Tutorial Letter 202/1/2016

Applied Statistics II

STA2601

Semester 1

Department of Statistics

Solutions to Assignment 02

BAR CODE

Learn without limits.

UNISA university of south africa

QUESTION 1

(a) We have to test

H_0 : The observations come from a normal distribution.

H_1 : The observations do not come from a normal distribution.

(3)

(b) $n = 30 \qquad \sum_{i=1}^{n} X_i = 41\,396,$

$$\widehat{\mu} = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} == \frac{41\,396}{30} \approx 1\,379.8667$$

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} \\ &= \frac{4\,622\,377.467}{30} \\ &= 154\,079.2489\end{aligned}$$

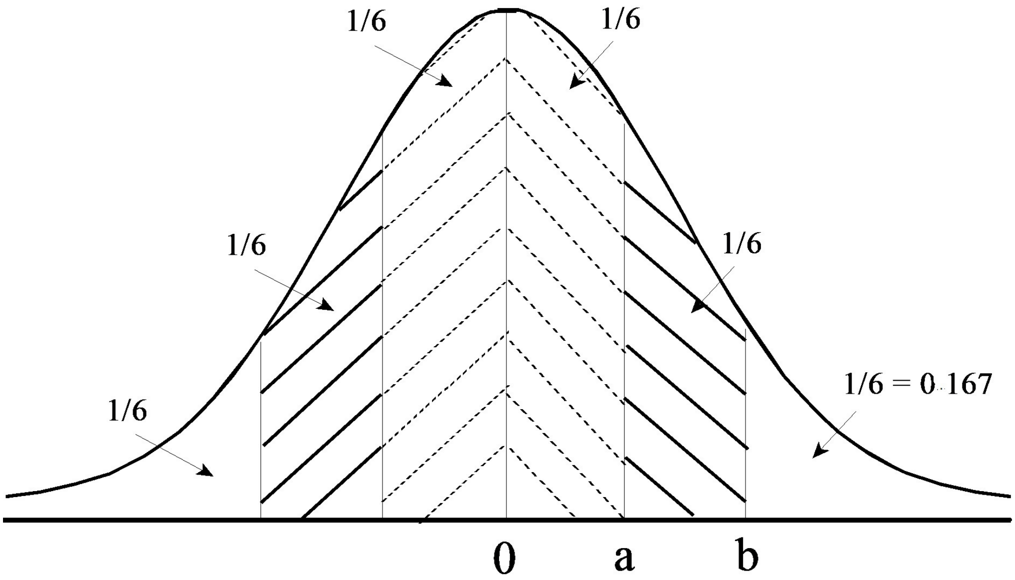
(5)

(c) If we divide the observations into 6 classes with equal expected frequencies, it means that $\pi_i = \frac{1}{6}$ for each interval $\Rightarrow n\pi_i = 5$.

The biggest problem is to *determine the interval limits* in terms of the X-scale such that each interval has a probability of $\frac{1}{6} = 0.167$.

We start with the standardised $n(0; 1)$ scale (as always) and transform back to the X-scale by making use of

$$Z = \frac{X - \widehat{\mu}}{\widehat{\sigma}} = \frac{X - 1\,379.8667}{\sqrt{154\,079.2489}}.$$



The 4th interval is where $0 \leq Z \leq a$

For the sketch above the value "a" is found from table II (Stoker) as

$$\begin{aligned} \Phi(0.432) &= P(Z \leq a) = 0.5 + 0.167 = 0.667. \\ \text{Thus } a &= 0.432. \end{aligned}$$

$$\begin{aligned} \text{i.e. } 0 \leq & \quad Z & \leq 0.432 \\ 0 \leq & \quad \frac{X - 1\,379.8667}{392.5293} & \leq 0.432 \\ 0 \leq & \quad X - 1\,379.8667 & 169.5726576 \\ 1\,379.8667 \leq & \quad X & \leq 1\,549.439358. \\ \Rightarrow 1\,379.87 \leq & \quad X & \leq 1\,549.44. \end{aligned}$$

(7)

(d) The normality assumption is not violated because from the JMP graphical output we see that the normal curve does fit the histogram very well. The box plot depicts an almost symmetric distribution. From the normal quantile plot we see no systematic deviation around the line. So we conclude from the graphical output that the sample comes from a normal distribution. and the points seem not to be deviating from the diagonal on the Normal Quantile Plot. The skewness is almost close to zero. (4)

(e) (i) **Test for skewness:**

H_0 : The distribution is normal $\left(\Rightarrow \beta_1 = 0\right)$.

H_1 : $\beta_1 \neq 0$.

(Please note: The alternative must be two-sided. There is no indication of a one-sided test.)

The critical value is 0.662. Reject H_0 if $\beta_1 < -0.662$ or $\beta_1 > 0.662$ or $|\beta_1| > 0.662$.

$$\text{Now } \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{30}\left(X_i - \overline{X}\right)^3}{\left[\frac{1}{n}\sum_{i=1}^{30}\left(X_i - \overline{X}\right)^2\right]^{\frac{3}{2}}} = \frac{\frac{1}{30}(146\,534\,638.8)}{\left[\frac{1}{30}(4\,622\,377.467)\right]^{\frac{3}{2}}}$$

$$= \frac{4\,884\,487.96}{(154\,079.2489)^{\frac{3}{2}}}$$

$$= \frac{4\,884\,487.96}{60\,480\,619,23}$$

$$\approx 0.0808.$$

Since $-0.662 < 0.0808 < 0.662$ we do not reject H_0 at the 10% level of significance level and conclude that the distribution is symmetric. (7)

(ii) **Test for kurtosis:**

We have to test:

H_0 : The distribution is normal $\left(\Rightarrow \beta_2 = 3\right)$.

H_1 : $\beta_2 \neq 3$.

Since we have a small sample, the test is based on A (page 113 in the study guide).

The size of the sample, $n = 30$, thus $n - 1 = 29$. Since 29 is between 25 and 30, we need to interpolate the critical values.

From table C (page 114 in the study guide):

The upper 5% percentage point for A is

$$0.8686 + \frac{(29-25)}{30-25}(0.8625 - 0.8686) = 0.8686 + \frac{4}{5}(-0.0061) = 0.8637.$$

The lower 5% percentage point for A is

$$0.7360 + \frac{(29-25)}{30-25}(0.7404 - 0.7360) = 0.7360 + \frac{4}{5}(0.0044) = 0.7395.$$

We reject H_0 at the 10% significance level if $A <$ lower 5% point or $A >$ upper 5% point in table C.

The critical values are 0.7395 and 0.8637. Reject H_0 if $A < 0.7395$ or $A > 0.8637$.

Now the value of the test statistic is

$$A = \frac{\frac{1}{n}\Sigma\left|X_i - \overline{X}\right|}{\sqrt{\frac{1}{n}\Sigma\left(X_i - \overline{X}\right)^2}} = \frac{\text{mean deviation}}{\text{standard deviation}}$$

$$= \frac{\frac{1}{30}(9\,771.3334)}{\sqrt{\frac{1}{30}(4\,622\,377.467)}}$$

$$= \frac{325.7111133}{392.5292969}$$

$$\approx 0.8298$$

Since $0.7395 < 0.8298 < 0.8637$, we do not reject H_0 at the 10% level of significance and conclude that the distribution does have the kurtosis of a normal distribution. (7)

(f) Yes, the distribution does originate from a normal distribution since it passed both tests. (2)

[35]

QUESTION 2

(a) H_0 : There is no relationship between treatment and improvement of arthritis.

H_1 : Arthritis improves when treated.

For this 2×2 table for the exact test is

	Knows about the product			
	Treated	**Untreated**	**Row total**	
Improved	6	2	8	
Not improved	$0(= x)$	4	4	$\leftarrow k$
Column total	6	6	12	
	$\uparrow$ n			

Now $k = 4$ (smallest marginal total)

$n = 6$ (treated patients)

$\therefore \; x = 0$ (cell corresponding to column 1, row 2)

The alternative hypothesis implies a small value of x. (There should be fewer "not improved" patients who were treated.)

Since the alternative hypothesis implies a small value of x, we reject H_0 if $P(X \leq 0) \leq \alpha$.

Now $x = 0$ and

$$\begin{aligned} P(X \leq x) &= P(X \leq 0) \\ &= 0.030 \text{ (from table D study guide p131)} \end{aligned}$$

Since $0.030 < 0.05$, we reject H_0 at the 5% level of significance and conclude that there is a significant positive relationship between treatment and improvement of arthritis.

(10)

(b) $n_X = 12 \qquad \Sigma X_i = 288 \qquad \Sigma\left(X_i - \overline{X}\right)^2 = 1\,114.62$

$n_Y = 10 \qquad \Sigma Y_i = 260 \qquad \Sigma\left(Y_i - \overline{Y}\right)^2 = 198.68$

(i) $H_0: \; \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 > \sigma_2^2$

$$\begin{aligned} n_X &= 12 & n_Y &= 10 \\ S_X^2 &= \frac{1}{n_X - 1}\Sigma\left(X_i - \overline{X}\right)^2 & S_Y^2 &= \frac{1}{n_Y - 1}\Sigma\left(Y_i - \overline{Y}\right)^2 \\ &= \frac{1}{12-1}(1\,114.62) & &= \frac{1}{10-1}(198.68) \\ &= \frac{1}{11}(1\,114.62) & &= \frac{1}{9}(198.68) \\ &\approx 101.3291 & &\approx 22.0756 \end{aligned}$$

The test statistic is

$$\begin{aligned} F &= \frac{\sigma_Y^2}{\sigma_X^2} \times \frac{S_X^2}{S_Y^2} \\ &= 1 \times \frac{101.3291}{22.0756} \\ &\approx 4.5901 \end{aligned}$$

The critical value is $F_{\alpha;n_x-1;n_y-1} = F_{0.05;11;9} = \frac{1}{2}(3.14+3.07) = 3.105$. Reject H_0 if $F > 3.105$.

Since $4.5901 > 3.105$, we reject H_0 at the 5% level of significance and conclude that the concentrations are **more variable** in the Kuruman area than in the Thabazimbi area.(10)

(ii) The test is based on the assumptions that:

- The samples are independent.
- Both samples are from normal populations.

(3)

(iii) The 95% confidence interval for $\frac{\sigma_Y^2}{\sigma_X^2}$ is

$$P\left(F_{1-\frac{\alpha}{2};n_1-1;n_2-1} < \frac{\sigma_Y^2}{\sigma_X^2}\frac{S_X^2}{S_Y^2} < F_{\frac{\alpha}{2};n_1-1;n_2-1}\right) = 1-\alpha$$

$$\left[\frac{F_{1-\frac{\alpha}{2};n_1-1;n_2-1}}{S_X^2/S_Y^2}; \frac{F_{\frac{\alpha}{2};n_1-1;n_2-1}}{S_X^2/S_Y^2}\right]$$

$\alpha = 0.05, \alpha/2 = 0.025$

$$F_{1-\frac{\alpha}{2};n_1-1;n_2-1} = F_{0.975;11;9} = \frac{1}{F_{0.025;9;11}} = \frac{1}{3.59} \approx 0.2786$$

$$F_{\frac{\alpha}{2};n_1-1;n_2-1} = F_{0.025;11;9} = \frac{1}{2}(3.96+3.87) = 3.915$$

$\therefore$The 95% confidence interval is

$$\left[\frac{F_{1-\frac{\alpha}{2};n_1-1;n_2-1}}{S_X^2/S_Y^2}; \frac{F_{\frac{\alpha}{2};n_1-1;n_2-1}}{S_X^2/S_Y^2}\right]$$
$$\left[\frac{0.2786}{101.3291/22.0756}; \frac{3.915}{101.3291/22.0756}\right]$$
$$\left[\frac{0.2554}{4.590094946}; \frac{3.915}{4.590094946}\right]$$
$$[0.0607; 0.8529].$$

(7)

[30]

QUESTION 3

(a) If $U = \dfrac{\Sigma\left(X_i - \overline{X}\right)^2}{\sigma^2}$ then $U \sim \chi^2_{n-1}$ (result 1.3).

Then

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

Thus the $100\,(1-\alpha)\,\%$ two-sided confidence interval for σ^2 is given by

$$\left[\frac{\Sigma\left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{1}{2}\alpha;n-1}};\quad \frac{\Sigma\left(X_i - \overline{X}\right)^2}{\chi^2_{1-\frac{1}{2}\alpha;n-1}}\right].$$

(5)

(b) $n = 20$ $\quad \Sigma X_i = 300$ $\quad \Sigma\left(X_i - \overline{X}\right)^2 = 778$ $\quad \alpha = 0.10, \alpha/2 = 0.05$

$\chi^2_{\frac{1}{2}\alpha;n-1} = \chi^2_{0.05;19} = 30.1435$ $\quad \chi^2_{1-\frac{1}{2}\alpha;n-1} = \chi^2_{0.95;19} = 10.117$

$\therefore$The 90% confidence interval is

$$\left[\frac{\Sigma\left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{1}{2}\alpha;n-1}};\quad \frac{\Sigma\left(X_i - \overline{X}\right)^2}{\chi^2_{1-\frac{1}{2}\alpha;n-1}}\right]$$

$$\left[\frac{778}{30.1435};\quad \frac{778}{10.117}\right]$$

$$[25.8099;\ 76.9003]\,.$$

(5)

(c) We have to test

$H_0: \quad \sigma^2 = 38$

$H_1: \quad \sigma^2 > 38$

The test statistic is

$$U = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{c} = \frac{778}{38} \approx 20.4737$$

The critical value is $\chi^2_{\alpha;n-1} = \chi^2_{0.1;\,19} = 27.2036$. Reject H_0 if $U > 27.2036$

Since $U = 20.4737 < 27.2036$, we can not reject H_0 at the 10% level of significance and conclude that $\sigma^2 = 38 \Longrightarrow$ **the variability is not greater than 38.**

(10)

[20]

QUESTION 4

(a) $n = 10$ $\quad \sum X_i = 37.1$ $\quad \sum X_i^2 = 151.91$

$\sum X_i Y_i = 172.56$ $\quad \sum Y_i = 45.1$ $\quad \sum Y_i^2 = 215.49$

$$R = \frac{\Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\sqrt{\left(\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right)\left(\Sigma Y_i^2 - \frac{(\Sigma Y_i)^2}{n}\right)}}$$

$$= \frac{172.56 - \frac{(37.1)(45.1)}{10}}{\sqrt{\left(151.91 - \frac{(37.1)^2}{10}\right)\left(215.49 - \frac{(45.1)^2}{10}\right)}}$$

$$= \frac{172.56 - 167.321}{\sqrt{(151.91 - 137.641)\,(215.49 - 203.401)}}$$

$$
\begin{aligned}
&= \frac{5.239}{\sqrt{(14.269)\,(12.089)}} \\
&= \frac{5.239}{\sqrt{172.497941}} \\
&= \frac{5.239}{13.13384715} \\
&\approx 0.3989
\end{aligned}
$$

(7)

(b) $H_0 : \rho = 0.65$ against $H_1 : \rho > 0.65$

$n = 10 \qquad R = 0.3989$

$$
\begin{aligned}
U &= \frac{1}{2}\log_e \frac{1+R}{1-R} & \eta &= \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} \\
&= \frac{1}{2}\log_e \frac{1+0.3989}{1-0.3989} & &= \frac{1}{2}\log_e \frac{1+0.65}{1-0.65} \\
&= \frac{1}{2}\log_e \frac{1.3989}{0.6011} & &= \frac{1}{2}\log_e \frac{1.65}{0.35} \\
&= \frac{1}{2}\log_e 2.327233405 & &= \frac{1}{2}\log_e 4.714285714 \\
&\approx 0.4223 & &\approx 0.7753
\end{aligned}
$$

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

The test statistic is

$$
\begin{aligned}
Z &= \sqrt{n-3}(U-\eta) \\
&= \sqrt{10-3}(0.4223 - 0.7753) \\
&= \sqrt{7} \times (-0.353) \\
&\approx -0.9340
\end{aligned}
$$

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

Since $-0.9340 < 1.645$, we do not reject H_0 at the 5% level of significance and conclude that $\rho = 0.65$.

(8)

[15]

[100]