COLLEGE OF CHARLESTON

Final Exam Review

Math 104-03: Elementary Statistics

Thursday December 5th

- Final Exam is on Hawkes taken at home
- No collaborating
- Opens at 6:00pm, Closes 8pm
- 2 hours to answer questions

Assignment: Final

Due in Weeks

This assignment is not available yet.

Details			
Start Date:	Dec 05, 2024 9:00AM ET		
Due:	Dec 05, 2024 9:00PM ET	**Grading Method:**	NA
Attempts Remaining:	1	**Max Points:**	90
Time Limit:	120 minutes		
Description:			

Chapter 1: Introduction to Statistics

- Understand the types of data
  - qualitative = labels or descriptions of traits
  - quantitative = counts or measurements
- Understand the difference between a population and a sample
  - population = particular group of interest
  - sample = subset of population from which data is collected
- Understand the difference between a population parameter and sample statistic
  - parameter = numerical description of population characteristic
  - sample statistic = numerical description of sample characteristic

Chapter 2: Graphical Descriptions of Data

- Be able to read and interpret a histogram
  - histogram = bar-graph depicting frequency distribution of quantitative data
- Identify if a distribution is
  - uniform
  - symmetric
  - skewed to the right
  - skewed to the left

Chapter 3: Numerical Descriptions of Data

- Measures of center. Compute:
  - sample mean, $\bar{x} = \frac{1}{n}\sum x_i$
  - median
- Identify mode.
  - identify if the dataset is unimodal, bimodal, multimodal, or has no mode
- Measures of dispersion. Compute:
  - sample variance, $s^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$
  - sample standard deviation, s
  - range
- Understand quartiles, interquartile range, and read a box and whisker plot

Chapter 4: Probability

- Probability = number of outcomes in event / the number of outcomes in sample space
- Compute probabilities of events for simple random experiments such as
  - rolling a die
  - picking cards from a 52 card deck
  - flipping a coin
- Understand the definition of probability
  - could 2.5 be a probability? NO!

Chapter 5: Discrete Distributions

- Calculate quantities from a given discrete random variable
  - expected value, $\mathbb{E}(X) = \sum x_i \mathbb{P}(X = x)_i$
  - variance, $\mathbb{V}(X) = \sum (x_i - \mu)^2 \mathbb{P}(X = x_i)$
  - standard deviation, $\sigma = \sqrt{\mathbb{E}(X)}$
  - probabilities, e.g. $\mathbb{P}(X < x)$
- Binomial distribution
  - find the probability, given the number of trials and the probability of obtaining a success: $\mathbb{P}(X = x) = nCx\ p^x(1 - p)^{n-x}$

Chapter 6: Continuous Distributions

- Describe the characteristics of the normal distribution
  - symmetric
  - bell-shaped
  - completely determined by μ and σ
  - total area under curve is 1
  - x-axis is horizontal asymptote
- Identify/compare normal distributions based on their means and standard deviation
- Find probability using a normal distribution & z-scores: $z = \dfrac{x - \mu}{\sigma}$

Chapter 7: Central Limit Theorem

- Sampling distribution = distribution of the values of a particular sample statistic for all possible samples of a given size, n
- What kind of distribution is the sampling distribution of sampling means as the sample size n is large?
- Find the mean and standard deviation of the sampling distribution of sample means
  - $\mu_{\bar{x}} = \mu$
  - $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Chapter 8: Confidence Intervals

- Understand difference between point estimates and interval estimates
- Compute confidence intervals for population mean (σ known) with the standard normal distribution: $(\mu - E, \mu + E)$ where $E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- Compute confidence intervals for population mean (σ unknown) with the Student's t-distribution: $(\mu - E, \mu + E)$ where $E = t_{\alpha/2}\frac{s}{\sqrt{n}}$

Chapter 10: Hypothesis Testing

- Be able to state H_0 and H_a for any hypothesis test
- Perform z-test for population mean (σ known) in two ways
  - Test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
  - Rejection regions
    - $z \leq z_\alpha$ (left-tail), $z \geq z_\alpha$ (right-tail), $|z| \geq z_{\alpha/2}$ (two-tailed)
  - P-values
    - p-value $\leq \alpha$ (reject null hypothesis)
    - p-value $\geq \alpha$ (fail to reject null hypothesis)
- Perform t-test for population mean (σ unknown)
  - Test statistic: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
- Perform z-test for population proportion
  - Test statistic: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

Chapter 12: Regression (more emphasis)

- Explanatory variable x and response variable y
- Understand the concept of linear correlation between variables
- Calculate the Pearson correlation coefficient

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

- Determine whether r is significant at α level of significance
- Find the slope and intercept of the least squares regression line $\hat{y} = b_0 + b_1 x$ given a small dataset

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n}$$

Artificial Intelligence

- Black box model: inputs = explanatory variables, outputs = response variable
- Myths of machine learning:
  - AI does NOT reason like humans
  - AI is only as good as the data used to train a model
- Identify the purpose and difference between training and testing data
  - training data is used to fit a model
  - testing data is used to evaluate a model on unseen data
- Parameters of a model
  - like b_0 and b_1 but there is not a formulate for them and there are **billions** of them
- Overfitting
  - if a model is overfit, it is unable to generalize on unseen data

Thank you so much for a great semester!
Please submit course evaluations.