Name (Last, First):

Student ID:

Assignment 6

METCS544A3A4_F2024

Instructions:

1. **This assignment has no specific R programming questions, but you're encouraged to use R to plot and graph the data and calculate relevant statistic summaries.**
2. For answering programming questions, please use Adobe Acrobat to edit the pdf file in two steps **[See Appendix: Example Question and Answer]**:
    a. Copy and paste your R code as text in the box provided (so that your teaching team can run your code);
    b. Screenshot your R console outputs, save them as a .PNG image file, and paste/insert them in the box provided.
    c. Show all work - credit will not be given for code without showing the code in action by including the screenshot of R console outputs.
3. To answer non-programming questions, please type or handwrite your final answers clearly in the boxes. Show all work - credit will not be given for numerical solutions that appear without explanation in the space above the boxes.
4. **[Total 117 pts = 54 + 27 + 24 pts + 12 Extra Credit pts]**

Grading Rubric

Each question is worth 3 points and will be graded as follows:

3 points: Correct answer with work shown

2 points: Incorrect answer but attempt shows some understanding (work shown)

1 point: Incorrect answer but an attempt was made (work shown), or **correct answer without explanation (work not shown)**

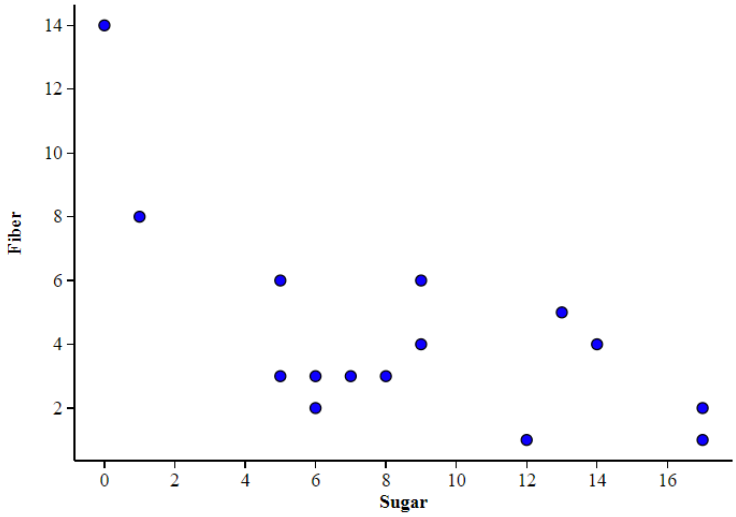
0 points: Left blank or made little to no effort/work not shown

Reflective Journal [3 pts]

(Copy and paste the link to your live Google doc in the box below)

Part I. Exploring Two Variable Data: Linear Regression (54 pts)

1. Fiber helps regulate the body's use of sugars, helping to keep hunger and blood sugar in check. In children's cereal, which is usually loaded with large amounts of sugar, scientists wanted to investigate if a larger sugar content led to less nutritional value overall, such as lower fiber. A scatter plot of 15 randomly selected children's cereals, along with some selected summary statistics, are given below. Both sugar and fiber are measured in grams per serving.



Sugar	Fiber
$\bar{x} = 8.6$	$\bar{y} = 4.33$
$S_x = 5.18$	$S_y = 3.31$
$r = -0.654$	

a) Does the scatterplot indicate that it is okay to create a linear model? Explain.

Answer:

b) Find the slope of the least-squares regression line. Interpret this value in context.

Answer:

c) What point must be on the least-squares regression line?

Answer:

d) Find the intercept of the least-squares regression line. Interpret this value in context.

Answer:

e) Write the equation of the linear model.

Answer:

f) If you pick up a cereal and see that it has 3 grams of sugar, what is the predicted fiber content?

Answer:

g) What is the residual for the cereal with 9 grams of sugar and 6 grams of fiber?

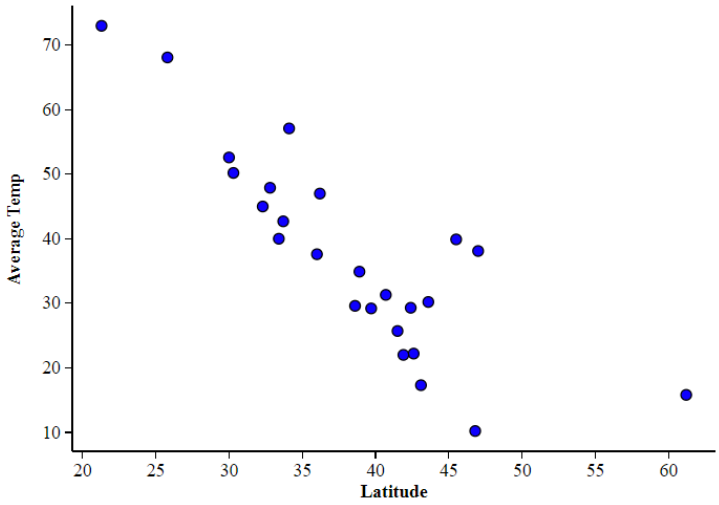
Answer:

h) Calculate the value of R^2 and interpret this value in context.

Answer:

2. We know that the further you get from the equator, the colder the climate becomes. For the following cities, we have assembled their latitude (angular distance from the equator, measured in degrees) and their average temperature in December 2021. The data and the scatterplot are given below.

City	Latitude (°N)	Dec Average Temp (°F)
Albany, NY	42.6	22.2
Anchorage, AK	61.2	15.8
Atlanta, GA	33.7	42.7
Austin, TX	30.3	50.2
Bismarck, ND	46.8	10.2
Boise, ID	43.6	30.2
Boston, MA	42.4	29.3
Charleston, SC	32.8	47.9
Chicago, IL	41.9	22
Cleveland, OH	41.5	25.7
Denver, CO	39.7	29.2
Honolulu, HI	21.3	73
Jackson, MS	32.3	45
Knoxville, TN	36	37.6
Las Vegas, NV	36.2	47
Los Angeles, CA	34.1	57.1
Madison, WI	43.1	17.3
Miami, FL	25.8	68.1
Newark, NJ	40.7	31.3
New Orleans, LA	30	52.6
Olympia, WA	47	38.1
Portland, OR	45.5	39.9
Roswell, NM	33.4	40
St. Louis, MO	38.6	29.6
Washington, DC	38.9	34.9



a) Describe the relationship between latitude and average December temperature in the US.

Answer:

b) Find the LSRL using R.

Answer:

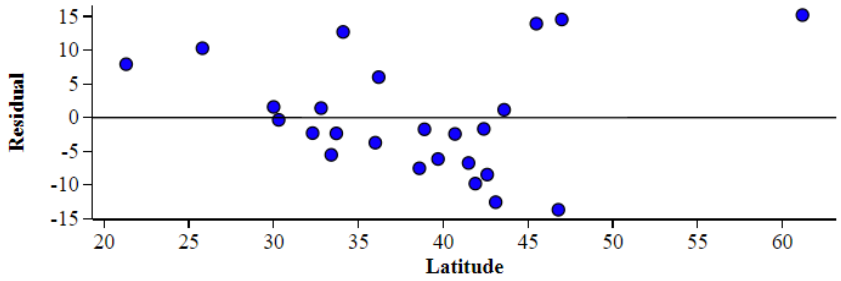
c) Find the following summary statistics using R:

$\bar{x} =$ _____ $\bar{y} =$ _____ $S_x =$ _____ $S_y =$ _____ r = _____

d) Using the summary statistics from (c), show that the LSRL you find matches what you found in (b).

Answer:

e) Below is the residual plot for latitude and average December temperature.



f) Given the residual plot above, is a linear regression model appropriate for this data set? Explain.

Answer:

g) Kansas City, MO is located at 39.1°N and had an average December 2021 temperature of 42°F. What is the residual for this point?

Answer:

Part II. Collecting Data: Sampling (27 pts)

Your local movie theater has worked hard to improve its customers' movie-going experience. To that end, it would like to take a random sample of its recent customers and email them a survey about their experience.

1. For the following situations, identify the bias and explain how they would introduce bias into the results.

a) You send out an email to the first ten customers who buy tickets to a movie.

Answer:

b) Of the ten movies playing in a single day, you randomly select one and email the survey to everyone who attended that movie.

Answer:

c) You set up a sign as patrons exit the theater and tell them to email support@movies.com if they would like to share their movie-going experience.

Answer:

d) You have the manager of the movie theater waiting outside the movie theater and interviewing every 10^{th} person who walks out about their movie-going experience.

Answer:

You obtain a list of the following data in a month at the movie theater: Order ID, Date, Customer Name, Customer Email, Movie Title, Show Time, Number of Tickets Purchased.

2. You would like to email out a survey to a sample of customers. For the following sampling methods, describe how you would use the information above to select a sample.

a) Simple Random Sample

Answer:

b) Stratified Random Sample

Answer:

c) Cluster Random Sample

Answer:

d) Systematic Random Sample

Answer:

e) A Census

Answer:

Part III. Collecting Data: Selecting Random Samples (24 pts + 9 Extra Credit pts)

1. The local university is interested in how many of its students plan on using some of the new facilities it installed over the summer. 3000 students attend the university, and they want to send out a questionnaire to 100 of them. They decide to take an SRS from an alphabetized list of students.

(a) Describe how you will select the sample.

Answer:

(b) Below is a sample of the table of random digits. Using the table and method described above, what are the first four numbers that will be part of your sample? **(12 pts)**

23328 99330 01231 42492 73831 02911 01524 32932 34334 74280 29357 29301

Answer:

2. Your school has a pep rally in the gymnasium. The following is a map of the gym, where each box represents a student's seat.

1	2	3	4	5	**G**	301	302	303	304	305
…	…	…	…	…	**y**	…	…	…	…	…
146	147	148	149	150	**m**	446	447	448	449	450
151	153	154	155	156	**F**	451	452	453	454	455
…	…	…	…	…	**l**	…	…	…	…	…
296	297	298	299	300	**o o r**	596	597	598	599	600

Freshman: Seats 1 – 150
Sophomores: Seats 151 – 300
Juniors: Seats 301 – 450
Seniors: Seats 451 – 600

Describe how you would use a random number generator to select 40 students to complete a survey with each of the following sampling methods. **NOTE: I want you to describe it, not actually do it!*

(a) Simple Random Sample

Answer:

(b) Stratified Random Sample

Answer:

(c) Cluster Sample

Answer:

3. **Extra Credit:** A study was done to determine if eating chocolate during pregnancy has an effect on infant temperament at the age of 6 months. The study asked about 300 pregnant women to report their chocolate consumption (Never/Seldom, Weekly, and Daily). Six months after they had given birth, the researchers then asked the mothers to report their child's overall temperament (activity, smiling, fear, etc.). They determined that "Maternal prenatal consumption of chocolate was associated with the infant temperament in this study. Mothers who reported consuming chocolate daily rated their infants at 6 months as more positively reactive and active". Was this an observational study or an experiment? Explain.

Answer:

4. **Extra Credit:** A large study in the UK was done to determine if there was a link between cell phone usage and brain cancer. The study set out to determine if using a cell phone more increased the likelihood of getting brain cancer. The study took about 800,000 women and asked them to report their cell phone use in 1999, in 2005, and again in 2009. In 2009, they recorded the proportion of women who had developed brain cancer. They concluded, "In this large prospective study, mobile phone use was not associated with increased incidence of glioma, meningioma, or non-CNS cancers." Is this an observational study or an experiment? Justify your answer.

Answer:

5. **Extra Credit:** McGraw Hill would like to test synchronous vs asynchronous learning for a new Algebra I classroom. They give a test to a group of 200 students who are brand new to the Algebra I classroom. They then assign half of the students to study Algebra I synchronously (with a classroom and traditional teacher) and half of the students to study Algebra I asynchronously (online and at their own pace). Students were given the same test at the end of the year, and the differences between the first and second tests were compared. Is this an observational study or an experiment? Justify your answer.

Answer:

THE END