

Exam

9 questions

-short answer

-computation (calculators are allowed)

-displaying distributions visually and numerically

covers chapters 1-6

test will be held in **Gilman 1002 @ 2:10** Friday the 11th

Chpt 2-Data

Two kinds of variables

categorical=used as labels, place into a category

Ex. Male or Female, New or Used, etc

quantitative=used as numbers with their context

Chpt 3-Displaying categorical data

Two ways to display categorical data

Pie Chart

i) all possible categories

ii) displays percent of whole for each category

Bar Graph

i) does not need to have all categories

ii) can display count/frequency or percent

Chpt 4-Displaying quantitative data

3 pictures-always well-labeled!!

1) Histogram

-choose equal size bins that cover the range of your data

-find the count/frequency for each bin

-plot bins on x-axis

-plot count/frequency on y-axis

2) Stem and Leaf Plot

-choose your stem

-draw a vertical line

-place stems (leading digits) on the left in ascending order

-place leaves (trailing digits) on the right in ascending order

-shows each individual observation

-used from smaller data sets

-can split stems if data are concentrated around one or two stems

-back-to-back stem and leaf display

3) Boxplot

-use five number summary

Describing Distributions

Always mention

i) **shape**

- mode
- symmetry vs. skewness
- outliers

ii) **center**

- designates the middle of the distribution

iii) **spread**

- variation in your data

Chpt 5-Describing Distributions Numerically

Either use Five Number Summary or Mean and Standard Deviation

Five Number Summary

- minimum, Q1, median, Q3, maximum

Median

- use $(n+1)/2$ to find the index/place of the median in the ordered data set
 - if n is odd, then $(n+1)/2$ is a whole number => easy
 - literally the middle number
 - if n is even, then $(n+1)/2$ is halfway between two numbers
 - average of the middle two numbers

Q1

- the first quartile, has 25% of the data values below it
- find the median, and split the ordered data into two halves
 - when n is odd, include the median in both halves
- Q1 is the median of the lower half

Q3

- the third quartile, has 75% of the data values below it
- find the median, and split the ordered data into two halves
 - when n is odd, include the median in both halves
- Q3 is the median of the upper half

Mean and Standard Deviation

Mean

- measure of the center of a distribution
- just the average, add up all the observations and divide the sum by n
- heavily influenced by outliers and skewed data
 - mean is always pulled in the direction of the outlier or skew

Standard Deviation

- measure of the spread of a distribution $s_x = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$
 - large standard deviation => data has a lot of variation
 - small standard deviation => data has little variation
 - be sure to define “large” and “small” relative to your data
- heavily influenced by outliers and skewed data
- standard deviation is always positive

Five Number Summary vs. Mean and Standard Deviation

- use five number summary for skewed data or data with outliers
 - mean and standard deviation are heavily influenced by outliers
 - mean is “pulled” in the direction of outliers or skew
 - skewed right mean > median
 - skewed left mean < median
 - standard deviation gets large in the presence of outliers and skew
 - median, along with Q1 and Q3, are more resistant to outliers and skew
- use mean and standard deviation for unimodal, symmetric data

Chpt 6-Normal Distribution and Standardizing

Standardizing Variables for a given data set

-allows us to compare different variables $z = \frac{x - \bar{x}}{s_x}$

- puts different variables on the same scale
- z = # of standard deviations a value (x) is from the mean
 - negative z means value (x) is below the mean
 - positive z mean value (x) is above the mean

Normal Distribution

- distributions that are unimodal and symmetric can be approximated by a Normal Distribution
- written as $N(\mu, \sigma)$, where μ = mean and σ = standard deviation
- Properties of a Normal Distribution
 - mean/median is the center of a Normal Distribution
 - median = mean
 - symmetric about the mean
 - area under the curve is 1
 - tails never touch the x-axis
 - 68-95-99.7 rule
 - 68% of the observations fall between $(\mu - \sigma, \mu + \sigma)$
 - 95% of the observations fall between $(\mu - 2\sigma, \mu + 2\sigma)$
 - 99.7% of the observations fall between $(\mu - 3\sigma, \mu + 3\sigma)$

Standard Normal Distribution

- $z = \frac{x - \mu}{\sigma}$
- puts all normal distributions on the same scale
 - converts them to a $N(0,1)$
 - mean = 0
 - standard deviation = 1

Z-table

- yields areas **below** a given z-value
- areas are in the middle of the table
- z-values are on the outside/perimeter
 - Converting a z-value to area/chance/probability
 - area below a z-value => look up z-value on the table
 - area above a z-value => look up z-value on table and then 1 - (that area)
 - Converting area to a z-value
 - given area below a z-value => look up area in the middle of table
 - given area above a z-value => do 1 - (that area), then look up area in the middle of table
 - Finding area between two z-values
 - find area below each z-value, then subtract the smaller area from the larger area

Formulas given on the test:

$$\bar{x} = \frac{\sum x}{n} \quad (\text{mean})$$

$$s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad (\text{standard deviation})$$

$$z = \frac{x - \mu}{\sigma} \quad (\text{use for normal distribution only})$$

$$z = \frac{x - \bar{x}}{s_x} \quad (\text{use when analyzing a data set only})$$

