Quiz 1, STATS 401 F18

In lab on 10/5

This document produces different random quizzes each time the source code generating it is run. The actual quiz will be a realization generated by this random process, or something similar.

Instructions. You have a time allowance of 40 minutes, though the quiz may take you less time and you can leave lab once you are done. The quiz is closed book, and you are not allowed access to any notes. Any electronic devices in your possession must be turned off and remain in a bag on the floor.

Formulas

The following formulas will be provided. To use these formulas properly, you need to make appropriate definitions of the necessary quantities.

(1) $\mathbf{b} = \left(\mathbb{X}^{\mathsf{T}}\mathbb{X}\right)^{-1}\mathbb{X}^{\mathsf{T}}\mathbf{y}$

(2) $\mathrm{Var}(X) = \mathrm{E}\big[(X-\mathrm{E}[X])^2\big] = \mathrm{E}[X^2] - \big(\mathrm{E}[X]\big)^2$

(3) The probability density function of the standard normal distribution is $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$

(4) Syntax from `?pnorm`:

```
pnorm(q, mean = 0, sd = 1)
qnorm(p, mean = 0, sd = 1)
q: vector of quantiles.
p: vector of probabilities.
```

Q1. Matrix exercises

Q1-1.

(a). Evaluate $\mathbb{A}\mathbb{B}$ when

$$\mathbb{A} = \begin{bmatrix} 2 & 3 \\ 1 & 3 \\ -1 & -2 \end{bmatrix}, \quad \mathbb{B} = \begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix}$$

Solution:

$$\mathbb{A}\mathbb{B} = \begin{bmatrix} 6 & 5 \\ 3 & 4 \\ -3 & -3 \end{bmatrix}$$

(b). For $\mathbb{A}$ as above, write down $\mathbb{A}^{\mathsf{T}}$.

Solution:

$$\mathbb{A}^{\mathsf{T}} = \begin{bmatrix} 2 & 1 & -1 \\ 3 & 3 & -2 \end{bmatrix}$$

(c). For $\mathbb{B}$ as above, find $\mathbb{B}^{-1}$ if it exists. If $\mathbb{B}^{-1}$ doesn't exist, explain how you know this.

Solution:

$$\mathbb{B}^{-1} = \frac{1}{3}\begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix}$$

Q1-2.

(a). Evaluate $\mathbb{AB}$ when

$$\mathbb{A} = \begin{bmatrix} -1 & -1 & 3 \\ 2 & 0 & 3 \end{bmatrix}, \quad \mathbb{B} = \begin{bmatrix} -1 & 1 & -2 \\ 0 & 0 & 0 \\ -2 & 3 & 0 \end{bmatrix}$$

Solution:

$$\mathbb{AB} = \begin{bmatrix} -5 & 8 & 2 \\ -8 & 11 & -4 \end{bmatrix}$$

(b). For $\mathbb{A}$ as above, write down $\mathbb{A}^\intercal$.

Solution:

$$\mathbb{A}^\intercal = \begin{bmatrix} -1 & 2 \\ -1 & 0 \\ 3 & 3 \end{bmatrix}$$

(c). For $\mathbb{A}$ as above, find $\mathbb{A}^{-1}$ if it exists. If $\mathbb{B}^{-1}$ doesn't exist, explain how you know this.

Solution:

Only square matrices can be invertible. $\mathbb{A}$ is 2×3 and so cannot have an inverse.

Q2. Summation exercises

Q2-1.

Calculate $\sum_{i=k}^{k+3}(i+3)$, where k is a whole number. Your answer should depend on k.

Solution:

TBD

Q2-2.

Evaluate $\sum_{i=1}^{30} 10 - \sum_{i=10}^{20} 20$.

Solution:

TBD

Q2-3.

Calculate $\sum_{k=m}^{n} a$, where m and n are whole numbers and a is a real number.

Solution:

TBD

Q2-4.

Evaluate $3\sum_{k=1}^{5} 2 - 0.5\sum_{i=2}^{11} 6$.

Solution:

$3\sum_{k=1}^{5} 2 - 0.5\sum_{i=2}^{11} 6 = 3 \times 10 - 0.5 \times 60 = 0$

Q3. R exercises

Q3-1.

(a) Which of the following is the output of `matrix(c(rep(0,times=4),rep(1,times=4)),ncol=2)`

$$\text{(i).} \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \text{(ii).} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \text{(iii).} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{(iv).} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Solution:

TBD

(b) Suppose we define an R vector by `y <- c(3,NA,-1,4,NA,-2)`. What will `y[y>0]` give you?

(i). A vector of the positive elements and NA values of `y`.

(ii). A vector of the negative elements of `y`.

(iii). A vector of all `NA`s.

(iv). A vector of `TRUE`s and `FALSE`s.

(v). A vector of `TRUE`s and `FALSE`s and `NA`s.

Solution:

TBD

Q3-2.

(a) Which of the following code successfully construct the matrix $\mathbb{A} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}$

(i). `A <- matrix(c(1,1,2,2,3,3) ,nrow=3)`

(ii). `A <- cbind(c(1,1),c(2,2),c(3,3))`

(iii). `A <- t(matrix(c(1,1,2,2,3,3) ,nrow=2))`

(iv). `A <- c(c(1:3),c(1:3))`

Solution:

TBD

(b) Suppose `X` is a matrix in R. Which of the following is NOT equivalent to `X`?

(i). `t(t(X))`

(ii). `X %*% matrix(1,ncol(X)`

(iii). `X*1`

(iv). `X%*%diag(ncol(X))`

Solution:

TBD

Q3-3.

(a) Which of the following is the matrix $\mathbb{A}$ generated by

```
A <- t(matrix(c(rep(1,times=2),rep(3,times=2), 6, 4),ncol=3))
```

(i) $\mathbb{A} = \begin{bmatrix} 1 & 1 \\ 3 & 3 \\ 6 & 4 \end{bmatrix}$

(ii) $\mathbb{A} = \begin{bmatrix} 1 & 3 & 6 \\ 1 & 3 & 4 \end{bmatrix}$

(iii) $\mathbb{A} = \begin{bmatrix} 1 & 3 \\ 1 & 6 \\ 1 & 3 \end{bmatrix}$

(iv) $\mathbb{A} = \begin{bmatrix} 1 & 1 & 3 \\ 3 & 6 & 4 \end{bmatrix}$

Solution:

TBD

(b) Which of the following successfully select the first five odd elements of the vector $x = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11)$? (List all that apply. Do not list commands that will give an error)

(i) `x[rep(c(TRUE,FALSE),each=5)]`

(ii) `x[rep(c(TRUE,FALSE),times=5)]`

(iii) `x[rep(c(TRUE,FALSE),length=9)]`

(iv) `x[rep(c(TRUE,FALSE)][1:5]`

(v) `x[rep(c("TRUE","FALSE"),5)]`

(vi) None of the above

(vii) All of the above

Solution:

TBD

Q3-4.

Define the matrix A as:

```
##      [,1] [,2]
## [1,]    0    3
## [2,]    1    3
## [3,]    1    2
```

What is the output of `apply(A,2,mean)`?

(i). A vector of length 3 corresponding to the average of each row of `A`.

(ii). A vector of length 2 corresponding to the average of each column of `A`.

(iii). The mean of all the values in `A`.

(iv). The mean of the second column of `A`.

(v). The mean of the second row of `A`.

Solution:

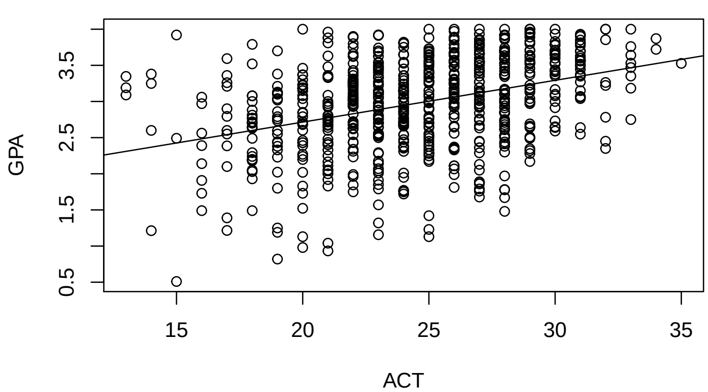
(ii). A vector of length 2 corresponding to the average of each column of `A`.

Q4. Fitting a linear model by least squares

Q4-1.

The admissions officer at a large state university wants to assess how well academic success can be predicted based on information available at admission. She collects data on freshman GPA and highschool ACT exam scores for 705 students in an R dataframe called `gpa`. The plot below shows a line fitted to a scatterplot of the points in the dataset.

```r
gpa_lm <- lm(GPA~ACT,data=gpa)
plot(GPA~ACT,data=gpa)
abline(coef(gpa_lm))
```



(a) Explain in words the criterion that is used to obtain the fitted line in the plot above.

Solution:

The line is fitted by least squares. This minimizes the sum of squared residuals, where the residual for each student is the difference between the value of GPA for that student and the value predicted by their ACT score.

(b) Defining appropriate notation, write an equation for the fitted model in subscript form. At this point, you don't have to explain how the coefficients are calculated.

Solution:

Let y_i be the freshman GPA for student i, $i = 1, \ldots, n$ with $n = 705$. Let x_i be the corresponding ACT score. The model in subscript form is

$$y_i = b_1 x_i + b_2 + e_i, i = 1, \ldots, n$$

where e_i is the residual for student i.

(c) Defining appropriate notation, write an equation for the fitted model in matrix form. You still don't have to explain how the coefficients are calculated.

Solution:

TBD

(d) Now, explain using matrix notation how the model coefficients are calculated.

Solution:

TBD

(e) Define the *fitted values* for this model and explain how they are calculated.

Solution:

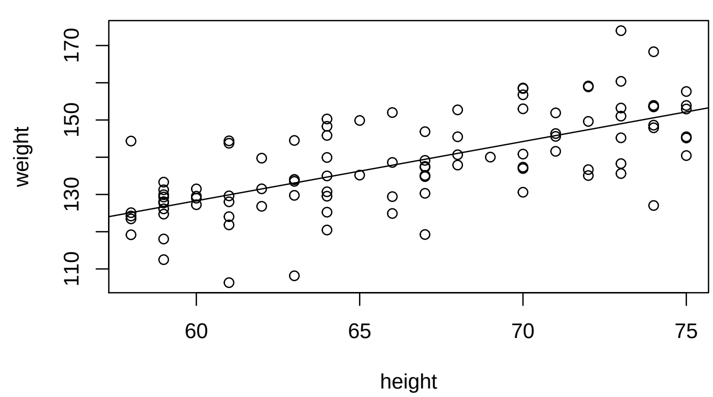
TBD

Q4-2.

Suppose we are interested in studying how well weight can be predicted by height. We collect the height and weight of 30 individuals and record them in the R vectors `height` and `weight`.

We fit a linear model and plot the data and the resulting fitted line using the following R code:

```
weight_lm = lm(weight ~ height)
plot(height,weight)
abline(coef(weight_lm))
```



(a) Write out the linear model in subscript notation including the following coefficients from `weight_lm`. Make sure to define appropriate notation.

```
## (Intercept)      height
##      32.822       1.591
```

(b) Use matrix notation to explain how these coefficients were calculated.

(c) The height for observation $i = 10$ is 66. Write out the formula for the fitted value for this observation. You do not need to simplify your calculation.

(d) Use matrix notation to write out an expression for the fitted values of the model. Make sure to define appropriate notation.

Solution:

(a) Let y_i be the weight of observation i, $i = 1, \ldots, 100$, and x_i be the corresponding height. The model in subscript form is

$$y_i = 1.591 \times x_i + 32.822 + e_i, \;\; i = 1, \ldots, 100$$

where e_i is the residual for observation i.

(b) Define the column vector of coefficients as $\mathbf{b} = (1.591, 32.822)$. Let $\mathbf{y}$ be the column vector $(y_1, \ldots, y_{100})$ and let

$$\mathbb{X} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_{100} & 1 \end{bmatrix}$$

We obtain $\mathbf{b}$ using the equation

$$\mathbf{b} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbf{y}$$

(c) We have the fitted value

$$\hat{y}_{10} = 1.591 \times 66 + 32.822$$

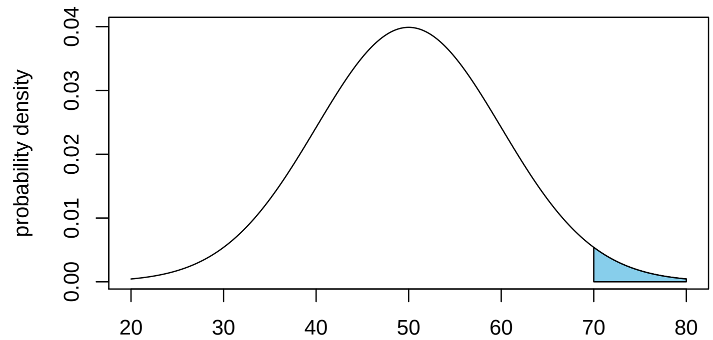
(d) Define $\mathbb{X}$ and $\mathbf{b}$ as above. Let $\hat{\mathbf{y}}$ be the column vector $(\hat{y}_1, \ldots \hat{y}_{100})$ where $\hat{y}_i$ is the fitted value corresponding to observation i. The fitted values are given by:

$$\hat{\mathbf{y}} = \mathbb{X}\mathbf{b}$$

Q5. Probability exercises

Q5-1.

The figure below shows the probability density function of a normal random variable X.



(a) By looking at the probability density function, estimate the mean and standard deviation of X. Use these estimates for the subsequent parts of this question.

Solution:

The center is at about 50. The points of inflection on the density are at about 40 and 60, which should be the mean plus/minus one standard deviation. It looks like about 95% of the area is between 30 and 70, which should be the mean plus/minus two standard deviations. These facts are consistent with a mean of 50 and an SD of 10.

(b) Write a probability statement about the random variable X that corresponding to the shaded area.

Solution:

The shaded area is $\mathrm{P}(X > 70)$.

(c) Write an integral corresponding to this shaded area.

Solution:

$$\int_{70}^{\infty} \frac{1}{\sqrt{2\pi 10^2}} \exp\left\{-\frac{(x-50)^2}{10^2}\right\} dx$$

(d) Write R code to evaluate this integral numerically.

Solution:

```
1-pnorm(70,mean=50,sd=10)
```

```
## [1] 0.02275013
```

It is acceptable not to label arguments, but then you have to get them in the right order!

Q5-2.

Let Y be a discrete random variable that takes values 0, 1, or 2 with probabilities 0.25, 0.5, and 0.25, respectively.

(a) What is the expected value of Y?

(b) What is the variance of Y?

Solution:

(a) The expected value is

$$0.25 \times 0 + 0.5 \times 1 + 0.25 \times 2 = 1$$

(b) The variance is

$$0.25 \times (0-1)^2 + 0.5 \times (1-1)^2 + 0.25 \times (2-1)^2 = 0.5$$

Alternatively, we can calculate the expected value of X^2 first:

$$\mathrm{E}(X^2) = 0.25 \times 0^2 + 0.5 \times 1^2 + 0.25 \times 2^2 = 0.5 \times 1 + 0.25 \times 4 = 1.5$$

and then

$$\mathrm{Var}(X) = 1.5 - 1^2 = 0.5$$

Q5-3.

The probability density function of the standard exponential distribution is e^{-x} for $x \geq 0$. Suppose X is a standard exponential random variable. Write out an integral corresponding to $\mathrm{P}(1 \leq X \leq 5)$.

Solution:

$$\int_1^5 e^{-x} dx$$
