Tutorial letter 201/2/2017

Applied Statistics II

STA2601

Semester 2

Department of Statistics

Solutions to Assignment 01

Define tomorrow.

UNISA
university of south africa

QUESTION 1

(a) (i) $f(1) = F(1) = \dfrac{1^3+k}{40} = \dfrac{1+k}{40}$

$$f(2) = F(2) - F(1) = \frac{2^3+k}{40} - \frac{1^3+k}{40} = \frac{8+k-1-k}{40} = \frac{7}{40}$$

$$f(3) = F(3) - F(2) = \frac{3^3+k}{40} - \frac{2^3+k}{40} = \frac{27+k-8-k}{40} = \frac{19}{40}$$

$$\begin{aligned}
\sum_{x \in A} f_X(x) &= 1 \\
f(1) + f(2) + f(3) &= 1 \\
\frac{1+k}{40} + \frac{7}{40} + \frac{19}{40} &= 1 \\
\frac{27+k}{40} &= 1 \\
27+k &= 40 \\
\Longrightarrow k &= 40-27 \\
\Longrightarrow k &= 13
\end{aligned}$$

(3)

(ii) The probability distribution of X is

X	1	2	3
$f(x)$	$\frac{7}{20}$	$\frac{7}{40}$	$\frac{19}{40}$

(4)

(iii) The mean value of X is calculated as:

$$\begin{aligned}
\mu &= E(x) = \sum_x x f(X=x) \\
&= 1\left(\frac{7}{20}\right) + 2\left(\frac{7}{40}\right) + 3\left(\frac{19}{40}\right) \\
&= \frac{7}{20} + \frac{7}{20} + \frac{57}{40} \\
&= 2\frac{1}{8} \\
&= 2.125
\end{aligned}$$

(3)

(iv) Now the *variance* of X is calculated as:

$$\begin{aligned}
\sigma^2 &= \sum_{x\in A}(x-\mu)^2\ f_X(x)\\
&= (1-2.125)^2\times\frac{7}{20}+(2-2.125)^2\times\frac{7}{40}+(3-2.125)^2\times\frac{19}{40}\\
&= (-1.125)^2\times\frac{7}{20}+(-0.125)^2\times\frac{7}{40}+(0.875)^2\times\frac{19}{40}\\
&= 0.44296875+0.002734375+0.363671875\\
&= 0.809375
\end{aligned}$$

OR

$$\begin{aligned}
\sigma^2 &= E\left(X^2\right)-(E(X))^2\\
&= \sum_x x^2\ f_X(x)-\mu^2\\
&= 1^2\left(\frac{7}{20}\right)+2^2\left(\frac{7}{40}\right)+3^2\left(\frac{19}{40}\right)-(2.125)^2\\
&= 0.35+0.7+4.275-4.515625\\
&= 5.325-4.515625\\
&= 0.809375
\end{aligned}$$

(4)

(v) The coefficient of skewness of X is calculated as $\beta_1=\frac{\mu_3}{\sigma^3}$. Now $\mu_3=$ the *third central moment* of X is calculated as:

$$\begin{aligned}
\mu_3 &= \sum_{x\in A}(x-\mu)^3\ P(X=x)\\
&= (1-2.125)^3\times\frac{7}{20}+(2-2.125)^3\times\frac{7}{40}+(3-2.125)^3\times\frac{19}{340}\\
&= (-1.125)^3\times\frac{7}{20}+(-0.125)^3\times\frac{7}{40}+(0.875)^3\times\frac{19}{40}\\
&= -0.498339843-0.000341796+0.31821289\\
&= -0.18046875
\end{aligned}$$

Now, Thus,

$$\begin{aligned}
\beta_1 &= \frac{\mu_3}{\sigma^3} \\
&= \frac{-0.18046875}{\left(\sqrt{(0.809375)}\right)^3} \\
&= \frac{-0.18046875}{0.728156412} \\
&\approx -0.2478
\end{aligned}$$

Data is negatively skewed$\Longrightarrow$ the distribution is not symmetrical.

(6)

(b) (i)

$$\begin{aligned}
E\left(T_1\right) &= E\left(\frac{4}{7}X_1 + \frac{3}{7}X_2\right) \\
&= \frac{4}{7}E\left(X_1\right) + \frac{3}{7}E\left(X_2\right) \\
&= \frac{4}{7}\mu + \frac{3}{7}\mu \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
E\left(T_2\right) &= E\left(\frac{2}{5}X_1 + \frac{3}{5}X_2\right) \\
&= \frac{2}{5}E\left(X_1\right) + \frac{3}{5}E\left(X_2\right) \\
&= \frac{2}{5}\mu + \frac{3}{5}\mu \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
E\left(T_3\right) &= E\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) \\
&= \frac{1}{4}E\left(X_1\right) + \frac{3}{4}E\left(X_2\right) \\
&= \frac{1}{4}\mu + \frac{3}{4}\mu \\
&= \mu
\end{aligned}$$

Thus T_1, T_2 and T_3 are unbiased estimators of μ. (7)

(ii) The most efficient of the two estimators will be the one with the smallest variance. Thus, the variance of each estimator should be computed. Do you recall that if X and Y are stochastically independent variables then $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$.

Now

$$\begin{aligned} Var(T_1) &= Var\left(\frac{4}{7}X_1 + \frac{3}{7}X_2\right) \\ &= \frac{16}{49}Var\,(X_1) + \frac{9}{49}Var\,(X_1) \\ &= \frac{16}{49}\sigma^2 + \frac{9}{49}\sigma^2 \\ &= \frac{25}{49}\sigma^2 \\ &= 0.5102\sigma^2 \end{aligned}$$

$$\begin{aligned} Var(T_2) &= Var\left(\frac{2}{5}X_1 + \frac{3}{5}X_2\right) \\ &= \frac{4}{25}Var\,(X_1) + \frac{9}{25}Var\,(X_1) \\ &= \frac{4}{25}\sigma^2 + \frac{9}{25}\sigma^2 \\ &= \frac{13}{25}\sigma^2 \\ &= 0.52\sigma^2 \end{aligned}$$

$$\begin{aligned} Var(T_3) &= Var\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) \\ &= \frac{1}{16}Var\,(X_1) + \frac{9}{16}Var\,(X_1) \\ &= \frac{1}{16}\sigma^2 + \frac{9}{16}\sigma^2 \\ &= \frac{5}{8}\sigma^2 \\ &= 0.625\sigma^2 \end{aligned}$$

$\therefore Var(T_1) < Var(T_2) < Var(T_3) \Longrightarrow T_1$ is a more efficient estimator of μ. In other words we would prefer T_1 because it has a smaller variance (which is a desirable property for an estimator since the $\sqrt{\text{variance}}$ is often called the "error of the estimation"). (8)

[35]

QUESTION 2

(a) True. The variance of any distribution is the sum of the squares of the deviation.

(b) False. There are some highly sophisticated method that can also be used and the method of least squares is used especially in problems where the unknown function are linear functions of known constants.

(c) False. A type I error is committed when H_0 is rejected when in actual fact it is true.

(d) False. There is a positive correlation between the length (in cm) and the mass (in kg) of a child.

(e) True. A correlation coefficient of zero, i.e., $r = 0$ indicates absence of a linear relationship.

(f) False. Correlation does not mean causation, i.e., correlation does not imply causality. Thus, one can not draw cause and effect conclusions based on correlation.

(g) False. The assumptions underlying a one-way analysis of variance are

- observations are independent (given).
- data comes from a normal population.
- equal population variances.

(h) True. The statistic T is called an unbiased estimator for the parameter θ if $E(T) = \theta$.

(i) False. A type II error is committed when we do not reject H_0 when in actual fact it is false.

(j) False. The larger the sample size, the larger the power.

[20]

QUESTION 3

(a) Yes. $f_{X_1}(x_1) = f_{X_3}(x_3)$ (3)

(b)

$$
\begin{aligned}
P(X_2 > 11) \text{ or } P(X_2 > 19) &= P\left(\frac{X_2-\mu}{\sigma} > \frac{11-\mu}{\sigma}\right) \text{ or } P\left(\frac{X_2-\mu}{\sigma} > \frac{11-\mu}{\sigma}\right) \\
&= P\left(Z > \frac{11-15}{3}\right) + P\left(Z > \frac{19-15}{3}\right) \\
&= P\left(Z > -\frac{4}{3}\right) + P\left(Z > \frac{4}{3}\right) \\
&= P(Z > -1.33) + P(Z > 1.33) \\
&= P(Z < 1.33) + P(Z > 1.33) \\
&= P(Z < 1.33) + 1 - P(Z < 1.33) \\
&= 1
\end{aligned}
$$

(5)

(c) From **property (iv) of result 1.1** it follows that if $n = 1$ then

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_1, \text{ and thus } \frac{(X_1-15)^2}{9} \text{does indeed have a } \chi^2_1 \text{ distribution.}$$

(2)

(d) $V_3 = \sum_{i=1}^{7} \frac{[X_i - 15]^2}{9}$ is defined as the sum of 7 independent squared $n(0;\ 1)$ variates. Using **result 1.2 in our study guide (page 29),** $V_3 \sim \chi^2_n \Longrightarrow V_3 \sim \chi^2_7$. Since $V_3 \sim \chi^2_7$, it follows from the properties of the chi-square distribution that $E(V_3) = 7$ using **result 1.1 in our study guide (page 28).**

(2)

(e) $Y = \sum_{i=1}^{9} \left[\frac{X_i - \overline{X}}{\sigma}\right]^2$ then Y is a $\chi^2_{n-1} \Longrightarrow$ a χ^2_8 variate and from the properties of the chi-square distribution in the study guide, it follows that $Var(Y) = 2d = 2 \times 8 = 16$. (2)

(f) Since $V_1 \sim \chi^2_5$ and $V_2 \sim \chi^2_7$, then $U = \dfrac{V_1/5}{V_2/7} \sim F_{5;\,7}$ using **definition 1.21**. Using **result 1.4**, if $U \sim F_{5;\,7}$, then $\dfrac{1}{U} \sim F_{7;\,5}$.

(4)

(g) The table gives $P(U > a) = \alpha$. $P(U < a) = 1 - P(U > a) = 1 - \alpha$. Since $U \sim F_{5;\,7} \Longrightarrow P(U < a) = 1 - P(U > a) = 1 - 0.05$and from Table V (Stoker) we find that $F_{0.05;5;7} = 3.97$ and $P(U > 3.97) = 0.05$. Hence $P(U < 3.97) = 0.95$ and thus $a = 3.97$

(2)

[20]

QUESTION 4

(a) $E(X_i) = \theta_1 + C_i\theta_2 \qquad i = 1, 2, \cdots, n$

$$\begin{aligned} Q(\theta_1, \theta_2) &= \sum_{i=1}^{n} (X_i - E(X_i))^2 \\ &= \sum_{i=1}^{n} (X_i - (\theta_1 + C_i\theta_2))^2 \\ &= \sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2)^2 \end{aligned}$$

$$\begin{aligned} \frac{dQ}{d\theta_1} &= 2\sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \times -1 \\ &= -2\sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \end{aligned}$$

Now $\dfrac{dQ}{d\theta_1} = 0$

$$\begin{aligned} \Longrightarrow \quad & 0 = -2\sum_{i=1}^{n} (X_i - \theta_1 - C_i\theta_2) \\ = \quad & -2\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \theta_1 - \theta_2 \sum_{i=1}^{n} C_i\right) \end{aligned}$$

Making θ_2 subject of the formula

$$\begin{aligned}\theta_2\sum_{i=1}^{n}C_i &= \sum_{i=1}^{n}X_i - n\theta_1 \\ \theta_2 &= \frac{\sum_{i=1}^{n}X_i - n\theta_1}{\sum_{i=1}^{n}C_i}\text{..........................(1)}\end{aligned}$$

Making θ_1 subject of the formula

$$\begin{aligned}n\theta_1 &= \sum_{i=1}^{n}X_i - \theta_2\sum_{i=1}^{n}C_i \\ \theta_1 &= \frac{\sum_{i=1}^{n}X_i - \theta_2\sum_{i=1}^{n}C_i}{n}\text{..........................(2)}\end{aligned}$$

Now

$$\begin{aligned}\frac{dQ}{d\theta_2} &= 2\sum_{i=1}^{n}(X_i-\theta_1-C_i\theta_2)\times -C_i \\ &= -2\sum_{i=1}^{n}\left(C_iX_i - C_i\theta_1 - C_i^2\theta_2\right) \\ &= -2\left(\sum_{i=1}^{n}C_iX_i - \theta_1\sum_{i=1}^{n}C_i - \theta_2\sum_{i=1}^{n}C_i^2\right)\end{aligned}$$

Then set $\frac{dQ}{d\theta_2}=0$

$$\begin{aligned}\implies\quad & 0 = -2\left(\sum_{i=1}^{n}C_iX_i - \theta_1\sum_{i=1}^{n}C_i - \theta_2\sum_{i=1}^{n}C_i^2\right) \\ = \quad & \sum_{i=1}^{n}C_iX_i - \theta_1\sum_{i=1}^{n}C_i - \theta_2\sum_{i=1}^{n}C_i^2\end{aligned}$$

Making θ_2 subject of the formula

$$\theta_2 \sum_{i=1}^{n} C_i^2 = \sum_{i=1}^{n} C_i X_i - \theta_1 \sum_{i=1}^{n} C_i$$

$$\theta_2 = \frac{\sum_{i=1}^{n} C_i X_i - \theta_1 \sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} C_i^2} \text{..........................(3)}$$

Making θ_1 subject of the formula

$$\theta_1 \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} C_i X_i - \theta_2 \sum_{i=1}^{n} C_i^2$$

$$\theta_1 = \frac{\sum_{i=1}^{n} C_i X_i - \theta_2 \sum_{i=1}^{n} C_i^2}{\sum_{i=1}^{n} C_i} \text{..........................(4)}$$

Finding θ_2 by equating equations 2 and 4.

$$\frac{\sum_{i=1}^{n} X_i - \theta_2 \sum_{i=1}^{n} C_i}{n} = \frac{\sum_{i=1}^{n} C_i X_i - \theta_2 \sum_{i=1}^{n} C_i^2}{\sum_{i=1}^{n} C_i}$$

$$\left(\sum_{i=1}^{n} C_i\right)\left(\sum_{i=1}^{n} X_i - \theta_2 \sum_{i=1}^{n} C_i\right) = n\left(\sum_{i=1}^{n} C_i X_i - \theta_2 \sum_{i=1}^{n} C_i^2\right)$$

$$\sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i - \theta_2 \left(\sum_{i=1}^{n} C_i\right)^2 = n \sum_{i=1}^{n} C_i X_i - n\theta_2 \sum_{i=1}^{n} C_i^2$$

$$
\begin{aligned}
n\theta_2 \sum_{i=1}^{n} C_i^2 - \theta_2 \left(\sum_{i=1}^{n} C_i \right)^2 &= n \sum_{i=1}^{n} C_i X_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i \\
\theta_2 \left(n \sum_{i=1}^{n} C_i^2 - \left(\sum_{i=1}^{n} C_i \right)^2 \right) &= n \sum_{i=1}^{n} C_i X_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i \\
\widehat{\theta}_2 &= \frac{n \sum_{i=1}^{n} C_i X_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i}{n \sum_{i=1}^{n} C_i^2 - \left(\sum_{i=1}^{n} C_i \right)^2} \\
\text{or } \widehat{\theta}_2 &= \frac{\sum_{i=1}^{n} C_i \sum_{i=1}^{n} X_i - n \sum_{i=1}^{n} C_i X_i}{\left(\sum_{i=1}^{n} C_i \right)^2 - n \sum_{i=1}^{n} C_i^2}
\end{aligned}
$$

Finding θ_1 by equating equations 1 and 3.

$$
\begin{aligned}
\frac{\sum_{i=1}^{n} X_i - n\theta_1}{\sum_{i=1}^{n} C_i} &= \frac{\sum_{i=1}^{n} C_i X_i - \theta_1 \sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} C_i^2} \\
\sum_{i=1}^{n} C_i^2 \left(\sum_{i=1}^{n} X_i - n\theta_1 \right) &= \sum_{i=1}^{n} C_i \left(\sum_{i=1}^{n} C_i X_i - \theta_1 \sum_{i=1}^{n} C_i \right) \\
\sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i - n\theta_1 \sum_{i=1}^{n} C_i^2 &= \sum_{i=1}^{n} C_i \sum_{i=1}^{n} C_i X_i - \theta_1 \left(\sum_{i=1}^{n} C_i \right)^2 \\
\theta_1 \left(\sum_{i=1}^{n} C_i \right)^2 - n\theta_1 \sum_{i=1}^{n} C_i^2 &= \sum_{i=1}^{n} C_i \sum_{i=1}^{n} C_i X_i - \sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i \\
\theta_1 \left(\left(\sum_{i=1}^{n} C_i \right)^2 - n \sum_{i=1}^{k} C_i^2 \right) &= \sum_{i=1}^{n} C_i \sum_{i=1}^{n} C_i X_i - \sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i \\
\widehat{\theta}_1 &= \frac{\sum_{i=1}^{n} C_i \sum_{i=1}^{n} C_i X_i - \sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i}{\left(\sum_{i=1}^{n} C_i \right)^2 - n \sum_{i=1}^{k} C_i^2}
\end{aligned}
$$

$$\text{or } \widehat{\theta}_1 = \frac{\sum_{i=1}^{n} C_i^2 \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} C_i X_i}{n \sum_{i=1}^{k} C_i^2 - \left(\sum_{i=1}^{n} C_i\right)^2} \tag{15}$$

(b) The probability density function is

$$f_X(x) = \pi^x (1-\pi)^{1-x} \quad \text{for } x = 0 \text{ or } x = 1.$$

The likelihood function

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \pi^{x_i} (1-\pi)^{1-x_i} \\ &= \pi^{\sum_{i=1}^{n} x_i} (1-\pi)^{\sum_{i=1}^{n}(1-x_i)} \\ &= \pi^{\sum_{i=1}^{n} x_i} (1-\pi)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

$$\begin{aligned} \therefore InL(\pi) &= \sum_{i=1}^{n} x_i In\pi + \left(n - \sum_{i=1}^{n} x_i\right) In(1-\pi) \\ \frac{dInL(\pi)}{d\pi} &= \frac{\sum_{i=1}^{n} x_i}{\pi} + \frac{\left(n - \sum_{i=1}^{n} x_i\right)(-1)}{(1-\pi)} \end{aligned}$$

Setting $\frac{dInL(\pi)}{d\pi} = 0$

$$\begin{aligned}
\Longrightarrow \qquad \frac{\sum_{i=1}^{n} x_i}{\pi} &= \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{(1-\pi)} \\
\therefore \qquad (1-\pi)\sum_{i=1}^{n} x_i &= \pi\left(n - \sum_{i=1}^{n} x_i\right) \\
\sum_{i=1}^{n} x_i - \pi\sum_{i=1}^{n} x_i &= n\pi - \pi\sum_{i=1}^{n} x_i \\
\therefore \sum_{i=1}^{n} x_i &= n\pi \\
\widehat{\pi} &= \frac{\sum_{i=1}^{n} x_i}{n} \\
\widehat{\pi} &= \overline{x}
\end{aligned}$$

(10)

[25]

[100]