Name (Last, First):	Student ID:
Kunjangada Arun Sushma	002473132

Assignment 5

NEU_COE_INFO6105_Fall2024

Instructions:

1. 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 or python code as text in the box provided (so that your teaching team can run your code);
    b. Screenshot your R or python 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 or python console outputs.
2. 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. **You're encouraged to use R or python to graph/plot the data and produce numerical summaries; please append your code and screenshot of the outputs at the end of your pdf submission.**
3. **[Total 111 pts = 27 + 33 + 33 + 15 pts + 3 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)

https://docs.google.com/document/d/1ptEhnYHniNtT1yxDPcvXK7LpJaPGzi80BCSGZZhom7Y/edit?usp=sharing

Part I. 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:

• This situation introduces convenience sampling bias and timing bias.

• Selecting only the first ten ticket buyers introduces convenience sampling bias, limiting the sample to readily available customers. This can lead to an unrepresentative sample, as early ticket buyers may have different characteristics or opinions than those who buy later.

• Additionally, focusing on the initial customers introduces timing bias, as it may capture responses from a specific time of day and exclude other customers.

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:

• This introduces selection bias based on movie preference.

• While randomly selecting one movie helps reduce some bias,it still doesn't represent the entire customer base. People who choose to watch a particular movie may have different tastes, or expectations compared to those who attend other movies.

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:

• This method introduces voluntary response bias.

• Only customers who are motivated enough to take the extra step of emailing will participate in the survey. This often results in responses from people who had either very positive or very negative experiences, potentially missing out on the opinions of those with more moderate views.

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:

• This method introduces systematic sampling bias.

• While interviewing every 10th person is an attempt at systematic sampling, it may not capture a truly random sample if there are patterns in how people exit the theater.

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:

1. Assign a unique number to each Order ID in your list.
2. Use a random number generator to select the desired number of samples.
3. Choose the Order IDs that correspond to the randomly generated numbers.
4. Use the Customer Email associated with those Order IDs to send out the survey.

b) Stratified Random Sample

Answer:

1. Divide the customers into strata (groups) based on Movie Title.
2. For each Movie Title, assign a unique number to each Order ID.
3. Determine how many samples you want from each movie.
4. Use a random number generator to select the desired number of samples from each movie stratum.
5. Use the Customer Email associated with the selected Order IDs to send out the survey.

c) Cluster Random Sample

Answer:

1. Define clusters based on a characteristic.
2. Assign a number to each cluster.
3. Randomly select a predetermined clusters using a random number generator.
4. Include all customers who are in selected clusters in your sample.

d) Systematic Random Sample

Answer:

1. Determine your sample size (n) and the total number of orders in the month (N).
2. Calculate the sampling interval k = N/n
3. Randomly select a starting point between 1 and k.
4. Starting from that point, select every kth order from your list.

e) A Census

Answer:

Email the survey to every customer on the list.

This would involve sending surveys to all individuals who purchased tickets during the specified month.

Part II. Collecting Data: Selecting Random Samples (33 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:

1. Assign each student a unique number from 0001 to 3000.
2. Use a random number generator or a table of random digits to select numbers within this range (0001-3000).
3. If a number outside the range (0001-3000) appears, ignore it and move to the next number.
4. If a number has already been selected, ignore it and move to the next number.
5. Continue this process until you have selected 100 unique numbers within the range.
6. The students corresponding to these 100 unique numbers form your sample.

(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: 23328 | 99330 | 01231 | 42492

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		301	302	303	304	305
…	…	…	…	…		…	…	…	…	…
146	147	148	149	150		446	447	448	449	450
151	153	154	155	156		451	452	453	454	455
…	…	…	…	…		…	…	…	…	…
296	297	298	299	300		596	597	598	599	600

Gym Floor

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:

Assign a unique number to each seat in the gym, from 1 to 600.

Use a random number generator to select numbers between 1 and 600.

If a number is generated that has already been selected, ignore it and generate another number.

Continue this process until you have 40 unique numbers.

The students sitting in the seats corresponding to these 40 unique numbers would form your sample.

(b) Stratified Random Sample

Answer:

1. Divide the students into four strata based on their grade level (Freshman, Sophomore, Junior, Senior).

2. Since we need to select 40 students in total, and we have four strata, select 10 students from each stratum to ensure equal representation.

3. For each stratum (grade level): a. Assign each student in the stratum a unique number. b. Use a random number generator to select 10 unique numbers within the range for that stratum. c. If a number is repeated, ignore it and generate another.

4. The 10 students corresponding to the selected numbers from each stratum will form your sample, giving you a total of 40 students (10 from each grade level)

(c) Cluster Sample

Answer:

• Define clusters as rows of seats in the gymnasium. Each row becomes a cluster.

• Assign each row (cluster) a unique number, from 1 to the total number of rows.

• Determine the average number of students per row to estimate how many rows you'll need to select.
• Use a random number generator to select row numbers.

• For each randomly selected row: a. Include all students in that row in your sample. b. Keep a running count of the total students selected.

• Continue selecting rows randomly and adding their students to your sample until you reach or just exceed 40 students.

3. 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:

This is an observational study.

The researchers did not impose any treatments on the pregnant women.

Instead, they observed and collected data on the women's existing chocolate consumption habits and later collected data on their infants' temperaments.

The researchers did not control or manipulate the chocolate consumption, which is a key characteristic of an observational study.

4. 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:

This is an observational study.

The researchers did not assign or control cell phone usage among the participants.

Instead, they collected data on the women's existing cell phone usage habits over time and observed the incidence of brain cancer.

The lack of imposed treatment or controlled variables classifies this as an observational study.

5. 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:

This is an experiment.

The researchers actively intervened by assigning students to two different learning methods (synchronous and asynchronous).

They controlled the independent variable (learning method) and measured its effect on the dependent variable (test scores).

The random assignment of students to treatment groups and the deliberate manipulation of the learning environment are key characteristics of an experimental design

Part III. Collecting Data: Experimental Design (33 pts)

1. A local hospital compiles data on the length of time a patient is in surgery and the length of their stay in the hospital after surgery. A newbie hospital worker notices that the length of time a patient is in surgery is highly correlated with their hospital stay after. Explain to the newbie why this correlation does not imply a causation.

Answer:

Correlation does not imply causation because there could be underlying factors affecting both variables.

For instance, more complex surgeries might take longer and require longer recovery times.

The severity of the patient's condition could influence both surgery duration and recovery time.

2. The good folks at Apples Inc want to create bags of pre-sliced apples for kids to easily put in their lunch boxes. They want to test different mixtures of preservatives (A, B, and C) on their Honeycrisp apples. They will treat all the apples in a bushel of Honeycrisps by randomly assigning them to each preservation treatment, and then comparing how long they are able to remain in the bag before they begin to brown.

(a) Identify the experimental units, the explanatory and response variables, and the treatments.

Answer:

Experimental units = Individual Honeycrisp apples

Explanatory variable = Preservation treatment (A, B, or C)

Response variable =Time before browning begins

Treatments = Preservative mixtures A, B, and C

(b) The researchers plan to use a completely randomized design. Describe how they should assign treatments to the 126 experimental units.

Answer:

1. Generate a random number between 1 and 126.
2. Assign the apple corresponding to that number to Treatment A.
3. Remove that number from the pool of available numbers.
4. Repeat steps 3-5 until Treatment A has approximately 42 apples.
5. Repeat the process for Treatments B and C, ensuring equal group sizes.

(c) The researchers suspect that the type of apple will influence how well the preservation method works. They want to repeat this experiment with 4 different types of apples. Describe how they should change the design of the experiment to account for this addition.

Answer:

Use a randomized block design

Create four blocks, one for each apple type.

Within each block, randomly assign apples to the three preservation treatments.

Analyze the browning time for apples in each treatment group within each block.

3. A study was conducted to determine if taking a daily dose of aspirin reduces the chance of catching the common cold. This study was conducted using 550 volunteers who were not already on an aspirin regimen. The subjects were randomly assigned to one of two groups: a treatment group who received a low dose of aspirin daily, or a control group who received a placebo. The subjects were unaware of what group they were in. At the end of the study, the subjects were asked to meet with a doctor to discuss if they had any symptoms pertaining to the common cold.

(a) Is this study an experiment or an observational study? Explain your answer.

Answer:

This is an experiment.

The researchers actively intervene by randomly assigning subjects to treatment (aspirin) and control (placebo) groups, and then measure the outcome (cold symptoms).

(b) What would be the advantage of having this study be double blind?

Answer:

The study prevents both researcher and subject bias.

Researchers can't unconsciously treat groups differently, and subjects can't be influenced by knowing their treatment, ensuring more objective results.

(c) Would blocking according to gender be worthwhile in this study? Explain your answer.

Answer:

Blocking by gender could be worthwhile if there's reason to believe men and women respond differently to aspirin or have different susceptibility to colds.

It would control for potential gender-based variations, increasing the precision of the treatment effect estimate.

(d) Describe what the “placebo effect” would look like in this experiment.

Answer:

The placebo effect would occur if subjects in the control group experience fewer cold symptoms simply because they believe they're taking medication.

4. Type 1 diabetes is thought to be caused by an autoimmune reaction (the body attacks itself by mistake) that destroys the cells in the pancreas that make insulin. It is a disease that currently has no cure. Researchers were to conduct a clinical trial on a promising new drug that helps the body produce its own insulin.

(a) You have 400 volunteers with Type 1 diabetes. Describe a completely randomized design for this experiment.

Answer:

Assign each volunteer a number from 1 to 400.

Use a random number generator to assign half (200) to the treatment group and half to the control group.

Administer the new drug to the treatment group and a placebo to the control group.

Monitor the volunteers for changes in insulin production and other relevant outcomes.

(b) You have 250 children (17 and younger) with Type 1 diabetes (120 females and 130 males), as well as 300 adults (18 and older) with Type 1 diabetes (190 females and 110 males). Describe a completely randomized block design.

Answer:

- Divide the volunteers into four blocks based on age and gender: Children-Female, Children-Male, Adults-Female, Adults-Male
- Assign a unique number to each volunteer within their respective block.
- Divide the total number of volunteers in each block by 2 to get the ideal group size for treatment and control.
- Within each block, generate random numbers to assign volunteers to the treatment or control group.
- Ensure that approximately equal numbers of volunteers are assigned to each group within each block.

(c) Explain why a matched pairs design would not work for this experiment.

Answer:

Matched pairs design typically involves each subject receiving both treatments at different times or paired subjects receiving different treatments.

For a drug trial on a chronic condition like diabetes, it's not feasible for subjects to switch between drug and placebo, nor is it ethical to withhold potential treatment from one member of a pair.

Part IV. Collecting Data: Scope of Inference (15 pts)

Directions: For the study below, answer the questions that follow. Then, in the graphic organizer, made a check on the correct scope of the study.

Lack of sleep is associated with increased risk of cardiovascular disease, depression, and other health concerns. A new study now shows that lack of sleep also affects social interactions, making people less willing to help others (empathy). Scientists placed 24 healthy volunteers in a functional magnetic resonance imager (fMRI) to scan their brains after eight hours of sleep and then again after a night of no sleep. The order that they received the treatment of sleep vs no sleep was randomly selected with the flip of a coin. They found that areas of the brain that form the theory of mind network, which is engaged when people empathize with others or try to understand other people's wants and needs, were less active after a sleepless night. "Here, we

found that a decrease in the quality of someone's sleep from one night to the next predicted a significant decrease in the desire to help other people from one subsequent day to the next." Source: https://www.sciencedaily.com/releases/2022/08/220823143827.htm

Is this an observational study or an experiment? Explain.

This is an experiment. The researchers actively manipulated the independent variable (sleep condition) by randomly assigning participants to either get a full night'ssleep or no sleep. They then measured the effect on brain activity and empathy.

Does this study involve random sampling or random assignment? Explain.

This study involves random assignment, not random sampling.

The 24 volunteers were not randomly selected from a larger population, but they were randomly assigned to the order in which they experienced the sleep and no-sleep conditions

Can this study make inferences about the population or only the study participants? Explain.

This study can only make inferences about the study participants, not the broader population. The 24 volunteers were not randomly sampled from a larger population, so the results may not be generalizable.

Can this study make inferences about cause and effect or just an association between variables? Explain.

This study can make inferences about cause and effect. The experimental design, with random assignment to treatment order and controlled manipulation of sleep, allows researchers to establish a causal relationship between lack of sleep and decreased empathy/willingness to help others.

Which box of inference does this study fall under?

	Inference about cause and effect	No Inference about cause and effect
Inference about the Population		
No Inference about the Population	yes	

THE END