Tutorial letter 202/1/2018

Applied Statistics II

STA2601

Semester 1

Department of Statistics

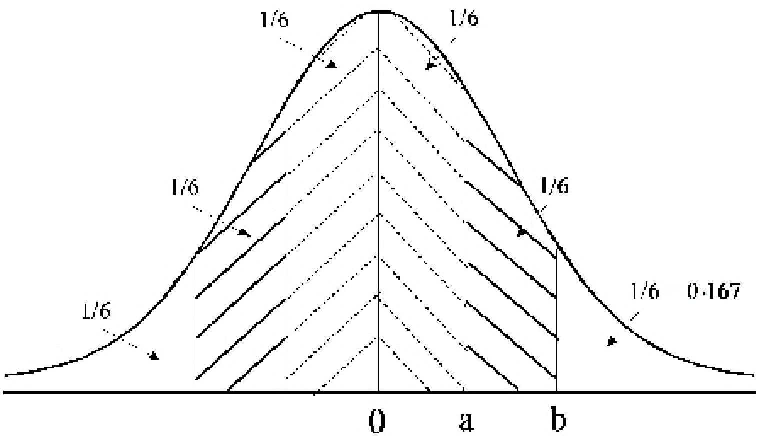
Solutions to Assignment 2

UNISA university of south africa

QUESTION 1

(a) Since μ and σ are unknown, we estimate them.

$$\begin{aligned}\overline{X} &= \frac{\sum_{i=1}^{n} X_i}{n} \\ &= \frac{2\,520}{42} \\ &= 60\end{aligned}$$



$$\begin{aligned}\sigma &= \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}} \\ &= \sqrt{\frac{27\,362}{42}} \\ &= \sqrt{651.4761905} \\ &\approx 25.524\end{aligned}$$

If we have to use six classes of equal intervals then the probability of each interval is $\pi_i = \frac{1}{6}$ for each interval $\Rightarrow n\pi_i = 7$.

The fifth interval is where $a \leq Z \leq b$.

From the sketch above the value "a" is found from table II as

$$\Phi(0.432) = P(Z \leq a) = 0.5 + 0.167 = 0.667.$$

Thus $a = 0.432.$

From the sketch above the value "b" is found from table II as

$$\Phi(0.966) = P(Z \leq b) = 0.5 + 0.333 = 0.833.$$

$$\text{Thus } a = 0.966.$$

Since $a \leq Z \leq b$. That is,

$$\begin{aligned}
0.432 &\leq Z \leq 0.966 \\
0.432 &\leq \frac{X-60}{25.524} \leq 0.966 \\
0.432 \times 25.524 &\leq X - 60 \leq 0.966 \times 25.524 \\
11.026368 &\leq X - 60 \leq 24.656184 \\
60 + 11.026368 &\leq X \leq 60 + 24.656184 \\
\Rightarrow 71.03 &\leq X \leq 84.66.
\end{aligned}$$

(7)

(b)

Equal probability intervals	Expected frequency	Count marks	Observed frequency
$-\infty < X \leq 35.34$	7	卌 II	7
$35.34 < X \leq 48.97$	7	卌 I	6
$48.97 < X \leq 60$	7	卌 IIII	9
$60 < X \leq 71.03$	7	卌 III	8
$71.03 < X \leq 84.66$	7	卌	5
$84.66 < X \leq \infty$	7	卌 II	7
Total	42		

(2)

(c) H_0 : The sample comes from a normal distribution.
H_1 : The sample does not come from a normal distribution.

$$\begin{aligned}
Y^2 &= \sum_{k=1}^{k} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \\
&= \frac{(7-7)^2}{7} + \frac{(6-7)^2}{7} + \frac{(9-7)^2}{7} + \frac{(8-7)^2}{7} + \frac{(5-7)^2}{7} + \frac{(7-7)^2}{7}
\end{aligned}$$

$$
\begin{aligned}
&= 0 + \frac{1}{7} + \frac{4}{7} + \frac{1}{7} + \frac{4}{7} \\
&= \frac{10}{7} \\
&\approx 1.4286
\end{aligned}
$$

From table IV (Stoker) $\chi^2_{\alpha;k-1} = \chi^2_{0.05;5} = 11.0705$ and $\chi^2_{\alpha;k-r-1} = \chi^2_{0.05;3} = 7.81473$. Reject H_0 if $Y^2 > 7.81473$.

Since $1.4286 < 7.18473$, we cannot reject H_0 at the 5% level of significance. We may conclude that the sample comes from a normal distribution.

FOR YOUR INFORMATION: For the degrees of freedom $k - r - 1$ where k is the number of classes and r is the number of estimated parameters.

Test	Value of r	Parameter unknown
$n\left(20, 5^2\right)$	$r = 0$	
$n\left(20, \sigma^2\right)$	$r = 1$	σ unknown
$n\left(\mu, 5^2\right)$	$r = 1$	μ unknown
$n\left(\mu, \sigma^2\right)$	$r = 2$	both μ and σ unknown

(6)

[15]

QUESTION 2

(a) H_0 : The lady has no discerning ability.

H_1 : The lady has discerning ability.

For this 2×2 table for the exact test is

		Lady says				
		Tea first	**Milk first**	**Total**		
Poured first	**Milk**	$5^* = x$	1	6	$\longleftarrow$	k
	Tea	1	5	6		
	Total	6 $\uparrow$ n	6	12	$\longrightarrow$	N

Now $k = 6$, $n = 6$ and $x = 5$

In this case

$$\begin{aligned} P(X \geq x) &= 1 - P(X < x - 1) \\ P(X \geq 5) &= 1 - P(X \leq 4) \\ &= 1 - 0.96 \\ &= 0.04 \end{aligned}$$

and $P(X \leq x) = P(X \leq 5) = 0.999$.

We can only reject H_0 in favour of the two-sided alternative if x is too large or too small and if it represents a "rare event", in other words only if

$$P(X \leq x) \leq \frac{\alpha}{2} \text{ or if } P(X \geq x) \leq \frac{\alpha}{2}$$

Since the test is two tailed we take the smaller of the two probabilities, i.e., we take 0.04. Since $0.04 > \frac{\alpha}{2} = 0.025$, we do not reject H_0 at the 5% level of significance and conclude that the lady has no discerning ability.

(10)

(b) (i) H_0 : The number of boys in a family with $n = 4$ children, has a binomial distribution with $p = \frac{1}{2}$.

H_1 : The data do not represent a binomial distribution.

If $X \sim b(n; p)$ then $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, ..., n.$

$$\begin{aligned} P(X = 0) &= \binom{4}{0} (0.5)^0 (0.5)^4 = 1 \times 0.0625 = 0.0625 = \pi_1 \\ P(X = 1) &= \binom{4}{1} (0.5)^1 (0.5)^3 = 4 \times 0.0625 = 0.25 = \pi_2 \\ P(X = 2) &= \binom{4}{2} (0.5)^2 (0.5)^2 = 6 \times 0.0625 = 0.375 = \pi_3 \\ P(X = 3) &= \binom{4}{3} (0.5)^3 (0.5)^1 = 4 \times 0.0625 = 0.25 = \pi_4 \\ P(X = 4) &= \binom{4}{4} (0.5)^4 (0.5)^0 = 1 \times 0.0625 = 0.0625 = \pi_5 \end{aligned}$$

The expected frequencies are $n\pi_i = 80 \times \pi_i$.

Number of boys in the family	Number of families	
	Observed N_i	Expected $(b\,(n;\,p))$ e_i
0	6	5
1	22	20
2	33	30
3	16	20
4	3	5
	80	80

$$
\begin{aligned}
\text{But } Y^2 &= \sum_{i=1}^{k} (N_i - n\pi_i)^2 / n\pi_i \\
&= \frac{(6-5)^2}{5} + \frac{(22-20)^2}{20} + \frac{(33-30)^2}{30} + \frac{(16-20)^2}{20} \\
&\quad + \frac{(3-5)^2}{5} \\
&= 0.2 + 0.2 + 0.3 + 0.8 + 0.8 \\
&= 2.3
\end{aligned}
$$

The critical value is $\chi^2_{\alpha;k-1} = \chi^2_{0.05;\,4} = 9.48773$. Reject H_0 if $Y^2 > 9.48773$

Since $2.3 < 9.48773$, we cannot reject H_0 at the 5% level of significance. It seems as if the binomial distribution gives a good fit.

(15)

(ii) If we are not given that $p = \frac{1}{2}$ we have to estimate p, say $\hat{p} = \theta$, from the observed data, by using the method of maximum likelihood.

Secondly we have to repeat the whole process of question (a(i)) with $\theta = \hat{p}$ and $\hat{\pi}_{r+1} = P(X = r) = \binom{4}{r} \theta^r (1-\theta)^{4-r}$ for $r = 0, 1, ..., 4$.

[If you are worried about the "$r + 1$" as subscript with $\hat{\pi}$ you have to keep in mind that we usually have k classes which start with the 1-st class. In other words, $\hat{\pi}_1$ is the expected probability for class 1 when there are $r = 0$ boys. Similarly, $\hat{\pi}_2$ is the expected probability for class 2 when there are $r = 1$ boys ..... $\hat{\pi}_5$ is the expected probability for class 5 when there are $r = 4$ boys.]

The expected frequencies are then $n\hat{\Pi}_i = 80 \times \hat{\Pi}_i$ for $i = 1, 2, ..., 5$.

$Y^2 = \sum_{i=1}^{5} \left(N_i - n\hat{\pi}_i\right)^2 / n\hat{\pi}_i$ is approximately a $\chi^2_{k-1-1} = \chi^2_3$ variable and the 5% critical value is $\chi^2_{0.05;\,3} = 7.81473$ (and not 9.48773 as in question (a(i))).
Reject $H_0 : X \sim b(4;\, p)$ with $\hat{p} = \theta$ if $Y \geq 7.81473$.

(5)

[30]

QUESTION 3

(a) $n = 10$ $\quad\sum X_i = 191$ $\quad\sum X_i^2 = 4\,956.2$

$\sum X_i Y_i = 5\,072.82$ $\quad\sum Y_i = 194.3$ $\quad\sum Y_i^2 = 5\,397.83$

(i)

$$
\begin{aligned}
R &= \frac{\Sigma X_i Y_i - \dfrac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\sqrt{\left(\Sigma X_i^2 - \dfrac{(\Sigma X_i)^2}{n}\right)\left(\Sigma Y_i^2 - \dfrac{(\Sigma Y_i)^2}{n}\right)}} \\
&= \frac{5\,072.82 - \dfrac{(191)\,(194.3)}{10}}{\sqrt{\left(4\,956.2 - \dfrac{(191)^2}{10}\right)\left(5\,397.83 - \dfrac{(194.3)^2}{10}\right)}} \\
&= \frac{5\,072.82 - 3\,711.13}{\sqrt{(4\,956.2 - 3\,648.1)\,(5\,397.83 - 3\,775.249)}} \\
&= \frac{1\,361.69}{\sqrt{(1\,308.1)\,(1\,622.581)}} \\
&= \frac{1\,361.69}{\sqrt{2\,122\,498.206}} \\
&= \frac{1\,361.69}{1\,456.879613} \\
&\approx 0.9347
\end{aligned}
$$

(5)

(ii) The 95% confidence for η is

$$U - \frac{1.96}{\sqrt{n-3}} < \eta < U + \frac{1.96}{\sqrt{n-3}}$$

where $U = \frac{1}{2}\log_e \frac{1+R}{1-R}$ and $\eta = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho}$

Now

$$\begin{aligned} U &= \frac{1}{2}\log_e \frac{1+R}{1-R} \\ &= \frac{1}{2}\log_e \frac{1+0.9347}{1-0.9347} \\ &= \frac{1}{2}\log_e \frac{1.9347}{0.0653} \\ &= \frac{1}{2}\log_e 29.62787136 \\ &= \frac{1}{2} \times 3.388715518 \\ &\approx 1.6944 \end{aligned}$$

Now

$$\begin{aligned} U - \frac{1.96}{\sqrt{n-3}} &< \eta < U + \frac{1.96}{\sqrt{n-3}} \\ 1.6944 - \frac{1.96}{\sqrt{10-3}} &< \eta < 1.6944 + \frac{1.96}{\sqrt{10-3}} \\ 1.6944 - \frac{1.96}{\sqrt{7}} &< \eta < 1.6944 + \frac{1.96}{\sqrt{7}} \\ 1.6944 - 0.7408 &< \eta < 1.6944 + 0.7408 \\ 0.9536 &< \eta < 2.4352 \end{aligned}$$

Now $\frac{e^{0.9536} - e^{-0.9536}}{e^{0.9536} + e^{-0.9536}} = \frac{2.5950 - 0.3854}{2.5950 + 0.3854} = \frac{2.2096}{2.9804} \approx 0.7414 \approx 0.74$

and $\frac{e^{2.4352} - e^{-2.4352}}{e^{2.4352} + e^{-2.4352}} = \frac{11.4181 - 0.0876}{11.4181 + 0.0876} = \frac{11.3305}{11.5057} \approx 0.9848 \approx 0.98$

i.e., 95% confidence interval for ρ is $(0.74; 0.98)$.

OR alternatively

Using Table X we have

for $\eta = 0.9505 : \rho = 0.74$ and $\eta = 0.9730 : \rho = 0.75$

Using linear interpolation for $\eta = 0.9536$

$$\begin{aligned}
\rho &= 0.74 + \frac{(0.9536 - 0.9505)}{(0.9730 - 0.9505)}(0.75 - 0.74) \\
&= 0.74 + \frac{0.0031}{0.0225} \times 0.01 \\
&= 0.74 + 0.001377777 \\
&= 0.741377777 \\
&\approx 0.74
\end{aligned}$$

for $\eta = 2.4101 : \rho = 0.984$ and $\eta = 2.4427 : \rho = 0.985$

Once more using linear interpolation for $\eta = 2.4352$

$$\begin{aligned}
\rho &= 0.984 + \frac{(2.4352 - 2.4101)}{(2.4427 - 2.4101)}(0.985 - 0.984) \\
&= 0.984 + \frac{0.0251}{0.0326} \times 0.001 \\
&= 0.984 + 0.000769938 \\
&= 0.984769938 \\
&\approx 0.98
\end{aligned}$$

Thus, the 95% confidence interval for ρ is $(0.734; 0.98)$. (6)

(iii) $H_0 : \rho = 0.9$ against $H_1 : \rho > 0.9$

$n = 10 \qquad R = 0.9347$

$$\begin{aligned}
U &= \frac{1}{2}\log_e \frac{1+R}{1-R} & \eta &= \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} \\
&= \frac{1}{2}\log_e \frac{1+0.9347}{1-0.9347} & &= \frac{1}{2}\log_e \frac{1+0.9}{1-0.9} \\
&= \frac{1}{2}\log_e \frac{1.9347}{0.0653} & &= \frac{1}{2}\log_e \frac{1.9}{0.1} \\
&= \frac{1}{2}\log_e 29.62787136 & &= \frac{1}{2}\log_e 19 \\
&\approx 1.6944 & &\approx 1.4722
\end{aligned}$$

Note: You can read the value of 0.9 from Table X Stoker.

The test statistic is

$$\begin{aligned} Z &= \sqrt{n-3}(U-\eta) \\ &= \sqrt{10-3}(1.6944-1.4722) \\ &= \sqrt{7}\times(0.2222) \\ &\approx 0.5879 \end{aligned}$$

$\alpha = 0.05$, and $Z_{0.05} = 1.645$. Reject H_0 if $Z > 1.645$.

Since $0.5879 < 1.645$, we do not reject H_0 at the 5% level of significance and conclude that $\rho = 0.9$.

(9)

(b) (i) $H_0 : \rho_1 = \rho_2$ against $H_1 : \rho_1 < \rho_2$

$r_1 = 0.5 \quad n_1 = 103$
$r_2 = 0.8 \quad n_2 = 52$

$$\begin{aligned} U_1 &= \frac{1}{2}\log_e \frac{1+r_1}{1-r_1} & U_2 &= \frac{1}{2}\log_e \frac{1+r_2}{1-r_2} \\ &= \frac{1}{2}\log_e \frac{1+0.5}{1-0.5} & &= \frac{1}{2}\log_e \frac{1+0.8}{1-0.8} \\ &= \frac{1}{2}\log_e \frac{1.5}{0.5} & &= \frac{1}{2}\log_e \frac{1.8}{0.2} \\ &= \frac{1}{2}\log_e 3 & &= \frac{1}{2}\log_e 9 \\ &\approx 0.5493 & &\approx 1.0986 \end{aligned}$$

(or just read the values for U_1 and U_2 from table X)

The test statistic is

$$\begin{aligned} z &= \frac{U_1 - U_2}{\sqrt{\dfrac{1}{n_1-3}+\dfrac{1}{n_2-3}}} \\ &= \frac{0.5493-1.0986}{\sqrt{\dfrac{1}{103-3}+\dfrac{1}{53-3}}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{-0.5493}{\sqrt{\frac{1}{100}+\frac{1}{50}}} \\
&= \frac{-0.5493}{\sqrt{0.03}} \\
&= \frac{-0.5493}{0.17320508} \\
&\approx -3.1714
\end{aligned}
$$

$\alpha = 0.05$ and $Z_{0.05} = 1.645$. We reject H_0 if $Z < -1.645$.

Since $-3.1714 < -1.645$, we reject H_0 at the 5% level of significance and conclude that $\rho_1 < \rho_2$, i.e., the correlation coefficient for population 1 is significantly smaller than that for population 2.

(8)

(ii) So

$$P\left[-1.645 \leq \frac{U_1 - U_2 - (\eta_1 - \eta_2)}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}} \leq 1.645\right] = 0.90$$

$$
\begin{aligned}
-1.645 &\leq \frac{U_1 - U_2 - (\eta_1 - \eta_2)}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}} \leq 1.645 \\
-1.645\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}} &\leq U_1 - U_2 - (\eta_1 - \eta_2) \leq 1.645\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}} \\
(U_1 - U_2) - 1.645\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}} &\leq \eta_1 - \eta_2 \leq (U_1 - U_2) + 1.645\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}} \\
(-0.5493) - 1.645\sqrt{0.03} &\leq \eta_1 - \eta_2 \leq (-0.5493) + 1.645\sqrt{0.03} \\
(-0.5493) - 0.284922357 &\leq \eta_1 - \eta_2 \leq (-0.5493) + 0.284922357 \\
-0.8342 &\leq \eta_1 - \eta_2 \leq -0.2644
\end{aligned}
$$

Therefore the 95% confidence interval for $\eta_1 - \eta_2$ is $(-0.8342; -0.2644)$.

Now $\frac{e^{-0.8342} - e^{0.8342}}{e^{-0.8342} + e^{0.8342}} = \frac{0.4342 - 2.3030}{0.4342 + 2.3030} = \frac{-1.8688}{2.7372} \approx -0.6827 \approx -0.68$

and $\frac{e^{-0.2644} - e^{0.2644}}{e^{-0.2644} + e^{0.2644}} = \frac{0.7677 - 1.3026}{0.7677 + 1.3026} = \frac{-0.5349}{2.0703} \approx -0.2584 \approx -0.26$

i.e., 95% confidence interval for ρ is $(-0.68; -0.26)$.

OR alternatively

Using Table X we have

for $\eta = 0.8291 : \rho = 0.68$ and $\eta = 0.8480 : \rho = 0.69$

Using linear interpolation for $\eta = 0.8342$

$$\begin{aligned}
\rho &= 0.68 + \frac{(0.8342 - 0.8291)}{(0.8480 - 0.8291)}(0.69 - 0.68) \\
&= 0.68 + \frac{0.0051}{0.0189} \times 0.01 \\
&= 0.68 + 0.002698412 \\
&= 0.682698412 \\
&\approx 0.68
\end{aligned}$$

for $\eta = 0.2554 : \rho = 0.25$ and $\eta = 0.2661 : \rho = 0.26$

Once more using linear interpolation for $\eta = 0.2644$

$$\begin{aligned}
\rho &= 0.25 + \frac{(0.2644 - 0.2554)}{(0.2661 - 0.2554)}(0.26 - 0.25) \\
&= 0.25 + \frac{0.009}{0.0107} \times 0.01 \\
&= 0.25 + 0.008411214 \\
&= 0.258411214 \\
&\approx 0.26
\end{aligned}$$

Thus, the 95% confidence interval for ρ is $(-0.68; -0.26)$. (10)

(iii) Yes. Since this upper bound (at the 90% level) will be the same as the 95% one-sided interval we may say we are 95% confident that $\rho_1 - \rho_2 \leq -0.26$. (This means we reject $H_0 : \rho_1 - \rho_2 = 0$ which confirms our conclusion.). (2)

[40]

QUESTION 4

(a) Let $U = \dfrac{\Sigma\,(X_i - \mu)^2}{\sigma^2}$ then $U \sim \chi^2_n$ (known value of μ).

$$
\begin{aligned}
1-\alpha &= P\left(\chi^2_{1-\frac{1}{2}\alpha;n} < U < \chi^2_{\frac{1}{2}\alpha;n}\right) \\
&= P\left[\chi^2_{1-\frac{1}{2}\alpha;n} < \frac{\Sigma\,(X_i-\mu)^2}{\sigma^2} < \chi^2_{\frac{1}{2}\alpha;n}\right] \\
&= P\left[\frac{1}{\chi^2_{1-\frac{1}{2}\alpha;n}} > \frac{\sigma^2}{\Sigma\,(X_i-\mu)^2} > \frac{1}{\chi^2_{\frac{1}{2}\alpha;n}}\right] \\
&= P\left[\frac{\Sigma\,(X_i-\mu)^2}{\chi^2_{\frac{1}{2}\alpha;n}} < \sigma^2 < \frac{\Sigma\,(X_i-\mu)^2}{\chi^2_{1-\frac{1}{2}\alpha;n}}\right]
\end{aligned}
$$

(6)

(b) (i)

$$
\begin{aligned}
\Sigma\,(X_i-\mu)^2 &= \Sigma x_i^2 - 2\mu\,\Sigma X_i + n\mu^2 \\
&= 380 - 2(4)(60) + 20(4)^2 \\
&= 380 - 480 + 320 \\
&= 220
\end{aligned}
$$

$$
\begin{aligned}
\chi^2_{\frac{1}{2}\alpha;n} &= \chi^2_{0.05;20} = 31.4104 \\
\chi^2_{1-\frac{1}{2}\alpha;n} &= \chi^2_{0.95;20} = 10.8508
\end{aligned}
$$

Thus, the 90% two-sided confidence interval for σ^2 now becomes

$$
\begin{aligned}
&\left[\frac{\Sigma\,(X_i-\mu)^2}{\chi^2_{\frac{1}{2}\alpha;n}} < \sigma^2 < \frac{\Sigma\,(X_i-\mu)^2}{\chi^2_{1-\frac{1}{2}\alpha;n}}\right] \\
&\left[\frac{220}{31.4104} < \sigma^2 < \frac{220}{10.8508}\right] \\
&\left[7.0040 < \sigma^2 < 20.2750\right] \\
&[7.004;\,20.275]\,.
\end{aligned}
$$

(7)

(ii) Since this interval includes 9, we cannot reject $H_0 : \sigma^2 = 9$ against $H_1 : \sigma^2 \neq 9$ at the 10% level of significance. (A two-sided confidence interval$\Longrightarrow$ two-sided hypothesis testing). (2)

[15]

[Total marks: 100]