Solution Exam TMA 4255, Spring 2023

Problem 1

a) Fitted models

$$\hat{y}_1 = 1168.1 - 4.72\,X1 \quad = \quad b_{01} - b_{11}\,X1$$

$$\hat{y}_2 = 1203.4 - 4.84\,X2 \quad = \quad b_{02} - b_{12}\,X2$$

b_{01} and b_{02} are the estimated expected run times when $X1$ and X_2 are zero. But the value 0 is far away from the observed values of X_1 and X_2 and the estimates may be of little value.

b_{11} and b_{12} are the estimated changes in expected run time if $X1$ and $X2$ are changed by 1, respectively.

b) $\varepsilon_1, \dots, \varepsilon_m$ are all assumed $N(0, \sigma^2)$ and independent.

$$T = \frac{\hat{\beta}_1}{\dfrac{S}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2}}}, \quad \text{where} \quad S = \sqrt{\frac{\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2}{m-2}} \quad \text{and}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})Y_i}{\sum_{i=1}^{m}(x_i - \bar{x})^2}.$$

$T \sim t$-distributed with 19 degrees of freedom

For model 1: $b_{11} = -4.72$, $\dfrac{s}{\sqrt{\sum_{i=1}^{m}(x_i-\bar{x})^2}} = 1.43 \Rightarrow T_{obs} = -3.3$

$2P(T_{19} \leq -3.3 \mid H_0) = 0.004 < 0.05 \Rightarrow H_0$ is rejected

For model 2: $b_{12} = -4.84$ and $\dfrac{s}{\sqrt{\sum_{i=1}^{m}(x_i-\bar{x})^2}} = 1.16 \Rightarrow T_{obs} = -4.18$

and $2 \cdot P(T_{19} \leq -4.18 \mid H_0) = 0.001 < 0.05 \Rightarrow H_0$ is rejected c)

	R^2	R^2_{adj}	R^2_{pred}
Model 1	36.4	33.1	25.6
Model 2	47.9	45.2	35.4
Model 3	48.6	42.9	30.8

Model 3 has the largest R^2, but R^2 will always increase when a regression variable is added.

Model 2 has the largest R^2_{adj}, but more important when prediction is concerned, it has the largest R^2_{pred}. Therefore Model 2 should be used for prediction

$X1 = 60$ and $X2 = 70$ gives the prediction

$\hat{y} = 1203.1 - 4.84 \cdot 70 = \underline{864.6}$

Both X1 and X2 are clearly significant when they are in the model alone, but together X1 is far from being significant and X2 is on the border. Such things may happen when regression variables are strongly correlated.

Problem 2

a) Estimated effects are calculated as $\bar{y}_H - \bar{y}_L$ where $\bar{y}_H$ is the average of the observations on the high level and $\bar{y}_L$ is the average on the low level.

Let $Var(Y_i) = \sigma^2$, $i = 1, 2, \dots, 16$

There are for each effect column 8 observations on the ~~the~~ high level and 8 on the low level.

Thereby: $Var(\bar{Y}_H - \bar{Y}_L) = \frac{\sigma^2}{8} + \frac{\sigma^2}{8} = \frac{\sigma^2}{4}$

b) $E = ABC$ and $F = ABD$ are the generators

This gives. $I = ABCE = ABDF = CDEF$ since $I^2 = ABCE\,ABDF = CDEF$.

The shortest word in the defining relation is 4 which means that the resolution is 4.

c) $AI = BCE = BDF = ACDEF$. Hence these 3 effects are aliased ~~with~~ with the main effect of A

$CI = ABE = ABCDF = DEF$. These three effects are aliased ~~with~~ with the main effect of C.

$BCDI = AED = ACF = ABEF$. These three effects are confounded with the block factor in addition to BCD

Problem 3

In this case since we have 2 populations, it is natural to use a χ^2-test for homogeneity.

For homogeneity we must have

$$\left.\begin{array}{l} p_1 = p_2 \\ 1-p_1 = 1-p_2 \end{array}\right\} \Longleftrightarrow p_1 = p_2$$

b) The four estimated expected numbers are

$$c_{11} = \frac{167 \cdot 74}{307} = 40.25 \qquad c_{12} = \frac{140 \cdot 74}{307} = 33.75$$

$$c_{21} = \frac{167 \cdot 233}{307} = 126.75 \qquad c_{22} = \frac{140 \cdot 233}{307} = 106.25$$

$$Q = \frac{(27-40.25)^2}{40.25} + \frac{(47-33.75)^2}{33.75} + \frac{(140-126.75)^2}{126.75} + \frac{(93-106.25)^2}{106.5} = 12.609.$$

$\chi(1)_{0.05} = 3.84 \Rightarrow$ Ho is rejected on a 5% significance level.

Problem 4

a) $\bar{X} \sim N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$ which means that by these levels we declare the process to be out of control if $|\bar{X} - \mu| \geq k\, SD(\bar{X})$ which is natural since the normal-distribution is symmetric.

$$P(\bar{X} - \mu \geq 3\tfrac{\sigma}{\sqrt{n}}) + P(\bar{X} - \mu \leq -3\tfrac{\sigma}{\sqrt{n}})$$

$$= 2P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq -3\right) = 2P(Z \leq -3) = 2 \cdot 0.0013 = 0.0026$$

b)

$$E[T] = \frac{1}{p} = \frac{1}{0.0026} = \underline{384.62}$$

$P(T > t) = P($ no out-of-control signal for the first t hours$)$

$$= (1-p)^t$$

$$\left.\begin{array}{l} P(t_1 \le T \le t_2) = 0.95 \\ \text{and} \quad P(T < t_1) = 0.025 \end{array}\right\} \Rightarrow \left\{\begin{array}{ll} P(T \le t_1 - 1) = 0.025 & (1) \\ P(T \le t_2) = 0.975 & (2) \end{array}\right.$$

From (1) $P(T \le t_1 - 1) = 1 - P(T > t_1 - 1) = 1 - (1-p)^{t_1 - 1} = 0.025$

$$\Rightarrow (1-p)^{t_1 - 1} = 0.975 \Rightarrow (t_1 - 1)\ln(1-p) = \ln(0.975)$$

$$\Rightarrow t_1 - 1 = \frac{\ln(0.975)}{\ln(0.9974)} \Rightarrow t_1 = 1 + 9.72 = 10.72 \approx \underline{11}$$

From (2) $P(T \le t_2) = 1 - P(T > t_2) = 1 - (1-p)^{t_2} = 0.975$

$$\Rightarrow (1-p)^{t_2} = 0.025 \Rightarrow t_2 \ln(1-p) = \ln(0.025) \Rightarrow t_2 = \frac{\ln(0.025)}{\ln(0.9974)}$$

$$= 1416.95 \approx \underline{1417}$$