Tutorial Letter 201/1/2016

Applied Statistics II

STA2601

Semester 1

Department of Statistics

Solutions to Assignment 1

BAR CODE

QUESTION 1

(a) The *mean* value of X is calculated as:

$$\begin{aligned}\mu &= E(x) = \sum_x x \; f_X(x) \\ &= 0\left(\frac{1}{10}\right) + 1\left(\frac{3}{10}\right) + 2\left(\frac{2}{5}\right) + 3\left(\frac{1}{5}\right) \\ &= 0 + \frac{3}{10} + \frac{4}{5} + \frac{3}{5} \\ &= 1.7\end{aligned}$$

(3)

(b) The *variance* of X is calculated as:

$$\begin{aligned}\sigma^2 &= \sum_{x\in A} (x-\mu)^2 \; f_X(x) \\ &= (0-1.7)^2\frac{1}{10} + (1-1.7)^2\frac{3}{10} + (2-1.7)^2\frac{2}{5} + (3-1.7)^2\frac{1}{5} \\ &= (-1.7)^2\frac{1}{10} + (-0.7)^2\frac{3}{10} + (0.3)^2\frac{2}{5} + (1.3)^2\frac{1}{5} \\ &= 0.289 + 0.147 + 0.036 + 0.338 \\ &= 0.81\end{aligned}$$

OR

$$\begin{aligned}\sigma^2 &= E\left(X^2\right) - (E(X))^2 \\ &= \sum_x x^2 \; f_X(x) - \mu^2 \\ &= \left(0^2 \times \frac{1}{10}\right) + \left(1^2 \times \frac{3}{10}\right) + \left(2^2 \times \frac{2}{5}\right) + \left(3^2 \times \frac{1}{5}\right) - 1.7^2 \\ &= 0 + \frac{3}{10} + \frac{8}{5} + \frac{9}{5} - 2.89 \\ &= 3.7 - 2.89 \\ &= 0.81\end{aligned}$$

(4)

(c) The *third central moment* of X is calculated as:

$$\begin{aligned}
\mu_3 &= \sum_{x\in A}(x-\mu)^3\ f_X(x) \\
&= (0-1.7)^3\frac{1}{10}+(1-1.7)^3\frac{3}{10}+(2-1.7)^3\frac{2}{5}+(3-1.7)^3\frac{1}{5} \\
&= (-1.7)^3\frac{1}{10}+(-0.7)^3\frac{3}{10}+(0.3)^3\frac{2}{5}+(1.3)^3\frac{1}{5} \\
&= -0.4913-0.1029+0.0108+0.4394 \\
&= -0.144
\end{aligned}$$

The coefficient of skewness of X is calculated as $\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{-0.144}{(\sqrt{0.81})^3} = -\frac{16}{81} \approx -0.1975$

No. The data is negatively skewed.

(8)

[15]

QUESTION 2

(a) Yes. $f_{X_5}(x_5) = f_{X_6}(x_6)$ (2)

(b)

$$\begin{aligned}
f_{X_6}(x_6) &= \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \\
&= \frac{1}{5\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_6-9}{25}\right)^2} \qquad -\infty < x_6 < \infty
\end{aligned}$$

(2)

(c)

$$\begin{aligned}
P(X_1<8)\text{ or }P(X_1>10) &= P(X_1<8)+P(X_1>10) \\
&= P\left(\frac{X_1-\mu}{\sigma}<\frac{8-9}{5}\right)+P\left(\frac{X_1-\mu}{\sigma}>\frac{10-9}{5}\right) \\
&= P\left(Z<-\frac{1}{5}\right)+P\left(Z>\frac{1}{5}\right)
\end{aligned}$$

$$\begin{aligned} P(X_1 < 8) \text{ or } P(X_1 > 10) &= P(Z < -0.2) + P(Z > 0.2) \\ &= 2P(Z > 0.2) \\ &= 2(1 - P(Z < 0.2)) \\ &= 2(1 - 0.5793) \\ &= 0.8414 \end{aligned}$$

(5)

(d) Yes, $Z = \dfrac{X_5 - 9}{5} \sim n(0;\ 1)$. If we subtract 9 from X_5 and divide by $\sqrt{25}$, we obtain the standard normal distribution. (3)

(e) $W = \sum_{i=1}^{7} \dfrac{(X_i - 9)^2}{25}$ is defined as the sum of 7 independent squared $n(0;\ 1)$ variates. Using **result 1.2 in our study guide (page 29),** $W \sim \chi_n^2 \Longrightarrow W \sim \chi_7^2$. Since $W \sim \chi_7^2$, it follows from the properties of the

chi-square distribution that $E(W) = 14$ using **result 1.1 in our study guide (page 28).** (2)

(f) If $T = \dfrac{\sqrt{7}\left(\overline{X} - \mu\right)}{\sqrt{\frac{\Sigma\left(X_i - \overline{X}\right)^2}{6}}} \sim t_6$ $P(T \geq 2.447) = 0.025$ from Table III (Stoker) with $\nu = 6$ and $p = 0.025$. (2)

(g) Since $V_1 \sim \chi_3^2$ and $V_2 \sim \chi_4^2$, then $U = \dfrac{V_1/3}{V_2/4} \sim F_{3;\ 4}$ using definition 1.21. (2)

(h) The table gives $P(U > a) = \alpha$. $P(U < a) = 1 - P(U > a) = 1 - \alpha$.

Since $U \sim F_{3;\ 4} \Longrightarrow P(U < a) = 1 - P(U > a) = 1 - 0.05$

and from Table 4 (Stoker) we find that $F_{0.05;3;4} = 6.59$ and $P(U > 6.59) = 0.05$.
Hence $P(U < 6.59) = 0.95$ and thus $a = 6.59$ (2)

[20]

QUESTION 3

(a)

$$\begin{aligned} A_1 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \\ &= \frac{\sigma^2}{n}\left[\sum_{i=1}^{n}\left(\frac{X_i-\overline{\mu}}{\sigma}\right)^2\right] \text{ (multiplying both sides by } \sigma^2) \\ &\Longrightarrow E(A_1) = \frac{\sigma^2}{n}E\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] \end{aligned}$$

From **result 1.2** in the study guide we know $\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2 \sim \chi^2_n$.

This also means (from **result 1.1)** that $E\left(\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2\right) = n$ and

$var\left(\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2\right) = 2n$

$$\begin{aligned} \Longrightarrow E\left(A_1\right) &= \frac{\sigma^2}{n}E\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] \\ &= \frac{\sigma^2}{n}\times n \\ &= \sigma^2 \end{aligned}$$

Now

$$\begin{aligned} A_2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 \\ &= \frac{\sigma^2}{n-1}\left[\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\right] \text{ (multiplying both sides by } \sigma^2) \\ &\Longrightarrow E(A_2) = \frac{\sigma^2}{n-1}E\left[\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\right] \end{aligned}$$

From **result 1.3** in the study guide we know $\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2 \sim \chi^2_{n-1}$.

This also means (from **result 1.1)** that $E\left(\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2\right) = n-1$ and

$$var\left(\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2\right)=2(n-1)$$

$$\begin{aligned}\implies E\left(A_2\right) &= \frac{\sigma^2}{n}E\left[\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\right]\\ &= \frac{\sigma^2}{n-1}\times n-1\\ &= \sigma^2\end{aligned}$$

Thus A_1 and A_2 are both unbiased estimators for σ^2. (10)

(b) The *most efficient* estimator is the one with the smallest variance (see definition 2.4).

$$\begin{aligned}\text{Now }\ var(A_1) &= \left[\frac{\sigma^2}{n}\right]^2\cdot var\left[\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2\right]\\ &= \frac{\sigma^4}{n^2}\times 2n\\ &= \frac{2\sigma^4}{n}\end{aligned}$$

$$\begin{aligned}\text{Now }\ var(A_2) &= \left[\frac{\sigma^2}{n-1}\right]^2\cdot var\left[\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2\right]\\ &= \frac{\sigma^4}{(n-1)^2}\times 2(n-1)\\ &= \frac{2\sigma^4}{(n-1)}\end{aligned}$$

It follows that

$var(A_1)\ < var(A_2)$ because $\dfrac{2\sigma^4}{n}<\dfrac{2\sigma^4}{n-1}$

$\implies A_1$ is the *most efficient* estimator.

(10)

[20]

QUESTION 4

$E(X_i) = c_i\theta_1 + w_i\theta_2$ for $i = 1, 2, 3 \cdots, n$

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

Differentiating with respect to θ_1 gives

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

Equating to zero gives

$$\begin{aligned} 0 &= -2\left(\sum_{i=1}^{n} \left(c_iX_i - c_i^2\theta_1 - c_iw_i\theta_2\right)\right) \\ &= \sum_{i=1}^{n} \left(c_iX_i - c_i^2\theta_1 - c_iw_i\theta_2\right) \\ &= \sum_{i=1}^{n} c_iX_i - \theta_1\sum_{i=1}^{n} c_i^2 - \theta_2\sum_{i=1}^{n} c_iw_i \end{aligned}$$

Making θ_1 subject of the formula gives

$$\begin{aligned} \theta_1\sum_{i=1}^{n} c_i^2 &= \sum_{i=1}^{n} c_iX_i - \theta_2\sum_{i=1}^{n} c_iw_i \\ \theta_1 &= \frac{\sum_{i=1}^{n} c_iX_i - \theta_2\sum_{i=1}^{n} c_iw_i}{\sum_{i=1}^{n} c_i^2} \ldots\ldots\ldots\ldots\ldots\ldots(1) \end{aligned}$$

Similary making θ_2 subject of the formula gives

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

Differentiating with respect to θ_2 gives

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

Equating to zero gives

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

Making θ_1 subject of the formula gives

$$\begin{aligned}\theta_1 \sum_{i=1}^{n} c_i w_i &= \sum_{i=1}^{n} w_i X_i - \theta_2 \sum_{i=1}^{n} w_i^2 \\ \theta_1 &= \frac{\sum_{i=1}^{n} w_i X_i - \theta_2 \sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} c_i w_i} \quad \text{....................(3)}\end{aligned}$$

Similary making θ_2 subject of the formula gives

$$\begin{aligned}
\theta_2 \sum_{i=1}^{n} w_i^2 &= \sum_{i=1}^{n} w_i X_i - \theta_1 \sum_{i=1}^{n} c_i w_i \\
\theta_2 &= \frac{\sum_{i=1}^{n} c_i X_i - \theta_1 \sum_{i=1}^{n} c_i w_i}{\sum_{i=1}^{n} w_i^2} \text{......................(4)}
\end{aligned}$$

Equating equations 1 and 3 gives

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

Similarly equating equations 2 and 4 gives

$$\begin{aligned}
\frac{\sum_{i=1}^{n} c_i X_i - \theta_1 \sum_{i=1}^{n} c_i^2}{\sum_{i=1}^{n} c_i w_i} &= \frac{\sum_{i=1}^{n} c_i X_i - \theta_1 \sum_{i=1}^{n} c_i w_i}{\sum_{i=1}^{n} w_i^2} \\
\sum_{i=1}^{n} w_i^2 \left(\sum_{i=1}^{n} c_i X_i - \theta_1 \sum_{i=1}^{n} c_i^2 \right) &= \sum_{i=1}^{n} c_i w_i \left(\sum_{i=1}^{n} c_i X_i - \theta_1 \sum_{i=1}^{n} c_i w_i \right)
\end{aligned}$$

$$\sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i X_i - \theta_1 \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i^2 = \sum_{i=1}^{n} c_i w_i \sum_{i=1}^{n} c_i X_i - \theta_1 \left(\sum_{i=1}^{n} c_i w_i \right)^2$$

$$\theta_1 \left(\sum_{i=1}^{n} c_i w_i \right)^2 - \theta_1 \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i^2 = \sum_{i=1}^{n} c_i w_i \sum_{i=1}^{n} c_i X_i - \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i X_i$$

$$\theta_1 \left(\left(\sum_{i=1}^{n} c_i w_i \right)^2 - \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i^2 \right) = \sum_{i=1}^{n} c_i w_i \sum_{i=1}^{n} c_i X_i - \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i X_i$$

$$\widehat{\theta}_1 = \frac{\sum_{i=1}^{n} c_i w_i \sum_{i=1}^{n} c_i X_i - \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i X_i}{\left(\sum_{i=1}^{n} c_i w_i \right)^2 - \sum_{i=1}^{n} w_i^2 \sum_{i=1}^{n} c_i^2}$$

[20]

QUESTION 5

(a)

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f_X(x_i;\ \theta) \\
&= \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} \\
&= \frac{1}{\theta} e^{-x_1/\theta} \times \frac{1}{\theta} e^{-x_2/\theta} \times \ldots \times \frac{1}{\theta} e^{-x_n/\theta} \\
&= \frac{1}{\theta^n} e^{-(x_1/\theta + x_2/\theta + \ldots + x_n/\theta)} \\
&= \theta^{-n} e^{-\sum x_i/\theta} \quad \text{(see definition 2.5)}
\end{aligned}$$

$$\Longrightarrow \ln L(\theta) = -n \ln \theta - \sum x_i \theta^{-1}$$

$$\Longrightarrow \frac{\partial \ln L(\theta)}{\partial \theta} = \frac{-n}{\theta} - \frac{\sum x_i}{\theta^2}(-1)$$

If we set $\frac{\partial \ln L(\theta)}{\partial \theta} = 0$ we get

$$\begin{aligned}
\frac{\sum x_i}{\theta^2} &= \frac{n}{\theta} \\
\Longrightarrow \widehat{\theta} &= \frac{\sum X_i}{n} \\
&= \overline{X}
\end{aligned}$$

Thus $\widehat{\theta} = \overline{X}$ (the maximum likelihood estimator(m.l.e.) of θ)

(6)

(b) To show that the m.l.e. is an unbiased estimator, we have to show that $E(\widehat{\theta}) = \theta$.

$E(\widehat{\theta}) = E\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum E(X_i) = \frac{1}{n}n\theta = \theta$ (q.e.d.). (4)

(c) $Var(\widehat{\theta}) = var\left[\frac{1}{n}\sum X_i\right] = \left(\frac{1}{n}\right)^2 \sum var(X_i) = \left(\frac{1}{n}\right)^2 \sum \theta^2 == \frac{1}{n_2}n\theta^2 = \frac{\theta^2}{n}$ (3)

[13]

QUESTION 6

(a) It follows from snapshot 1 that we have a data table with the name "Untitled 2". (2)

(b) From snapshot 1 it follows that we have activated the first column with the name "Column 1" and selected the Formula option under the menu of Column Properties. (2)

(c) From snapshot 2 it follows that we have selected the random function to create simulated data form a normal population. (2)

(d) From snapshot 2 it follows that we have edited the random function to create simulated data form a **normal population** with $\mu = 73$ and $\sigma^2 = 8$. (2)

(e) From snapshot 3 it follows that we have a data table with the name "**STA2601Sem1Ass12016Q6**". (2)

(f) From snapshot 3 it follows that we will have a first column with the name "normal$(73, 8)$" which will contain 20 rows if the OK button is pressed. (2)

[12]

[**100]**