《統計學概要》

試題評析
第一題：有關Poisson分配之計算題，只要注意單位時間之倍數，滿分不難。
第二題：本題是迴歸分析基本計算題，要注意題目缺少線性迴歸模型之基本假設，故需要特別提醒。
第三題：有關母體成功比例 p 之估計，要注意題目缺少隨機樣本之假設，只要小心計算，滿分不難。
第四題：有關兩獨立母體成功比例 $p_1 - p_2$ 之估計。

註：$z_{0.05}=1.645$, $z_{0.025}=1.96$, $z_{0.158}=1.0$, z_α 為右尾百分位數，均需說明理由。

一、若某塑化公司之工安事故次數服從卜瓦松分配（Poisson Distribution），平均每半年發生一次，則：
(一)在未來一年內發生工安事故次數之平均數及變異數別為何？
(二)在未來半年內不會發生工安事故的機率為何？
(三)請概算在未來18年內發生工安事故總次數介於30到42的機率為何？

答：

令 $N(t)$ 表在 t 倍半年內工安事故之次數，$N(t) \sim Poisson(1t)$

(一) 由題意，$N(t=2) \sim Poisson(2)$

$$E[N(t=2)] = V[N(t=2)] = 2$$

(二) 由題意，$N(t=1) \sim Poisson(1)$

$$P(N(t=1)=0) = \frac{e^{-1}(1)^0}{0!} = e^{-1} = 0.3679$$

(三) 由題意，$N(t=36) \sim Poisson(36)$

$$P(30 \le N(t=36) \le 42) \approx P\left(\frac{30-0.5-36}{\sqrt{36}} \le Z \le \frac{42+0.5-36}{\sqrt{36}}\right)$$

$$= P(-1 \le Z \le 1) = 1 - 2(0.158) = 0.684$$

【高分閱讀】
趙治勳老師，統計學上課講義5.5。

二、考慮一個簡單線性迴歸問題：
(一)若判定係數 $r^2=1$，則 SSE（Sum of Square Error）之值為何？
(二)若 $r^2=0.64$，則自變數與因變數之相關係數之值為何？
(三)若迴歸式為 $\hat{Y}=12+1.8x$，$n=32$、$SSR=225$、$SSE=75$、$S_b=0.2683$（斜率之估計量的標準誤），可得假設檢定 H_0：斜率 $=0$ 之 t 值為何？
(四)在(三)中，SSE/σ^2 的分配為何？
(五)在(三)中，H_0：斜率 $=0$，在 $\sigma=0.05$ 下，你如何下結論？

答：

(一) $r^2 = 1 - \frac{SSE}{SST} = 1 \Rightarrow SSE = 0$

(二) 相關係數: $sign(\hat{\beta}_1)\sqrt{r^2} = sign(\hat{\beta}_1)0.8$

(三) $t = \dfrac{b-0}{S_b} = \dfrac{1.8-0}{0.2683} = 6.7089$

(四) 假設滿足線性迴歸模型之基本假設，$\dfrac{SSE}{\sigma^2} \sim \chi^2_{(32-1-1=30)}$

(五) 假設滿足線性迴歸模型之基本假設

$H_0 : \beta_1 = 0 vs H_1 : \beta_1 \neq 0$

T.S.: $T = \dfrac{b-0}{S_b} \sim t_{(30)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $|T^*| > t_{(30)0.025} \approx z_{0.025} = 1.96$

$\because T^* = 6.7089 \qquad \because reject H_0$

我們有足夠證據去推論模型是適當的。

【高分閱讀】
趙治勳老師，《分析熱門題庫》第二篇第二章。

三、考慮一個母體比例 p 的估計問題，假設蒐集200個觀測值，得樣本比例 $\hat{p}=0.4$。則：

（一）$\hat{p}$ 的實際分配為何？

（二）$\hat{p}$ 的平均數及變異數分別為何？

（三）$\hat{p}$ 的近以分配為何？

（四）$\hat{p}$ 的95%信賴區間為何？

答：

母體: $X \sim Ber(p)$

樣本: $X_1, X_2, \cdots, X_{200} \overset{iid}{\sim} Ber(p)$ 假設隨機樣本

(一) $\hat{p} = \dfrac{\sum_{i=1}^{200} X_i}{200}$，其中 $\sum_{i=1}^{200} X_i \sim Bin(200, p)$

(二) $E(\hat{p}) = E(\dfrac{\sum_{i=1}^{200} X_i}{200}) = \dfrac{200p}{200} = p$，$V(\hat{p}) = V(\dfrac{\sum_{i=1}^{200} X_i}{200}) = \dfrac{200p(1-p)}{200^2} = \dfrac{p(1-p)}{200}$

(三) $\hat{p} = \dfrac{\sum X_i}{200} \underset{by C.L.T.}{\sim} N(p, \dfrac{p(1-p)}{200})$

(四) 題目有誤，應該是問 p 之95%信賴區間

樞紐量: $\dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{200}}} \underset{by C.L.T.}{\sim} N(0,1)$

機率區間: $P(-z_{0.025} \leq \dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{200}}} \leq z_{0.025}) = 0.95$

信賴區間: $P(\hat{p}-z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{200}}\le p\le \hat{p}+z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{200}})=0.95$

結論: p 之95%信賴區間

$$(\hat{p}\mp z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{200}})=(0.4\mp 1.96\sqrt{\frac{0.4(1-0.4)}{200}})=(0.3321,0.4679)$$

【高分閱讀】
趙治勳老師，統計學上課講義11.5五。

四、檢定兩個伯努力（Bernoulli）母體的平均數 p_1、p_2 是否相等時，若兩組獨立隨機樣本數分別為 n_1 與 n_2，且兩組樣本之成功總數分別為 X 與 Y，則：

(一)在 $p_1=p_2$ 時，$X\cdot Y$ 的分配為何？

(二)當 p_1、p_2 相等時，如何估計這個相等的平均數？

(三)若想檢定 H_0：$p_1=p_2$ 時，檢定統計量為何？

(四)在 H_0：$p_1=p_2$ 時，檢定統計量的分配何為？

(五)H_1：$p_1\ne p_2$，若檢定統計量的計算值為 -0.2，則P-Value約是何值？

下列4個答案取1個或自行填寫正確表示法：

(A) $P(Z<-0.2)$　(B) $P(t<-0.2)$，其中 t 的自由度為 n_1-n_2-2

(C) $2P(Z>0.2)$　(D) $2P(t<-0.2)$，其中 t 的自由度為 n_1-n_2-2

答：

令X1,X2分別表該兩個伯努力母體

母體: $X_1\sim Ber(p_1)\perp X_2\sim Ber(p_2)$

樣本: $X_{11},X_{12},\cdots,X_{1n_1}\overset{iid}{\sim}Ber(p_1)\perp X_{21},X_{22},\cdots,X_{2n_2}\overset{iid}{\sim}Ber(p_2)$

設 $X=\sum_{i=1}^{n_1}X_{1i}$，$Y=\sum_{i=1}^{n_2}X_{2i}$

(一) 當 $p_1=p_2=p$ 下，$X=\sum_{i=1}^{n_1}X_{1i}\sim Bin(n_1,p)\perp Y=\sum_{i=1}^{n_2}X_{2i}\sim Bin(n_2,p)$

由二項分配加成性可得，$X+Y\sim Bin(n_1+n_2,p)$

(二) 當 $p_1=p_2=p$ 下，$\hat{p}=\frac{X+Y}{n_1+n_2}$

(三) T.S.: $Z=\frac{\hat{p}_1-\hat{p}_2-(0)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{950}+\frac{1}{940})}}\overset{d}{\sim}N(0,1)$

(四) (C) $p-value=2P(Z>|-0.2|)=2P(Z>-0.2)$

【高分閱讀】
趙治勳老師，統計學上課講義11.5六。