\sqrt{\sigma^2_{\bar{P}}}} \text{.......... } (34)$$

سيتبع توزيعا قريبا من التوزيع الطبيعي المعياري.

مثال (12): إذا علمت أن نسبة الأسر التي تقيم في شقق في ولاية ما 58.27 %، فإذا سحبنا عينة عشوائية من هذه الولاية تشمل 40 أسرة.

فما هو احتمال أن تكون نسبة الأسر التي تقيم في شقق في هذه العينة تتراوح بين 55 % و 70 % ؟

الحل:

البيانات المتوافرة لدينا هي:

نسبة المجتمع: P = 58.27 %   حجم العينة: n = 40

والاحتمال المطلوب هو:

$$\boldsymbol{P}(0.55 \leq \bar{p} \leq 0.70) = ?$$

بما أن:

$$np = 40\,(0.5827) = 23.31 \quad , \quad nq = 40\,(0.4173) = 16.69$$

أي أن كلا من np و nq أكبر من 5، وبالتالي فإن توزيع المعاينةلنسبة $\bar{p}$ سيكون قريبا من التوزيع الطبيعي بمتوسط وتباين قدرهما على التوالي كما يلي:

$$\mu_{\bar{p}} = P = 0.5827$$

$$\sigma^2_{\bar{P}} = \frac{\mathbf{Pq}}{\mathbf{n}} = [(0.5827)(1-0.5827)] / 40 = 0.0061$$

نستطيع التعبير عن الاحتمال المطلوب كما يلي:

$$P(0.55 \leq \bar{p} \leq 0.70) = P(z_1 \leq Z \leq z_2)$$

حيث:

$$z_1 = \frac{\bar{p}_1 - P}{\sqrt{\sigma^2_{\bar{P}}}} = \frac{0.55 - 0.5827}{\sqrt{0.0061}} = -0.4$$

$$z_2 = \frac{\bar{p}_2 - P}{\sqrt{\sigma^2_{\bar{P}}}} = \frac{0.70 - 0.5827}{\sqrt{0.0061}} = 1.5$$

إذن:

$$P(0.55 \leq \bar{p} \leq 0.70) = P(-0.42 \leq Z \leq 1.50)$$

$$= 0.1628 + 0.4332 = 0.5960$$

3-6. توزيع المعاينة للفرق بين نسبتي عينتين $(\bar{p}_1 - \bar{p}_2)$

إذا كانت دراستنا خاصة بمقارنة نسبة ظاهرة معينة في مجتمعين مختلفين، أي محاولة معرفة الفرق بين نسبتين $(P_1 - P_2)$، حيث P_1 ترمز لنسبة الظاهرة في المجتمع الأول، وP_2 ترمز لنسبة الظاهرة في المجتمع الثاني، وعند عدم توافر بيانات عن مفردات كل من المجتمعين، نقوم بالاستدلال على المعلمة $(P_1 - P_2)$ أي استنتاجها باستخدام الفرق بين نسبتي العينتين العشوائيتين المسحوبتين من هذين المجتمعين، أي $(\bar{p}_1 - \bar{p}_2)$، حيث $\bar{p}_1$ هي نسبة الظاهرة في العينة العشوائية المسحوبة من المجتمع الأول، و$\bar{p}_2$ هي نسبة الظاهرة في العينة العشوائية المسحوبة من المجتمع الثاني، ولذلك يجب دراسة توزيع المعاينة لهذه الإحصائية، والذي يطلق عليه "**توزيع المعاينة للفرق بين نسبتي عينتين**".

فإذا سحبنا من المجتمع الأول كل العينات العشوائية ذات الحجم 1 ، وحسبنا نسبة الظاهرة المدروسة $\bar{p}_1$ لكل عينة، وسحبنا من المجتمع الثاني كل العينات العشوائية ذات الحجم n_2 ، وحسبنا نسبة الظاهرة المدروسة $\bar{p}_2$ لكل عينة، وكانت العينات المسحوبة من المجتمع الأول مستقلة عن العينات المسحوبة من المجتمع الثاني، وحسبنا كل **الفروق** بين نسب عينات المجتمع الأول ونسب عينات المجتمع الثاني، أي حسبنا كل قيم $(\bar{p}_1 - \bar{p}_2)$، فسنحصل على توزيع المعاينة للفرق بين

نسبتي عينتين $(\bar{p}_1 - \bar{p}_2)$ ، وإذا حسبنا الوسط الحسابي $\mu_{\bar{p}_1-\bar{p}_2}$ ، والتباين $\sigma^2_{\bar{p}_1-\bar{p}_2}$ لهذا التوزيع، فنجد أن هناك علاقات تربط هذين المقياسين مع نسبة المجتمع الأول ونسبة المجتمع الثاني، وذلك كما يلي:

$$\mu_{\bar{p}_1-\bar{p}_2} = \mathbf{P_1 - P_2} \quad \text{......................} \quad (35)$$

فإذا كان المجتمع غير محدود أو كان السحب مع الإرجاع، و 1 / n_1 ، N_2 / n_2 كليهما أقل من أو يساوي 0.05 فإن:

$$\sigma^2_{\bar{p}_1-\bar{p}_2} = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2} \quad \text{..................} \quad (36)$$

أما إذا كان المجتمع محدود أو كان السحب دون إرجاع، وn_1/N_1 ، n_2/N_2 كليهما أو احدهما أكبر من 0.05 ، فإن:

$$\sigma^2_{\bar{p}_1-\bar{p}_2} = \frac{P_1 q_1}{n_1}\left(\frac{N_1-n_1}{N_1-1}\right) + \frac{P_2 q_2}{n_2}\left(\frac{N_2-n_2}{N_2-1}\right) \quad \text{.........} \quad (37)$$

ومن هنا نصل إلى النظرية التالية:

نظرية (11): إذا كان لدينا عينتان مستقلتان كبيرتا الحجم تم سحبهما من مجتمعين، ووفقا لنظرية النهاية المركزية، يكون توزيع المعاينة للفرق بين نسبتي العينتين $(\bar{p}_1 - \bar{p}_2)$ توزيعا قريبا من التوزيع الطبيعي بوسط حسابي وتباين تم توضيحهما في العلاقتين (35) و (36) على الترتيب. ومن ثم فإن المتغير العشوائي Z حيث:

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} \quad \text{...............} (38)$$

سيتبع توزيعا قريبا من التوزيع الطبيعي المعياري.

مثال (13): عن الكتاب الإحصائي الصادر عن الديوان الوطني للإحصاء، نجد أن العدد الكلي للسكان الذين أعمارهم تتراوح بين 10 سنوات و30سنة، موزعين حسب الجنس كما يلي: 2157136 ذكرا منهم 221914 يحملون شهادة جامعية، و 2067508 أنثى منهن 144423 يحملن شهادة جامعية، فإذا سحبنا من هذين المجتمعين عينتين عشوائيتين مستقلتين، الأولى من الذكور حجمها 2000 ذكر، والثانية من الإناث حجمها 1500 أنثى.

المطلوب:

أوجد احتمال أن يكون الفرق بين نسبتي العينتين أكبر أو يساوي 5 %.

الحل:

بافتراض أن المجتمع الأول يمثل مجتمع الذكور، والمجتمع الثاني يمثل مجتمع الإناث، نجد أن:

P_1: تمثل نسبة حاملي شهادة جامعية في المجتمع الأول $= \frac{221914}{2157136} = 0.10$

P_2: تمثل نسبة حاملي شهادة جامعية في المجتمع الثاني $= \frac{144423}{2067508} = 0.07$

n_1 : حجم العينة الأولى $= 2000$ ذكر.

n_2 : حجم العينة الثانية $= 1500$ أنثى.

والاحتمال المطلوب هو : $\boldsymbol{P}\,[(\bar{p}_1 - \bar{p}_2) \geq 0.05] = ?$

بما أن n_1 و n_2 كبيرتان، فإن توزيع المعاينة للإحصائية $(\bar{p}_1 - \bar{p}_2)$ سيكون قريبا من التوزيع الطبيعي، وبالتالي فإن الإحتمال المطلوب يتم حسابه كما يلي:

$$\boldsymbol{P}\,[(\bar{p}_1 - \bar{p}_2) \geq 0.05\,] = P(Z \geq z)$$

حيث أن:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{(0.05) - (0.10 - 0.07)}{\sqrt{\frac{0.10\text{x}0.90}{2000} + \frac{0.07\text{x}0.93}{1500}}}$$

$$= 2.13$$

إذن:

$$P\,[(\bar{p}_1 - \bar{p}_2) \geq 0.05\,] = P(Z \geq 2.13) = 0.5 - 0.4834 = 0.0166$$