The Hong Kong Polytechnic University
AMA1501 Introduction to Statistics for Business
Exam 2016/17 Semester 1 Outline Suggested Solution

Question 1.

(a) Classmarks: 15,25,32.5,37.5,42.5,47.5,55
$\sum f = 120, \sum fx = 4730, \sum fx^2 = 194125$, Mean $= \frac{4730}{120} = 39.4167\%$
Mode $= 35 + \frac{36-20}{(36-20)+(36-28)}(40-35) = 38.3333\%$
Standard Deviation $= \sqrt{\frac{120(194125)-4730^2}{120(119)}} = 8.0357\%$

(b) $\frac{45-42}{45-40}(28) + 16 + 10 = 42.8$

(c) CV of this survey is $\frac{8.0357}{39.4167}100\% = 20.3866\%$
CV of the previous survey is $\frac{15}{35}100\% = 42.8571\%$
Since the coefficient of variation of the previous survey is larger than that of this survey, the set of data for the survey conducted 10 yaers ago has a larger variability.

(d) Let X be the percentage of monthly household income used for mortgage payment. Assume that random samples are drawn with replacement or infinite populations. By central limit theorem, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ approximately. Furthermore, $\sigma \cong s$.
$H_0 : \mu \leq 35$
$H_1 : \mu > 35$
$\alpha = 0.025$
Critical region: $z > 1.96$
Under H_0, test statistic $z = \frac{39.4167-35}{8.0357/\sqrt{120}} = 6.0209$
Reject H_0. It is concluded that the mean percentage of monthly household income for mortgage payment of all customers with mortgage loan is significant greater than 35% at 2.5% level of significance.

Question 2.

(a) $\begin{pmatrix} 4 \\ 2 \end{pmatrix} \begin{pmatrix} 5 \\ 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} 6! = 129600$

(b)(i) A: selected credit card holder sets automatic payment instructions to settle credit card payments

B: selected credit card holder uses Octopus Automatic Add Value Service

$P(A) = 0.45, P(B) = 0.6, P(A|B) = 0.55$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.45 + 0.6 - 0.6 * 0.55 = 0.72$

(b)(ii) $P(\bar{B}|\bar{A}) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{A})} = \frac{1-P(A \cup B)}{1-P(A)} = \frac{1-0.72}{1-0.45} = \frac{28}{55}$

(c)(i) A: selected price proposal under-estimate the cost

B_1, B_2, B_3: selected price proposal is prepared by Joe, Jenny and Jack, respectively

$P(B_1) = 0.2, P(B_2) = 0.3, P(B_3) = 0.5$

$P(\bar{A}|B_1) = 0.95, P(\bar{A}|B_2) = 0.98, P(\bar{A}|B_3) = 0.99$

$P(B_2|\bar{A}) = \frac{P(B_2)P(\bar{A}|B_2)}{\sum_{i=1}^{3} P(B_i)P(\bar{A}|B_i)} = \frac{0.3*0.98}{0.2*0.95+0.3*0.98+0.5*0.99} = 0.3003$

(c)(ii) X: number of price proposal that are prepared by Jenny out of 20, given that none of them under-estimate the cost

$X \sim B(20, 0.3003)$

$P(X = 6) = \begin{pmatrix} 20 \\ 6 \end{pmatrix} (0.3003)^6 (1 - 0.3003)^{14} = 0.1916$

Question 3.

(a)(i) X: time spent by students in handling final year projects in hours
$X \sim N(220, 50^2)$
$P(205 < X < 245) = P(-0.3 < Z < 0.5) = 1 - 0.3821 - 0.3085 = 0.3094$

(a)(ii) Let a be the required time spent
$P(X > a) = P(Z > \frac{a-220}{50}) = 0.15 \cong P(Z > 1.04)$
$\frac{a-220}{50} \cong 1.04, a \cong 220 + 1.04 * 50 = 272$ hours.

(a)(iii) $P(X \geq 205) = P(Z > -0.3) = 0.6179$
Y: number of students out of 50, spent more than 205 hours in handling their final year project $Y \sim N(50, 0.6179)$
Since $np > 5, nq > 5, 0.1 < p < 0.9$, normal approximation to binomial distribution is used, $\mu = 50(0.6179) = 30.895, \sigma^2 = 50(0.6179)(1-0.6179) = 11.8050$
$P(Y \geq 25) = P\left(Z > \frac{24.5-30.895}{\sqrt{11.8050}}\right) = P(Z > -1.86) = 0.9686$

(a)(iv) $\bar{X} \sim N(220, 50^2/16)$
$P(\bar{X} < 245) = P(Z < 2) = 0.9772$

(b) X: number of VIP customers arriving at the customer services counter in a randomly selected 15-minute period
$X \sim Po\left(12 * \frac{15}{60} = 3\right)$
$P(X > 4) = 1 - \sum \frac{e^{-3}3^x}{x!} = 0.1847$

Question 4.

(a)(i) Let $\hat{p}_A, \hat{p}_E$ be the sample proportion of adults and elderlies, respectively, who prefer a tour visiting Country X would have a duration of at most 7 days. Assume that random samples are drawn with replacement or infinite populations and by central limit theorem, $\hat{p}_A - \hat{p}_E \sim N\left(p_A - p_E, \frac{p_A(1-p_A)}{n_A} + \frac{p_E(1-p_E)}{n_E}\right)$ approximately.
$\hat{p}_A = \frac{120}{200}, \hat{p}_E = \frac{50}{150}, \hat{p} = \frac{120+50}{200+150} = \frac{170}{350}$
$H_0 : p_A = p_E$
$H_1 : p_A > p_E$
$\alpha = 0.01$
Critical region: $z > 2.33$
Under H_0, test statistic $z = \frac{120/200-50/150-0}{\sqrt{\frac{170}{350}\frac{180}{350}\left(\frac{1}{200}+\frac{1}{150}\right)}} = 4.94$
Decision: reject H_0 Conclusion: there is sufficient evidence indicating that the population proportion of all adults who prefer a tour visiting Country X would have a duration of at most 7 days is higher than that of elderlies at 1% level of significance.

(a)(ii) Assume a large sample is drawn and random sample is drawn with replacement or infinite population.
$z_{0.025}\sqrt{\frac{pq}{n}} \leq$ error which is maximized when $p = 0.5$.
$n \geq \left(\frac{1.96*0.5}{0.03}\right)^2 = 1067.111$

(b) Assume the random sample of 36 students is selected with replacement or there is an infinite population, and by central limit theorem the sample mean weekly hours in using the e-learning platform to perform learning related activities $\bar{X} \sim N(\mu, \sigma^2/36)$. Furthermore, $\sigma \cong s$.
A 95% confidence interval for μ is $38 \pm 1.96\frac{8}{6}$, that is $35.3867 < \mu < 40.6133$ hours.

(c) Let X_M, X_J be the evaluation score of staff member of management level and junior level, respectively. Assume $X_M \sim N(\mu_M, \sigma_M^2)$, $X_J \sim N(\mu_J, \sigma_J^2)$, $\sigma_M^2 = \sigma_J^2$ and random samples are selected with replacement or infinite populations.
$s_p^2 = \frac{11(9^2)+17(10^2)}{12+18-2} = \frac{2591}{28} = 92.5357$
$H_0 : \mu_M = \mu_J$
$H_1 : \mu_M < \mu_J$
$\alpha = 0.01$

Critical region: $t < -t_{0.01,12+18-2} = -2.467$
Under H_0, test statistic $t = \frac{68-74-0}{\sqrt{92.5357\left(\frac{1}{12}+\frac{1}{18}\right)}} = -1.6736$
Decision: do not reject H_0.
Conclusion: The mean score of staff members of management level is not significant less than that of junior level at 1% level of significance. Staff members of junior level, on average, do not have a more positive response to the new reward system than staff member of management level.

Question 5.

(a) Let D be the difference in performance score, which is computed by the score of customer services team - score of sales team. Assume $D \sim N(\mu_D, \sigma_D^2)$ and random sample is drawn with replacement or infinite population.

Project	1	2	3	4	5	6	7	8
d	8	4	-2	4	7	6	9	3

$n = 8, \sum d = 39, \sum d^2 = 275, \bar{d} = \frac{39}{8}, s_d = \sqrt{\frac{8*275-39^2}{8*7}} = 3.4821$
$H_0 : \mu_D = 0, H_1 : \mu_D > 0, \alpha = 0.05$
Critical region: $t > t_{0.05,7} = 1.895$
Under H_0, test statistics $t = \frac{4.875}{3.4821/\sqrt{8}} = 3.9599$
Decision: Reject H_0
Conclusion: the mean score of sales team is significantly lower than that of customer services team at 5% level of significance.

(b) H_0: number of customers arriving at the bank follow the Poisson distribution.
H_1: number of customers arriving at the bank do not follow the Poisson distribution.
$\alpha = 0.025, \hat{\lambda} = \bar{x} = 316/108$

x	0	1	2	3	4	5	6	≥ 7
O_i	5	14	25	28	18	12	6	0
$P(X = x)$	0.0536	0.1569	0.2295	0.2238	0.1637	0.0958	0.0467	0.0299
E_i	5.79	16.94	24.79	24.17	17.68	10.35	5.05	3.23

Critical region: $\chi^2 > 12.832, \nu = 7 - 1 - 1 = 5$
Under H_0, test statistics $\chi^2 = \frac{(5-5.79)^2}{5.79} + \frac{(14+16.94)^2}{16.94} + \cdots + \frac{(6-8.28)^2}{8.28} = 2.12$
Do not reject H_0 and conclude that the number of customers arriving at the bank follow the Poisson distribution at 2.5% level of significance.

(c) H_0: students' academic performance and employer's evaluation are independent
H_0: students' academic performance and employer's evaluation are not independent
$\alpha = 0.01$
Critical region: $\chi^2 > 13.277, \nu = (3-1)(3-1) = 4$
Expected frequencies:

	Very satisfactory	Satisfactory	Unsatisfactory
Excellent	31.51	37.13	9.36
Good	52.12	61.40	15.48
Fair	17.37	20.47	5.16

Under H_0, test statistics $= \sum\sum \frac{(O-E)^2}{E} = \frac{(42-31.51)^2}{31.51} + \cdots + \frac{(15-5.16)^2}{5.16} = 44.73$
Reject H_0 and conclude that students' academic performance and employer's evaluation are not independent at 1% level of significance.

Question 6.

(a)(i) $b = \frac{9(25558030)-15870(13993)}{9(29170500)-15870^2} = 0.7449$
$a = \frac{13993}{9} - (0.7449)\frac{15870}{9} = 241.3342$
$\hat{y} = 241.3342 + 0.7449x$

(a)(ii) $\hat{y} = 241.3342 + 0.7449(3500) = 2848.36$ kilowatt-hours.
The observed value of x do not cover $x = 3500$ and hence the fitted model in (i) may not be applicable to explain the relationship between x and y when the value of x is outside the observable range, that is, extrapolation. On the other hand, the proposed model may not be the same as the true model.

(b)(i) Let y be the household food consumption in a month (\$0000) and x_1, x_2 be monthly household income (\$0000) and household size, respectively.
$\hat{y} = 1.4326 + 0.00999x_1 + 0.37929x_2$.

(b)(ii) $H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$
$\alpha = 0.05$
Critical region: $t < -t_{0.025,22} = -2.074$ and $t > 2.074$.
Under H_0, test statistic $t = \frac{0.00999062}{0.00316806} = 3.1535$
Decision: H_0 is rejected
Conclusion: Monthly household income is a significant independent in explaining household food consumption in a month at 5% level of significance.

(b)(iii) $H_0 : \beta_1 = \beta_2 = 0$
H_1: at least one $\beta_i \neq 0$
$\alpha = 0.01$
Critical region: $f > F_{0.01;2,22} = 5.72$
Under H_0, test statistic $f = \frac{(1-0.0984)*SST/2}{(0.0984)SST/(25-2-1)} = 100.79$
Decision: reject H_0
Conclusion: the regression equation of household food consumption in a month on monthly household income (\$0000) and household size is significant at 1% level of significance.

(b)(iv) If the monthly household income is increased by 1 unit (that is \$10000) while the household size remains unchanged, the household food consumption in a month is expected to increase by 0.00999 unit (that is \$99.9).