# 《統計學概要》

一、某家日本電器公司在臺灣設廠，臺灣廠的員工薪資均是以新臺幣支付。總公司為了瞭解臺灣廠員工薪資狀況，做了一些統計分析，考慮下列的統計量及分析

(一)薪資之變異係數。

(二)薪資與年齡之相關係數。

(三)為了檢定平均薪資$\mu$係數是否為新臺幣50000元，針對檢定$H_0$: $\mu$=50000，計算$t$ 檢定統計量。

(四)為了建立員工年齡與薪資之簡單直線迴歸模型，以年齡為自變數，以薪資為應變數，計算迴歸線斜率估計量。

(五)為了瞭解員工薪資是否受教育程度之影響，以員工教育程度為因子作單因子變異數分析，計算$F$檢定統計量。

前一陣子日圓匯率降至低點，引發日本廠員工抱怨。總公司欲比較臺灣廠和日本廠之員工薪資狀況，故將臺灣廠員工薪資統一轉為以日圓計算。試問上述題(一)至題(五)統計量是否會受計算貨幣為新臺幣或日圓不同之影響？請分別就上述題(一)至題(五)之統計量，說明影響是變大、變小、或不變。（題(一)至題(五)每小題4分，共20分）

| | |
|---|---|
| **試題評析** | 本題涉及資料單位改變對統計量之影響。 |
| **考點命中** | 高點高上《迴歸分析熱門題庫》，趙治勳著，頁2-26第二篇；第二章第七節一。 |

**答**：

為了說明,假設新臺幣對日圓匯率為$a$，以目前現實狀況$a>1$

令$X$ 表以新臺幣計算之員工薪資(元)

$X'=aX$ 表以日圓計算之員工薪資(元)

$Y$ 表員工年齡(歲)

員工教育程度分為高,中,低三個水準

(一)不變

$$\overline{X}=\frac{\sum X_i}{n}, S_X^2=\frac{\sum(X_i-\overline{X})^2}{n-1} \text{ 故 } C.V._X=\frac{S_X}{\overline{X}}$$

$$\overline{X}'=\frac{\sum X'_i}{n}=\frac{\sum aX_i}{n}=a\frac{\sum X_i}{n}=a\overline{X}$$

$$S_{X'}^2=\frac{\sum(X'_i-\overline{X}')^2}{n-1}=\frac{\sum(aX_i-a\overline{X})^2}{n-1}=a^2\frac{\sum(X_i-\overline{X})^2}{n-1}=a^2S_X^2$$

$$C.V._{X'}=\frac{S_{X'}}{\overline{X}'}=\frac{aS_X}{a\overline{X}}=\frac{S_X}{\overline{X}}=C.V._X$$

(二)不變

$$r_{XY}=\frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}}=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(X_i-\overline{X})^2}\sqrt{\sum(Y_i-\overline{Y})^2}}$$

$$r_{X'Y}=\frac{SS_{X'Y}}{\sqrt{SS_{X'}}\sqrt{SS_Y}}=\frac{\sum(X'_i-\overline{X}')(Y_i-\overline{Y})}{\sqrt{\sum(X'_i-\overline{X}')^2}\sqrt{\sum(Y_i-\overline{Y})^2}}=\frac{\sum(aX_i-a\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(aX_i-a\overline{X})^2}\sqrt{\sum(Y_i-\overline{Y})^2}}$$

$$=\frac{a\sum(X_i-\overline{X})(Y_i-\overline{Y})}{a\sqrt{\sum(X_i-\overline{X})^2}\sqrt{\sum(Y_i-\overline{Y})^2}}=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(X_i-\overline{X})^2}\sqrt{\sum(Y_i-\overline{Y})^2}}=r_{XY}$$

(三)不變

$H_0:\mu=50000$ 之檢定統計量 $T_X=\dfrac{\overline{X}-50000}{S_X/\sqrt{n}}$

$H_0:\mu'=a50000$ 之檢定統計量 $T_{X'}=\dfrac{\overline{X}'-a50000}{S_{X'}/\sqrt{n}}=\dfrac{a\overline{X}-a50000}{aS_X/\sqrt{n}}=\dfrac{\overline{X}-50000}{S_X/\sqrt{n}}=T_X$

(四)當 $\hat{\beta}_1>0$ 下，$\hat{\beta}'_1$ 變大；當 $\hat{\beta}_1<0$ 下，$\hat{\beta}'_1$ 變小

模型 $X_i=\beta_0+\beta_1 Y_i+\varepsilon_i$，可得 $\hat{\beta}_1=\dfrac{SS_{YX}}{SS_Y}=\dfrac{\sum(Y_i-\overline{Y})(X_i-\overline{X})}{\sum(Y_i-\overline{Y})^2}$

模型 $X'_i=\beta'_0+\beta'_1 Y_i+\varepsilon'_i$，可得 $\hat{\beta}'_1=\dfrac{SS_{YX'}}{SS_Y}=\dfrac{\sum(Y_i-\overline{Y})(X'_i-\overline{X}')}{\sum(Y_i-\overline{Y})^2}=\dfrac{\sum(Y_i-\overline{Y})(aX_i-a\overline{X})}{\sum(Y_i-\overline{Y})^2}$

$$=a\frac{\sum(Y_i-\overline{Y})(X_i-\overline{X})}{\sum(Y_i-\overline{Y})^2}=a\hat{\beta}_1$$

(五)不變

新臺幣計算下一因子 ANOVA 之檢定統計量為

$$F=\frac{SSR/2}{SSE/n-3}=\frac{\sum n_i(X_{i\bullet}-\overline{\overline{X}})^2/2}{\sum(n_i-1)S_{Xi}^2/n-3}$$

日圓計算下一因子 ANOVA 之檢定統計量為

$$F'=\frac{SSR'/2}{SSE'/n-3}=\frac{\sum n_i(X'_{i\bullet}-\overline{\overline{X}}')^2/2}{\sum(n_i-1)S_{X'i}^2/n-3}=\frac{\sum n_i(aX_{i\bullet}-a\overline{\overline{X}})^2/2}{\sum(n_i-1)a^2S_{Xi}^2/n-3}$$

$$=\frac{a^2\sum n_i(X_{i\bullet}-\overline{\overline{X}})^2/2}{a^2\sum(n_i-1)S_{Xi}^2/n-3}=\frac{\sum n_i(X_{i\bullet}-\overline{\overline{X}})^2/2}{\sum(n_i-1)S_{Xi}^2/n-3}=F$$

二、兄弟三人依老大、老二、老三，大小順序由大到小，先後輪流投擲三個銅板，看誰先投出剛好兩個正面誰就獲勝。假設兄弟三人約定一定要分出勝負遊戲才停。（每小題10分，共20分）

(一)試求老大獲勝的機率。

(二)試求老三獲勝的機率。

| | |
|---|---|
| **試題評析** | 本題涉及事件之機率計算。 |
| **考點命中** | 高點高上《統計學講義題型加強課試題》第一回，趙治勳編撰，例18。 |

**答**：

(一)P(老大獲勝) $=\dfrac{3}{8}+(\dfrac{5}{8})^3\dfrac{3}{8}+(\dfrac{5}{8})^6\dfrac{3}{8}+\cdots=\dfrac{\frac{3}{8}}{1-(\frac{5}{8})^3}=0.4961$

(二)P(老三獲勝) $= (\frac{5}{8})^2\frac{3}{8} + (\frac{5}{8})^5\frac{3}{8} + (\frac{5}{8})^8\frac{3}{8} + \cdots = \dfrac{(\frac{5}{8})^2\frac{3}{8}}{1-(\frac{5}{8})^3} = 0.1938$

三、一袋中放入編號1、2、3、4的大小、形狀、重量完全相同的4顆球。若以不歸還方式（取出不放回）由此袋中抽出3顆球為一樣本，令S表所抽3球中最大球號和最小球號的差（大減小），令$T$表所抽3球中最大的2個球號的和。（每小題10分，共20分）

(一)試求$S$ 與$T$ 的相關係數。

(二)試求給定$T=7$之下$S$ 的條件變異數$Var(S|T=7)$。

| 試題評析 | 本題涉及二元間斷型隨機變數之相關係數與條件變異數。 |
|---|---|
| 考點命中 | 高點高上《統計學講義》第一回，趙治勳編撰，頁86例題。 |

**答**：

| 可能樣本 | 機率 | S | T |
|---|---|---|---|
| 1,2,3 | 1/4 | 2 | 5 |
| 1,2,4 | 1/4 | 3 | 6 |
| 1,3,4 | 1/4 | 3 | 7 |
| 2,3,4 | 1/4 | 2 | 7 |

| $f_{ST}(s,t)$ | | T | | | |
|---|---|---|---|---|---|
| | | 5 | 6 | 7 | $f_S(s)$ |
| S | 2 | 1/4 | 0 | 1/4 | 1/2 |
| | 3 | 0 | 1/4 | 1/4 | 1/2 |
| | $f_T(t)$ | 1/4 | 1/4 | 1/2 | 1 |

(一) $E(S) = \sum_s sf_S(s) = 2(\frac{1}{2}) + 3(\frac{1}{2}) = \frac{5}{2}$, $E(S^2) = \sum_s s^2 f_S(s) = 2^2(\frac{1}{2}) + 3^2(\frac{1}{2}) = \frac{13}{2}$

$$V(S) = E(S^2) - [E(S)]^2 = \frac{13}{2} - (\frac{5}{2})^2 = \frac{1}{4}$$

$$E(T) = \sum_t tf_T(t) = 5(\frac{1}{4}) + 6(\frac{1}{4}) + 7(\frac{1}{2}) = \frac{25}{4}$$

$$E(T^2) = \sum_t t^2 f_T(t) = 5^2(\frac{1}{4}) + 6^2(\frac{1}{4}) + 7^2(\frac{1}{2}) = \frac{159}{4}$$

$$V(T) = E(T^2) - [E(T)]^2 = \frac{159}{4} - (\frac{25}{4})^2 = \frac{11}{16}$$

$$E(ST) = \sum_t \sum_s stf_{ST}(s,t) = (2)(5)(\frac{1}{4}) + (2)(7)(\frac{1}{4}) + (3)(6)(\frac{1}{4}) + (3)(7)(\frac{1}{4}) = \frac{63}{4}$$

$$\rho_{ST}=\frac{E(ST)-E(S)E(T)}{\sqrt{V(S)}\sqrt{V(T)}}=\frac{\frac{63}{4}-\frac{5}{2}\frac{25}{4}}{\sqrt{\frac{1}{4}}\sqrt{\frac{11}{16}}}=0.3015$$

(二) $V(S|T=7)=E(S^2|T=7)-[E(S|T=7)]^2=\frac{13}{2}-(\frac{5}{2})^2=\frac{1}{4}$

其中 $E(S|T=7)=\sum_s sf_{S|T}(s|t=7)=\sum_s s\frac{f_{ST}(s,t=7)}{f_T(t=7)}=(2)\frac{1/4}{1/2}+(3)\frac{1/4}{1/2}=\frac{5}{2}$

$$E(S^2|T=7)=\sum_s s^2 f_{S|T}(s|t=7)=\sum_s s^2\frac{f_{ST}(s,t=7)}{f_T(t=7)}$$
$$=(2^2)\frac{1/4}{1/2}+(3^2)\frac{1/4}{1/2}=\frac{13}{2}$$

四、令隨機變數X具有機率分配 $f(x)=\begin{cases}4x^2e^{-2x}, 0<x<\infty \\ 0,其他\end{cases}$，且令 $F(x)$ 為 $X$ 之累積機率分配函數（cumulative distribution function）（每小題10分，共20分）

(一)令隨機變數 $T=F(X)$，試求 $T$ 之機率分配。

(二)令 $F^{-1}(x)$ 為F(X)之反函數，且設 $U$ 為具有連續型均等分配 $U(0,1)$ 之隨機變數。令隨機變數 $Y=F^{-1}(U)$，試求 $Y$ 之機率分配。

| | |
|---|---|
| **試題評析** | 本題涉及機率積分轉換定理。 |
| **考點命中** | 高點高上《統計學講義加強課程》，趙治勳編撰，頁10加強五。 |

**答**：

(一) $F_T(t)=P(T\le t)=P(F_X(X)\le t)=P(X\le F_X^{-1}(t))=F_X(F_X^{-1}(t))=\begin{cases}0, t<0\\ t, 0\le t<1\\ 1, 1\le t\end{cases}$

$$f_T(t)=\frac{dF_T(t)}{dt}=1, 0<t<1,\quad \therefore T\sim Uniform(0,1)$$

(二)由(一)得知 $Y=F^{-1}(U)\sim f_Y(y)=f_X(y)=\begin{cases}4y^2e^{-2y}, 0<y<\infty\\ 0 \quad , o.w.\end{cases}$

五、已知兩母體資料相依且服從常態分配。今蒐集兩母體之成對變數資料如下：

| | | | | | |
|---|---|---|---|---|---|
| 母體Ⅰ($X_i$) | 0 | 4 | 2 | 1 | 3 |
| 母體Ⅱ($Y_i$) | 2 | 4 | 3 | 1 | 5 |

（每小題10分，共20分）

(一)試以顯著水準 $\alpha$=0.05檢定兩變數之母體相關係數是否為零。

(二)試求出 $Y$ 對 $X$ 的迴歸模型 $Y_i=\beta X_i+\varepsilon_i$ 之參數 $\beta$ 的最小平方估計值。

| | |
|---|---|
| **試題評析** | 本題涉及簡迴歸之參數估計與檢定。 |
| **考點命中** | 高點高上《迴歸分析熱門題庫》，趙治勳著，頁2-17,2-24。 |

**答**：

(一)假設模型：$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1,2,3,4,5$

$$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

$$r_{XY} = \frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}} = 0.8$$

$$H_0 : \rho_{XY} = 0 \quad vs \quad H_1 : \rho_{XY} \neq 0$$

T. S.： $\mathrm{T} = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}} \sim t_{(5-2=3)}$

R.R.: Reject $H_0$ at $\alpha = 0.05$ if $|\mathrm{T}^*| > t_{(3)0.025} = 3.1824$

$\because |\mathrm{T}| = \left|\frac{0.8\sqrt{5-2}}{\sqrt{1-0.8^2}}\right| = 2.3094$ $\quad \therefore$ don't reject $H_0$

我們沒有足夠證據去推論兩母體之相關係數不為零。

(二)假設模型：$Y_i = \beta X_i + \varepsilon_i, i = 1,2,3,4,5$

$$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{38}{30} = 1.2667$$