Two hours

Statistical tables to be provided

THE UNIVERSITY OF MANCHESTER

INTRODUCTION TO STATISTICS (SOLUTIONS)

?? May 2018

??.?? – ??.??

Answer **ALL FOUR** questions in Section A (10 marks each) and **TWO** of the **THREE** questions in Section B (20 marks each). If more than **TWO** questions from Section B are attempted, then credit will be given for the best **TWO** answers.

Electronic calculators may be used, provided that they cannot store text.

SECTION A

Answer **ALL** four questions

A1.

(i) ν_k is the number of observations in bin B_k. The bin width is $h = a_{k+1} - a_k$ for all k. (2 marks; BOOKWORK)

(ii)

$$
\begin{aligned}
\hat{\mu}_{\text{Hist}} &= \int_{a_1}^{a_{K+1}} x\text{Hist}(x)dx = \sum_{k=1}^{K} \int_{a_k}^{a_{k+1}} x\text{Hist}(x)dx = \sum_{k=1}^{K} \int_{a_k}^{a_{k+1}} x\frac{\nu_k}{nh}dx \\
&= \sum_{k=1}^{K} \frac{\nu_k}{nh} \int_{a_k}^{a_{k+1}} xdx = \sum_{k=1}^{K} \frac{\nu_k}{2nh}[x^2]_{a_k}^{a_{k+1}} = \sum_{k=1}^{K} \frac{\nu_k}{2nh}(a_{k+1}^2 - a_k^2) \\
&= \sum_{k=1}^{K} \frac{\nu_k}{2nh}(a_{k+1} - a_k)(a_{k+1} + a_k) = \sum_{k=1}^{K} \frac{\nu_k}{2nh}h(a_{k+1} + a_k) \\
&= \frac{1}{2n} \sum_{k=1}^{K} \nu_k(a_{k+1} + a_k)
\end{aligned}
$$

(5 marks; BOOKWORK - Example Sheet question)

(iii) First note $n = \sum_{k=1}^{6} \nu_k = 1 + 11 + 39 + 38 + 10 + 1 = 100$. Hence

$$
\begin{aligned}
\hat{\mu}_{\text{Hist}} &= \frac{1}{2 \times 100}[1 \times (5 + 10) + 11 \times (10 + 15) + 39 \times (15 + 20) \\
&\qquad + 38 \times (20 + 25) + 10 \times (25 + 30) + 1 \times (30 + 35)] \\
&= 19.9
\end{aligned}
$$

(3 marks; BOOKWORK)

A2.

(i) $\hat{\theta}$ is unbiased if $\mathrm{E}(\hat{\theta}) = \theta$. (3 marks; BOOKWORK - done in lectures, Chapter 5)

(ii) $\mathrm{E}(\bar{X}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}\,X_i = \frac{1}{n}(n\mu) = \mu$ so $\bar{X}$ is unbiased for μ. By the same argument, $\bar{Y}$ is unbiased for μ also.

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\,\mathrm{Var}\left(\sum_i X_i\right) = \frac{1}{n^2}\sum_i \mathrm{Var}(X_i) = n\sigma^2/n^2 = \sigma^2/n\,.$$

Similarly, $\mathrm{Var}\,\bar{Y} = \tau^2/m$. (3 marks; BOOKWORK - this is essentially the proof of Thm 1.4. Fine to just write down expressions or quote theorem)

(iii) First note that

$$v(w) = \mathrm{Var}(w\bar{X} + (1-w)\bar{Y}) = w^2\sigma^2/n + (1-w)^2\tau^2/m \text{ as the } X_i \text{ and } Y_j \text{ are independent.}$$

To minimize $v(w)$, solve $dv/dw = 0$, i.e.

$$\begin{aligned} 0 = dv/dw &= 2w\sigma^2/n - 2(1-w)\tau^2/m \\ \implies w\left(\sigma^2/n + \tau^2/m\right) &= \tau^2/m \\ \implies w = \frac{\tau^2/m}{\sigma^2/n + \tau^2/m} &= \frac{n\tau^2}{m\sigma^2 + n\tau^2} \end{aligned}$$

(4 marks: if most of main ideas but couldn't quite do it exactly, 4 marks; if one key idea but not both, 2 marks.)

UNSEEN (but only just, as the method is identical to Ex Sheet 8, Q2).

[Total 10 marks]

A3.

(i) A Type I error occurs if we reject H_0 when it is true. A Type II error occurs if we do not reject H_0 when H_1 is true. The significance level, α, is the probability of rejecting H_0 when H_0 is true, i.e. the probability of a Type I error.

(3 marks. BOOKWORK. Award 2 marks if they mix up Type I and Type II but otherwise correct.)

Some data are obtained with $n = 8$, $\sum_{i=1}^n x_i = 113.6627$, and $\sum_{i=1}^n x_i^2 = 1621.391$. For the following questions, show your working by clearly stating the general formula for the appropriate test statistic and the critical values of the rejection region, as well as their numerical values in this specific case.

(ii) Here $\bar{X} = 113.6627/8 = 14.20784$. The test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{14.20784 - 13.5}{1/\sqrt{8}} = 2.002074$$

For a two-tailed test, we reject H_0 if $|Z| > z_{\alpha/2}$. Here, $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$. As $|Z| = 2 > 1.96 = z_{\alpha/2}$, we reject H_0.

(3 marks; BOOKWORK)

(iii) $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right) = \frac{1}{7}(1621.391 - 8 \times 14.20784^2) = 0.927$. The test statistic is

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} = 2.07941\,.$$

We reject H_0 if $|T| > t_{\alpha/2}$ where $t_{\alpha/2} = t_{0.025} = 2.3646$ is the upper $\alpha/2$ point of a t distribution on $n - 1 = 7$ degrees of freedom. Here $|T| < t_{0.025}$ and so H_0 is not rejected.

(4 marks; BOOKWORK)

[Total 10 marks]

A4.

(i) $I(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence interval if

$$\mathrm{P}[a(\mathbf{X}) \le \theta \le b(\mathbf{X})] = 1-\alpha$$

(2 marks; BOOKWORK - Chapter 7)

(ii)

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}(\bar{X}_1 - \bar{X}_2) = \mathrm{E}(\bar{X}_1) - \mathrm{E}(\bar{X}_2) = \mu_1 - \mu_2$$

$$\mathrm{Var}\,\hat{\theta} = \mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \mathrm{Var}\,\bar{X}_1 + \mathrm{Var}\,\bar{X}_2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$$

$$\hat{\theta} \sim N\left[\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right]$$

(4 marks; BOOKWORK - Chapter 4/5/7.)

(iii) Note that $\bar{x}_1 = 96.08/10 = 9.608$ and $\bar{x}_2 = 237.09/20 = 11.8545$. The 95% CI has end-points

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\theta})} = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$
$$= (9.608 - 11.8545) \pm 1.96 \times \sqrt{\frac{2}{10} + \frac{4}{20}}$$

the interval is $(-3.486, -1.007)$. As the interval does not contain zero, it is not plausible that $\mu_1 = \mu_2$ i.e. $\mu_1 - \mu_2 = 0$.

(4 marks; BOOKWORK, Chapter 7)

[Total 10 marks]

SECTION B

Answer **TWO** of the three questions

B5.

(i)

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i-\bar{X})^2 &= \sum_{i=1}^{n}((X_i-\mu)-(\bar{X}-\mu))^2 \\
&= \sum_{i=1}^{n}\left\{(X_i-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2\right\} \\
&= \sum_{i=1}^{n}(X_i-\mu)^2-2\sum_{i=1}^{n}(X_i-\mu)(\bar{X}-\mu)+\sum_{i=1}^{n}(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2-2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)+\sum_{i=1}^{n}(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2-2(\bar{X}-\mu)(n\bar{X}-n\mu)+n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2-2n(\bar{X}-\mu)^2+n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2-n(\bar{X}-\mu)^2\,.
\end{aligned}
$$

(6 marks: BOOKWORK, Chapter 4)

(ii) By the previous part

$$
\begin{aligned}
\mathrm{E}\ \mathrm{SSD}\ &= \mathrm{E}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = \mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2-n(\bar{X}-\mu)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2\right)-n\,\mathrm{E}\left((\bar{X}-\mu)^2\right) \\
&= \sum_{i=1}^{n}\mathrm{E}(X_i-\mu)^2-n\,\mathrm{E}\left((\bar{X}-\mu)^2\right) \\
&= n\,\mathrm{Var}\,X_i-n\,\mathrm{Var}(\bar{X}) = n\sigma^2-n(\sigma^2/n)=(n-1)\sigma^2
\end{aligned}
$$

Hence $\mathrm{E}\,S^2=\mathrm{E}[\mathrm{SSD}/(n-1)]=\mathrm{E}(\mathrm{SSD})/(n-1)=(n-1)\sigma^2/(n-1)=\sigma^2$.

(4 marks; BOOKWORK Chapter 4)

(iii)

$$
\begin{aligned}
&\mathrm{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right) = \mathrm{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}} < \sigma^2 \text{ AND } \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right)\\
&= \mathrm{P}\left(\frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\frac{\alpha}{2}} \text{ AND } \chi^2_{n-1;1-\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma^2}\right)\\
&= \mathrm{P}\left(\chi^2_{n-1;1-\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\frac{\alpha}{2}}\right)\\
&= \mathrm{P}\left(\chi^2_{n-1;1-\frac{\alpha}{2}} < \chi^2(n-1) < \chi^2_{n-1;\frac{\alpha}{2}}\right)\\
&= (1-\alpha/2) - \alpha/2 = 1-\alpha
\end{aligned}
$$

hence the interval is a $100(1-\alpha)\%$ confidence interval for σ^2.

(6 marks. Essentially BOOKWORK, though a little backwards compared to the description in Chapter 7)

(iv) For a 95% CI, $\alpha = 0.05$, hence

$$\chi^2_{n-1;\frac{\alpha}{2}} = \chi^2_{9;0.025} = 19.023\,, \qquad \chi^2_{n-1;1-\frac{\alpha}{2}} = \chi^2_{9;0.975} = 2.700$$

Here $\bar{x} = 104.334/10 = 10.4334$, and

$$s^2 = \frac{1}{9}\left(\sum_{i=1}^{10} x_i^2 - 10 \times \bar{x}^2\right) = \frac{1}{9}(1132.207 - 10 \times 10.4334^2) = 4.849849\,.$$

Hence the 95% CI is

$$\left(\frac{9 \times 4.849849}{19.023}, \frac{9 \times 4.849849}{2.700}\right) = (2.295, 16.166)\,.$$

(4 marks; BOOKWORK from Chapter 7, plugging numbers into given formula).

[Total 20 marks]

B6.

(i)

$$L(p) = \prod_{i=1}^{n} f(X_i) = \prod_{i=1}^{n} (1-p)^{X_i} p = (1-p)^{\sum_{i=1}^n X_i} p^{\sum_{i=1}^n 1} = (1-p)^{\sum_{i=1}^n X_i} p^n$$

(2 marks, BOOKWORK)

(ii) The log-likelihood is

$$\ell(p) = \left(\sum_{i=1}^{n} X_i \right) \log(1-p) + n \log p$$

Differentiating,

$$\frac{d\ell}{dp} = -\frac{\sum_{i=1}^n X_i}{1-p} + \frac{n}{p}$$

If $\sum_{i=1}^n X_i > 0$ then we can solve $\left.\frac{d\ell}{dp}\right|_{\hat{p}} = 0$ to give

$$\begin{aligned} 0 &= -\frac{\sum_{i=1}^n X_i}{1-\hat{p}} + \frac{n}{\hat{p}} \\ &\implies \hat{p} \sum_{i=1}^{n} X_i = n(1-\hat{p}) \implies \hat{p}(n + \sum_{i=1}^{n} X_i) = n \\ &\implies \hat{p} = \frac{n}{n + \sum_{i=1}^n X_i} = \frac{1}{1+\bar{X}} \end{aligned}$$

To check it is a maximum, we consider the second derivative at $\hat{p}$,

$$\left.\frac{d^2\ell}{dp}\right|_{\hat{p}} = -\frac{\sum_{i=1}^n X_i}{(1-\hat{p})^2} - \frac{n}{\hat{p}^2} < 0\,.$$

When $\sum_{i=1}^n X_i = 0$, we have that $\bar{X} = 0$ and $L(p) = p^n$ which is monotonic increasing in $p \in [0,1]$ and so is $L(p)$ maximized when $\hat{p} = 1 = 1/(1+\bar{X})$. Thus $\hat{p} = 1/(1+\bar{X})$ whether $\sum_{i=1}^n X_i = 0$ or $\sum_{i=1}^n X_i > 0$.

(8 marks. UNSEEN example, but method is bookwork from Chapter 7).

(iii)

$$\begin{aligned} &\mathrm{P}(a < \hat{p} < b) = \mathrm{P}\left(a < \frac{1}{1+\bar{X}} \text{ and } \frac{1}{1+\bar{X}} < b\right) = \mathrm{P}\left(a\bar{X} + a < 1 \text{ and } 1 < b + b\bar{X}\right) \\ &= \mathrm{P}\left(\bar{X} < \frac{1-a}{a} \text{ and } \frac{1-b}{b} < \bar{X}\right) \end{aligned}$$

(5 marks: UNSEEN, similar to a previous exam)

(iv) Note that $E(\bar{X}) = \frac{1-p}{p} = 0.75/0.25 = 3$ and $\mathrm{Var}(\bar{X}) = \mathrm{Var}(X_1)/n = (0.75)/(0.25^2 \times 100) = 0.12$. Using the central limit theorem, as $n = 100$ is large, $\bar{X} \sim N(3, 0.12)$ approximately.

$$\mathrm{P}(0.22 < \hat{p} < 0.26) = \mathrm{P}\left(\frac{1-0.26}{0.26} < \bar{X} < \frac{1-0.22}{0.22}\right) \quad \text{using the result given in part (iii)}$$

$$= \mathrm{P}\left(2.846154 < \bar{X} < 3.545455\right) = \mathrm{P}\left(\frac{2.846154-3}{\sqrt{0.12}} < \frac{\bar{X}-3}{\sqrt{0.12}} < \frac{3.545455-3}{\sqrt{0.12}}\right)$$

$$\approx \mathrm{P}(-0.4441 < Z < 1.5746) \approx \Phi(1.57) - \Phi(-0.44)$$

$$= \Phi(1.57) - (1 - \Phi(0.44)) = 0.9418 - 1 + 0.6700 = 0.6118$$

(5 marks. UNSEEN, but once you have spotted how to use the trick from the previous part, this is a simple application of a standard technique from Chapter 3)

[Total 20 marks]

B7.

(i)

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

If H_0 is true and n is large, then $\hat{p} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$ approximately.

(5 marks; easy bookwork).

(ii)

$$\begin{aligned}
&\mathrm{P}(\text{reject } H_0) = \mathrm{P}\left(\hat{p} > \gamma\right) = \mathrm{P}\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} > \frac{\gamma-p}{\sqrt{p(1-p)/n}}\right) \\
&\approx \mathrm{P}\left(Z > \frac{\gamma-p}{\sqrt{p(1-p)/n}}\right), \quad Z \sim N(0,1) \\
&\qquad \text{since } \hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \text{ approx for large } n \text{ by central limit theorem} \\
&= 1 - \Phi\left(\frac{\gamma-p}{\sqrt{p(1-p)/n}}\right) = 1 - \Phi\left(\frac{\sqrt{n}(\gamma-p)}{\sqrt{p(1-p)}}\right).
\end{aligned}$$

(5 marks; UNSEEN extension of idea from Example Sheet 11, reasonably similar to a past exam question, reasonably simple combination of ideas from Chapters 9 and 4. However I am sure some students will find it tricky)

(iii) First note $z_\alpha = 1.6449$, and so

$$\begin{aligned}
&a = 1 + 1.6449^2/200 = 1.013528 \\
&b = -(2 \times 0.3 + 1.6449^2/200) = -0.6135285 \\
&c = 0.09 \\
&\gamma = \frac{0.6135285 + \sqrt{0.6135285^2 - 4 \times 1.013528 \times 0.09}}{2 \times 1.013528} = 0.3556814 \\
&\mathrm{P}(\text{reject } H_0) \approx 1 - \Phi\left(\frac{\sqrt{n}(\gamma-p)}{\sqrt{p(1-p)}}\right) = 1 - \Phi\left(\frac{\sqrt{200}(0.3556814 - 0.35)}{\sqrt{0.35 \times 0.65}}\right) \\
&= 1 - \Phi(0.1684) \approx 1 - \Phi(0.17) = 0.4325.
\end{aligned}$$

(5 marks; UNSEEN - plugging numbers into (somewhat complex) formulae, but needs skills with z_α values and Φ values as practised in Chapters 3, 7 and 9/10.)

(iv) Using the hint,

$$
\begin{aligned}
Z_2^2 = z_\alpha^2 &\iff \frac{(\hat{p}-p_0)^2}{\hat{p}(1-\hat{p})/n} = z_\alpha^2 \\
&\iff (\hat{p}-p_0)^2 = z_\alpha^2 \hat{p}(1-\hat{p})/n \\
&\iff \hat{p}^2 - 2p_0\hat{p} + p_0^2 = \frac{z_\alpha^2}{n}\hat{p} - \frac{z_\alpha^2}{n}\hat{p}^2 \\
&\iff \left(1 + \frac{z_\alpha^2}{n}\right)\hat{p}^2 - \left(2p_0 + \frac{z_\alpha^2}{n}\right)\hat{p} + p_0^2 = 0 \\
&\iff a\hat{p}^2 + b\hat{p} + c = 0 \\
&\iff z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\end{aligned}
$$

As Z_2 is a strictly increasing function of $\hat{p}$, the rejection region $Z_2 > z_\alpha$ is of the form $\hat{p} > \gamma$ with γ the value of $\hat{p}$ satisfying $Z_2 = z_\alpha$.

The two roots of $Z_2^2 = z_\alpha^2$ correspond to $Z_2 = \pm z_\alpha$. Since Z_2 is a strictly increasing function of $\hat{p}$, the larger of the two roots of $Z_2^2 = z_\alpha^2$ corresponds to $Z_2 = z_\alpha$, and the smaller root corresponds to $Z_2 = -z_\alpha$. Hence γ is the larger of the two roots, i.e. $(-b + \sqrt{b^2 - 4ac})/(2a)$.

(5 marks, UNSEEN)

[5 marks]

[Total 20 marks]

END OF EXAMINATION PAPER