UNIVERSITY OF TORONTO

Faculty of Arts and Science

August 2018 EXAMINATIONS

STA261H1S

Duration - 3 Hours

Aids Allowed: Non-programmable calculator

First Name: SOLUTIONS.

Last Name:

Student Number:

This exam booklet contains ?? 21 pages. Write all final answers in this exam booklet, on the front of the same page on which the question appears. Use the 3 pages at the end, or the backs of the pages, for rough work. Fill in multiple choice answers using the bubble sheet at the end of the exam, in pencil. Answer all other questions in pen.

Questions:

Question	Marks Achieved	Total Possible
1		
2		
3		
4		
5		
Total		

FORMULA SHEET

You may use results on this sheet without proof.

If $Z \sim N(0,1)$ then $P(Z < -1.96) = 0.025$ and $P(Z < 1.96) = 0.975$.

If $\hat{\theta}$ is the MLE for θ and θ_0 is the true value then

$$\frac{\hat{\theta} - \theta_0}{1/\sqrt{i(\theta_0)}} \xrightarrow{d} N(0,1)$$

and

$$\frac{\hat{\theta} - \theta_0}{1/\sqrt{j(\hat{\theta})}} \xrightarrow{d} N(0,1)$$

If $\bar{X}$ is the sample mean then

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} \xrightarrow{d} N(0,1)$$

If there are d free parameters under H_0, and $p > d$ free parameters under H_1, then as $n \to \infty$, for a likelihood ratio test of H_0 against H_1,

$$-2\log\Lambda \xrightarrow{d} \chi^2_{p-d}$$

If $X \sim N(\mu, \sigma^2)$, then

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$
$$E(X) = \mu$$
$$Var(X) = \sigma^2$$
$$\frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$$

If $W \sim \chi^2_n$ then $E(W) = n$ and $Var(W) = 2n$.

1. (16 marks) Let X be a random variable taking values on the whole real line, with density depending on parameter $\theta \in \Omega \subset \mathbb{R}$ given by

$$f(x;\theta) = \exp\left(\theta x - b(\theta)\right) h(x)$$

where $h(x)$ is a function depening on x but not θ and $b(\theta)$ is a function depending on θ but not x.

(a) (4 marks) Let $x_i, i = 1 \dots n$ be an IID random sample from this distribution. Find the log-likelihood and the score statistic for θ.

$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \exp(\theta \sum x_i - nb(\theta)) \prod_{i=1}^{n} h(x_i)$ (Likelihood)

2 $\ell(\theta) = \theta \sum_{i=1}^{n} x_i - nb(\theta) + \sum_{i=1}^{n} \log h(x_i)$ (log-likelihood)

2 $S(\theta) = \partial \ell / \partial \theta = \sum_{i=1}^{n} x_i - n b'(\theta)$ (score stat.)

(b) (2 marks) Find a sufficient statistic for θ

1 $f(\underline{x};\theta) = \underbrace{\exp(\theta \sum x_i - nb(\theta))}_{g(\sum x_i, \theta)} \cdot \underbrace{\prod_{i=1}^{n} h(x_i)}_{h(\underline{x})}$ (from above)

1 By the factorization theorem, $T(\underline{x}) = \sum_{i=1}^{n} x_i$ is sufficient for θ.

(c) (6 marks) Show that $E(X) = b'(\theta)$ (where the derivative is with respect to θ). Hint:

$$\int_{-\infty}^{\infty} f(x;\theta)dx = 1$$
$$\implies \int_{-\infty}^{\infty} \exp(x\theta)\, h(x)dx = \exp\big(b(\theta)\big)$$

You may assume any necessary mathematical conditions required to exchange the order of differentiation and integration.

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} \exp(x\theta)h(x)dx = \frac{\partial}{\partial\theta}\exp(b(\theta)) \quad 2$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\exp(x\theta)h(x)dx = b'(\theta)\exp(b(\theta)) \quad 2$$

$$\Rightarrow \underbrace{\int_{-\infty}^{\infty} x\exp(x\theta - b(\theta))h(x)dx}_{=E(X)} = b'(\theta) \quad 2$$

(d) (4 marks) Find the Method of Moments estimator of θ, and show that it equals the Maximum Likelihood Estimator. You may assume that $b'(\theta)$ has a unique inverse in Ω.

MoM: set $E(X) = \bar{X}$

1 $\Rightarrow b'(\hat{\theta}) = \bar{X}$

MLE: set $S(\hat{\theta}) = 0$

1 $\Rightarrow \sum X_i - nb'(\hat{\theta}) = 0$

$b'(\hat{\theta}) = \frac{1}{n}\sum X_i = \bar{X}$

2 As $b'(\theta)$ assumed uniquely invertible, the MoM and MLE estimators satisfy same equation and are hence equal.

2. (20 marks) The amount of rainfall in inches was recorded for 227 storms in Illinois from 1960 - 1964. We wish to fit a probability model to these data. We have two candidates in mind, a $Gamma(\alpha, \lambda)$ distribution and a simpler $Exponential(\lambda)$ distribution. Recall $Gamma(1, \lambda) \stackrel{d}{=} Exponential(\lambda)$, with pdfs as follows:

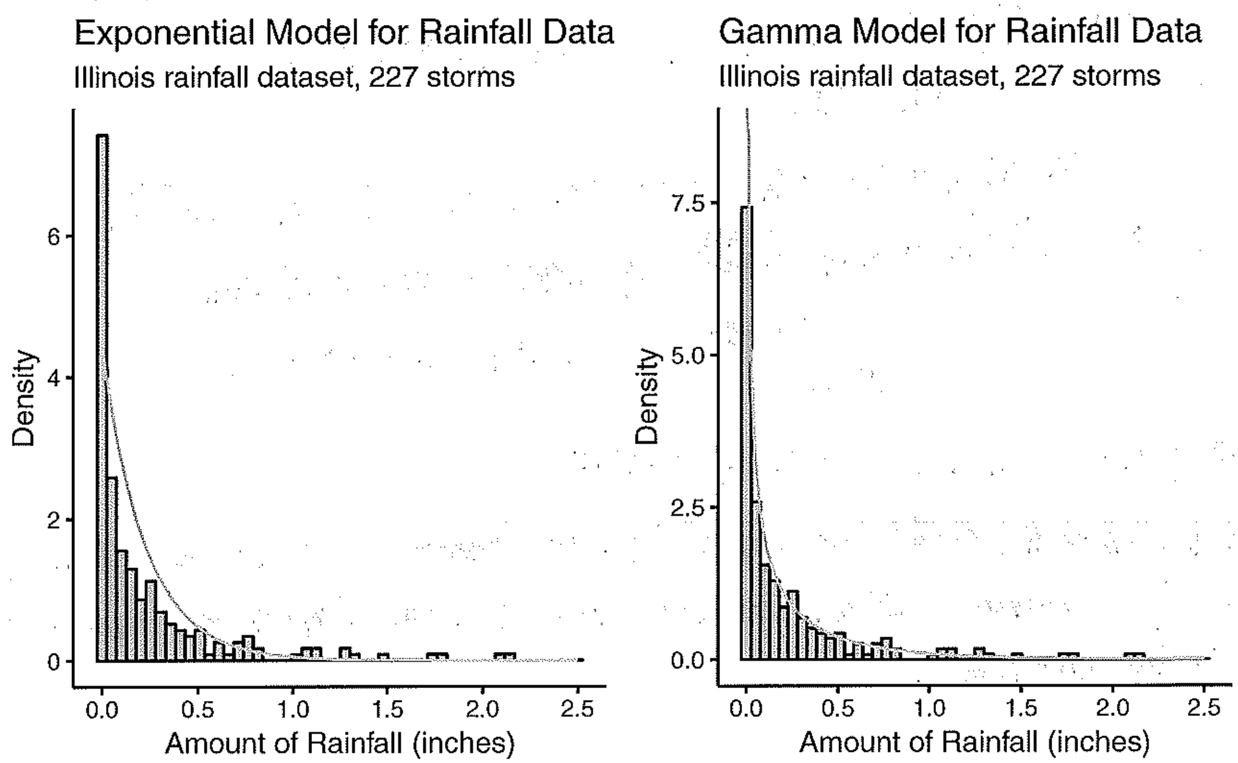
$$\text{Gamma: } f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x)$$

$$\text{Exponential: } f(x; \lambda) = \lambda \exp(-\lambda x)$$

We are interested in whether the exponential model fits the data well enough, or whether the gamma model is necessary.

We fit both curves by maximum likelihood, obtaining the result shown in the histograms entitled "Exponential Model for Rainfall Data" and "Gamma Model for Rainfall Data".

```
## Exponential distribution, MLE for lambda =  4.456485

## Gamma distribution, MLE for alpha =  0.4407903

## Gamma distribution, MLE for lambda =  1.964367

## Rainfall data, sum of x =  50.937

## Rainfall data, sum of log(x) =  -672.87

## Gamma(.441) =  2.008587
```



(20 marks) Perform a Likelihood Ratio Test to investigate whether the Exponential model is appropriate for these data. Fully derive an expression for the LRT statistic, state its asymptotic distribution and what assumptions and conditions must be satisfied for this to be valid, compute the statistic for these data, and give a reasonable conclusion in plain language that would be understandable by a non-statistician.

2 Exponential likelihood: $L_1(\lambda_1) = \lambda_1^n \exp(-\lambda_1 \Sigma X_i)$

Gamma likelihood : $L_2(\alpha, \lambda_2) = \left(\frac{\lambda_2^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} \exp(-\lambda_2 \Sigma x_i)$

2 $\Lambda = \frac{L_1(\hat{\lambda}_1)}{L_2(\hat{\alpha}, \hat{\lambda}_2)}$; likelihood ratio, evaluated at MLE.

2 If exponential model appropriate, $-2\log\Lambda \mathrel{\dot\sim} \chi^2_1$

2 - assume MLE's of both models are in interior of parameter space, that n is large enough, that both likelihoods are C^3

2 Compute: $\ell_1(\lambda_1) = n\log\lambda_1 - \lambda_1 \Sigma X_i$

$\Rightarrow \ell_1(\hat{\lambda}_1) = (227)\log(4.457) - 4.457 \times 50.937.$

$= 112.23.$

$\ell_2(\alpha, \lambda_2) = n\alpha\log\lambda_2 - n\log\Gamma(\alpha) + (\alpha - 1)\Sigma\log x_i - \lambda_2 \Sigma X_i$

2 $\Rightarrow \ell_2(\hat{\alpha}, \hat{\lambda}_2) = (227)(0.441)\log(1.965) - (227)\log\Gamma(0.441)$

$+ (0.441 - 1)(-672.87) - 1.965 \times 50.937$

$= 186.32$

2 $-2\log\Lambda = -2(112.23 - 186.32) = 148.18.$

6 If the exponential model were appropriate, we observed a value of 148.18 from a χ^2_1 - extremely unlikely. Conclude that the data supports the Gamma model over the exponential.

(Question 2 continued)

3. (24 marks) Let $X_i \overset{IID}{\sim} N(\mu, \sigma^2)$. We wish to estimate the variance $Var(X) = \sigma^2$. We consider two estimators:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$$s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

We saw before that s_n^2 is the maximum likelihood estimator of σ^2. The estimator s_{n-1}^2 is often taught in statistics courses as being "correct", because it "corrects bias" in s_n^2. This being a math stats course, let's go a bit further and compare the properties of these estimators.

(a) (4 marks) Compute the bias of s_n^2 and s_{n-1}^2. Is either estimator unbiased?

1 $\begin{cases} \dfrac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2 \Rightarrow E\left(\dfrac{(n-1)S_n^2}{\sigma^2}\right) = n-1 \Rightarrow E(S_n^2) = \sigma^2 \\ S_n^2 = \dfrac{n-1}{n} S_{n-1}^2 \Rightarrow E(S_n^2) = \dfrac{n-1}{n}\sigma^2 \end{cases}$

1 $\text{Bias}(S_{n-1}^2) = E(S_{n-1}^2) - \sigma^2 = 0$ $\quad S_{n-1}^2$ is unbiased

1 $\text{Bias}(S_n^2) = E(S_n^2) - \sigma^2 = -\sigma^2/n$ $\quad S_n^2$ is biased. 1

(b) (4 marks) Compute the variance of s_n^2 and s_{n-1}^2. Is one always higher than the other, or is it impossible to say?

$$Var\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) = 2(n-1) \Rightarrow Var(S_{n-1}^2) = \frac{\sigma^4}{(n-1)^2}\cdot 2(n-1) = \frac{2\sigma^4}{n-1} \quad 2$$

$$Var(S_n^2) = Var\left(\frac{n-1}{n}S_{n-1}^2\right) = \left(\frac{n-1}{n}\right)^2 Var(S_{n-1}^2) < Var(S_{n-1}^2) \ \forall n \in \mathbb{N} \quad 1$$

$$= \frac{(n-1)^2}{n^2}\cdot\frac{2\sigma^4}{n-1}$$

$$= \frac{(n-1)2\sigma^4}{n^2}. \quad 1$$

$I(\sigma^2)$

(c) (6 marks) Find the Fisher information $\cancel{i(\sigma^2)}$ in the sample ~~(you may treat μ as a fixed, known constant).~~ Is either estimator efficient?

$$L(\sigma^2,\mu) = (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\sum(x_i-\mu)^2\right)$$

① $$\ell(\sigma^2,\mu) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

① $$S(\mu) = \frac{1}{\sigma^2}\sum(x_i-\mu) \Rightarrow \hat{\mu} = \bar{x}$$

$$S(\sigma^2) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum(x_i-\hat{\mu})^2$$

Note: finding MLEs not necessary for full marks.

① $$J(\sigma^2) = \frac{-n}{2\sigma^4} + \frac{1}{\sigma^6}\sum(x_i-\mu)^2$$

① $$I(\sigma^2) = \frac{-n}{2\sigma^4} + \frac{1}{\sigma^6}\underbrace{E\left(\sum(x_i-\mu)^2\right)}_{=n\sigma^2}$$

$$= \frac{n}{2\sigma^4}$$

The cramer-rao lower bound on the variance of any unbiased estimator of σ^2 is

② $Var(\hat{\sigma}^2) \geq \frac{1}{I(\sigma^2)} = \frac{2\sigma^4}{n}$

$Var(s_{n-1}^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$, so s_{n-1}^2 is not efficient.

$Var(s_n^2) = \frac{n-1}{n} \times \frac{2\sigma^4}{n} < \frac{2\sigma^4}{n}$, so it actually breaks the CRLB - but, as s_n^2 is biased, the CRLB and concept of efficiency do not apply.

(d) (4 marks) Compute the Mean Squared Error of both estimators. Is one always lower than the other, or can you not say?

$MSE(\hat{\sigma}^2) = E((\hat{\sigma}^2 - \sigma^2)^2)$

1 Because $E(S_{n-1}^2) = \sigma^2$, $MSE(\cancel{a}S_{n-1}^2) = Var(S_{n-1}^2) = \frac{2\sigma^4}{n-1}$

For $\cancel{a}$ S_n^2, note $MSE(\hat{\sigma}^2) = E((\hat{\sigma}^2 - E(\hat{\sigma}^2))^2 + (\cancel{E}E(\hat{\sigma}^2) - \sigma^2)^2 + 2(E(\hat{\sigma}^2) - \hat{\sigma}^2)(\hat{\sigma}^2 - \sigma^2))$

$= Var(\hat{\sigma}^2) + bias(\hat{\sigma}^2)^2$

2 So $MSE(S_n^2) = \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2}$

$= (2(n-1)+1)\sigma^4 / n^2$

$= (2n-1)\sigma^4 / n^2$

1 As $\frac{2n-1}{n^2} > \frac{1}{n-1}$, $MSE(S_n^2) > MSE(S_{n-1}^2)$

(e) (6 marks) Discuss the relative merits of each estimator. Give at least one positive and one negative aspect of each. You will be marked on the clarity and thoroughness of your discussion.

3 S^2_{n-1} : unbiased, so more correct on average across samples

Higher variance, so less likely to be correct in any given sample.

3 S^2_n : Higher MSE, so less likely to be correct in any given sample.

Biased, so will be wrong on average across samples

Lower variance, more likely to be correct in any given sample.

Both: as $n \to \infty$, the differences between the two vanish, as $\frac{n-1}{n} \to 1$.

Anything reasonable here is fine, as long as they attempt an honest discussion.
Don't award marks if all they do is randomly write buzzwords from class.

4. (20 marks) Data on the annual temperature measured in Ann Arbor, Michigan, ~~(or "Ann Arbour" on the Canadian side)~~ is shown in the plots entitled "Histogram of Annual Temperature" and "Temperature by Year". Overlayed on the histogram is a Normal density curve; the normal model seems to fit quite well. We know how to estimate the ~~mena~~ μ and standard deviation σ of this distribution. However, scientists want to know: is the average temperature increasing across years?

You propose the model

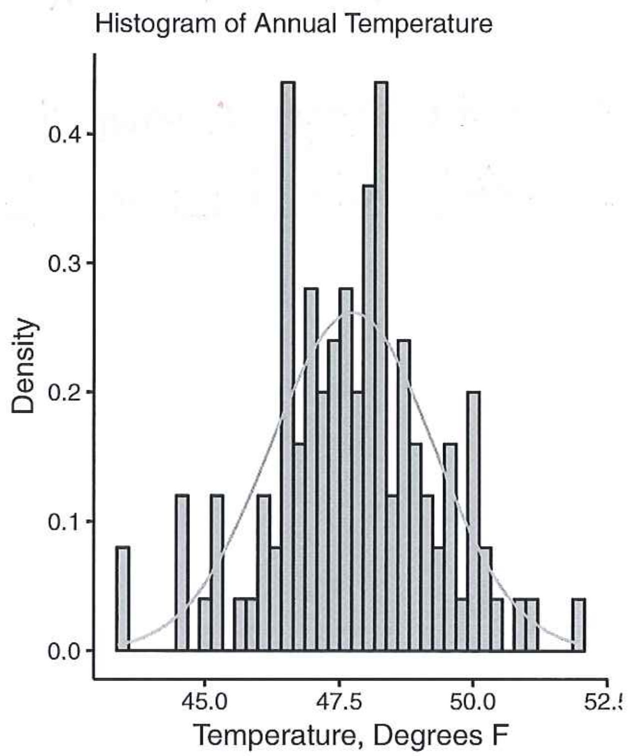
$$Y_i \overset{ind}{\sim} N(\mu_i, \sigma^2)$$
$$\mu_i = \alpha + \beta x_i$$

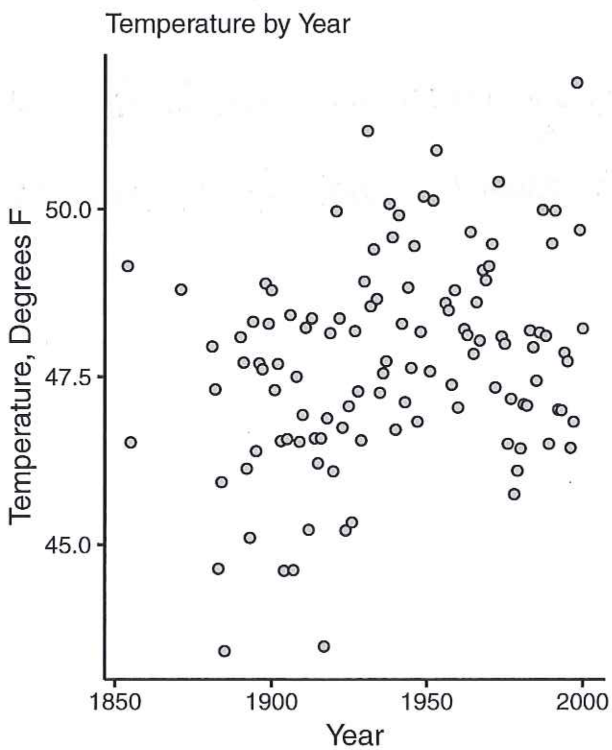
where Y_i is the rainfall in the ith year, with year 1 being 1854, $x_i = 1, 2, \ldots 115$ is the i^{th} year (i.e. $x_1 = 1854, x_{115} = 2000$), and α, β, σ^2 are parameters to be estimated. Under this model, it is clear that β has the interpretation as the expected increase in temperature in any pair of consecutive years, since

$$\mu_{i+1} - \mu_i = \beta$$

hence if we can estimate β, we can tell the scientists whether their hypothesis is reasonable.

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```





(a) (2 marks) Write down the log-likelihood for α, β. You can treat σ^2 as a fixed, known constant for this question.

$\mu_i = \alpha + \beta x_i$

1 $L(\alpha,\beta) = \prod_{i=1}^{n} f(y_i;\mu_i) = (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\sum(y_i-\mu_i)^2\right)$

$= (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\sum(y_i-(\alpha+\beta x_i))^2\right)$

1 $\ell(\alpha,\beta) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\alpha+\beta x_i))^2$

(b) (2 marks) Find the score statistics for α, β.

$S(\alpha) = \partial\ell/\partial\alpha = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-(\alpha+\beta x_i))$ 1

$S(\beta) = \partial\ell/\partial\beta = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i-(\alpha+\beta x_i))$ 1

(c) (8 marks) Find the Maximum Likelihood Estimators for α, β, and compute them for these data. I have centred and scaled the data, so you may use

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 0$$

$$\sum_{i=1}^{n} x_i^2 = 114; \sum_{i=1}^{n} x_i y_i = 33.31$$

1 $$S(\alpha) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i)) = 0$$

$$S(\beta) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i(y_i - (\alpha + \beta x_i)) = 0$$

$$\Rightarrow \sum y_i = n\hat{\alpha} + \hat{\beta} \sum x_i$$

$$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$$

2 $$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\Rightarrow \sum x_i y_i = \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2$$

3 $$\sum x_i y_i = \frac{1}{n} \sum y_i \sum x_i + \hat{\beta}\left(\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right)$$

Any of these are fine.

$$\hat{\beta} = \frac{\sum x_i y_i - \frac{1}{n} \sum y_i \sum x_i}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}$$

OR

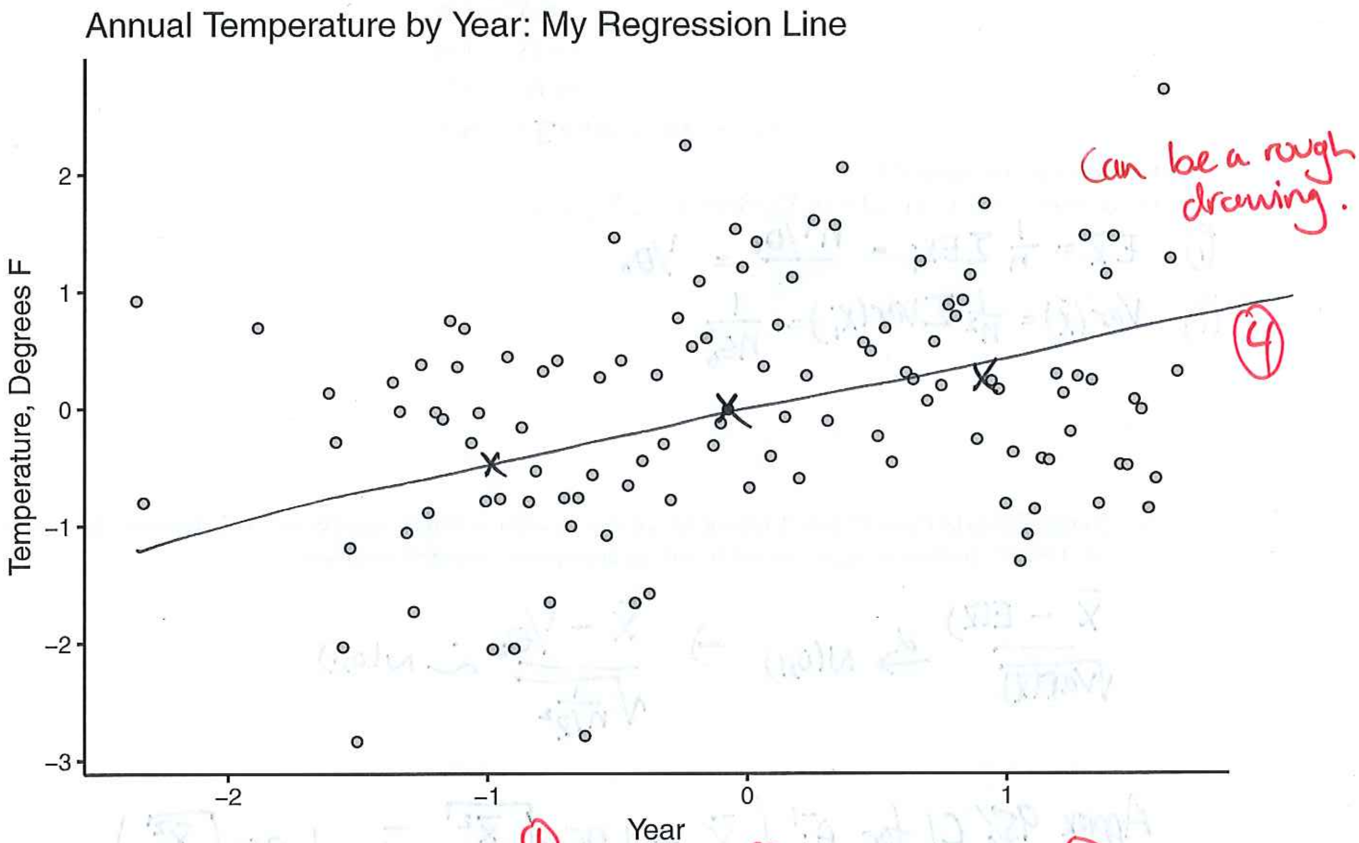
$$= \frac{\frac{1}{n} \sum x_i y_i - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2}$$

OR

$$= \frac{\frac{1}{n} \sum x_i y_i - \bar{y}\bar{x}}{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

2 For these data, $\hat{\alpha} = 0$, $\hat{\beta} = \frac{33.31}{114.} = 0.292.$

(a) (8 marks) Consider the plot entitled "Annual Temperature in Ann Arbor by Year: My Regression Line". You just fit a linear regression- modelling $E(Y_i|x_i) = \alpha + \beta x_i$. Draw a line on the plot with the slope and intercept you found in the previous steps. The data in the plot has been centred and scaled as described in the previous part.



1 1 1

Should pass through (0,0), (−1, −0.292), (1, 0.292)

Line is $Y = 0.292X$. 1

5. (16 marks) Let $X_i \sim Exp(\theta), i = 1 \ldots n$ be an IID random sample from an exponential distribution with the parametrization

$$f(x|\theta_0) = \theta_0 e^{-x\theta_0}$$
$$E(X) = 1/\theta_0$$
$$Var(X) = 1/\theta_0^2$$
$$\text{MLE: } \hat{\theta} = 1/\bar{X}$$
$$\text{Observed Information: } \mathcal{J}(\theta) = n/\theta^2$$

where θ_0 is the true value of θ.

(a) (2 marks) Find $E(\bar{X})$ and $Var(\bar{X})$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

1 $E\bar{X} = \frac{1}{n}\sum EX_i = \frac{n\,1/\theta_0}{n} = 1/\theta_0$

1 $Var(\bar{X}) = \frac{1}{n^2}\sum Var(X_i) = \frac{1}{n\theta_0^2}$

(b) (3 marks) Use the Central Limit Theorem for the sample mean to find an approximate 95% confidence interval for $1/\theta$. $Var(\bar{X})$ depends on θ_0, so replace it with an appropriate consistent estimator.

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} \xrightarrow{d} N(0,1) \Rightarrow \frac{\bar{X} - 1/\theta_0}{\sqrt{\frac{1}{n/\bar{X}^2}}} \dot\sim N(0,1)$$

Approx 95% CI for θ^{-1}: $\left(\bar{X} - 1.96 \cdot \sqrt{\frac{\bar{X}^2}{n}}, \bar{X} + 1.96\sqrt{\frac{\bar{X}^2}{n}}\right)$

$= \left(\bar{X} - 1.96\, \bar{X}/\sqrt{n}, \bar{X} + 1.96\, \bar{X}/\sqrt{n}\right)$

(c) (3 marks) Use your answer to the previous question to find an approximate 95% confidence interval for θ. Call this interval V_n

$$V_n = \left(\left(\bar{x} + 1.96\, \bar{x}/\sqrt{n}\right)^{-1}, \left(\bar{x} - 1.96\, \bar{x}/\sqrt{n}\right)^{-1} \right)$$

(d) (3 marks) Use the Central Limit Theorem for the MLE to find an approximate 95% confidence interval for θ. Call this interval W_n.

$$\frac{\hat{\theta} - \theta_0}{j(\hat{\theta})^{1/2}} \dot{\sim} N(0,1)$$

$$W_n = \left(\frac{1}{\bar{x}} - 1.96\sqrt{1/j(\hat{\theta})}, \frac{1}{\bar{x}} + 1.96\sqrt{1/j(\hat{\theta})} \right)$$

$$= \left(\frac{1}{\bar{x}} - 1.96\, \bar{x}/\sqrt{n}, \frac{1}{\bar{x}} + 1.96\, \bar{x}/\sqrt{n} \right)$$

(e) (5 marks) These intervals for θ are different for any finite n. Show that as $n \to \infty$, both intervals converge in probability to the singleton set $\{\theta_0\}$:

$$V_n \xrightarrow{p} \{\theta_0\}$$
$$W_n \xrightarrow{p} \{\theta_0\}$$

If you use a familiar theorem, be sure to state it.

$V_n = \left(\left(\bar{X} + 1.96\, \bar{X}/\sqrt{n}\right)^{-1}, \left(\bar{X} - 1.96\, \bar{X}/\sqrt{n}\right)^{-1} \right)$

$W_n = \left(\frac{1}{\bar{X}} - 1.96\, \bar{X}/\sqrt{n},\ \bar{X} + 1.96\, \bar{X}/\sqrt{n} \right)$

1 $\bar{X} \xrightarrow{P} E(X) = 1/\theta_0$ by LLN

1 $\frac{1}{\bar{X}} \xrightarrow{P} \frac{1}{E(X)} = \theta_0.$ ($g(x) = 1/x$ continuous)

1 ~~Hence~~ Also, $\frac{\bar{X}}{\sqrt{n}} \xrightarrow{P} 0$ (can state without proof)

1 So $V_n \xrightarrow{P} \left(\frac{1}{\frac{1}{\theta_0} + 0}, \frac{1}{\frac{1}{\theta_0} + 0} \right) = (\theta_0, \theta_0) = \{\theta_0\}.$

1 $W_n \xrightarrow{P} (\theta_0 - 0, \theta_0 + 0) = \{\theta_0\}.$

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED