IT 223 -- Takehome Midterm Exam

February 11, 2026

Submit your answers to these questions in an MS Word or PDF file. At the beginning of your document, include your name, and submission date. Also include this honor statement at the top of your document: "I completed this midterm exam by myself without the help of any other persons."

Part A – Short Essay Question. 15 points.

For full credit, answer in complete sentences and paragraphs with an introduction and conclusion. Only answer one out of the three questions. About one third to one half of a page.

1. What are some statistics that measure the center of a (not necessarily normal) histogram? What are some statistics that measure the spread of such a histogram? How do you use R to compute these statistics?

2. What do these R functions do? **`dnorm pnorm qnorm rnorm`**
   Give examples of how to use each of these functions.

3. What is a controlled study for medical research? What is done in such a study to reduce the influence of lurking variables, also called confounding variables. If you look up any information for this question on the web, cite your sources.

Part B – Analyze Univariate Dataset. 10 points each.

This dataset consists of blood pressure readings from patients in supine position, which means lying down.

```
> print(bloodPressure)
[1]  136 145 140 147 142 132 160 137 136 149 158 120 135 150
```

1. For the BloodPressure data vector, use R to perform the following computations Q0, Q1, Q2, Q3, Q4, sample mean, and sample standard deviation. Include your R source code, R output, and R comments to show what you are trying to a do. An R comment starts with a **#** character like this:
   `> # This is an R comment`

2. Use R to create these graphs of the BloodPressure data vector. Include your R source code and the graphs that it produces.
   (a) histogram (b) boxplot (c) scatterplot of data vs. observation number.
   Supply labels for the horizontal and vertical axes and a main title for the graphs.
   What does each graph tell you?

3. Compute the z-scores of the BloodPressure data vector. Are there any outliers according to the z-scores?

Continued on Page 2

Part C: Analyze a Bivariate Dataset. 10 points each.

This dataset consists of two data vectors Years of Job Experience (e) and Salary (s) in thousands of dollars:

```
> # Print Years of Job Experience (e).
> print(e)
[1]  5  3 15  7 20  2 12  4  1 10  3 18  6 14  2
> # Print Salary (s) in thousands of dollars.
> print(s)
[1]  90  65 150  60 200  55 120  80  45 110  75 140  65 130  40
```

Include your R source code, R output, and R comments to show what you are trying to a do. An R comment starts with a **#** character like this:

```
> # This is an R comment
```

1. (10 pts) Use R to calculate the sample mean and sample SD for the data vectors e and s. Also use R to calculate the correlation r between e and s.

2. (10 pts) Use R to obtain the linear regression model for predicting s from e. Type the regression equation.

3. (10 pts) Use the sample statistics in Part D, Problem 1 to calculate the estimated regression equation by hand. Verify that it is the same as the equation you found in 2.

4. (5 pts) For an employee with 8 years of job experience, what is his or her predicted salary, according to the estimated regression equation.

5. (5pts) Calculate the R-squared value. What does this tell you?

6. (10 pts) Use R to create a plot of s vs. e. Include labels for each axis and a title for your plot. Use the R **`lines`** function to add the regression line to the scatterplot. Does the scatterplot indicate that the residuals are well behaved? Explain your answer.

7. (10 pts) Use R to create a plot of residuals vs. predicted values. Are the residuals unbiased or biased? Are they homoscedastic or heteroscedastic?

8. (10 pts) Use R to create a normal plot of the residuals. Include a title for this normal plot. Use the R function **`qqline`** to add a normal line to the normal plot. Interpret this normal plot.