Quantitative Analysis Cheat Sheet
Graduate Resource Center, University of New Mexico

1 UNDERSTANDING DATA

Useful Definitions and Terms

Types of studies:
Observational - *Data without an intervention e.g, cohort study*
Experimental - *Data with an intervention e.g, Randomized controlled trail*

Population: Entire collection of individuals or objects about which information is desired.

Sample: A subset of the population that is selected for analysis.

Random sampling: Every possible sample of a certain size has the same chance of being selected.

Types of variable:
Qualitative - *the information is non-numeric*
Quantitative - *the information is numeric*

Variable Measurement
Quantitative variables:
Discrete - *Counts of individual items or values*
Continuous - *Measurements of continuous or non-finite values*

Qualitative variables:
Binary/Dichotomous - *Occurs only in groups*
Nominal - *Groups with no rank or order*
Ordinal - *Groups with rank or order*

Levels of Measuring Variables
Nominal: Groups with no rank or order (e.g., academic department- Economics, Math, Biology)

Ordinal: Groups with rank or order (e.g., letter grade- A, B, C)

Interval: Has order, distance between rank is measurable, lacks natural zero (e.g., GRE scores)

Ratio: Has order, distance between rank is measurable, has natural zero (e.g., height)

2 DESCRIPTIVE STATISTICS

Describe Variables
Frequency Tables:
Absolute - *Count/tally of occurrence of each individual category*
Relative - *Percentage of each category relative to the total*
Cumulative - *Summation of relative frequencies*
Cross tabulation - *Grouped absolute frequency table for two or more variables*

Central Tendency:
Mode - *The most commonly occurring value*
Median - *Value that lies in the middle*
Mean - *The mathematical average*

Measure of Dispersion:
Range - *Highest value minus lowest value*
Standard Deviation - *how much each individual value is dispersed around the mean*
Variance - *Standard deviation squared*

Methods and Formulas
Mean:
$$Mean\ (\bar{a}) = \frac{Sum\ of\ all\ values\ in\ the\ sample}{Number\ of\ values\ in\ the\ sample}$$

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

Standard Deviation:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})^2}$$

INFERENCE FOR THE MEAN

Useful Concepts
Confidence Interval: Interval where we believe with C% confidence the population mean lies. Typically C = 95%, 99%, 90%

Margins of Error: How many percentage points will differ from the real population value

3 INFERENTIAL STATISTICS

Definitions and Terms
Inferential Statistics: To find relationship/association between two or more groups

Parameter: Unknown population characteristics such as population mean (μ) mean and variance (σ^2)

Statistical hypothesis: Statement about the parameters of one or more population.
Null hypothesis (H_0) - *Usually, no relationship between two variables being tested*
Alternate hypothesis (H_a) - *Usually, two or more variables are somehow related*

Statistically Significance: Result of statistical test being performed - the relationship is not due to random chance

α **level** or **p-value**: The probability or percentage of rejecting the null when it should not have been rejected

Independent variable: The presumed cause in an analysis

Dependent variable: The presumed effect in an analysis

Comparing Between Two or More Groups
t-test:
Paired t-test:
Analysis of Variance (ANOVA):

Association and Correlation
Chi Square
Pearson

Explaining a Dependent Variable
Simple vs. Multiple Regressions
Linear vs. Non-linear Regressions
E.g., - *ordinary least square, logit/probit, multinomial logit*

4 IMPORTANT ISSUES

Data is Messy
Data Cleaning: Correcting or removing inaccurate or non-uniform entries from the data set
Inspect: *Detect unexpected, incorrect, and inconsistent data*
E.g., Check for outliers, duplicates

Clean: *Fix or remove the anomalies discovered*
E.g., Transpose data structure, delete duplicates and inconsistent values

About Assumptions
All of the tests or models have certain assumptions

All models assume that the observations are independent and are randomly sampled from the entire population of interest

The relationship between two variable depends on the method used to fit the data

Reporting and Discussing Results
Figures:
Graphs of a variable - *E.g., histogram, bar graph*
Graphs of two variables - *E.g., scatter plot*

Tables:
Descriptive - *Usually reports mean, standard deviation, sample size, or frequency*
Hypothesis testing - *Report α level or p-value, group mean, sample size*
Regression - *Report regression coefficient, p-value, sample size*

NEED HELP?

Contact Us
Graduate Resource Center
Mesa Vista Hall, Suite 1057
Phone: 505-277-1407
Email: unmgrc@unm.edu
Website: https://unmgrc.unm.edu/