المحاضرة الخامسة:

توزيع الفرق بين متوسطي عينتين $(\overline{X}_1 - \overline{X}_2)$

وتوزيع النسبة للعينة $(\hat{p})$

أولا: توزيع الفرق بين متوسطي عينتين $\overline{X}_1$ و $\overline{X}_2$

إذا كان $X_1 \sim N(\mu_1,\sigma_1^2)$ وسحبت من مجتمعه عينة حجمها n_1 و $X_2 \sim N(\mu_2,\sigma_2^2)$ وسحبت من مجتمعه عينة حجمها n_2 وكان المجتمعين مستقلين وعرفنا المتغيرين العشوائيين $\overline{X}_1$ و $\overline{X}_2$ فإن: $\overline{X}_1-\overline{X}_2$ يمثل متغيراً عشوائياً للفرق بين متوسطي العينتين، ومنه نحصل على الوسط الحسابي والتباين له كالتالي:

$$\mu_{\overline{X}_1-\overline{X}_2}=\mu_{\overline{X}_1}-\mu_{\overline{X}_2}=\mu_1-\mu_2$$

$$\sigma^2_{\overline{X}_1-\overline{X}_2}=\sigma^2_{\overline{X}_1}+\sigma^2_{\overline{X}_2}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

وسنتحدث الآن عن حالات مختلفة لإيجاد توزيع الفرق بين وسطي العينتين.

1- توزيع الفرق بين متوسطي عينتين مستقلتين:

وهنا نفرق بين ثلاث حالات:

1-1- عند معلومية σ_1^2 و σ_2^2:

$$\overline{X}_1-\overline{X}_2 \sim N(\mu_1-\mu_2\,,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$

وبالتالي:

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0\,,1)$$

مثال: إذا كان $X_1 \sim N(30,25)$ وسحبت من مجتمعه عينة حجمها 30 مشـــاهدة و $X_2 \sim N(20,16)$، وسُحبت من مجتمعه عينة حجمها 35 مشاهدة.

المطلوب: أوجد توزيع الفرق بين متوسطي العينتين؟، ثم أحسب الإحتمال $P(\overline{x}_1-\overline{x}_2<12)$؟

الحل:

بالتطبيق المباشر للعلاقة: $\overline{X}_1-\overline{X}_2 \sim N(\mu_1-\mu_2\,,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$

نحصل على: $\overline{X}_1-\overline{X}_2 \sim N(30-20\,,\frac{25}{30}+\frac{16}{35})=N(10\,,1.29)$

وبالتالي فإن: $Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{1.29}}=\frac{(\overline{X}_1-\overline{X}_2)-10}{\sqrt{1.29}}$

فيكون الإحتمال المطلوب هو:

$$P(\overline{x}_1-\overline{x}_2<12)=P\left(\frac{(\overline{X}_1-\overline{X}_2)-10}{\sqrt{1.29}}<\frac{12-10}{\sqrt{1.29}}\right)$$

$$= P\left(Z < \frac{12 - 10}{\sqrt{1.29}} \right)$$
$$= P(Z < 1.76) = 0.9608$$

2-1- عند مجهولية σ_1^2 و σ_2^2: هناك حالتين:

1-2-1- إذا كان على الأقل حجم أحد العينتين $(n_1, n_2 \geq 30)$:

$$\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2 \,, \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})$$

<u>**مثال**</u>: مصنع ينتج 700 كغم من المكرونة كمعدل يومي، سحبت منه عينة عشوائية تمثل انتاج 40يوماً فبلغ وسطها الحسابي 740 كغم بانحراف معياري 40 كغم، و مصنع آخر ينتج 500 كغم كمعدل يومي، سحبت منه عينة عشوائية تمثل انتاج 35 يوماً فبلغ وسطها الحسابي480 كغم بانحراف معياري20 كغم.

<u>**المطلوب**</u>: ما هو التوزيع الإحتمالي للفرق بين متوسطي العينتين؟؛ ثم أحسب الإحتمال التالي:

$$P(180 \leq \overline{X}_1 - \overline{X}_2 \leq 210)$$

<u>**الحل**</u>: لدينا من المسألة $S_1^2 = 1600$ و $S_2^2 = 400$ وحيث أن $n_1 = 40$ ، $n_1 = 35$ وكليهما أكبر من 30 إذن يكون توزيع الفرق بين وسطي العينتين هو:

$$\overline{X}_1 - \overline{X}_2 \sim N(700 - 500 \,, \frac{1600}{40} + \frac{400}{35})$$

وبالتالي:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - 200}{\sqrt{51.43}} \sim N(0\,,1)$$

لذلك يكون الإحتمال المطلوب هو:

$$P(180 \leq \overline{X}_1 - \overline{X}_2 \leq 210) = P\left(\frac{180 - 200}{\sqrt{51.43}} \leq \frac{(\overline{X}_1 - \overline{X}_2) - 200}{\sqrt{51.43}} \leq \frac{210 - 200}{\sqrt{51.43}} \right)$$
$$= P(-2.79 \leq Z \leq 1.39) = P(Z \leq 1.39) - P(Z \leq -2.79)$$
$$= 0.9177 - 0.0029 = 0.9148$$

<u>**ملاحظة**</u>: التوزيع أعلاه يصلح أيضاً عند السحب من مجتمعين غير طبيعيين أو مجهولي التوزيع بشرط كبر حجم العينة، وذلك استناداً لنظرية النهاية المركزية.

1-2-2- إذا كان على الأقل حجم أحد العينتين صغير $(n_1, n_2 < 30)$:

$$\overline{X}_1 - \overline{X}_2 \sim t(n_1 + n_2 - 2)$$

و نميز هنا حالتين:

أ-إذا كان $\sigma_1^2 = \sigma_1^2 = \sigma^2$

لاستنباط العلاقة التي سنستخدمها في هذه الحالة سنحسب أولاً تباين توزيع الفرق بين المتوسطين في هذه الحالـــة كمايلي:

من نظرية النهاية المركزية نعلم أن:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

ومنه:

$$\frac{S_1^2(n_1-1)}{\sigma^2} \sim \chi^2_{(n_1-1)} \quad \text{و كذلك:} \quad \frac{S_2^2(n_2-1)}{\sigma^2} \sim \chi^2_{(n_2-1)}$$

$$\Rightarrow \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2}$$

ومنه نحصل على:

$$T = \frac{\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\left(\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2}\right) \Big/ (n_1 + n_2 - 2)}}$$

وبالتبسيط وبوضع:

$$S_P^2 = (n_1-1)S_1^2 + (n_2-1)S_2^2 \Big/ (n_1 + n_2 - 2)$$

نحصل على:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

وبوضع:

$$S^2_{\overline{X}_1 - \overline{X}_2} = S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

نحصل في الأخير على الصيغة المبسطة التالية:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S^2_{\overline{X}_1 - \overline{X}_2}}} \sim t_{(n_1+n_2-2)}$$

مثال: في المثال السابق بفرض أن حجم العينة الأولى 20 يوما بانحراف معياري 30 كغم، وحجم العينة الثانية 18 يوماً بانحراف معياري 25 كغم وبفرض تجانس تباين المجتمعين.

المطلوب: أحسب: $P(\overline{X}_1 - \overline{X}_2 \geq 215)$؟

الحل:

حيث أن: $n_1 = 20$ ،$n_2 = 18$ وكليهما أصغر من 30 إذن الفرق بين وسطي العينتين هو:

$$\overline{X}_1 - \overline{X}_2 \sim t(20+18-2)$$

$$\overline{X}_1 - \overline{X}_2 \sim t(36)$$

ويكون تباين توزيع الفرق بين وسطي العينتين هو:

$$S^2_P = \frac{(20-1)30^2 + (18-1)25^2}{20+18-2} = 770.14$$

$$S^2_{\overline{X}_1 - \overline{X}_2} = 770.14\left(\frac{1}{20} + \frac{1}{18}\right) = 81.29$$

وبالتالي:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - 200}{\sqrt{81.29}} \sim t(36)$$

$$P(\overline{X}_1 - \overline{X}_2 > 215) = P\left(T > \frac{215-200}{\sqrt{81.29}}\right) = P(T > 1.6637) = 0.05$$

ثانيا: توزيع النسبة للعينة:

في المجتمعات الإحصائية التي تتبع توزيع ذي الحدين يتم رصد نسبة لخاصية معينة ندرسها في المجتمع الإحصائي ونرمز لها بالرمز (P)، مثل نسبة المدخنين ونسبة الصالح من إنتاج مصنع معين...إلخ، حيث يتم حساب النسبة بقسمة مجموع مفردات الخاصية (X) على حجم المجتمع (N) أي أن $P=\frac{X}{N}$، فلو أخذنا جميع العينات التي حجمها (n) من مجتمع حجمه قد تختلف النسب لهذه العينات، مما يعني أن سلوك هذه النسب هو سلوك متغير عشوائي، وبالتالي هذا المتغير العشوائي له توزيع احتمالي مستمداً من توزيع المجتمع الذي سحبت منه العينات، وسنرمز لنسبة الظاهرة في العينة بالرمز($\hat{p}$) و بعدد مفردات الخاصية فيها (x) فتكون نسبة الخاصية في العينة ($\hat{p}=\frac{x}{n}$).

إن هذه النسبة تمثل متغير عشوائي له توزيع احتمالي يسمى توزيع النسبة حيث يكون وسطه الحسابي:

$$\mu_{\hat{p}}=E(\hat{p})=E\left(\frac{X}{n}\right)=\frac{1}{n}E(X)=P$$

ويكون تباينه هو:

$$\sigma^2_{\hat{p}}=V\left(\frac{X}{n}\right)=\frac{1}{n^2}V(X)=\frac{1}{n^2}\times np(1-p)=\frac{P(1-P)}{n}$$

في حالة العينات الكبيرة ($n\geq 30$) "موضوع دراستنا" يقرب التوزيع الطبيعي وذلك استناداً لنظرية النهائية المركزية، ونعبر عن ذلك بالصيغة:

$$Z=\frac{\hat{P}-\mu_{\hat{p}}}{\sigma_{\hat{p}}}\sim N(0\,,1)$$

مثال: مصنع ينتج عادة 25% عبوات كبيرة الحجم، سحبت من إنتاجه عينة حجمها 2200 عبوة تبين أن منها 500 عبوة كبيرة الحجم.

المطلوب: أوجد توزيع النسبة للعبوات الكبيرة ثم احسب احتمال أن المصنع ينتج أقل من 26% من العبوات الكبيرة في فترة إجراء البحث.

الحل:

لاحظ أن المجتمع هنا هو مجتمع ذي الحدين بنسبة نجاح ($P=0.25$)، ولاحظ أن حجم العينة كبير وبحساب وسط التوزيع وتباينه نجد أن: $\mu_{\hat{p}}=P=0.25$

$$\sigma^2_{\hat{p}}=\frac{P(1-P)}{n}=\frac{0.25\times 0.75}{500}=0.0004$$

$$\hat{P}\sim N(0.25\,,\,0.0004)$$

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - 0.25}{0.02} \sim N(0\,,1)$$

$$P(p < 0.26) = P\left(Z < \frac{0.26 - 0.25}{0.02} \right) = P(Z < 0.5) = 0.6915$$

ملاحظة: في حالة مجهولية P نعوض بدلاً منها $\hat{P}$.

بفرض أن $X_1 \sim b(N_1\ ,\ P_1)$ وسحبت من مجتمعه عينة حجمها $n_1 \geq 30$ و $X_2 \sim b(N_2\ ,\ P_2)$، وسُحبت من مجتمعه عينة حجمها $n_2 \geq 30$، وكان المجتمعين مستقلين فإن: $\hat{p}_1 - \hat{p}_2$ يمثل متغيراً عشوائياً للفرق بين نسبتي العينتين، ومنه نحصل على الوسط الحسابي والتباين له، وذلك كمايلي:

$$\mu_{\hat{p}_1 - \hat{p}_2} = \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = P_1 - p_2$$

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}$$

ويكون توزيع فرق النسب:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}} \sim N(0\ ,\ 1)$$

مثال: إذا علمت أن نسبة الذكور في مؤسسة A تبلغ 0.3 وفي مؤسسة B تبلغ 0.2، فإذا سحبت عينتين عشوائياً الأولى من المؤسسة A بحجم 100 والثانية من المؤسسة B بحجم 200.

المطلوب: ما هو احتمال أن يكون الفرق بين نسبتي الذكور في العينتين أكبر من 6%؟.

الحل:

لاحظ أن التوزيع هو توزيع فرق بين نسبتين بوسط حسابي وتباين:

$$\mu_{\hat{p}_1 - \hat{p}_2} = P_1 - p_2 = 0.3 - 0.2 = 0.1$$

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{0.3 \times 0.7}{100} + \frac{0.2 \times 0.8}{200} = 0.0029$$

ويكون توزيع فرق النسب:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0.1}{0.054} \sim N(0\,,1)$$

فيكون الإحتمال المطلوب هو:

$$\begin{aligned} P(\hat{p}_1 - \hat{p}_2 > 0.06) &= P\left(Z > \frac{0.06 - 0.1}{0.054} \right) \\ &= P(Z > -0.74) = 1 - P(Z < -0.74) \\ &= 1 - 0.2296 = 0.7704 \end{aligned}$$