LEE TONG MING A19EC0069
SAHARULNIZAM HAKIMY BIN SHOBRI A19EC0205

PSDA (Assignment 4)

Question 1

(a) i) $n = 1100, \ x = 990$

$\therefore \hat{p} = \frac{990}{1100} = 0.90$

ii) $1-\alpha = 0.99$

$\alpha = 0.01, \ \frac{\alpha}{2} = 0.005$

$Z_{\alpha/2} = Z_{0.005} = 2.575$

$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.9)(0.1)}{1100}} = 9.045 \times 10^{-3}$

$99\% \ CI = \hat{p} \pm Z_{\alpha/2}(SE)$

$= 0.90 \pm 2.575(9.045 \times 10^{-3})$

$= 0.90 \pm 0.0233$

$99\% \ CI = (0.8767, 0.9233)$

$\therefore$ We are 99% confident that the proportion of all drivers who have engaged in careless or aggresive driving in the previous 6 months is between 0.8767 and 0.9233.

(b) $n = 16, \ \bar{x} = 18.6, \ \sigma = 9.486$

$1-\alpha = 0.80$

$\alpha = 0.20, \ \frac{\alpha}{2} = 0.10$

$Z_{\alpha/2} = Z_{0.1} = 1.282$

80% CI:

$\bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 18.6 \pm 1.282\left(\frac{9.486}{\sqrt{16}}\right)$

$= 18.6 \pm 3.0403$

$80\% \ CI = (15.5597, 21.6403)$

$\therefore$ We are 80% confident that the mean time is between 15.5597 seconds and 21.6403 seconds.

Question 2

(a) null hypothesis, $H_0: p = 30\% = 0.30$

alternative hypothesis, $H_1: p > 0.30$

$\therefore$ A type I error is the mistake of rejecting a true null hypothesis, so this is a type I error:

The allergist concludes that there is sufficient evidence to support $p > 0.30$, when in reality $p = 0.30$.

(b) i) $H_0: \mu = 35$ minutes, $H_1: \mu < 35$ minutes

ii) $\alpha = 0.05$

$n = 20$

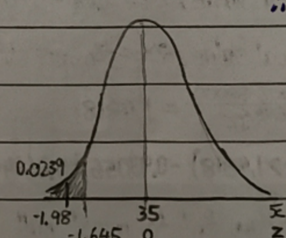
$\bar{x} = 33.1$

$\sigma = 4.3$

$C.V. = -Z_{\alpha} = -Z_{0.05} = -1.645$

$\therefore z = \frac{\bar{x} - \mu_x}{\sigma/\sqrt{n}} = \frac{33.1 - 35}{4.3/\sqrt{20}} = -1.98$

P-value $= P(z < -1.98) = 0.0239$



$\therefore$ Since P-value $0.0239 < 0.05$, the null hypothesis H_0 is rejected. There is sufficient evidence to support alternative hypothesis $H_1: \mu < 35$. This means average time taken to complete the test is less than 35 minutes.

(c) i) $H_0: \sigma = 0.470 kg$, $H_1: \sigma \neq 0.470 kg$

ii) $\bar{x} = \frac{2.47+2.54+3.02+3.22+3.39+3.43+3.52+3.68+2(3.73)+4.09+4.13+4.33+4.37+4.47+4.68}{16}$

$\bar{x} = \frac{58.8}{16} = 3.675$

$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{(2.47-3.675)^2+(2.54-3.675)^2+(3.02-3.675)^2+(3.22-3.675)^2+(3.39-3.675)^2+(3.43-3.675)^2+(3.52-3.675)^2+(3.68+3.675)^2+2(3.73-3.675)^2+(4.09-3.675)^2+(4.13-3.675)^2+(4.33-3.675)^2+(4.37-3.675)^2+(4.47-3.675)^2+(4.68-3.615)^2}{15}$

$s^2 = \frac{6.481}{15} = 0.4321$

$n = 16, \ s^2 = 0.4321, \ \sigma = 0.470$

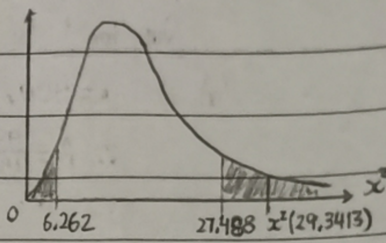
$\therefore$ Test statistics: $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{15(0.4321)}{(0.47)^2} = 29.3413$

iii) $\alpha = 0.05, \ \frac{\alpha}{2} = 0.025 \quad 1 - \frac{\alpha}{2} = 0.975, \ df = 16-1 = 15$

Upper tail critical value, $\chi^2_{0.025,15} = 27.488$

Lower tail critical value, $\chi^2_{0.975,15} = 6.262$

(c) iv) Conclusion: Since test statistics, χ^2: 29.3413 > upper tail cv: 27.488, reject null hypothesis H_0. There is sufficient evidence to support the alternative hypothesis, $H_1: \sigma \neq 0.470kg$. This means the vitamin supplement does not affect variation among birth weights.



Question 3

(a) $n = 350$, $x = 130$

$\hat{p} = \frac{130}{350} = 0.3714$

$1-\alpha = 95\% = 0.95$

$\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$

$Z_{\alpha/2} = Z_{0.025} = 1.96$

$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.3714)(0.6286)}{350}} = 0.0258$

$95\%\ CI = \hat{p} \pm Z_{\alpha/2}(SE)$

$= 0.3714 \pm 1.96(0.0258)$

$= 0.3714 \pm 0.0506$

$95\%\ CI = (0.3208, 0.4220)$

∴ We are 95% confident that the population of all food shops with special offers is between 0.3208 and 0.4220.

(b) Proportion from (a): 0.3208 to 0.4220

$1-\alpha = 95\%$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $Z_{0.025} = 1.96$

Margin of error: 0.04, $Z_{\frac{\alpha}{2}}(SE) = 0.04$ $SE = \frac{0.04}{1.96} = 0.0204$

Min proportion: 0.3208

$\hat{p} - 0.04 = 0.3208$

$\hat{p} = 0.3608$, $1-\hat{p} = 0.6392$

$SE = \sqrt{\frac{(0.3608)(0.6392)}{n}} = 0.0204$

$n = \frac{(0.3608)(0.6392)}{0.0204^2} = 554.17$

$n \approx 555$

Max proportion: 0.4220

$\hat{p} + 0.04 = 0.4220$

$\hat{p} = 0.3820$, $1-\hat{p} = 0.6180$

$SE = \sqrt{\frac{(0.3820)(0.6180)}{n}} = 0.0204$

$n = \frac{(0.3820)(0.6180)}{0.0204^2} = 567.27$

$n \approx 568$

∴ For proportion 0.3208 to 0.4220, the size of random sample required is 555 to 568 food shops for an approximate 95% CI with margin of error 0.04.

Question 4

(a) $n = 6$

$\bar{x} = 58392$

$\sigma = 648$

$\alpha = 0.05$

$H_0: \mu = 58000psi$, $H_1: \mu \neq 58000psi$

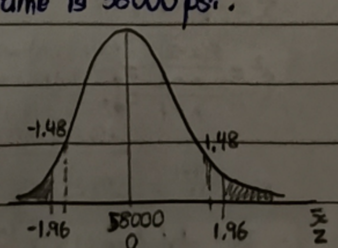
Upper tail critical value, $Z_{\alpha/2} = Z_{0.025} = 1.96$

Lower tail critical value, $-Z_{\alpha/2} = -Z_{0.025} = -1.96$

∴ $z = \frac{\bar{x}-\mu_x}{\sigma/\sqrt{n}} = \frac{58392-58000}{648/\sqrt{6}} = 1.48$

P-value $= 2(P>1.48) = 2(1-0.9306) = 0.1388$

∴ Since P-value 0.1388 > 0.05, we fail to reject null hypothesis H_0. This shows that the true average compressive strength of steel from which the sample came is 58000 psi.



(b) $n = 101$

$s^2 = 0.13$

$\sigma^2 = 0.18$

$\alpha = 0.05$

$df = 101-1 = 100$

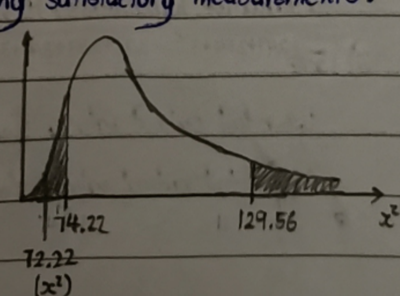
$H_0: \sigma^2 = 0.18\ inch^2$, $H_1: \sigma^2 \neq 0.18 inch^2$

Upper tail critical value, $\chi^2_{0.025,100} =$ 129.56

Lower tail critical value, $\chi^2_{0.975,100} = 74.222$

∴ Test statistics: $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{100(0.13)}{0.18} = 72.222$

∴ Since test statistics χ^2: 72.222 < lower tail critical value: 74.222, reject null hypothesis H_0. There is insufficient evidence to support the inspector making satisfactory measurements.



Question 5

(a) $n=418$, $x=280$

$\hat{p}=\frac{280}{418}=0.6699$

$1-\alpha=95\%=0.95$

$\alpha=0.05$, $\frac{\alpha}{2}=0.025$

$Z_{\alpha/2}=Z_{0.025}=1.96$

$SE=\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}=\sqrt{\frac{(0.6699)(0.3301)}{418}}=0.0230$

$95\%\ CI=\hat{p}\pm Z_{\alpha/2}(SE)$

$=0.6699\pm 1.96(0.023)$

$=0.6699\pm 0.0451$

$95\%\ CI=(0.6248, 0.7150)$

$\therefore$ We are 95% confident that the proportion of all insured patients who blame rising insurance rates on large court settlements against doctors, p is between 0.6248 and 0.7150.

(b) $n=126$

$\bar{x}=29.2$

$\sigma=7.5$

$1-\alpha=0.90$, $\alpha=0.10$, $\frac{\alpha}{2}=0.05$, $Z_{0.05}=1.645$

$90\%\ CI=\bar{x}\pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$=29.2\pm 1.645\left(\frac{7.5}{\sqrt{126}}\right)$

$=29.2\pm 1.0991$

$90\%\ CI=(28.1009, 30.2991)$

$\therefore$ We are 90% confident that the average blood lead level is between 28.1009 µg/dl and 30.2991 µg/dl.

Question 6

(a) $n=64$

$\bar{x}=38$

$\sigma=5.8$

$\alpha=0.05$

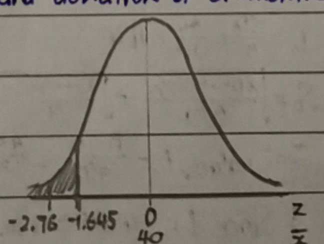
$H_0: \mu=40$ months, $H_1: \mu<40$ months

$C.V.=-Z_\alpha=-Z_{0.05}=-1.645$

$\therefore Z=\frac{\bar{x}-\mu_{\bar{x}}}{\sigma/\sqrt{n}}=\frac{38-40}{5.8/\sqrt{64}}=-2.76$

$P_{value}=P(z<-2.76)=0.0029$

$\therefore$ Since P-value $0.0029<0.05$, the null hypothesis H_0 is rejected. There is sufficient evidence to support alternative hypothesis $H_1: \mu<40$ if 64 mice placed on this diet have average life of 38 months with a standard deviation of 5.8 months.



(b) $n=10$, $\alpha=0.01$, $H_0: \mu=10$ litres, $H_1: \mu\neq 10$ litres

$\bar{x}=\frac{9.7+2(9.8)+9.9+2(10.1)+10.2+2(10.3)+10.4}{10}=\frac{100.6}{10}=10.06$

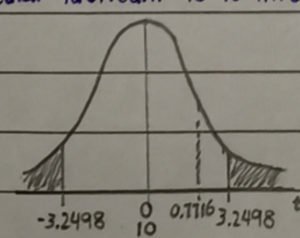
$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}=\sqrt{\frac{(9.7-10.06)^2+2(9.8-10.06)^2+(9.9-10.06)^2+2(10.1-10.06)^2+(10.2-10.06)^2+2(10.3-10.06)^2+(10.4-10.06)^2}{10-1}}=\sqrt{\frac{0.544}{9}}=0.2459$

Upper tail critical value, $t_{\alpha/2,n-1}=t_{0.005,9}=3.2498$

Lower tail critical value, $-t_{\alpha/2,n-1}=-t_{0.005,9}=-3.2498$

Test statistic, $t=\frac{\bar{x}-\mu}{s/\sqrt{n}}=\frac{10.06-10}{0.2459/\sqrt{10}}=0.7716$

$\therefore$ Since $t: 0.7716<$ critical value $: 3.2498$, we fail to reject null hypothesis H_0. There is sufficient evidence to support that the average content of container of a particular lubricant is 10 litres.



Question 7

(a) Confidence interval (49.8, 50.6) has larger sample size. From CI formula: $\bar{x}\pm t\left(\frac{s}{\sqrt{n}}\right)$, the interval would be smaller difference if $t\left(\frac{s}{\sqrt{n}}\right)$ is smaller. Since s is same, n needs to be large for $t\left(\frac{s}{\sqrt{n}}\right)$ to be small. A smaller $t\left(\frac{s}{\sqrt{n}}\right)$ produce a smaller difference in interval (49.8, 50.6): difference 0.8 compared to interval (50.3, 52.7): difference 2.4.

(b) Statistic II is recommended. Firstly, Statistic I is not chosen because it is biased, the distribution is not centered at true value. Both Statistic II and III are unbiased, but Statistic II has smaller standard deviation, Statistic III has larger standard deviation, so Statistic II is more precise.

(c) $n = 500 \quad \hat{p} = 15\% = 0.15 \qquad SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.15(0.85)}{500}} = 0.01597$

$1-\alpha = 90\% = 0.90 \qquad 90\%\ CI = \hat{p} \pm Z_{\alpha/2}(SE)$

$\alpha = 0.10,\ \frac{\alpha}{2} = 0.05 \qquad = 0.15 \pm 1.645(0.01597)$

$Z_{\alpha/2} = Z_{0.05} = 1.645 \qquad = 0.15 \pm 0.0263$

$90\% CI = (0.1237, 0.1763) = (12.37\%, 17.63\%)$

∴ We are 90% confident that the market share for the new brand is between 12.37% and 17.63%.

Question 8

(a) Let μ_1 = mean from category (18-34) age group

μ_2 = mean from category 35+ age group

$H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$, $\alpha = 0.05$

$n_1 = 150,\ n_2 = 250$ Test statistic, Z_0:

$\bar{x}_1 = 75.5,\ \bar{x}_2 = 65.2$

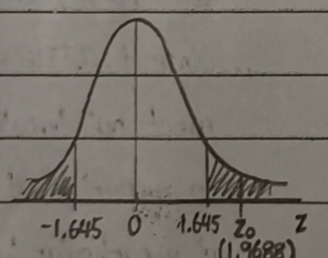
$\sigma_1 = 53.2,\ \sigma_2 = 46.1$

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{75.5 - 65.2}{\sqrt{\frac{53.2^2}{150} + \frac{46.1^2}{250}}} = 1.9688$$

Upper tail critical value, $Z_\alpha = Z_{0.05} = 1.645$

Lower tail critical value, $-Z_\alpha = -Z_{0.05} = -1.645$

∴ Since $Z_0 : 1.9688 >$ c.v. : 1.645, we reject null hypothesis $H_0: \mu_1 = \mu_2$ and conclude that there is difference in the mean spending between the two populations.



(b) $H_0: \mu_D = 3$, $H_1: \mu_D > 3$

$n = 10,\ df = n-1 = 9,\ \alpha = 0.05$, critical value = 1.833

1 hour after (x_1)	24 hours later (x_2)	$D = x_1 - x_2$	$D^2 = (x_1 - x_2)^2$
14	10	4	16
12	4	8	64
18	14	4	16
7	5	2	4
11	8	3	9
9	5	4	16
16	11	5	25
15	12	3	9
13	9	4	16
17	10	7	49
		$\Sigma D = 44$	$\Sigma D^2 = 224$

$\bar{D} = \frac{\Sigma D}{n} = \frac{44}{10} = 4.4$

$S_D = \sqrt{\frac{\Sigma D^2 - (\Sigma D)^2/n}{n-1}} = \sqrt{\frac{224 - 44^2/10}{9}} = \sqrt{\frac{30.4}{9}} = 1.8379$

Test value, $t = \frac{\bar{D} - \mu_0}{S_D/\sqrt{n}} = \frac{4.4 - 3}{1.8379/\sqrt{10}} = 2.4088$

∴ Since $t = 2.4088 >$ c.v. : 1.833, we reject the null hypothesis $H_0: \mu_0 = 3$. There is sufficient evidence to suggest that the mean number of words recalled after 1 hour exceeds mean recall after 24 hours by more than 3, that is $\mu_D > 3$.