Statisics Formula

1 Descriptive Statistics

相對次數

$$\frac{\text{該組別的次數}}{n}$$

近似組距

$$\frac{\text{最大資料值} - \text{最小資料值}}{\text{組數}}$$

樣本平均數

$$\bar{x} = \frac{\sum x_i}{n}$$

母體平均數

$$\mu = \frac{\sum x_i}{N}$$

加權平均數

$$\bar{x} = \frac{\sum w_i x_i}{\sum x_i}$$

幾何平均數

$$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\dots(x_n)}$$

pth 百分位置的位置

$$L_{\mathrm{p}} = \frac{p}{100}(n+1)$$

四分位距

$$\mathrm{IQR} = Q_3 - Q_1$$

母體變異數

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

樣本變異數

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

標準差

$$\text{樣本標準差} = s = \sqrt{s^2}$$

$$\text{母體標準差} = \sigma = \sqrt{\sigma^2}$$

變異係數

$$\left(\frac{\text{標準差}}{\text{平均數}} \times 100\right)\%$$

z 分數

$$z = \frac{x_i - \bar{x}}{s}$$

樣本共變異數

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

母體共變異數

$$\sigma_{xy} = \frac{\sum(x_i - \mu_x)(y_i - \mu_y)}{N}$$

皮爾森相關係數：樣本資料

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

皮爾森相關係數：母體資料

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

2 Probability

組合的計數法則

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

排列的計數法則

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!}$$

餘集

$$P(A) = 1 - P(A^c)$$

加法律

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

條件機率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

乘法律

$$P(A \cap B) = P(B)P(A|B)$$

獨立事件的乘法律

$$P(A \cap B) = P(A)P(B)$$

貝氏定理

$$P(A_i \cap B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

3 Discrete Probability Distributions

離散型均勻機率函數

$$f(x) = \frac{1}{n}$$

離散型機率分布的期望值

$$E(X) = \mu = \sum xf(x)$$

離散型隨機變數的變異數

$$Var(x) = \sigma^2 = \sum (x-\mu)^2 f(x)$$

隨機變數 x 與 y 的相關係數

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

隨機變數 x 與 y 的期望值

$$E(ax+by) = aE(x) + bE(y)$$

兩隨機變數的線性組合的變異數

$$Var(ax+by) = a^2 Var(x) + b^2 Var(y) + 2ab\sigma_{xy}$$

二項機率函數

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

二項機率分配的期望值

$$E(x) = \mu = np$$

二項機率分配的變異數

$$Var(x) = \sigma^2 = np(1-p)$$

卜瓦松機率分布

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}$$

超幾何機率函數

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

超幾何分配的期望值

$$E(x) = \mu = n(\frac{r}{N})$$

超幾何分配的變異數

$$Var(x) = \sigma^2 = n(\frac{r}{N})(1-\frac{r}{N})(\frac{N-n}{N-1})$$

負二項機率函數

$$f(x) = \binom{x-1}{r-1} p^r q^{x-r}$$

負二項隨機變數的平均數

$$E(x) = \mu = \frac{r}{p}$$

負二項隨機變數的變異數

$$Var(X) = \sigma^2 = \frac{rq}{p^2}$$

4 Continuous Probability Distributions

常態機率密度函數

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

轉換成標準常態隨機變數

$$z = \frac{x-\mu}{\sigma}$$

伽馬函數

$$\Gamma(\alpha) = \int_0^x x^{\alpha-1} e^{-x} \, \mathrm{d}x$$

伽馬機率密度函數

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{x/\beta}$$

指數型機率密度函數

$$f(x) = \frac{1}{\mu} e^{-x/\mu}$$

指數分配：累績機率

$$P(x \leq x_0) = 1 - e^{-x_0/\mu}$$

5 Sampling and Sampling Distribution

$\bar{x}$ 的期望值

$$E(\bar{x}) = \mu$$

$\bar{x}$ 的標準差（標準誤）

- 有限母體

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} (\frac{\sigma}{\sqrt{n}})$$

- 無限母體

$$\sigma_{\bar{x}} (\frac{\sigma}{\sqrt{n}})$$

$\hat{p}$ 的期望值

$$E(\hat{p}) = p$$

$\hat{p}$ 的標準差（標準誤）

- 有限母體

$$\sigma_{\hat{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$$

- 無限母體

$$\sigma_{\hat{p}}\sqrt{\frac{p(1-p)}{n}}$$

6 Interval Estimation

母體平均數的區間估計值：σ 已知

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

母體平均數的區間估計值：σ 未知

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

母體平均數區間估計值所需的樣本數

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2}$$

母體比例的區間估計值

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

母體比例區間估計值所需的樣本數

$$n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$$

7 Hypothesis Tests of Single Populations

母體平均數假設檢定的檢定統計量：σ 已知

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

母體平均數假設檢定的檢定統計量：σ 未知

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

母體比例假設檢定的檢定統計量

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

母體平均數檢定單尾檢定的樣本大小

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2}$$

單一母體變異數的區間估計

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

單一母體變異數假設檢定之檢定統計量

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

當 $\sigma_1^2 = \sigma_2^2$ 時，母體變異數的假設檢定之檢定統計量

$$F = \frac{s_1^2}{s_2^2}$$

8 Hypothesis Tests of Two Populations

兩母體平均數之差的點估計量

$$\bar{x}_1 - \bar{x}_2$$

$\bar{x}_1 - \bar{x}_2$ 的標準誤

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

兩母體平均數之差的區間估計：σ_1 與 σ_2 已知

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$\mu_1 - \mu_2$ 之假設檢定的統計檢定量：σ_1 與 σ_2 已知

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

兩母體平均數之差的區間估計：σ_1 與 σ_2 未知

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

自由度：有兩個獨立隨機樣本之 t 分配

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{1}{n_1-1}(\frac{s_1^2}{n_1})^2 + \frac{1}{n_2-1}(\frac{s_2^2}{n_2})^2}$$

$\mu_1 - \mu_2$ 之假設檢定的統計檢定量：σ_1 與 σ_2 未知

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

配對樣本之假設檢定的檢定統計量

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$$

兩母體之差的點估計量

$$\hat{p}_1 - \hat{p}_2$$

$\hat{p}_1 - \hat{p}_2$ 的標準誤

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

兩母體比例之差的區間估計

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-p_1)}{n_1} + \frac{\hat{p}_2(1-p_2)}{n_2}}$$

當 $p_1 = p_2 = p$ 時之 $\hat{p}_1 - \hat{p}_2$ 的標準誤

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}$$

當 $p_1 = p_2 = p$ 時之 p 的混合估計量

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$p_1 - p_2$ 之假設檢定的檢定統計量

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

9 The Analysis of Variance for Designed Experiments

完全隨機設計

第 j 個處理的樣本平均數

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j}}{n_j}$$

第 j 個處理的樣本變異數

$$s_j^2 = \frac{\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2}{n_j - 1}$$

總樣本平均數

$$\bar{\bar{x}} = \frac{\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n_j} x_{ij}}{n_T}$$

$$n_T = n_1 + n_2 + \cdots + n_k$$

處理間均方

$$\text{MSTR} = \frac{\text{SSTR}}{k-1}$$

處理間平方和

$$\text{SSTR} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2$$

誤差均方

$$\text{MSE} = \frac{\text{SSE}}{n_T - k}$$

誤差平方和

$$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2$$

k 個母體平均數是否相等的檢定統計量

$$F = \frac{\text{MSTR}}{\text{MSE}}$$

總平方和

$$\text{SST} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{\bar{x}})^2$$

總平方和之分割

$$\text{SST} = \text{SSTR} + \text{SSE}$$

多重比較程序

費雪 LSD 程序的檢定統計量

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\text{MSE}(\frac{1}{n_i} + \frac{1}{n_j})}}$$

費雪 LSD

$$\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}(\frac{1}{n_i} + \frac{1}{n_j})}$$

隨機區集設計

總平方和

$$\text{SST} = \sum_{i=1}^{b}\sum_{j=1}^{k}(x_{ij} - \bar{\bar{x}})^2$$

處理間平方和

$$\text{SSTR} = b\sum_{j=1}^{k}(x_{.j} - \bar{\bar{x}})^2$$

區集造成的平方和

$$\text{SSBL} = k\sum_{i=1}^{b}(x_{i.} - \bar{\bar{x}})^2$$

誤差平方和

$$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL}$$

因子實驗

總平方和

$$\text{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(x_{ijk} - \bar{\bar{x}})^2$$

因素 A 之平方和

$$\text{SSA} = br\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{\bar{x}})^2$$

因素 B 之平方和

$$\text{SSB} = ar\sum_{j=1}^{b}(\bar{x}_{.j} - \bar{\bar{x}})^2$$

交互作用之平方和

$$\text{SSAB} = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}})^2$$

誤差造成的平方和

$$\text{SSE} = \text{SST} - \text{SSA} - \text{SSB} - \text{SSAB}$$

10 Simple Linear Regression

簡單線性迴歸模型

$$y = \beta_0 + \beta_1 x + \epsilon$$

簡單線性迴歸方程式

$$E(y) = \beta_0 + \beta_1 x$$

估計簡單線性迴歸方程式

$$\hat{y} = b_0 + b_1 x$$

最小平方法準則

$$\min\sum(y_i - \hat{y}_i)^2$$

估計迴歸方程式的斜率與 y 截距

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

誤差平方和

$$\text{SSE} = \sum(y_i - \hat{y}_i)^2$$

總平方和

$$\text{SST} = \sum(y_i - \bar{y}_i)^2$$

迴歸平方和

$$\text{SST} = \sum(\hat{y}_i - \bar{y}_i)^2$$

SST、SSR 與 SSE 的關係

$$\text{SST} = \text{SSR} + \text{SSE}$$

判定係數

$$r^2 = \frac{\text{SSR}}{\text{SST}}$$

樣本相關係數

$$r_{xy} = (b_1\text{之符號})\sqrt{\text{判定係數}} = (b_1\text{之符號})\sqrt{r^2}$$

誤差均方 (σ 的估計值)

$$s^2 = \text{MSE} = \frac{\text{SSE}}{n-2}$$

估計值的標準誤

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n-2}}$$

b_1 的標準差

$$\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2}}$$

b_1 的估計標準差

$$s_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$$

t 檢定統計量

$$t = \frac{b_1}{s_{b_1}}$$

迴歸均方

$$\text{MSR} = \frac{\text{SSR}}{\text{自變數個數}}$$

F 檢定統計量

$$F = \frac{\text{MSR}}{\text{MSE}}$$

$\hat{y}^*$ 的估計標準差

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

$E(\hat{y}^*)$ 的信賴區間

$$\hat{y}^* \pm t_{\alpha/2s_{\hat{y}^*}}$$

預測 y^* 個別值時的估計標準差

$$s_{\mathrm{p}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

y^* 的預測區間

$$\hat{y}^* \pm t_{\alpha/2} s_{\mathrm{p}}$$

第 i 個觀察值的殘差

$$y_i - \hat{y}_i$$

第 i 個殘差的標準差

$$s_{y_i - \hat{y}_i} = s\sqrt{1 - h_i}$$

第 i 個觀察值的標準化殘差

$$\frac{y_i - \hat{y}_i}{s_{y_i - \hat{y}_i}}$$

第 i 個觀察值的槓桿作用

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}$$

11 Multiple Linear Regression

複迴歸模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

複迴歸方程式

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

估計複迴歸方程式

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

最小平方法準則

$$\min \sum (y_i - \hat{y}_i)^2$$

SST、SSR 與 SSE 的關係

$$\text{SST} = \text{SSR} + \text{SSE}$$

複判定係數

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

調整複判定係數

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

迴歸均方

$$\text{MSR} = \frac{\text{SSR}}{p}$$

誤差均方

$$\text{MSE} = \frac{\text{SSE}}{n-p-1}$$

F 檢定統計量

$$F = \frac{\text{MSR}}{\text{MSE}}$$

t 檢定統計量

$$t = \frac{b_1}{s_{b_1}}$$

第 i 個觀察值的標準化殘差

$$\frac{y_i - \hat{y}_i}{s_{y_i - \hat{y}_i}}$$

第 i 個殘差的標準差

$$s_{y_i - \hat{y}_i} = s\sqrt{1 - h_i}$$

庫克距離度量

$$D_i = \frac{(y_i - \hat{y}_i)^2}{(p-1)s^2}\left[\frac{h_i}{(1-h_i)^2}\right]$$

羅吉斯迴歸方程式

$$E(y) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p}}$$

估計羅吉斯迴歸方程式

$$\hat{y} = P(y = 1 | x_1, x_2, \ldots, x_p)\text{的估計值} = \frac{e^{b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p}}{1 + e^{b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p}}$$

勝算比

$$\text{勝算比} = \frac{\text{odd}s_1}{\text{odd}s_0}$$

logit

$$g(x_1, x_2, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

估計的 logit

$$\hat{g}(x_1, x_2, \ldots, x_p) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

12 Regression Analysis:Model Building

一般線性模型

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_p z_p + \epsilon$$

增加或刪減 $p-q$ 個自變數的 F 檢定

$$F = \frac{\frac{\text{SSE}(x_1,x_2,\ldots,x_q)-\text{SSE}(x_1,x_2,\ldots,x_q,x_{q+1},\ldots,x_p)}{p-q}}{\frac{\text{SSE}(x_1,x_2,\ldots,x_q,x_{q+1},\ldots,x_p)}{n-p-1}}$$

一階自身相關

$$\epsilon_t = \rho\epsilon_{t-1} + z_t$$

杜賓—華生檢定量

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Reference

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2016). Statistics for business & economics. Cengage Learning.

Mendenhall, W. M., Sincich, T. L., & Boudreau, N. S. (2016). Statistics for engineering and the sciences student solutions manual. Chapman and Hall/CRC.