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

Question 1.

(a) The frequency distribution:

Class mark (x)	Frequency, f	Cum. Freq
300.5	4	4
450.5	9	13
550.5	15	28
650.5	28	56
750.5	56	112
850.5	32	144
950.5	6	150

$\sum f = 150, \sum fx = 106675, \sum fx^2 = 78691637.5$ Mean $= \frac{106675}{150} = \$711.17$.

Standard deviation $= \sqrt{\frac{150(7869163.5) - 106675^2}{150(150-1)}} = \$137.77$

(b) $D_9 = 800.5 + \frac{135-112}{32}(900.5 - 800.5) = \$872.375$

(c) $\frac{1}{150}\left(\frac{700.5-680}{700.5-600.5}(28) + 56 + 32 + \frac{920-900.5}{1000.5-900.5}(6)\right) = 0.6327$

(d) Let p_1 and p_2 be the population proportion of invoices with amount of expenses between \$680 and \$920 in the last year and in this year, respectively. It is assumed that random samples are drawn with replacement, or there are infinite populations. By central limit theorem, $\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$ approximately.

$H_0 : p_1 = p_2$

$H_1 : p_1 < p_2$

$\alpha = 0.05$

Critical region: $z < -1.645$

$\hat{p}_1 = \frac{30}{100}, \hat{p}_2 = 0.6327, \hat{p} = \frac{30+94.91}{100+150} = \frac{124.91}{250} = 0.49964$

Under H_0, test statistics $z = \frac{0.3-0.6327-0}{\sqrt{0.49964(1-0.46694)\left(\frac{1}{100}+\frac{1}{150}\right)}} = -5.15$

Reject H_0. It is concluded that the proportion of invoices with amount of expenses between \$680 and \$920 in this year is significantly larger than that of the last year at 5% level of significance.

Question 2.

(a) $\binom{3}{2}\binom{4}{2}\binom{5}{2}6! = 129600$

(b)(i) A: selected student prefers Western European countries
B: selected student prefers North American countries
$P(A) = 0.55, P(B) = 0.8, P(\bar{A} \cap \bar{B}) = 0.05$
$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}) = 0.95$

(b)(ii) $P(A \cup B) = P(A) + P(B) - P(A \cap B), P(A \cap B) = 0.4, P(A|B) = \frac{0.4}{0.8} = 0.5$

(b)(iii) $P(B|\bar{A}) = \frac{P(\bar{A} \cap B)}{P(\bar{A})} = \frac{P(B) - P(A \cap B)}{1 - P(A)} = \frac{8}{9}$

(c) A: 2 out of 5 participants rate the course as useful
B_1, B_2, B_3, B_4: HR, marketing, R& D and other department, respectively, is chosen
$P(B_1) = 0.25 = P(B_2) = P(B_3) = P(B_4)$
$P(A|B_1) = \binom{5}{2} 0.75^2 (0.25^3) = P(A|B_2) = P(A|B_3) = P(A|B_4)$ $P(B_3|A) = \frac{P(B_3)P(A|B_3)}{\sum_i P(B_i)P(A|B_i)} = 0.1731$

Question 3.

(a)(i) X: score of generic competence test
$X \sim N(78, 8^2)$
$P(72 < X < 92) = P(-0.75 < Z < 1.75) = 1 - 0.2266 - 0.0401 = 0.7333$

(a)(ii) Y: number of candidates having a score between 72 and 92, out of 10 students.
$Y \sim B(10, 0.7333)$
$P(Y > 3) = 1 - P(Y = 0) - P(Y = 1) - P(Y = 2) - P(Y = 3) = 0.9948$

(a)(iii) $P(X > 75) = P(Z > -0.375) = 0.64615$
Y_2: number of candidates out of 200, will be invited to interview
$Y_2 \sim N(200, 0.64615)$
Since $np > 5, nq > 5, 0.1 < p < 0.9$, normal approximation to binomial distribution is used, $\mu = 129.23, \sigma^2 = 45.7280355$
$P(Y_2 \geq 150) = P\left(Z > \frac{149.5-129.23}{\sqrt{45.7280355}}\right) = P(Z > 3) = 0.00135$

(b)(i) X: number of vehicles arriving at the maintenance depot in a 30-minute period
$X \sim Po(4)$
$P(X > 5) = 1 - \sum \frac{e^{-4}4^x}{x!} = 0.2149$

(b)(ii) Expected number of technical staff $= 6(0.2149) + 4(0.7851) = 4.4298$

Question 4.

(a) X: monthly rate of return of a company (%)
$X \sim N(12, 5^2)$ and random sample is drawn with replacement or infinite population.
$P(\bar{X} > 10) = P(Z > \frac{10-12}{5/\sqrt{4}}) = 0.7881$

(b) X: assembly time of a product by trainees (minutes)
Assume that $X \sim N(\mu, \sigma^2)$ and random sample is drawn with replacement or infinite population.
$\sum x = 98.7, \sum x^2 = 1225.11, n = 8, \bar{x} = 12.3375, s = 1.0281$
A 95% confidence interval for the population mean assembly time of a product by trainees is $12.337 \pm 2.365(\frac{1.0281}{\sqrt{8}})$, that is $11.4779 < \mu < 13.1971$.

(c) Let p be the population proportion of VIP customers who spent more than \$20000 during annual mega sales. It is assumed that random sample of size 400 is selected with replaement or infinite population. By central limit theorem, $\hat{p} \sim N(p, pq/n)$ approximately.

$H_0 : p = 0.23$, $H_1 : p > 0.23$, $\alpha = 0.01$, $\hat{p} = 0.84$
critical region: $z > z_{0.01} \cong 2.33$
Under H_0, test statistic $z = \frac{84/400-0.23}{\sqrt{0.23(1-0.23)/400}} = -0.95$
Decision: do not reject H_0.
Conclusion: the population proportion of VIP customers who spent more than \$20000 during annual mega sales is not more than 23% at 1% level of significance.

(d) Let X_U and X_R be the weekly sales revenue of branch located at urban area and rural area, respectively.
Assume $X_U \sim N(\mu_U, \sigma_U^2), X_R \sim N(\mu_R, \sigma_R^2), \sigma_U = \sigma_R$ and random samples are selected with replacement or infinite populations.
$S_P^2 = \frac{(10-1)12000^2+(8-1)*11000^2}{10+8-2} = 133937500$
$H_0 : \mu_U = \mu_R$, $H_1 : \mu_U > \mu_R$, $\alpha = 0.05$
Critical region: $t > t_{0.05;10+8-2} = 1.746$
Under H_0, $t = \frac{280000-200000}{\sqrt{133937500(\frac{1}{10}+\frac{1}{8})}} = 14.57$ Decision: reject H_0
Conclusion: The mean weekly sales revenue of branches locating at rural area is significantly lower than that of urban area at 5% level of significance.

Question 5.

(a) Let D be the difference in estimated duration estimated by manager B and manager A. 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
d	2	-1	3	-1	3	3	-1

$n = 7, \sum d = 8, \sum d^2 = 34, \bar{d} = \frac{8}{7}, s_d = \sqrt{\frac{7*34-64}{7*6}} = 2.0354$
$H_0 : \mu_D = 0, H_1 : \mu_D \neq 0, \alpha = 0.05$
Critical region: $t < -t_{0.025,6} = -2.447, t > t_{0.025,6} = 2.447$
Under H_0, test statistics $t = \frac{1.1429}{2.0354/\sqrt{7}} = 1.4856$
Decision: Do not reject H_0
Conclusion: the mean project duration estimated by the two managers do not have significance difference at 5% level of significance.

(b) H_0: number of telephone enquiries received in a 5-minute period follow the uniform distribution.
H_1: number of telephone enquiries received in a 5-minute period do not follow the uniform distribution.
$\alpha = 0.05$

x	0	1	2	3	4	5
O_i	23	35	62	48	24	8
E_i	33.33	33.33	33.33	33.33	33.33	33.33

Critical region: $\chi^2 > 11.07, \nu = 6 - 1 = 5$
Under H_0, test statistics $\chi^2 = \frac{1}{33.33}((23 - 33.33)^2 + (35 - 33.33)^2 + \cdots + (8 - 33.33)^2) = 56.26$.
Reject H_0 and conclude that the number of telephone enquiries received in 5-minute period does not follow the uniform distribution at 5% level of significance.

(c) H_0: employee's level of satisfaction on medical benefits and staff banding are independent
H_0: employee's level of satisfaction on medical benefits and staff banding are dependent
$\alpha = 0.01$
Critical region: $\chi^2 > 13.277, \nu = (3-1)(3-1) = 4$
Expected frequencies:

	Unsatisfactory	Neural	Satisfactory
Senior	24	48	48
Middle	28	56	56
Subordinate	18	36	36

Under H_0, test statistics $= \sum\sum \frac{(O-E)^2}{E} = \frac{(33-24)^2}{24} + \cdots + \frac{(39-36)^2}{36} = 21.37$
Reject H_0 and conclude that employee's level of satisfaction on medical benefits and staff banding are not independent at 1% level of significance.

Question 6.

(a)(i) $b = \frac{12(6093)-(533)(132)}{12(24529)-533^2} = 0.2690$
$a = \frac{132}{12} - 0.2690\left(\frac{533}{12}\right) = -0.9495$
$\hat{y} = -0.9495 + 0.2690x$

(a)(ii) $\hat{y} = -0.9495 + 0.2690(58) = 14.65\%$

(a)(iii) $1 - R^2 = 1 - \frac{(12(6093)-(533)(132))^2}{[12(24529)-(533)^2][12(1526)-132^2]} = 1 - 0.8362 = 0.1638$
16.38% variation in moisture content cannot be explained by the fitted equation.

(b)(i) $\hat{y} = 0.002 + 0.0204x_1 - 0.0231x_2 + 0.0765x_3 + 0.276x_4 + 0.0018x_5$

(b)(ii) If the percentage of female in labour force is increased by 1 unit, the suicide rate is expected to decrease 0.0231 unit.

(b)(iii) $H_0 : \beta_1 = \beta_2 = \cdots = \beta_5 = 0$
H_1: at least one $\beta_i \neq 0$
$\alpha = 0.05$
Critical region: $f > F_{0.05;5,46-5-1} = 2.45$
Under H_0, test statistic $f = \frac{0.45 * SST/5}{(1-0.45)SST/(46-5-1)} = 6.5455$
Decision: reject H_0
Conclusion: the fitted regression equation of suicide rate on the five independent variables is significant at 5% level of significance.

(b)(iv) $H_0 : \beta_i = 0, H_1 : \beta_i \neq 0, i = 1, 2, 3, 4, 5,\ \alpha = 0.1$
From the output, only the first two p-value are less than 0.1 (0.002 for x_1 and 0.02 for x_2) and the remaining three p-value are greater than 0.1. The independent variables unemployment rate x_1 and percentage of female in the labour force x_2 are significant at 10% level of significance.