Quiz 1, STATS 401 W18

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.

This version lists all the questions currently in the quiz generator. The quiz will have one question drawn at random from each of the five categories. No new questions will be added after Wednesday 10/3. Small changes may be made.

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

(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

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

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

(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.

Q1-2.

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

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

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

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

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.

Q2-2.

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

Q2-3.

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

Q2-4.

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

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)`

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

(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.

Q3-2.

(a) Which one of the following lines of code successfully constructs 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))`

(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))`

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

(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

Q3-4.

(a) 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`.

(b) For each of the lines of code below, say whether it will correctly make 50 draws from the normal$(100, 20)$ distribution. Among the correct answers, comment briefly on some strengths and weaknesses from the perspective of writing good R code. Which answer do you think is the best code, and why?

(i) `rnorm(50,20,100)`

(ii) `rnorm(100,20,50)`

(iii) `rnorm(100,20,n=50)`

(iv) `rnorm(mean=100,sd=20,n=50)`

(v) `rnorm(n=50,mean=100,sd=20)`

(vi) `replicate(rnorm(100,20),50)`

(vii) `replicate(rnorm(n=1,mean=100,sd=20),n=50)`

(viii) `rnorm(50)*20+100`

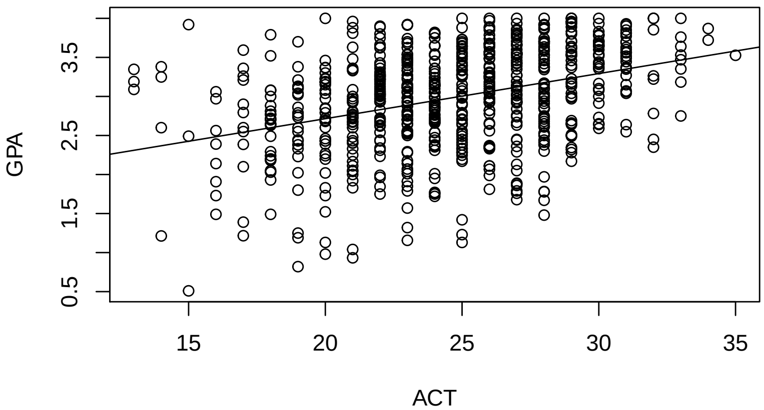
(ix) `100+sqrt(20)*rnorm(50)`

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.

```
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.

(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.

(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.

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

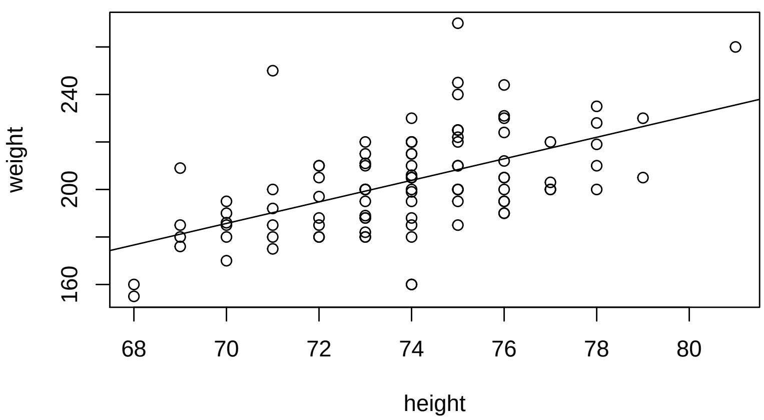
(e) Write an equation using subscript notation for the *fitted value* for the ith baseball player. Write a sentence to explain the interpretation of this fitted value.

Q4-2.

A statistician employed by a major league baseball team is asked to assess the range of typical weights for major league baseball players of a given height. She obtains data from http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_MLB_HeightsWeights and reads them into R as a dataframe including variables 'Height' (in inches) and 'Weight' (in pounds) for each of 1035 Major League Baseball players. She starts by analyzing just the first 100 players.

She fits a linear model and plots 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 fitted linear model using subscript notation, including the following coefficients from `weight_lm`. This means you are asked to use actual numbers, rather than letters, for the model coefficients. Make sure to define any notation you introduce.

```r
round(coef(weight_lm),3)
```

```
## (Intercept)      height
##    -131.652       4.534
```

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

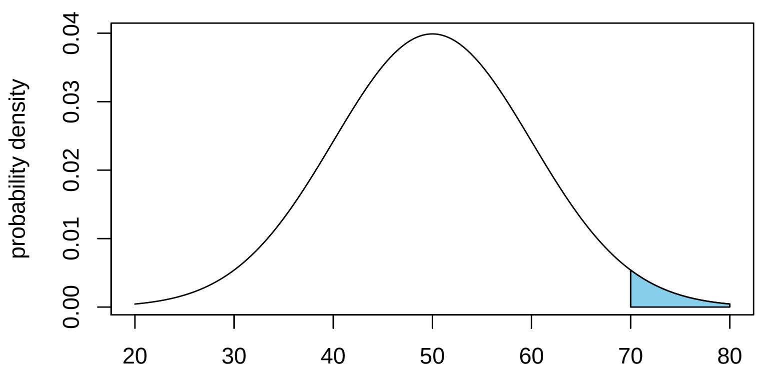
(c) The tenth observation corresponds to Adam Stern, an outfielder for the Baltimore Orioles. His recorded height is 71 inches. Write out the formula for the fitted value for this observation. You do not need to simplify your calculation.

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

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.

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

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

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

Q5-2.

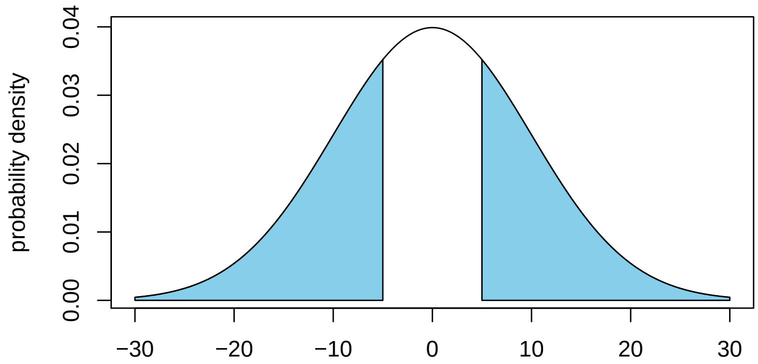
Let Y be a discrete random variable that takes values 0, 1, or 2 with probabilities 0.25, 0.5, and 0.25, respectively.

(a) What is the expected value of Y?

(b) What is the variance of Y?

Q5-3.

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.

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

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

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