The Hong Kong Polytechnic University

AMA1501 Introduction to Statistics for Business

Exam 2017/18 Semester 2 Outline Suggested Solution

1. (a) Class marks: 2.5, 7.5, 12.5, 17.5, 22.5, 27.5, 32.5

$\sum f = 100, \quad \sum fx = 1980, \quad \sum fx^2 = 44275$

$$\text{Mean} = \frac{1980}{100} = 19.8 \text{ hours}$$

$$\text{Mode} = 15 + \frac{35-13}{(35-13)+(35-19)} \times (20-15) = 17.8947 \text{ hours}$$

$$\text{Standard deviation} = \sqrt{\frac{100(44275) - 1980^2}{100(100-1)}} = 7.1570$$

(b) $P_{85} = 25 + \dfrac{85-75}{16} \times 5 = 28.125$ hours

(c) $50\Big(3+5+13+35+\dfrac{24-20}{25-20} \times 19\Big)\Big/100 = 33.1$

(d) Let X be the weekly total amount of time (hours) spent at the centre in the last week.

Assumed that random sample is drawn with replacement or infinite populations. By central limit theorem, $\bar{X} \sim N(\mu, \sigma^2/n)$ approximately. Furthermore, $\sigma \approx s$.

A 98% confidence interval for μ is

$19.8 \pm 1.96 \times \dfrac{7.1570}{\sqrt{100}}$

$= (18.397, 21.203)$

2. (a) $2! \times 5! = 240$

(b) A: selected staff is rated 'Very Satisfactory' by his/her supervisor

B: selected staff is rated 'Very Satisfactory' by his/her sub-ordinates

$P(A) = 0.65, P(B) = 0.7, P(B|A) = 0.6$

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

$= P(A) + P(B) - P(B|A)P(A)$

$= 0.65 + 0.7 - (0.6)(0.65) = 0.96$

ii. $P(B|\bar{A}) = \dfrac{P(\bar{A} \cap B)}{P(\bar{A})} = \dfrac{P(B) - P(A \cap B)}{1 - P(A)} = \dfrac{0.7 - 0.65 \times 0.6}{1 - 0.65} = \dfrac{31}{35}$

(c) i. D: project has generated revenue above \$200,000

A, B, C: project relate to research interest of A, B and C, respectively

$P(A) = 0.3, \quad P(B) = 0.38, \quad P(C) = 0.32$

$P(D|A) = 0.3$

$P(D|B) = 0.5$

$P(D|C) = 0.4$

$$P(B|D) = \frac{P(D|B)P(B)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)}$$

$= 0.2206$

ii. X: number of projects relate to research interest of A out of 5, given that the projects have generated revenue above \$200,000

$X \sim B(5, 0.2206)$

$$P(X \geq 3) = \sum_{x=3}^{5} \binom{5}{x} (0.2206)^x (1 - 0.2206)^{5-x} = 0.00975$$

3. (a) i. X: Food delivery time (minutes)

$X \sim N(30, 8^2)$

$P(24 < X < 34) = P(-0.75 < Z < 0.5) = 1 - 0.2266 - 0.3085 = 0.4649$

ii. Let a be the required delivery time (minutes)

$$P(X < a) = P\Big(Z < \frac{a - 30}{8}\Big) = 0.025 \approx P(Z < -1.96)$$

$$\frac{a - 30}{8} \approx -1.96$$

$a \approx 30 - 1.96 \times 8 = 14.32$

iii. $P(X \leq 32) = P(Z < 0.25) = 1 - 0.4013 = 0.5987$

Y: number of delivery jobs out of 200, having delivery time within 32 minutes

$Y \sim B(200, 0.5987)$

Since $np > 5$, $nq > 5$, $0.1 < p < 0.9$, normal approximation to binomial distribution is used, $\mu = 200 \times 0.5987 = 119.74$ and

$\sigma^2 = 200 \times 0.5987 \times (1 - 0.5987) = 48.051662$

$$P(Y \geq 110) \approx P\Big(Z > \frac{109.5 - 119.74}{\sqrt{48.051662}}\Big) \approx P(Z > -0.035)$$

$= 1 - (0.488 + 0.484)/2 = 0.514$

iv. $\bar{X} \sim N(30, 8^2/100)$

$P(28 < \bar{X} < 33) = P(-2.5 < Z < 3.75) = 1 - 0.00621 - 0.00009 = 0.9937$

(b) X: number of patients suffered from the illness again after completed the treatment of a certain psychological illness, among 1000 patients

$X \sim B(1000, 0.008)$

As n is large and p is close to 0, Poisson Approximation to binomial is used.

$$P(X = 5) = \frac{e^{-8}8^5}{5!} = 0.0916$$

4. (a) Let $\hat{p}$ be the sample proportion of employees who support the change. Assume that random sample is drawn with replacement or infinite populations and by central limit theorem, $\hat{p} \sim N\Big(p, \frac{p(1-p)}{n}\Big)$ approximately.

A 95% confidence interval for p is $\frac{185}{250} \pm 1.96\sqrt{\left(\frac{185}{250}\right)\left(\frac{65}{250}\right)\Big/250}$, i.e., (0.6856, 0.7944).

(b) Assume population distribution of daily sales of two new stores are normally distributed, and random samples are drawn with replacement or there are infinite population sizes.

$$z_{0.025}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} \le \text{error}$$

$$1.96\sqrt{\frac{500^2}{n} + \frac{480^2}{n}} \le 100 \Rightarrow n \ge \left(\frac{1.96 \times \sqrt{500^2 + 480^2}}{100}\right)^2 = 184.55$$

(c) Let X_1, X_2 be the traffic flow (number of vehicles passing through a certain junction) in a 15-minute period in weekdays and weekends, respectively.

Assume $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $\sigma_1^2 = \sigma_2^2$ and random samples are selected with replacement or infinite populations

$$s_p^2 = \frac{(22-1)6^2 + (20-1)7^2}{22+20-2} = 42.175$$

$H_0 : \mu_1 = \mu_2,\quad H_1 : \mu_1 > \mu_2,\quad \alpha = 0.05$

Critical region: $t > t_{0.05,22+20-2} = 1.684$

Under H_0, test statistic $t = \dfrac{40-32}{\sqrt{42.175\left(\frac{1}{22} + \frac{1}{20}\right)}} = 3.987$.

Decision: Reject H_0.

Conclusion: The mean number of vehicles passing through the junction in 15-minute periods during weekdays is significant higher than that during weekends at 5% level of significance.

5. (a) Let D be the difference in total score, which is computed by the score of caterer A rated by customer − score of caterer B rated by customer.

Assume $D \sim N(\mu_D, \sigma_D^2)$ and random sample is drawn with replacement or infinite population size.

Customer	1	2	3	4	5	6	7	8	9
d	8	−2	−3	5	8	4	5	5	4

$n = 9,\quad \sum d = 34,\quad \sum d^2 = 248$

$$\bar{d} = \frac{34}{9} = 3.778,\quad s_d = \sqrt{\frac{9(248) - (34)^2}{9(9-1)}} = 3.8658$$

$H_0 : \mu_D = 0,\quad H_1 : \mu_D > 0,\quad \alpha = 0.025$

Critical region: $t > t_{0.025,8} = 2.306$

Under H_0, test statistic $t = \dfrac{3.778}{3.8658/\sqrt{9}} = 2.9317$.

Decision: Reject H_0.

Conclusion: The mean score of caterer A is significance higher than caterer B at 2.5% level of significance.

(b) H_0: number of special edition products in a box follows the Binomial distribution

H_1: number of special edition products in a box do not follow the Binomial distribution

$\alpha = 0.05$

$\hat{p} = \bar{x}/n = (253/200)/5 = 0.253$

x	0	1	2	3	4	5
O_i	45	84	50	16	4	1
$P(X = x)$	0.2326	0.3939	0.2668	0.0904	0.0153	0.0010
E_i	46.5	78.8	53.4	18.1	3.1	0.2
					21.3	

Critical region: $\chi^2 > 5.991, \quad \nu = 4 - 1 - 1 = 2$

Under H_0, test statistic $\chi^2 = \frac{(45-46.5)^2}{46.5} + \frac{(84-78.8)^2}{78.8} + \frac{(50-53.4)^2}{53.4} + \frac{(21-21.3)^2}{21.3} =$ 0.613.

Do not reject H_0 and conclude that the number of special edition products in a box follows the Binomial distribution at 5% level of significance.

(c) H_0: students' preference for the proposed study tour and level of study are independent

H_1: students' preference for the proposed study tour and level of study are not independent

$\alpha = 0.025$

Critical region: $\chi^2 > 11.143, \quad \nu = (3-1)(3-1) = 4$

Expected frequencies:

	Students' preference		
Level of study	Not preferred	Neutral	Preferred
Freslunen	14.63	24.75	35.63
Sophomores	8.97	15.18	21.85
Seniors	15.41	26.07	37.53

Under H_0, test statistic

$$\chi^2 = \sum\sum \frac{(O-E)^2}{E} = \frac{(15-14.63)^2}{14.63} + \cdots + \frac{(34-37.53)^2}{37.53} = 3.00$$

Do not reject H_0 and conclude that students' preference for the proposed study tour and level of study are independent at 2.5% level of significance.

6. (a) i. $b = \dfrac{8(253.62) - (75)(27.77)}{8(1051) - (75)^2} = -0.01933,$

$$a = \frac{27.77}{8} - (-0.01933) \times \frac{75}{8} = 3.6525$$

$\hat{y} = 3.6525 - 0.01933x$

ii. $r = \dfrac{\left[8(253.62) - (75)(27.77)\right]}{\sqrt{\left[8(1051) - (75)^2\right]\left[8(96.6477) - (27.77)^2\right]}} = -0.719$

There is a moderate negative linear association between academic perfonnance of students and time spent on part-time job per week.

(b) i. Let y be daily traffic flow on a certain highway and x_1 be rainfall forecast for the day.

$\hat{y} = 779 + 0.1039x_1 - 0.00000222x_2 + 98x_3 + 0.00291x_4$

ii. $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.01$

Critical region: $t < -t_{0.005,65-4-1} = -2.66$ and $t > 2.66$

Under H_0, test statistic $t = \dfrac{0.1039}{0.0136} = 7.64$

Decision: Reject H_0.

Conclusion: Rainfall forecast for the day is a significant independent in explaining daily traffic flow at 5% level of significance.

iii. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1 : at least one $\beta_i \neq 0$

$\alpha = 0.05$

Critical region: $f > F_{0.05;4,60} = 2.53$

Under H_0, test statistic $f = 520$

Decision: Reject H_0.

Conclusion: The regression equation of traffic flow on rainfall forecast for the day together with three more independent variables is significant at 5% level of significance.

iv. $\dfrac{SST \times R^2/4}{SST \times (1 - R^2)/60} = 520 \quad \Rightarrow \quad R^2 = \dfrac{520}{535} = 0.972$

97.2% variation in traffic flow on a certain highway can be explained by the regression equation on rainfall forecast for the day together with three more independent variables.