THE UNIVERSITY OF MELBOURNE

Student Number

Semester 2 Final Exam, 2018

School of Mathematics and Statistics

MAST90058 Elements of Statistics

Writing time: 3 hours

Reading time: 15 minutes

This is NOT an open book exam

Common content with: MAST20005 Statistics

This paper consists of 6 pages (including this page)

Authorised Materials

- Mobile phones, smart watches and internet or communication devices are forbidden.
- Only Casio FX82 (any suffix) calculators may be used.
- Students may bring one double-sided A4 sheet of handwritten notes.

Instructions to Students

- You must NOT remove this question paper at the conclusion of the examination.
- Some useful R output is given in an appendix on the last page.
- You should attempt all questions. Show full working for each of your answers.
- There are 8 questions with marks as shown. The total number of marks available is 90.

Instructions to Invigilators

- Students must NOT remove this question paper at the conclusion of the examination.
- All graphics or CAS calculators should be confiscated.
- Students may use one double-sided A4 sheet of handwritten notes.

Blank page (ignored in page numbering)

Question 1 (10 marks) You have random samples from two groups, $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$. Some R output from analysing these data are below. The variable x contains the observations of X and the variable y contains the observations of Y.

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.620   3.320   4.040   4.193   4.515   6.300
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  4.150   4.450   5.355   5.212   5.595   6.720
> sd(x)
[1] 1.200121
> sd(y)
[1] 0.8368034
> sort(y)
[1] 4.15 4.25 4.34 4.78 5.30 5.41 5.46 5.64 6.07 6.72
```

(a) For each of the following quantities, state or calculate its value if possible, or otherwise explain why it is not possible.

(i) $x_{(1)}$

(ii) $y_{(4.5)}$

(iii) $\bar{x}$

(iv) σ_2

(b) For each of the following statements, state whether they are true, false or if it is not possible to know from the given information. In each case state the values of the quantities, if possible.

(i) $x_{(1)} > y_{(1)}$

(ii) $\bar{x} > \bar{y}$

(iii) $\sigma_1 > \sigma_2$

(c) For each of the following pairs of hypotheses, carry out the test if it is possible, using a 5% significance level, or otherwise explain what further information you need in order to do it.

(i) $H_0 \colon \mu_1 = \mu_2$ versus $H_1 \colon \mu_1 \neq \mu_2$

(ii) $H_0 \colon \sigma_2 = 2$ versus $H_1 \colon \sigma_2 \neq 2$

Single variable

Question 2 (9 marks) A random sample on X produced the following observations:

5.5 5.8 6.0 6.6 6.8 6.9 7.1 7.3 7.5 8.7

For these data, we have $\bar{x} = 6.82$ and $s = 0.932$.

(a) Let $\mu = \mathbb{E}(X)$. Calculate a 95% confidence interval for μ, assuming that X is normally distributed.

(b) Let $p = \Pr(X > 6.5)$. Calculate a 95% confidence interval for p.

(c) Let m be the median of X. Calculate a distribution-free confidence interval for m, with an approximate confidence level of 90%.

Question 3 (11 marks) Consider a random sample of size n on X which has a geometric distribution with parameter θ. Its pmf is:

$$p_X(x) = \theta(1-\theta)^x, \quad x \in \{0, 1, 2, \dots\}$$

and it has mean $(1-\theta)/\theta$.

(a) Determine a sufficient statistic for θ.

(b) Find the method of moments estimator of θ.

(c) Find the maximum likelihood estimator (MLE) of θ.

(d) Find the Cramér–Rao lower bound for unbiased estimators of θ.

(e) Derive an expression for the standard error of the MLE.

(f) A random sample of size $n = 20$ produced the following observations:

3 2 0 1 1 3 0 0 0 0 0 3 0 3 1 1 0 2 2 0

Estimate θ and calculate an approximate 90% confidence interval.

Question 4 (12 marks) Laleh has bought a new toaster. It has a dial that allows her to set the 'strength' of toasting. She is not sure if it works very well and decides to run some experiments. She sets the dial to various values, x, and measures how long the toaster cooks the bread, Y, before it pops the bread out. She does a simple linear regression analysis of these data using the model $\mathbb{E}(Y \mid x) = \alpha + \beta x$. Some partial R output from her analysis is shown below.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.7323     1.8852
x             1.1556     0.3818
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.686 on 28 degrees of freedom
Multiple R-squared:  0.2465,     Adjusted R-squared:  0.2196
F-statistic: 9.159 on 1 and 28 DF,  p-value: 0.005262
```

(a) How many experiments did Laleh carry out?

(b) Carry out the following hypothesis tests, using a 5% significance level.

  (i) $H_0\colon \alpha = 0$ versus $H_1\colon \alpha \neq 0$

  (ii) $H_0\colon \beta = 0$ versus $H_1\colon \beta \neq 0$

(c) Is there evidence that the dial is having an effect on the toasting time?

(d) Write out the ANOVA table for this regression model fit.
(*Hint:* the F statistic is shown in the above R output.)

Question 5 (13 marks) On his way to work in the mornings, Damjan records how long he has to wait for his train at the station. Over a series of days, he observes the following times (in minutes):

4.8 1.2 3.7 0.9 0.7 0.3 0.9 3.2 1.4 ms

For these data we have $\bar{x} = 1.9$ and $s = 1.58$. Damjan decides to use an exponential distribution with mean θ as a model for these data. He would like to estimate his median waiting time, m.

(a) Express m in terms of θ.

(b) Damjan decides to use the sample median, $\hat{M}$, as his estimator.

(i) What is the asymptotic sampling distribution of $\hat{M}$?

(ii) What is Damjan's estimate of m for this dataset?

(iii) Calculate a standard error for Damjan's estimate.

(c) Damjan now considers using $\bar{X}$ as his estimator of m.

(i) Show that this estimator is biased.

(ii) Let $T = c\bar{X}$ be an adjusted estimator. Find c so that T is unbiased.

(iii) Determine $\text{var}(T)$.

(iv) Which of $\hat{M}$ and T is the better estimator?

(v) What is the estimate, t, based on the data above?

(vi) Calculate a standard error for this estimate.

Question 6 (10 marks) On his way to the office early every morning, Allan walks past South Lawn and counts how many students he sees there. He decides to model these counts using a Poisson distribution with pmf,

$\frac{e^{-\theta}\theta^{x}}{x!}$ $E(x)=Var(x)=\theta$.

$$p(x \mid \theta) = \frac{e^{-\theta}\theta^{x}}{x!}, \quad (x = 0, 1, \ldots).$$

$\theta = 3.8$

Across the first 45 days of semester, he observes on average 3.8 students per day. Robert says that he did a similar survey last year and got on average about 3.0 students per day, although he cannot remember over how many days he observed them. Allan would like to estimate θ by combining this information appropriately.

(a) Show that the gamma distribution is a conjugate prior for θ.
Note that the pdf of $\theta \sim \text{Gamma}(\alpha, \beta)$ is,

$$f(\theta|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \quad (\theta > 0)$$

and it has mean $\mathbb{E}(\theta) = \alpha/\beta$.

(b) Allan decides to treat Robert's information as if it came from 10 days of sampling (i.e. as pseudodata). Determine the parameters of the prior that encode this information appropriately.

(c) Using this prior, what is the posterior distribution of θ?

(d) Calculate the posterior mean.

Question 7 (13 marks) Ben runs a sports program for high school students. Each student that enters his program chooses one of the following three sports: volleyball, basketball and netball. The heights of the first 15 students, and the sports they chose, were:

Volleyball	183, 176, 170, 179
Basketball	165, 169, 171, 154, 165, 159
Netball	167, 160, 177, 173, 170

Ben runs an analysis of variance with these data. A partially complete ANOVA table from his analysis is given below.

Source	df	SS	MS	F
Treatment (sport)	2	416.4	208.2	5.48
Error	12	456	38	
Total	14	872.4		

$\sigma^2 = 38$

$F = \frac{208.2}{38}$

$F_{2,12} > \Phi^{-1}(0.975)$

(a) Is there evidence of a relationship between the students' heights and their choice of sport?

$\Phi^{-1}(0.975) \approx 5.09$

(b) What assumption has Ben made about the variance of the heights?

$\alpha = 0.05$ reject H0 at $\alpha = 0.025$ / 5%

(c) What is the sampling distribution of $\hat{\sigma}^2$?

(d) Let μ_1 be the population mean of students who choose volleyball. What is the sampling distribution of $\hat{\mu}_1$?

σ . $n-K$

(e) Calculate a 95% confidence interval for μ_1.

Question 8 (12 marks) You are studying a particular gene, for which the possible genotypes that an individual could have are AA, Aa and aa. In a sample of 600 individuals from the population, you observe 153 individuals of type AA, 286 individuals of type Aa and 161 individuals of type aa.

$n = 600$

(a) Use these data to test, with $\alpha = 0.05$, the hypothesis that the ratio of the genotypes is 1:2:1 respectively.

(b) You realise you made an error when determining the genotypes of the individuals of the sample, and that 15 of those you originally thought were of type Aa are actually of type aa. Repeat the hypothesis test with the corrected data.

End of exam questions—Total Available Marks = 90
Turn the page for appended material

Appendix (R output)

```
> p1 <- c(0.01, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975, 0.99)

> qnorm(p1)
[1] -2.326 -1.960 -1.645 -1.282  1.282  1.645  1.960  2.326

> qt(p1, df = 8)
[1] -2.896 -2.306 -1.860 -1.397  1.397  1.860  2.306  2.896
> qt(p1, df = 9)
[1] -2.821 -2.262 -1.833 -1.383  1.383  1.833  2.262  2.821
> qt(p1, df = 10)
[1] -2.764 -2.228 -1.812 -1.372  1.372  1.812  2.228  2.764
> qt(p1, df = 11)
[1] -2.718 -2.201 -1.796 -1.363  1.363  1.796  2.201  2.718
> qt(p1, df = 12)
[1] -2.681 -2.179 -1.782 -1.356  1.356  1.782  2.179  2.681
> qt(p1, df = 13)
[1] -2.650 -2.160 -1.771 -1.350  1.350  1.771  2.160  2.650
> qt(p1, df = 14)
[1] -2.624 -2.145 -1.761 -1.345  1.345  1.761  2.145  2.624
> qt(p1, df = 28)
[1] -2.467 -2.048 -1.701 -1.313  1.313  1.701  2.048  2.467
> qt(p1, df = 29)
[1] -2.462 -2.045 -1.699 -1.311  1.311  1.699  2.045  2.462

> qchisq(p1, df = 1)
[1] 0.0001571 0.0009821 0.0039321 0.0157908 2.7055435 3.8414588 5.0238862 6.6348966
> qchisq(p1, df = 2)
[1] 0.02010 0.05064 0.10259 0.21072 4.60517 5.99146 7.37776 9.21034
> qchisq(p1, df = 3)
[1]  0.1148  0.2158  0.3518  0.5844  6.2514  7.8147  9.3484 11.3449
> qchisq(p1, df = 8)
[1]  1.646  2.180  2.733  3.490 13.362 15.507 17.535 20.090
> qchisq(p1, df = 9)
[1]  2.088  2.700  3.325  4.168 14.684 16.919 19.023 21.666
> qchisq(p1, df = 10)
[1]  2.558  3.247  3.940  4.865 15.987 18.307 20.483 23.209

> qf(p1, 1, 13)
[1] 0.0001632 0.0010206 0.0040868 0.0164196 3.1362051 4.6671927 6.4142543 9.0738057
> qf(p1, 1, 14)
[1] 0.0001628 0.0010178 0.0040756 0.0163739 3.1022134 4.6001099 6.2979386 8.8615927
> qf(p1, 2, 12)
[1] 0.01006 0.02537 0.05151 0.10629 2.80680 3.88529 5.09587 6.92661
> qf(p1, 2, 13)
[1] 0.01006 0.02537 0.05150 0.10622 2.76317 3.80557 4.96527 6.70096
> qf(p1, 3, 11)
[1] 0.03686 0.06957 0.11411 0.19148 2.66023 3.58743 4.63002 6.21673
> qf(p1, 3, 12)
[1] 0.03697 0.06975 0.11436 0.19173 2.60552 3.49029 4.47418 5.95254

> pbinom(0:5, 10, 0.5)
[1] 0.0009766 0.0107422 0.0546875 0.1718750 0.3769531 0.6230469
> pbinom(0:5, 9, 0.5)
[1] 0.001953 0.019531 0.089844 0.253906 0.500000 0.746094
```

.975

Author/s:
Mathematics and Statistics

Title:
Elements of Statistics, 2018, Semester 2, MAST90058

Date:
2018

Persistent Link:
http://hdl.handle.net/11343/220973

Revision

① 2015. $\boxed{2005}$ Q7.

$Y \sim Bi(n,\theta)$ $H_0: p = p_0$ $H_1: p > p_0$. is $\{y : y/n > c\}$. Then find ⓒ
$\alpha = 0.05$ when n is large.

$L(\theta) = \binom{n}{y}\theta^y \cdot (1-\theta)^{n-y}$

$L_0(\theta) = L(\theta_0) = \binom{n}{y}\theta_0^y \cdot (1-\theta_0)^{n-y}$

$\lambda = \dfrac{L_0(\theta)}{L_1(\theta)}$

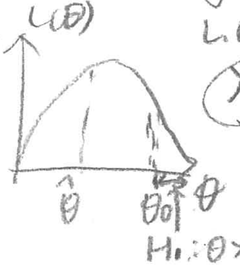
$L_1(\theta) = \max\limits_{\theta \in A_1} L(\theta)$

$\boxed{\theta \in A_1 \Rightarrow \theta > \theta_0}$

when to achieve maximum for $L_1(\theta)$. for $(L(\theta) : \theta > \theta_0)$

$\hat{\theta} = \frac{y}{n} \Rightarrow$ maximum $L(\theta)$

$\Rightarrow L_1(\theta) = L(\hat{\theta})$.

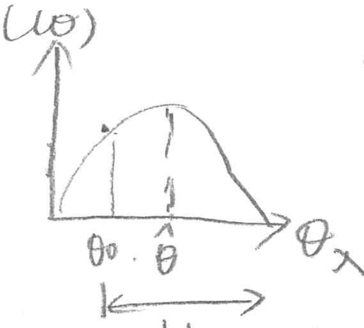


$L_1(\theta) = L(\theta_0)$

$\lambda = 1$

If $\frac{y}{n} < \theta_0$,

we will never reject H_0.



$\boxed{\text{If } y/n > \theta_0}$

$L_1(\theta) = L(\hat{\theta})$

$\lambda < 1$

when $\frac{y}{n} = \hat{\theta} > \theta_0$

$L_1(\theta) = \binom{n}{y}\hat{\theta}^y(1-\hat{\theta})^{n-y}$

$\lambda = \dfrac{L_0(\theta)}{L_1(\theta)} = \dfrac{\binom{n}{y}\theta_0^y(1-\theta_0)^{n-y}}{\binom{n}{y}\hat{\theta}^y(1-\hat{\theta})^{n-y}}$

$\boxed{\text{General rule}}$: If λ is small, we will reject H_0.

$\boxed{\text{Problem}}$ $\{y : y/n > c\}$.

when y/n is big $\Leftrightarrow \lambda$ is small.

$\Leftrightarrow$ when $\hat{\theta}$ increase $\Rightarrow \lambda$ decrease

$\Leftrightarrow \lambda =$ a decreasing function of $\hat{\theta}$

$g(y) = \ln(\lambda) = y\ln\theta_0 + (n-y)\ln(1-\theta_0)$
$\quad - y\log\frac{y}{n} - (n-y)\ln\frac{n-y}{n}$

$= \ldots\ldots$

$\dfrac{\partial g(y)}{\partial y} = \log\theta_0 - \log(1-\theta_0) - \log\hat{\theta} + \log\frac{n-y}{n}$

$= \log\left(\frac{\theta_0}{1-\theta_0} \times \frac{1-\hat{\theta}}{\hat{\theta}}\right) < 0$.

$\therefore$ when $\hat{\theta} > \theta_0$ $(y/n > \theta_0)$.

$\therefore \frac{y}{n} > \theta_0 \Rightarrow \boxed{\frac{\partial g(y)}{\partial y} < 0}$

$\boxed{\text{CLT}}$ $\frac{Y}{n} \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$.

$Z = \dfrac{\frac{Y}{n} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \approx (0,1)$ under H_0.

$Z > \Phi^{-1}(0.95)$

$\frac{Y}{n} > \underbrace{\theta_0 + \Phi^{-1}(0.95)\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}_{= c}$

2018. 9.25.8

① (a) (i) $X(1)$ ~~not possible to do~~. 2.620

(ii) $y(4.5) = y(4) + 0.5\,(y(5) - y(4)) = 4.78 + 0.5\,(5.30 - 4.78) = 5.04$

(iii) $\bar{x} = 4.193$

(iv) σ_2 = ~~samp~~ population sd, is a parameter, unknown

(b) (i) $x(1) > y(1)$. $x(1) = 2.62 < y(1) = 4.15$

$\boxed{\text{false}}$

(ii) $\bar{x} = 4.193$ $\bar{y} = 5.292$. false

(iii) $\sigma_1 > \sigma_2$, not possible

(c) (i) $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

x can't

$\boxed{\text{sample size of } x \text{ and } y}$

$$T = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

(ii) $H_0: \sigma_2 = 2$ $H_1: \sigma_2 \neq 2$

under H_0

$$T = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} = \chi^2_9$$

$\alpha = 0.05$

$$= \frac{9 \times 0.8368034^2}{2^2} = 1.5755$$

$< \Phi^{-1}(0.025)$

$\Phi^{-1}(0.025) = 2.7$ ~~3.325~~

$\Phi^{-1}(0.975) =$ ~~16.919~~ 19.023

∴ ~~fail to reject H_0~~

∴ reject H_0.

② $X \sim N(6.82, \frac{0.932^2}{10})$ $\sim t_9$

95% CI

(a) $6.82 \pm$ ~~1.96~~ $2.306 \times \frac{0.932}{\sqrt{10}} =$ ~~(7.3977, 6.24, 7.40)~~ $(6.15, 7.49)$

(b). $P = Pr(X > 6.5)$

$$Z = \frac{6.5 - 6.82}{0.932/\sqrt{10}} = -1.0857$$

$P = Pr(X > 6.5) = Pr(Z > -1.0857) = \Phi^{-1}(-1.0857)$

$P = Pr(X > 6.5)$. 7 observations of 10 : $\hat{p} = 0.7$

~~1.96~~ t_9 $c = \Phi^{-1}(0.975) = 2.262$

$$\hat{p} \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.7 \pm 2.262\sqrt{\frac{0.7 \times 0.3}{10}}$$

$(6.0, 7.3)$

(c) $Pr(X(i) < m < X(j)) = Pr(i \leq W \leq j-1)$

$Bi(10, 0.5)$

$\Rightarrow$ 90% CI is $(X(3), X(8))$

∴ we can chose $(X(1), X(10))$, $(X(2), X(9))$, ...

$Pr(X(1) < m < X(10)) = Pr(1 \leq W \leq 9) = 1 - 2Pr(W \leq 0) = 1 - 2 \times 0.0009766 = 0.998$

$Pr(X(2) < m < X(9)) = Pr(2 \leq W \leq 8) = 1 - 2Pr(W \leq 1) = 1 - 2 \times 0.0107422 = 0.9785$

$Pr(X(3) < m < X(8)) = Pr(3 \leq W \leq 7) = 1 - 2Pr(W \leq 2) = 1 - 2 \times 0.054875 = 0.891$

$\boxed{3}$ $P(x) = \theta(1-\theta)^x$

$L(\theta) = \prod_{i=1}^{n} (\theta(1-\theta)^x) = \theta^n \cdot (1-\theta)^{\sum_{i=1}^{n} x_i}$

(a) $\therefore Y = \sum_{i=1}^{n} X_i$ is sufficient statistic for θ

(b) $\mu_K = E(X) = \frac{1-\theta}{\theta}$ $M_K = \frac{1}{n}\sum X_i = \bar{X}$ $\mu_K = M_K$ $\therefore \bar{X} = \frac{1-\theta}{\theta}$

$\theta\bar{x} = 1-\theta$ $\theta(\bar{x}+1) = 1$ $\hat{\theta} = \frac{1}{\bar{x}+1}$

取ln 不要变号 求导才变号

(c) $\ln L(\theta) = n\ln\theta + \sum_{i=1}^{n} x_i \ln(1-\theta)$

$\frac{\partial(\ln L(\theta))}{\partial\theta} = \frac{n}{\theta} - \frac{1}{1-\theta} \cdot \sum_{i=1}^{n} x_i = 0$

$\cancel{\frac{n}{\theta} = \frac{\sum x_i}{1-\theta}}$ $\cancel{n\theta - n = \theta\sum x_i}$ $\theta = \frac{n}{n - \sum x_i}$

$\cancel{n\theta - \theta\sum x_i = n}$

$\cancel{\theta(n - \sum x_i) = n}$

$\frac{n}{\theta} = \frac{\sum_{i=1}^{n} x_i = n\bar{x}}{1-\theta}$

$\theta\bar{x} = 1-\theta$ $\theta\bar{x} + \theta = 1$ $\hat{\theta} = \frac{1}{\bar{x}+1}$

$\theta(\bar{x}+1) = 1$

(d) $V(\theta) = \frac{\partial^2 \ln L(\theta)}{\partial\theta^2} = -\frac{n}{\theta^2} - \frac{1}{(1-\theta)^2}\sum_{i=1}^{n} x_i$

$(-1)(1-\theta)^{-2}$

$(1-\theta)^{-1}$ $(-1)^{(-1)}(1-\theta)^{-2}$ θ

$= -\frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} X_i}{(1-\theta)^2}$

$I(\theta) = -E(V(\theta)) = +\frac{n}{\theta^2} + \frac{\sum x_i}{(1-\theta)^2} = \frac{n}{\theta^2} + \frac{n\bar{x}}{(1-\theta)^2} = \frac{n}{\theta^2} + \frac{n\frac{1-\theta}{\theta}}{(1-\theta)^2}$

$= \frac{n}{\theta^2} + \frac{n}{\theta(1-\theta)}$

$= \frac{n(1-\theta) + n\theta}{\theta^2(1-\theta)} = \frac{n}{\theta^2(1-\theta)}$

CR Lower bound $\frac{1}{I(\theta)} = \frac{\theta^2(1-\theta)}{n}$

(e) $se(\hat{\theta}) = \sqrt{\frac{\theta^2(1-\theta)}{n}}$

(f)

x	0	1	2	3
number	9	4	3	4

$\therefore \hat{\theta} = \frac{1}{1.1+1} = 0.47619 = \frac{10}{21}$

$\bar{x} = \frac{0\times9 + 1\times4 + 2\times3 + 3\times4}{20} \rightarrow 1.1$

$se(\hat{\theta}) = \sqrt{\frac{\theta^2(1-\theta)}{n}} = 0.07706$

$\hat{\theta} \pm c \cdot se(\hat{\theta}) = \cancel{\hat{\theta} \pm 1.645}$

normal $\Phi^{-1}(0.95)$

$0.476 \pm 1.645 \times 0.07706 = (0.35, 0.60)$

14 (a) $n = df + 2 = 28 + 2 = 30$ ✓

(b) (i) $H_0: \alpha = 0 \quad H_1: \alpha \neq 0$

$\alpha = 5\% = 0.05$. t_{28}

☆ $t = \frac{\hat{\alpha} - 0}{se(\hat{\alpha})} = \frac{-0.7323}{1.8852} = -0.388 \sim t_{n-2}$

$\alpha = \Phi^{-1}(1 - \frac{\alpha}{2}) = \Phi^{-1}(0.975)$

2.048

$|t| > 2.048$

fail to reject H_0.

(ii) $H_0: \beta = 0 \quad H_1: \beta \neq 0$.

$t = \frac{\hat{\beta} - 0}{se(\hat{\beta})} = \frac{1.1556}{0.388} = |3.03| > 2.048$ reject H_0

(c) ~~$R^2 = 0.2465$~~ ✗ $Pr t = 0.0052 < 0.01$

$\therefore$ reject H_0 at 1% significance level.

(d)

Source	df	SS	MS	F
Treatment →	1	26.01	26.01	9.159
Error →	28	79.52	2.84 $MS(E) = \sigma^2$	
Total →	29	105.53		

✓

$MS(E) = \sigma^2 = 1.686^2 = 2.84$

$F = \frac{MS(T)}{MS(E)}$

5 exponential $\Rightarrow f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}$. $F(x) = 1 - e^{-\frac{1}{\theta}x} = 0.5$

(a)

$e^{-\frac{1}{\theta}x} = 0.5$

$\frac{m^*}{\theta} = \ln 2$

$m = \ln 2 \cdot \theta$ ✓

(b) $\hat{m} \sim N(m, \frac{1}{4n\, f(m)^2})$ $\Rightarrow f(m) = f(\ln 2\, \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}\ln 2\,\theta}$

$= \frac{1}{\theta} e^{-\ln 2}$

$\hat{m} \sim N(\ln 2\,\theta, \frac{1}{4n \times \frac{1}{4\theta^2}})$ $\quad \theta = \frac{m}{\ln 2}$ $\quad \frac{1}{2\theta}$ ✓ $\frac{1}{4\theta^2}$

(i) $\boxed{N(\theta \ln 2\,\theta, \frac{\theta^2}{n})}$ ✓ $= \boxed{N(m, \frac{m^2}{n (\ln 2)^2})}$

large data

(ii) $\boxed{m = \theta \cdot \ln 2 = \bar{x} \cdot \ln 2 = 1.9 \times \ln 2 = 1.317}$ ✗

$\boxed{m = X_5 = 1.2}$ ✓

(iii)

~~$se(\hat{m}) = \sqrt{\frac{\theta}{n}} = \frac{\theta}{\sqrt{n}} = \frac{\bar{x}}{\sqrt{9}} = \frac{1.9}{3} = 0.633$~~

$\frac{m}{\sqrt{n} \cdot \ln 2} = \frac{1.2}{3 \times \ln 2} = 0.577$ ✓

5. C

(i) $E(\bar{X}) = E(X) = \theta = \frac{m}{\ln 2} \neq m$

(ii) $T = \ln 2\, \bar{X}$ $\quad E(T) = E(\ln 2\, \bar{X}) = \ln 2 \cdot \theta = \frac{m}{\ln 2} \cdot \ln 2 = m$, unbiased

$\therefore c = \ln 2$

(iii) $\boxed{Var(\bar{X}) = \theta^2}$ $\quad \therefore Var(T) = Var(\ln 2\, \bar{X}) = (\ln 2)^2 Var(\bar{X})$

$= (\ln 2)^2 \frac{\theta^2}{n} = (\ln 2)^2 \cdot \frac{m^2}{(\ln 2)^2} = \frac{m^2}{n}$

$Var(X) = \theta^2$

不是 $Var(\bar{X}) = \frac{\theta^2}{n}$

(iv)

$Var(\hat{m}) = \frac{m^2}{n \cdot (\ln 2)^2} > Var(T) = \frac{m^2}{n}$

0.69

$\therefore$ ~~M better~~ T better $\quad (+\frac{m^2}{n} \ln 2)$

(v) $\hat{t} = \ln 2\, \bar{x} = \ln 2 \times 1.9 = 1.317$

(vi) $se(t) = \sqrt{\frac{t^2}{n}} = \frac{t}{\sqrt{n}} = 0.44$

6. prior: Gamma(α, β).

$f(\theta|x) \propto f(x|\theta) \cdot f(\theta)$ $\quad f(\theta) = $ Gamma

$f(x|\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x}}{x!} = \frac{e^{-n\theta} \cdot \theta^{\sum x_i}}{\prod x!}$

(a) $\propto e^{-n\theta} \cdot \theta^{\sum x_i} \cdot \theta^{\alpha - 1} \cdot e^{-\beta\theta}$

$\propto \theta^{\sum x_i + \alpha - 1} e^{-\theta(n+\beta)}$ $\quad \alpha - 1 = \sum x_i + \alpha - 1$

$\therefore$ posterior $\theta \sim$ Gamma$(n\bar{x} + \alpha,\ n+\beta)$ $\Rightarrow$ conjugate prior

n days

(b) $\begin{cases} \alpha + \beta = 10 \\ \cancel{\beta = 3} \end{cases}$ $\quad 3\beta + \beta = \frac{10}{4} = 2.5$ $\quad \begin{cases} \alpha = 7.5 \\ \beta = 2.5 \end{cases}$ $\quad$ ~~prior $\theta \sim$ Gamma(7.5, 2.5)~~

$\therefore 10 = \beta$ $\quad \therefore \alpha = 30$. $\quad \therefore$ Gamma$(30, 10)$

(c) posterior $\sim$ Gamma$(30 + 3.8 \times 45,\ 45 + 10)$ $\quad (201, 55)$

(d) mean $= \frac{\alpha}{\beta} = \frac{201}{55} = 3.65$.

$E(\theta|x) =$

improper uniform prior

$f(\theta) = 1$.

$f(\theta|x) \propto e^{-n\theta} \theta^{\sum x_i}$

Gamma$(\sum x_i + 1, n)$

7 (a) $F_{2,12} = 5.48 > \Phi^{-1}(0.975) = 5.09$.
evidence

(b) Same ~~va~~ variance

(c) ~~$\hat{\sigma}^2 \sim X^2_{12}$~~ $\frac{12\hat{\sigma}^2}{\sigma^2} \sim X^2_{12}$

(d) $\hat{\mu}_1 \sim \bar{X}_1 \sim N(\mu_1, \frac{\sigma^2}{4})$

(e) $\Phi^{-1}(0.975)$ from t_{12} 2.179.

$\hat{\mu}_1 \pm 2.179 \cdot \frac{\sigma}{\sqrt{n}}$ = $\bar{X}_1 = \frac{183+176+170+179}{4} = 177$

$177 \pm 2.179 \times \frac{\sqrt{38}}{2} = (170.3 \quad 183.71)$

8 Goodness of fit test

	AA	Aa	aa
O	153	286	161
E	$600 \times \frac{1}{4}$ 150	$600 \times \frac{1}{2}$ 300	$600 \times \frac{1}{4}$ 150

$\therefore Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(153-150)^2}{150} + \frac{(286-300)^2}{300} + \frac{(161-150)^2}{150} = 1.52$

$\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95)$ from $X^2_2 = 5.99$

$Q < \Phi^{-1}$ fail to reject

b

	AA	Aa	aa
O	153	271	176
E	150	300	150

$X^2 = 7.37 > 5.99$

reject H_0