Week 1: Introduction

Types of Data and Measurement Scales

- Measurement and data are not the same thing
  - Measurement is the process
  - Data is the output
  - Items/Units we measure

Types of data

- Quantitative data are data measures along a numerical scale (continuous)
- Qualitative Data Descriptions that fall into categories
  - Often referred to as discrete
  - Frequencies, proportions or rates (discrete)
  - Items/units we count
- Measurement & Measurement Scales
  - The numbers, or symbols, are assigned such that the relationships amongst the numbers or symbols reflect relationships in the attribute studied
  - measurements

Type of Measurement Scales

Interval Scale

- Numbers are assigned to observations such that differences between any two numbers correspond to proportional differences in the underlying property studied, and there are equal intervals along the scale
- Examples
  - Temperature measuresed in degrees Farenheit
  - Directional distance from a reference point
  - Calendar date
  - Height above sea level
- differences in the measurement scale are the same as the underlying property
- Permissible transformations include any linear transformation
- Statistics such as the mean and standard deviation may be used
- Most of the parametric statistical tests may be employed as long as their underlying assumptions are met
- Interval scale values can be used to determine $=$, not equal, $>$, or $<$. Can also use sums and differences with meaning

Ratio Scale

- numbers are assigned to observations such that an interval scale has been reached and there is a zero ppoint which corresponds to zero point which corresponds to a zero, null state or absence of the underlying property
- Examples
  - Length
  - height
  - weight
  - distance
  - volume, weight
  - Cycle-time and time-to-repair
- basically the same as interval scale but 0 means that there is nothing there. OR an absence of data
- Good Transformations
  - Typically involve multiplying by a constant
- Statistics Available are
  - The geometric mean
  - Coefficient of variation
  - Any statistical test as long as their underlying assumptions are met

Ratio Scale vs Interval Scale

- Typically, when looking at a scale we think the set point is 0. Is 200 pounds twice as heavy as 100 pounds?
  - For ratio data the answer is yes
  - What if the scale was set at 50 to begin with?
    - Then our scale is interval and the answer is no

Absolute Scale

- Numbers are assigned to observations such that the number directly correspond to the underlying property being studied.
- Examples are
  - The number of defects
  - The number of scratches observed
  - The number of parts made
  - The number of safety accidents
  - The number of customer complaints
- The underlying property is the same as the measurement count (or just say data)
- Transformations
  - Some are permissible
  - Data on an absolute scale have some of the properties of ratio data
  - Mean, median and mode may be used
- Various statistical procedures may be used depending on resolution
  - In many cases, standard parametric methods may be used. In other cases nonparametric methods must be employed

Measurement as a Process: The 5 Big Aspects of Data

- We need to consider
  - Data costs money. We must make data both efficient and effective
  - How do we make data efficient and effective?
    - Make certain that the measurement process itself is effective
      - Capable
      - Acceptable

Remember is a process, data is the output

- Common Myths of Measurement
  - Is expensive
  - Has a digital readout
  - Uses radiation
  - Is made in Switzerland
  - Is only used by a Supervisor
  - Has a recent calibration sticker
- Measure measurement systems must demonstrate
  - Stability through time, or control
  - The ability to generate reliable or repeatable and reproducible measures
  - The ability to generate valide measures

Reliability in measurement is a measure of the PRecision of the Device / methodf

Validity in metrology is a measure of the Accuracy of the device/methods

- Precision vs Accuracy
  - Accuracy is the closeness of the desired target
  - Precision is the closeness of the predictions

Tools for Understanding Data

- There are many tools to help us understand data
  - Probability and statistics: to quanitfy and summarize data
  - Control charts: to determine if a process is stable
  - Experimental design: a;;s us to identify root causes of a problem so we can eliminate it or to identify and properly manipulate Special and Common Causes of Variability to improve optimize the process output

Big 5 Aspects of Data

Location

- Measures of location describe a middle or center point tendency of a distribution
  - Mean
  - Median
  - Mode

Spread

- Measures of dispersion reflect the variation or spread in a data set or distribution
- Range
- Standard Deviation
- Variance

Shape

- Type of distribution samples
- Skewness
  - degrees of departure from symmetry of a distribution
- Kurtosis is concerned with the peakedness of the distribution
  - Leptokurtic is a high peak. Positive Kurtosis.
  - Mesokurtic is a normal peak
  - Platykurtic is a lower peak

Time Sequence

Measures of Relationship

- Correlation
- Association
  - Same as correlation but on of the variables is discrete

Sampling Concepts

Population and Samples

- Population and Samples are not the same

General Approaches to sampling

- Nonrandom or judgement sampling
- Random or Probability Sampling
  - Simple random Sampling
  - Systematic random sampling
  - Stratified random sampling
  - Cluster sampling

Nonrandom Sampling

- Specimens or items are selected using personal judgement, reasoning, opinion or convenience

Random or probability sampling

- All specimens or items have a probability of being included in the sample

Simple Random Sampling

- Every possible sample of size n has an equal change of being selected

Systemic Random Sampling

- Specimens or items are selected at an interval

Stratified Random Sampling

- Specimens or items are divided into homogenous subsets or strata

Cluster Sampling

- Specimens or items are divided into groups that are homogenous between each other, but hetereogenous within

A **statistic** is a measure calculated from sample data that may be used to make inferences about a population - the average is a "statistic" - The range is another "statistic" - Ther are many more

descriptice statistics - Describe a characteristic of a sample - Frequently used to make inferences about population parameters - Represented by letters in english **Popoulation Parameters** - Describes a characteristic of the population - Represented by Greek letters (with a few exceptions)

```
# install.packages("devtools")
# require(devtools)
# install_github("burrm/lolcat")
```

```
library(lolcat)
```

```
## lolcat 2.0.0
```

Weekly Quiz: Practice Assessment

Were working with thet airline dataset

1. Ticket Price is Continuous-Ratio
2. Number of Cancelled Fligths per Dat per 500 *

- Discrete Nominal

3. Baggage Delivery Time to Carousel

- Continuous Interval

4. Number of Lost Bages Per 100 Passenger *

- Discrete-Ordinal-Count(Absolute)

Assessment: Data and Measurement