The Hong Kong Polytechnic University

AMA1501 Introduction to Statistics for Business

Exam 2017/18 Semester 1 Outline Suggested Solution

1. (a)

Class mark (x)	Frequency, f	Cum. Freq
100	4	4
175	7	11
225	15	26
275	24	50
325	17	67
375	11	78
425	6	84

$\sum f = 84, \quad \sum fx = 23800, \quad \sum fx^2 = 7255000$

$$\text{Mean} = \frac{23800}{84} = \$283\frac{1}{3}$$

$$\text{Mode} = 250 + \frac{24-15}{(24-15)+(24-17)} \times (300-250) = \$278.125$$

$$\text{Standard deviation} = \sqrt{\frac{84(7255000) - 23800^2}{84(84-1)}} = \$78.5153$$

(b) $$Q_1 = 200 + \frac{21-11}{15} \times (250-200) = \$233\frac{1}{3}$$

$$Q_3 = 300 + \frac{63-50}{17} \times (350-300) = \$338\frac{4}{17}$$

$$\text{Interquartile range} = Q_3 - Q_1 = \$104\frac{46}{51}$$

(c) $$\Big(\frac{250-240}{250-200} \times 15 + 24 + 17 + \frac{380-350}{400-350} \times 11\Big)\Big/84 = 0.6024$$

(d) Let p be the population proportion of customers who spent between \$240 and \$380 in the last month. It is assumed that random sample is drawn with replacement, or there is infinite population size. By central limit theorem,
$\hat{p} \sim N\Big(p, \frac{pq}{n}\Big)$ approximately.

A 98% confidence interval for p is
$$\frac{50.6}{84} \pm 2.326\sqrt{\frac{50.6}{84}\Big(1 - \frac{50.6}{84}\Big)\Big/84}$$
$= (0.4782, 0.7266)$

2. (a) $\binom{2}{1}\binom{4}{2}\binom{6}{3} = 240$

(b) C: selected customer purchased clothes

E: selected customer purchased small electricity appliances

$P(C) = 0.65, P(E) = 0.38, P(E|C) = 0.42$

i. $P(C \cup E) = P(C) + P(E) - P(C \cap E)$

$= P(C) + P(E) - P(E|C)P(C)$

$= 0.65 + 0.38 - (0.42)(0.65) = 0.757$

ii. $P(C|E) = \dfrac{P(C \cap E)}{P(E)} = \dfrac{P(E|C)P(C)}{P(E)} = \dfrac{(0.42)(0.65)}{0.38} = 0.7184$

$P(C'|E) = 1 - 0.7184 = 0.2816$

iii. $P(E'|C') = \dfrac{P(E' \cap C')}{P(C')} = \dfrac{1 - P(E \cup C)}{1 - P(C)} = \dfrac{1 - 0.757}{1 - 0.65} = 0.6943$

(c) D: one defective product out of 20

A, B, C: production line of City A, B and C, respectively, is chosen

$P(A) = \dfrac{3}{12}, \quad P(B) = \dfrac{4}{12}, \quad P(C) = \dfrac{5}{12}$

$$P(D|A) = \binom{20}{1}(0.05)^1(0.95)^{19} = 0.3774$$

$$P(D|B) = \binom{20}{1}(0.045)^1(0.955)^{19} = 0.3752$$

$$P(D|C) = \binom{20}{1}(0.04)^1(0.96)^{19} = 0.3683$$

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

$= 0.3354$

3. (a) i. X: percentage of revenue spend on R&D

$X \sim N(60, 12^2)$

$P(X \geq 45) = P(Z > -1.25) = 0.8944$

ii. $P(X < 65.4\%) = P(Z < 0.45) = 0.6736$

Y: number of companies out of 100, spend at most 65.4% of revenue on R&D

$Y \sim B(100, 0.6736)$

Since $np > 5$, $nq > 5$, $0.1 < p < 0.9$, normal approximation to binomial distribution is used,

$\mu = 100 \times 0.6736 = 67.36, \quad \sigma^2 = 100 \times 0.6736 \times (1 - 0.6736) = 21.986304$

$Y \sim B(100, 0.6736) \approx N(67.36, 21.986304)$

$$P(Y \geq 50) = P\Big(Z > \frac{49.5 - 67.36}{\sqrt{21.986304}}\Big) = P(Z > -3.81) = 0.99993$$

iii. $$P\big(\min_i X_i < 42\big) = 1 - P\big(\min_i X_i \geq 42\big) = 1 - P(X_1 \geq 42, \ldots, X_5 \geq 42)$$

$$= 1 - \prod_{i=1}^{5} P(X_i \geq 42) = 1 - \big[P(X \geq 42)\big]^5 = 1 - \big[P(Z \geq -1.5)\big]^5$$

$$= 1 - (1 - 0.0668)^5 = 0.2923$$

(b) X: number of click-throughs received by a website in a 5-minute period

$X \sim Po(8)$

i. $$P(X \geq 5) = 1 - \sum_{x=0}^{5} \frac{e^{-8} 8^x}{x!} = 0.900368$$

ii. $$P(X > 8 | X \geq 5) = \frac{P(X > 8 \cap X \geq 5)}{P(X \geq 5)} = \frac{P(X > 8)}{P(X \geq 5)}$$

$$= \frac{0.407453}{0.900368} = 0.45254$$

4. (a) X_A: IQ score of student of school A

X_B: IQ score of student of school B

$X_A \sim N(120, 12^2)$, $X_B \sim N(132, 13^2)$ and random samples are drawn with replacement or infinite population sizes.

$\bar{X}_A - \bar{X}_B \sim N\Big(120 - 132, \frac{12^2}{10} + \frac{13^2}{10}\Big)$

$P(-10 \leq \bar{X}_A - \bar{X}_B \leq 10) \approx P(0.36 \leq Z \leq 3.93) = 0.3594 - 0.00004 = 0.35936$

(b) X: pollutant concentration of NO_2 $(\mu g/m^3)$

Assume that $X \sim N(\mu, \sigma^2)$ and random sample is drawn with replacement or infinite population.

$\sum x812.8$, $\sum x^2 = 56055.14$, $n = 12$

$\bar{x} = \frac{812.8}{12} = 67.7333$, $s = \sqrt{\frac{12(56055.14) - 812.8^2}{12(12-1)}} = 9.541711$

A 95% confidence interval for the population mean pollutant concentration of NO_2 $(\mu g/m^3)$ is

$67.7333 \pm 2.201 \times \frac{9.541711}{\sqrt{12}} = (61.67078, 73.79589)$

(c) $1.96\sqrt{\frac{pq}{n}} \leq 0.08$ and the standard error of $\hat{p}$ is maximized when $p = 0.5$. Then

$1.96\sqrt{\frac{(0.5)(0.5)}{n}} \leq 0.08 \Rightarrow n \geq \Big(\frac{1.96}{0.08}\Big)^2 \times 0.5^2 = 150.0625$. Take $n = 151$.

(d) Let X_1, X_2 be the job satisfaction index of employee who worked for the company at most 3 years and worked for the company more than 15 years, 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 population sizes.

$s_p^2 = \frac{(12-1)8^2 + (14-1)7^2}{12 + 14 - 2} = 55.875$

$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$, $\alpha = 0.01$

Critical region: $t > t_{0.01, 12+14-2} = 2.492$

Under H_0, test statistic $t = \frac{82 - 75}{\sqrt{55.875\Big(\frac{1}{12} + \frac{1}{14}\Big)}} = 2.3804$.

Decision: Do not reject H_0.

Conclusion: The mean job satisfaction level of employees who worked for the company at most 3 years is not significant higher than those worked for the company more than 15 years at 1% level of significance.

5. (a) Let D be the difference in assessment scores of social responsibility which is computed by the score after the completion of programme score before joining the programme. Assume $D \sim N(\mu_D, \sigma_D^2)$ and random sample is drawn with replacement or infinite population size.

Student	1	2	3	4	5	6	7	8
d	14	24	19	26	28	10	19	18

$n = 8, \quad \sum d = 158, \quad \sum d^2 = 3378$

$\bar{d} = \dfrac{158}{8} = 19.75, \quad s_d = \sqrt{\dfrac{8(3378) - (158)^2}{8(8-1)}} = 6.0651$

$H_0 : \mu_D = 0, \quad H_1 : \mu_D > 0, \quad \alpha = 0.05$

Critical region: $t > t_{0.05,7} = 1.895$

Under H_0, test statistic $t = \dfrac{19.75 - 0}{6.065123/\sqrt{8}} = 9.21.$

Decision: Reject H_0.

Conclusion: The mean assessment score of students has increased after joining the programme at 5% level of significance.

(b) H_0: preference of customers on the design of bed linen set remain the same as before

H_1: preference of customers on the design of bed linen set have changed

$\alpha = 0.01$

Design	Floral pattern	Cartoon figures	Animal pattern	Classic design	Sum
O_i	28	20	24	48	120
E_i	24	18	18	60	

Critical region: $\chi^2 > 11.345, \quad \nu = 4 - 1 - 0 = 3$

Under H_0, test statistic $\chi^2 = \dfrac{(28-24)^2}{24} + \dfrac{(20-18)^2}{18} + \dfrac{(24-18)^2}{18} + \dfrac{(48-60)^2}{60} = 5.2889.$

Do not reject H_0 and conclude that the preference of customers on the design of bed linen set remain the same as before at 5% level of significance.

(c) H_0: employee's level of satisfaction on medical benefits and staff banding are independent

H_1: employee's level of satisfaction on medical benefits and staff banding are not independent

$\alpha = 0.05$

Critical region: $\chi^2 > 9.488, \quad \nu = (3-1)(3-1) = 4$

Expected frequencies:

Ranking of employee	Views on new promotion policy		
	Disagree	Neutral	Disagree
Junior	34.78	37.45	34.78
Middle	53.63	57.75	53.63
Senior	41.60	44.80	41.60

Under H_0, test statistic

$$\chi^2 = \sum\sum \frac{(O-E)^2}{E} = \frac{(58-34.78)^2}{34.78} + \cdots + \frac{(56-41.60)^2}{41.60} = 37.576$$

Reject H_0 and conclude that the views of employees on the new promotion policy and their ranking are not independent at 1% level of significance.

6. (a) i. $b = \dfrac{10(668.8) - (23.6)(228)}{10(71.02) - (23.6)^2} = 8.5304,$

$a = \dfrac{228}{10} - (8.5304) \times \dfrac{23.6}{10} = 2.6682$

$\hat{y} = 2.6682 + 8.5304x$

ii. $r = \dfrac{\left[10(668.8) - (23.6)(228)\right]}{\sqrt{\left[10(71.02) - (23.6)^2\right]\left[10(6376) - (228)^2\right]}} = 0.9731$

iii. If the age of machines are increased by 1 year, the mean machine down time per month is expected to increase by 8.5304 hours.

(b) i. Let y and x_1 be the weekly sales of the cafe and weekly pedestrian flow on the street where the cafe is located

$\hat{y} = 1.0834 + 0.1037x_1 - 1.2158x_2 - 0.5308x_3 - 1.0765x_4$

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

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

$\alpha = 0.05$

Critical region: $f > F_{0.05;4,24-4-1} = 2.90$

Under H_0, test statistic $f = 222.173$

Decision: Reject H_0.

Conclusion: The fitted regression equation of weekly sales of the cafe on the four independent variables is significant at 5% level of significance.

iii. $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.01$

From the output, the p-value is $0.0001 < 0.01$ so H_0 is rejected. The independent variables 'weekly pedestrian flow on the street where the cafe is located' (x_1) is significant at 1% level of significance.

iv. $\dfrac{SSR/4}{(119.1496 - SSR)/19} = 222.173$

$SSR = 116.6555$

$R^2 = \dfrac{116.6555}{119.1496} = 0.9791$

97.91% variation in weekly sales of the cafe can be explained by the fitted regression line.