Midterm Exam (217 points)

TYPE YOUR NAME HERE

Invalid Date

## Instructions

Complete each exercise either in this qmd file or on paper. Include your name as the author and the date completed in the YAML code at the top of this file. Any parts completed on paper should either (a) be handed to David Housman, (b) placed in SC 117, or (c) scanned into a pdf file. This qmd file should be rendered to an html, docx, or pdf file. Zip together all relevant files: Rproj, qmd, rendered file, any data or image files, and (optionally) the pdf file containing your answers completed on paper. Upload the zip file in Moodle. Points will be taken off if these instructions are not followed.

## Academic Integrity

In completing this exam, you may use any inanimate resource associated with this course: Moodle and the texts and class notes linked there, handouts and your notes, R, RStudio, and spreadsheet software such as Excel or Google Sheets. You may not use other resources nor communicate with any person except to ask the instructor clarification questions.

By submitting this assignment, the author attests to abiding by the above statement and the *Collaboration and Academic Integrity* policy stated in the course syllabus.

## Exercise 1 (23 points)

A manufacturer randomly selected 101 resistors and measured their resistances. The data (in ohms) is presented in ascending order in the table.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 213.0 | 215.2 | 225.2 | 225.3 | 227.3 | 227.6 | 227.6 | 228.8 | 230.4 | 230.5 |
| 230.6 | 231.7 | 231.7 | 233.2 | 233.2 | 234.6 | 234.6 | 234.7 | 234.8 | 235.2 |
| 235.5 | 236.0 | 236.1 | 236.1 | 236.5 | 236.7 | 237.0 | 237.2 | 237.4 | 237.6 |
| 237.7 | 237.7 | 237.8 | 238.0 | 238.1 | 238.2 | 238.3 | 238.5 | 239.0 | 239.2 |
| 239.2 | 239.4 | 239.4 | 239.5 | 239.5 | 239.5 | 239.7 | 240.0 | 240.3 | 240.5 |
| 240.6 | 240.6 | 241.1 | 241.2 | 241.6 | 242.0 | 242.1 | 242.2 | 242.2 | 242.3 |
| 242.7 | 242.8 | 242.8 | 243.3 | 243.4 | 243.4 | 243.4 | 244.6 | 244.8 | 245.4 |
| 245.6 | 245.6 | 245.7 | 245.8 | 245.8 | 246.1 | 246.8 | 246.9 | 247.2 | 247.3 |
| 248.3 | 248.3 | 248.7 | 248.9 | 249.3 | 249.8 | 250.0 | 250.3 | 250.4 | 251.8 |
| 252.0 | 252.1 | 252.1 | 252.5 | 252.9 | 254.9 | 255.2 | 256.5 | 257.9 | 258.1 |
| 258.2 |  |  |  |  |  |  |  |  |  |

1. (9 points) State with explanation the first quartile, median, and third quartile. Include appropriate units.
2. (8 points) Sketch a box plot of the data. Use appropriate axis label(s).
3. (6 points) Estimate with explanation the mean and standard deviation resistance. Use appropirate units in your answers.

## Special Instruction for Exercises 2-4

When asked to calculate something in these exercises, use only simple arithmetic ($+,−,×,÷$, and power). For full credit, show the arithmetic perfomed and give a final answer in decimal form rounded to four decimal places, e.g., $1/3≈0.3333$. If an exact answer can be expressed in decimal form with fewer than four decimal places, it is okay to drop the zeros, e.g., $2/5=0.4$.

## Exercise 2 (25 points)

Calculate the following probabilities.

1. (5 points) A fair 10-sided die is rolled until a 3 appears. What is the probability that the 3 appears on the fifth roll?
2. (5 points) You will draw a ball from a mystery box. You are told that the probability of drawing a green ball is 0.7, the probability of drawing a spotted ball is 0.6, and the probability of drawing a ball that is both green and spotted is 0.4. What is the probability of drawing a ball that is green or spotted?
3. (5 points) You offer 5 fiction and 10 nonfiction books of interest to a friend. If she chooses four books at random, what is the probability that they are all fiction?
4. (5 points) Suppose that 70% of the women who use the Acme early pregnancy test are actually pregnant and that the accuracy of the test is 90% (that is, a pregnant woman has a 90% chance of testing positive and a woman who is not pregnant has a 90% change of testing negative). If a woman tests positive, what is the probability that she is pregnant?
5. (5 points) Suppose $A$ and $B$ are independent events for which $P\left(A\right)=0.8$ and $P\left(B\right)=0.6$. Find $P\left(A∩B\right)$ and $P\left(A∪B\right)$.

## Exercise 3 (20 points)

Let $X$ and $Y$ be random variables with the joint probability distribution function given in the table.

| $f\left(x,y\right)$ | $x=0$ | $x=1$ |
| --- | --- | --- |
| $y=1$ | 0.6 | 0.3 |
| $y=2$ | 0.1 | 0.0 |

1. (4 points) Calculate the marginal distributions $f\_{X}\left(x\right)$ and $f\_{Y}\left(y\right)$.
2. (8 points) Find the marginal distribution means and variances.
3. (4 points) Find the covariance and correlation coefficient.
4. (4 points) Calculate the mean and variance for the random variable $W=2X+3Y$.

## Exercise 4 (22 points)

Consider the small data set given in the table.

|  |  |  |  |
| --- | --- | --- | --- |
| $x$ | 0 | 2 | 4 |
| $y$ | 0 | 2 | 7 |

1. (8 points) Calculate the mean and standard deviation for each of the two variables.
2. (2 points) Calculate the correlation coefficient.
3. (6 points) Obtain the regression equation in the form $y=mx+b$.
4. (6 points) Graph the data and regression line on a piece of graph paper.

## Exercise 5 (54 points)

Consider the data set Bears.csv and documentation in Bears.pdf provided with this exam.

1. (4 points) Read the bear data into the work space environment. Suppress messages.
2. (2 points) Change the values of the Sex variable from “1” and “2” to “Male” and “Female”, respectively.
3. (6 points) Obtain boxplots for weight as a function of sex. Include appropriate labels.
4. (4 points) What conclusions about the real-world situation are revealed by two striking features of the part (c) graphic?
5. (6 points) Obtain boxplots of bear weight as a function of the month the measurement was taken (thought of as a factor variable). Include appropriate labels.
6. (2 points) The documentation states, “Since bears hibernate in the winter, their body shape probably depends on the season.” Does the graphic obtained in part (e) support this statement? Why or why not?
7. (2 points) Create a data set containing only the female bears.
8. (8 points) Obtain a scatter plot of female bear weight as a function of bear length with models of the forms $y=mx+b$ and $y=ax^{3}$ overlaid. Include appropriate labels.
9. (4 points) Which of the two models appears to be a better fit to the data. Why?
10. (8 points) Obtain a symbolic form for the model chosen in part (i) and the model’s standard error and coefficient of determination.
11. (8 points) Provide interpretations for the model, standard error, and coefficient of determination found in part (j).

## Exercise 6 (20 points)

Consider the geyser data set found in the MASS package and the corresponding documentation found in Help.

1. (8 points) Obtain a histogram of duration with bin widths of 0.2. Include appropriate labels.
2. (4 points) Interpret the vertical bar nearest to the horizontal number 4 in the real-world context.
3. (2 points) Describe the most striking feature of the histogram.
4. (6 points) Based upon this striking feature, add a new variable to geyser that groups the data appropriately. Obtain the count, median, and interquartile range of duration grouped by your new variable.

## Exercise 7 (18 points)

Consider the data sets beaver1 and beaver2 found in the datasets package and the corresponding documentation found in Help.

1. (6 points) Create a single data set beaver that binds the two together and adds a factor column that indicates which beaver and a numeric column that indicates which reading. The table below shows what four of the rows of beaver should look like.

| row | day | time | temp | activ | animal | reading |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 346 | 840 | 36.33 | 0 | 1 | 1 |
| 2 | 346 | 850 | 36.34 | 0 | 1 | 2 |
| 115 | 307 | 930 | 36.58 | 0 | 2 | 1 |
| 116 | 307 | 940 | 36.73 | 0 | 2 | 2 |

1. (6 points) Obtain a time series plot of the two beavers temperatures by placing reading on the horizontal axis, temp on the vertical axis, and animal as the color. Include appropriate labels.
2. (6 points) What can you conclude about the real-world situation based upon the three most striking features of the graphic?

## Exercise 8 (10 points)

An urn contains 5 blue, 6 green, and 3 red balls. If 4 balls are randomly drawn from the urn without replacement, approximate with a simulation the probability of obtaining at least 1 ball of each color. For example, (blue, red, blue, green) would qualify but (blue, blue, green, blue) would not. Set the seed, write a simulation to answer this question, run the simulation at least 10,000 times, and answer the question in a sentence.

## Exercise 9 (10 points)

A four-sided die is rolled until an odd number appears or until it has been rolled three times. For example, one outcome could be rolls of 4 and 1, and a second outcome could be rolls of 4, 2, and 4. Let $X$ be the maximum number rolled.

1. (4 points) Determine the probability distribution function.
2. (4 points) Calculate the mean and standard deviation of $X$.
3. (2 points) Draw a graph of the cumulative distribution function.

## Exercise 10 (10 points)

Consider the experiment and random variable described in the previous exercise. Write and run a simulation at least 1000 times to estimate the probability distribution function, mean, and standard deviation of X. Draw a spike graph or histogram of the empirical probability distribution function. Compare your results with those obtained in the previous exercise.

## Exercise 11 (5 points)

The mean and standard deviation length of a sample of 100 widgets are 30 and 2 millimeters, respectively. Assuming it is bell shaped, sketch by hand a histogram of the lengths for this sample.