STAT 4033

Exam 1 Notes Name: Cooper Morris

C1S1:

Population: the entire set of all potential measurement

Sample: any subset of a population

Simple Random Sample: A sample of size n taken in such a way that any group of size n has the same chance of being selected

Sampling Variability: different samples from the same population can lead to differences

Stratified Random Sampling: the population is broken into groups based off a characteristic. Then a SRS is taken from each group

Cluster Sampling: target population has many groups, groups are selected by SRS of the groups. All elements of each group are selected

Systematic Sample: A listing is generated over time, every k^{th} member is included in the sample.

Tangible Population: A population composed of members/individuals that exist.

Conceptual Population: A population composed of all values that can potentially be observed. They do not necessarily exist at any point in time.

Observations: The measurement, or set of measurements recorded from any individual in a sample.

Variables: The characteristics being observed from individuals.

Quanitative Variables: Possible values that represent *quantiles of something.* Numbers of things.

Ratio Variables: Inherent zero value and ratios between values make sense.

Interval Variables: No meaningful ratios and arbitrary zero

Qualitative: A variable that takes a category of possible values.

Nominal: Ordering of categories makes sense.

Ordinal: No inherent ranking in categories.

Observational Study: Observe a sample from a population with minimal interaction.

Experimental Study: A study performed where the environment of subjects is strictly controlled.

Response Variable(s): The variable(s) of interest in a study.

Explanatory Variable: Variables to explain changes in the response variable.

Confounding Variable(s): Variables unaccounted for i a study that may explain changes in the response variable.

C1S2:

Measures of Central Tendency: Values that represent where the"center" of a dataset is located.

Measures of Variability: Values that indicate how spread out the data are.

Mode: The measurement that occurs most often.

Median: The middle value in an ordered set.

Mean: The sum of all measurements divided by the total number of measurements.

p% Trimmed Mean: The p% lowest values and p% of the highest values are removed from data, mean is taken.

p^{th} percentile: Value such that p% of observations are at or below and (100-p)% are above.

Range: difference between largest and smallest data points.

More relative variation is higher CV, less relative variation is lower CV.

C1S3:

Histogram: Number of classes should be smallest whole number K that makes $2^K \geq$ number of measurements. For large data sets either $\log_2(n)$ or $2n^{1/3}$

Unimodal: One major peak

Bimodal: Two major peaks

Symmetric: Symmetric

Right Skewed: Long right tail, short left tail

Boxplots: Outliers are outside 1.5×IQR. Box goes from Q_1 to Q_3, horizontal line at median, whiskers to largest data point inside 1.5×IQR, X's for outliers

C2S1:

Probability: the chance that something happens.

Experiment: A process with an uncertain outcome

Sample Space ($\mathcal{S}$): The set of all possible outcomes in an experiment.

Outcome: Each individual and non-reducible element of a sample space

Event: A set of 1 or more outcomes

Union: For events A and B the union is all outcomes in A, B, or Both. A∪B

Intersection: The set of outcomes that are in both A and B. A∩B

Complement: The set of outcomes in the sample space not in A. A^C

Mutually Exclusive Events: Events that share no outcomes in common.

Sample Mean: $\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}$

First Quartile: $Q_1 = y_{25\%}$

Second Quartile: $Q_2 = \text{median}$

Third Quartile: $Q_3 = y_{75\%}$

Interquartile Range: $IQR = Q_3 - Q_1$

Sample Variance: $s^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$

Sample Variance: $s^2 = \frac{\sum_{i=1}^{n} y_i - \frac{(\sum y_i)^2}{n}}{n-1}$

Sample Standard Deviation: $s = \sqrt{s^2}$

Coefficient of Variation: $CV = \frac{\sigma}{|\mu|}$

Histogram Classes: $2^K \geq$ Measurements

Histogram Large Set Classes: $\log_2(n)$ or $2n^{1/3}$

Histogram Class Length: $\frac{Max-Min}{K}$

$P(\mathcal{S}) = 1$

$0 \leq P(A) \leq 1$

If A & B are mutually exclusive:

$P(A \cup B) = P(A) + P(B)$

$P(A^C) = 1 - P(A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$