Topic 7: Inference for the Proportion Solutions

Q6

a) Let p be the population proportion of stores carried the brand

$H_0 : p \geq 0.19$

$H_1 : p < 0.19$

$\because n = 85$ $\quad n\hat{p}=11.9>5$ $\quad n(1-\hat{p})=73.1>5$

$\therefore$ Sampling distribution of $\hat{p}$ is approximately normal.

$\alpha = 0.05$ $\quad$ Critical Value $= -Z_{\alpha} = -Z_{0.05} = -1.645$

Reject H_0 if $Z < -1.645$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.14 - 0.19}{\sqrt{\frac{0.19 \times 0.81}{85}}} = -1.1751$$

$\because Z = -1.1751 > -1.645$

$\therefore$ We do not reject H_0. There is insufficient evidence that Grant has poorer distribution in Mainland China than it does in Hong Kong.

b) There is insufficient evidence that Grant has poorer distribution in Mainland China than it does in Hong Kong.

c) p-value=$P(Z \leq -1.18)$ =0.1190

Reject H_0 if p-value <0.05

$\because$ p-value = 0.1190> 0.05

$\therefore$ We do not reject H_0 and making the same decision as in (a)

d) Choose $\alpha = 0.1$

For a fixed sample size, a larger value of α would correspond to a smaller value of β, that can decrease the penalty of committing type II error.

Q7

a) Standard error of sample proportion=$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.42 \times 0.58}{300}}$ =0.0285

b) $\because n = 300$ $\quad np=300 \times 0.42 = 126 \geq 5$ $\quad n(1-p) = 300 \times 0.58 = 174 \geq 5$

$\therefore$ Sampling distribution of $\hat{p}_s$ is approximately normal.

$P(0.4 \leq \hat{p}_s \leq 0.45) = P(\frac{0.4-0.42}{0.0285} \leq Z \leq \frac{0.45-0.42}{0.0285}) = P(-0.70 \leq Z \leq 1.05)$

$= 0.8531 - 0.242 = 0.6111$

c) The range of 0.41 to 0.43 is more likely to lie because this range contains the population proportion that is 0.42

d) In the sample size 300, 61.11% of sample will be expected to have the sample proportions between 0.4 and 0.45.

Q8

a) Let $\hat{p}$ be the proportion of unemployment rate of Hong Kong in 2002

$n = 8500$

$n\hat{p} = 8500 \times (\frac{618}{8500}) = 618 \geq 5$

$n(1-\hat{p}) = 85000 \times \left(1 - \frac{618}{85000}\right) = 7882 \geq 5$

$\therefore$ sampling distribution $\hat{p}$ is normal

assume population follows binomial.

For 95% Confidence Interval,

$= 0.0727 \pm 1.96\sqrt{\frac{0.0727(0.9273)}{8500}}$

$= [0.0672, 0.0782]$

$\therefore$ We are 95% confident that the population proportion of Hong Kong unemployment rate is estimated to be between 0.0672 and 0.0782.

b) Sample size

$n = \frac{{Z_{\frac{0.05}{2}}}^2 \hat{p}(1-\hat{p})}{E^2}$=64750

c) Let p be the proportion of unemployment rate of Shatin in 2002

$H_0 : p \geq 0.073$

$H_1 : p < 0.073$

$\because n$ =620

$n\hat{p}$ = 34 >5

$n(1-\hat{p})$ = 586 >5

$\therefore$ the sampling distribution of $\hat{p}$ is approximately normal

$\hat{p} = \frac{34}{620} = 0.0548$

$\hat{p} \sim N(0.073, \sqrt{\frac{0.073(1-0.073)}{620}}^2)$

Reject H_0 if $Z <$ -1.645

$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.0548 - 0.073}{\sqrt{\frac{0.073(1-0.073)}{620}}}$ = -1.738<-1.645

Therefore we reject H_0. There is sufficient evidence that the unemployment in Shatin is lower than Hong Kong in 2002.

Q9

a) $p = \frac{1}{4} = 0.25$

b) Possible value of sample proportion of preferring the brand:

	Not prefer	Not prefer	Not prefer	prefer
Not prefer	0	0	0	0.5
Not prefer	0	0	0	0.5
Not prefer	0	0	0	0.5
prefer	0.5	0.5	0.5	1

Probability distribution of the sample proportion:

$\hat{p}$	0	0.5	1
$P(\hat{p})$	$\frac{9}{16} = 0.5625$	$\frac{6}{16} = 0.375$	$\frac{1}{16} = 0.0625$

Possible value of $\hat{p}$ from the above table: 0, 0.5, 1

c) $\mu_{\hat{p}}$ =0*0.5625+0.5*0.375+1*0.0625=0.25

From part (a), p =0.25

=> $\mu_{\hat{p}} = p$

=> sample distribution of $\hat{p}$ is an unbiased estimator for p

d) n=2, $n\hat{p}$=2*0.25=0.5<5, $n(1-\hat{p})$= 2*(1-0.25)=1.5<5

=> sample distribution of $\hat{p}$ does not follow a normal distribution

Q10

a) H_0: $p \leq 5\%$ vs H_1: $p > 5\%$

Type I error: let unhealthy patient with $\leq$5% white blood cells leave, resulting in unhealthy patients are not treated

Type II error: refer healthy patient with > 5% white blood cells to doctor, resulting in more cost is incurred or more patients are sent to doctor

Comparing the above type I and II error, type I error is more serious. We would rather make a type II error.

b) If there is no information available from past data,

α=1-90%=0.10, $Z_{\frac{0.10}{2}} = 1.645$

E=1.95%=0.0195

$p \approx 0.5$

Sample size, n=$\frac{1.645^2 \times 0.5 \times (1-0.5)}{0.0195^2}$=1779.11$\cong$1780 (round up)

Q11

X is the number of passengers responded to the survey

p is the population proportion of passenger responded to the survey

a) $n = \dfrac{(1.96)^2(0.13)(1-0.13)}{(0.06)^2} = 120.69 = 121$ (round-up)

b) H_0: $p \geq 0.13$

H_1: $p < 0.13$

As n = 350 $n\hat{p} = 28 > 5$; $n(1-\hat{p}) = 322 > 5$

→ $\hat{p} \sim N$

→ use Z test

At $\alpha = 0.05$, reject H_0 if $Z < -1.645$

$$\hat{p} = \frac{28}{350} = 0.08$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0\ (1\text{-}p_0)}{n}}} = \frac{0.08 - 0.13}{\sqrt{\dfrac{0.13(1-0.13)}{350}}} = -2.7815$$

As $Z = -2.7815 < -1.645$, reject H_0.

There is sufficient evidence that the response rate has been dropped.

c) 95% CI for p

$$= \hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.08 \pm 1.96\sqrt{\frac{0.08(1-0.08)}{350}} = 0.08 \pm 0.0284 = [0.0516,\ 0.1084]$$

We are 95% confident that the true unknown population proportion of passengers who are willing to response to the survey is between 0.0516 and 0.1084 (i.e. 5.16% or 10.84%).