Zhe Liu

Solutions to MOCK Final Exam

Exam Duration: 2 Hours
Total Marks Available: 100

Instructions

1. Except for the `Base R` cheat sheet, this is a closed-book exam.
2. All answers and explanations must be provided in the answer book.
3. Keep your answers succinct and to the point. Long rambling answers with irrelevant details may work against you.

Advice: Read the entire exam before starting on your answers.

Question 1. (18 marks)

Answer each of the following questions.

(a) (**5 marks**) Let $\mathbf{A}$ be an $m \times n$ matrix with column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and row vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m$. Let $\mathbf{x} = (x_1 \ldots x_m)^\top$. Give an expression for $\mathbf{A}^\mathrm{T}\mathbf{x}$ in terms of the column or row vectors.

Solution: The rule of matrix multiplication implies

$$\mathbf{A}^\mathrm{T}\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \ldots + x_m\mathbf{b}_m.$$

(b) (**5 marks**) Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathrm{rank}(\mathbf{A}) = n$. Is there always a solution to the equation $\mathbf{A}^\mathrm{T}\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^n$? Justify your answer.

Solution: Yes, there is: if $\mathbf{A}$ is of full rank, so is $\mathbf{A}^\mathrm{T}$, which is therefore invertible. The (unique) solution is $\mathbf{x} = (\mathbf{A}^\mathrm{T})^{-1}\mathbf{b}$.

(c) (**8 marks**) A projection matrix is a matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ such that $\mathbf{P}^2 = \mathbf{P}$. What are the possible eigenvalues of $\mathbf{P}$? Justify your answer.

Solution: Let λ be an eigenvalue of $\mathbf{P}$ with corresponding eigenvector $\mathbf{u}$ so that $\mathbf{Pu} = \lambda\mathbf{u}$. Multiplying across by $\mathbf{P}$ and noting $\mathbf{P}^2 = \mathbf{P}$ yields

$$\begin{aligned} \mathbf{P}^2\mathbf{u} &= \lambda\mathbf{Pu} \\ \Leftrightarrow \mathbf{Pu} &= \lambda^2\mathbf{u}. \end{aligned}$$

But also $\mathbf{Pu} = \lambda\mathbf{u}$ and so $\lambda^2\mathbf{u} = \lambda\mathbf{u}$. It must therefore be the case (why?) that $\lambda = 0$ and $\lambda = 1$ are the only possible eigenvalues.

Question 2. (12 marks)

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 13 & 0 & -15 \\ -3 & 4 & 9 \\ 5 & 0 & -7 \end{pmatrix}$$

and answer the following questions.

(a) (**4 marks**) Use `R` to compute and then report the eigenvalues of $\mathbf{A}$?

Solution: The eigenvalues of $\mathbf{A}$ are $\lambda = 8$, 4 and -2.

(b) (**3 marks**) True or false: $\mathbf{B}$ is a symmetric matrix where $\mathbf{B} := \mathbf{A}^\top \mathbf{A}$.

Note: Please use the properties from the module to get full marks. If you arrive at the correct answer using `R`*, you receive 2 marks.*

Solution: This is true since $\mathbf{B}^\top = \mathbf{B}$.

(c) (**5 marks**) Use `R` to compute and then report (to two decimal places) the eigenvalues of $\mathbf{B}$?

Solution: The eigenvalues of $\mathbf{B}$ are $\lambda = 549.44$, 24.26 and 0.31.

Question 3. (10 marks)

A group of 20 people go out to dinner. 10 go to an Italian restaurant, 6 to a Chinese restaurant and 4 to a French restaurant. The fractions of people satisfied with their meals were 4/5, 2/3 and 1/2 respectively. The next day the person you are talking to (one of the group of 20) was satisfied with their meal. What is the probability that they went to the Italian restaurant?

Solution: Let I denote the event that the person went to an Italian restaurant and let S denote the event that the person was satisfied with their meal. Then we are asked to find $\mathrm{P}(I \mid S)$. From Bayes' Theorem we have

$$\begin{aligned} \mathrm{P}(I \mid S) &= \frac{\mathrm{P}(S \mid I)\mathrm{P}(I)}{\mathrm{P}(S)} \\ &= \frac{(4/5)(1/2)}{(7/10)} \\ &= 4/7. \end{aligned}$$

Question 4. (20 marks)

Suppose that the lifespan of a piece of equipment is normally distributed with mean 10 years and standard deviation 3 years. A particular piece of equipment is now aged 11 years. The firm is considering two options:

- Replace it now;
- Wait until age 13 to replace it (unless it dies before then, in which case an unscheduled replacement is necessary).

Suppose that an immediate replacement will cost the company £a. Replacement at age 13 – if the equipment lasts that long – will cost £b, where $b < a$. An unscheduled replacement between now and age 13 – if the equipment dies suddenly during the next two years – will cost £c, where $c > a$.

(a) (**8 marks**) What is the probability that, barring replacement now, the equipment will last until age 13?

Solution: Let $X \sim \mathrm{N}(\mu, \sigma)$ denote the natural lifetime of the piece of equipment with $\mu = 10$ and $\sigma = 3$. The question is asking for $\mathrm{P}(X \geq 13 \mid X \geq 11)$. We can compute this as

$$\begin{aligned}
\mathrm{P}(X \geq 13 \mid X \geq 11) &= \frac{\mathrm{P}(X \geq 13,\, X \geq 11)}{\mathrm{P}(X \geq 11)} \\
&= \frac{\mathrm{P}(X \geq 13)}{\mathrm{P}(X \geq 11)} \\
&= \frac{1 - \mathrm{P}(X \leq 13)}{1 - \mathrm{P}(X \leq 11)} \\
&= 0.4294.
\end{aligned}$$

(b) (**8 marks**) What is the expected cost of replacement if the equipment is not replaced now?

Solution: If the equipment is not replaced now then

$$\begin{aligned}
\text{Expected cost} &= b \times \mathrm{P}(X \geq 13 \mid X \geq 11) + c \times (1 - \mathrm{P}(X \geq 13 \mid X \geq 11)) \\
&= 0.4294b + 0.5706c.
\end{aligned}$$

(c) (**4 marks**) Under what condition on a, b and c is the expected cost lower if the equipment is replaced now than under the age 13 policy?

Solution: The expected cost is lower if we replace now when $a < 0.4294b + 0.5706c$.

Question 5. (25 marks)

As a production manager, you are investigating the processing times of machines in the assembly line. In the production of the most essential product of your company, there is a specific machine of high importance. The processing time of the machine follows a normal distribution with an assumed standard deviation of 10 minutes according to the chief production engineer. A test run of the machine produces 100 items with an averaging processing time of 90 minutes per item. Answer the following questions.

(a) (**5 marks**) Give a 95% confidence interval for the mean processing time.

Solution: For $\bar{X} = 90, \sigma = 10, n = 100$, the 95% confidence interval is

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = 90 \pm 1.96 \times \frac{10}{\sqrt{100}} = (88.04, 91.96).$$

(b) (**5 marks**) Is it true that with 95% probability the true (population) mean processing time lies in the interval from part (a)?

Solution: No. The right interpretation is that if we take many samples of the processing time and create a confidence interval form each sample mean, then 95% of the confidence intervals will contain the true (population) mean processing time.

(c) (**5 marks**) How many items should the test run produce if we want the half width of the confidence interval (i.e., the margin of error) to be no more than 1 minute for a 95% confidence interval?

Solution: Solving

$$z\frac{\sigma}{\sqrt{n}} = 1,$$

we get

$$n = \left(\frac{z\sigma}{1}\right)^2 = \left(\frac{1.96 \times 10}{1}\right)^2 = 384.16.$$

Hence the test run should produce 385 items to reach a margin of error of 1 minute.

(d) (**10 marks**) Would you agree with the engineer's further claim that the average processing time is 88 minutes at a $\alpha = 0.05$ significance level?

Solution: The null and alternative hypotheses are:

$$H_0 : \mu_0 = 88,$$

$$H_\alpha : \mu_0 \neq 88.$$

Using z-test for the population mean, the test statistic is

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{90 - 88}{10/\sqrt{100}} = 2.$$

Hence the p-value for this **two-sided test** is given by

$$2 \times \mathrm{P}(z > z_0) = 2(1 - \Phi(2)) = 0.0456.$$

Since the p-value (0.0456) is less than the significance level ($\alpha = 0.05$), we can reject the null hypothesis that the average processing time is 88 minutes.

Question 6. (15 marks)

Data for 51 US states (50 states plus the District of Columbia) was used to examine the relationship between violent crime rate (violent crimes per 100,000 persons per year) and the predictor variables of urbanization (percentage of the population living in urban areas) and poverty rate. A predictor variable indicating whether or not a state is classified as a Southern state (1 = Southern, 0 = not) was also included. Some `R` output for the linear regression analysis is shown below.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -321.90     148.20   -2.17    0.035 .
urban           4.689      1.654    2.83    0.007 *
poverty         39.34      13.52    2.91    0.006 *
south(s=1)    -649.30     266.96   -2.43    0.019 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(a) (**5 marks**) What is the estimated linear model?

Solution: The model is

$$\text{crime rate} = -321.90 + 4.689 \times \text{urban} + 39.34 \times \text{poverty} - 649.30 \times \text{south}.$$

(b) (**5 marks**) How to interpret the estimated coefficient for south, -649.30?

Solution: It indicates that being a Southern state will decrease the violent crime rate by 649.30 per 100,000 persons per year on average.

(c) (**5 marks**) From the model summary, which predictor variables are significant at 0.01 level? Explain your reasoning.

Solution: Urban and poverty are significant since their p-values are less than the 0.01 significance level.