In Depth: Descriptive Research
Measure

Three types of studies:

1. Descriptive: What is the level of 1 variable?
   Ex: What is the president’s overall approval rating?

2. Correlational: How are 2 variables related?
   Ex: How does survey respondent’s age relate to approval rating? [Predictor is measured]

3. Experimental: Does one variable cause the other?
   Ex: Does darkness or lightness of Barack Obama’s skin in photos influence ratings of the president? [The independent variable is manipulated]
Descriptive Research Example

Gallup Daily Poll

Barack Obama's Presidential Job Approval Ratings
% Approve, weekly aggregates


65  62  64  60  51  53  50  49  49  49  45  45  44
Descriptive Research

: what is the level of 1 variable?

Ex: What is the president’s overall approval rating?

Statistical indicators:

central tendency (mean, median, mode)

variability (variance, standard deviation, standard error)
Central Tendency

: a score indicating the center of the distribution of a variable

Indicators of Central Tendency:

Mode: the most frequently occurring score
Median: the 50\textsuperscript{th} percentile (middle score)

Mean: average of the scores
\[
\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}
\]
Mean

Mean: average of the scores:

\[ \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} \]

President approval: \( x_1=65 \ x_2=62 \ x_3=64 \ x_4=60 \ x_5=51 \ x_6=53 \)

Mean = \( \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{N} \)

\[ = \frac{65 + 62 + 64 + 60 + 51 + 53}{6} \]

\[ = 59.16 \]
Variability: Normal Distribution

Applies to many variables, particularly when there are large samples: height, weight, IQ, SAT Scores...
Three Common Distributions

- Symmetrical Distribution
- Positive Skew or Right Skew
- Negative Skew or Left Skew
Skews

Floor and ceiling effects create skews
Due to sampling, physiology, the measuring instrument

Problem: Severe skew violates assumptions of parametric statistics (e.g., ANOVA, regression).

Solutions:
1. A measure that has wider range of sensitivity.
2. Transform data prior to analysis (e.g., log, inverse, or exponent).
3. Convert to ranks and use non-parametric stats.
Van Camp, Barden & Sloan (2010)

What type of distributions are these?
Bimodal Distribution

Law School graduates salary distribution (similar to pro sports and entertainment):

Estimate:
Mean?
Mode?
Variability describes the extent to which scores in a distribution differ from each other (i.e. how spread out they are).

Same Mean: Higher Variability

Lower Variability
Indicators of Variability

sum of squares (SS): take the deviations from the mean, square them and sum them

variance ($s^2$) or mean square (MS)

\[ = \frac{SS}{\text{degrees of freedom (df)}} \]

standard deviation ($s$ or SD): take the square root of the variance
### Definitional Formulas

<table>
<thead>
<tr>
<th>Definitional Formulas</th>
<th>Computational Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>( SS = \sum (X_i - \bar{X})^2 )</td>
<td>( SS = \sum X^2 - \frac{(\sum X)^2}{N} )</td>
</tr>
<tr>
<td>( s^2 = MS = \frac{\sum (X_i - \bar{X})^2}{N - 1} )</td>
<td>( s^2 = MS = \frac{SS}{N - 1} = \frac{SS}{df} )</td>
</tr>
<tr>
<td>( s = SD = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}} )</td>
<td>( s = SD = \sqrt{\frac{SS}{N - 1}} = \sqrt{\frac{SS}{df}} )</td>
</tr>
</tbody>
</table>

- \( N \) = sample size
- \( SS \) = sum of squares (squared differences from mean)
- \( \Sigma \) = summation sign
- \( MS \) = mean square
- \( s^2 \) = variance
- \( X \) = score on variable
- \( \bar{X} \) = sample mean of scores
Variance \( (s^2) \) = Sum of squares/\( df \)

\[
SS = \sum (X_i - \bar{X})^2 \quad df = N-1 = 6-1 = 5
\]

President approval: \( x_1=65 \ x_2=62 \ x_3=64 \ x_4=60 \ x_5=51 \ x_6=53 \)

Mean = 59

\[
SS = (65-59)^2 + (62-59)^2 + (64-59)^2 + (60-59)^2 + (51-59)^2 + (53-59)^2
\]

\[
= 6^2 + 3^2 + 5^2 + 1^2 + 8^2 + 6^2 = 171
\]

Variance \( (s^2) \) = \( SS/df \) = \( 171/5 = 34.2 \)

SD \( (s) \) = 5.85
What SD says about variability

If the variable is normally distributed, SD’s (x-axis) tell us a lot:
Degrees of freedom (df)
: the number of scores that are free to vary

- For one sample, \( df = N - 1 \)
- the number of separate pieces of info that you have about variability

Ex: \( N = 3, \bar{X} = 7 \) and \( X_1 = 7, X_2 = 3, \) then \( X_3 = ? \)

So, since we took out one statistic already, \( \bar{X} \), knowing any two of the values automatically tells us the third.

So, two of the three \((N-1)\) are free to vary but the final \( X_3 \) is fixed

The more statistics that you pull out of a sample in a simultaneous analysis, the fewer the \( df \) that are left
Central Tendency and Variability

Analyze → descriptive statistics → descriptives
Select from list of variables and move to right
Options and select as required
Continue and ok

OR

Analyze → descriptive statistics → frequencies
Select from list of variables and move to right
Statistics and select as required
Continue and ok
SPSS

Histogram Graph

Graphs → legacy diagrams → histogram
Select from list of variables and move to right
Select Superimpose Normal Curve (if you like)
Select other options as required
Continue and ok

Try with HUTerms and CentralityofRace
Van Camp, Barden & Sloan (2010)

Histogram output

![Histogram 1](image1)

- **Frequency**
- **huterm**
- **Mean**: 2.17
- **Std. Dev.**: 1.852
- **N**: 110

![Histogram 2](image2)

- **CentralityOfRace**
- **Mean**: 5.14
- **Std. Dev.**: .94
- **N**: 109
END
Types of Variables

Experimental Design
Independent (IV): the cause of the observed variation
Dependent (DV): the effect (variation) of interest
Control: try to hold all other variables constant

Correlational Design
Predictor: seen as exogenous (x-axis of graph)
Criterion: seen as endogenous (y-axis of graph)
Types of Scales of Measurement

Nominal (categorical) naming only
Examples: dogs, cats, birds, and fish; political party; religious affiliation

Ordinal categories have order but not equal distance
Examples: finishers in a race, class ranking, small, medium, and large size

Interval order and equal distance, but no true zero
Zero is not the absence of the property
Examples: Intelligence, degrees F or C

Ratio order, equal distance and a true zero
Examples: weight, height, time, degrees Kelvin

Parametric statistics reserved for these scales, otherwise use non-parametric.
Example: Central Tendency

Which indicator of central tendency best reflects these data?

<table>
<thead>
<tr>
<th>Income</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$14,000</td>
<td>3</td>
</tr>
<tr>
<td>$19,000</td>
<td>1</td>
</tr>
<tr>
<td>$20,000</td>
<td>1</td>
</tr>
<tr>
<td>$200,000</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Mean**: $46,833
- **Median**: $16,500
- **Mode**: $14,000

![Histogram of Income of 6 Employees](image)
<table>
<thead>
<tr>
<th>Scale of measurement</th>
<th>Measure of central tendency</th>
<th>Measure of variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal</td>
<td>mode</td>
<td>index of qualitative variation</td>
</tr>
<tr>
<td>ordinal</td>
<td>median</td>
<td>range and SIQR</td>
</tr>
<tr>
<td>interval &amp; ratio</td>
<td>mean</td>
<td>variance and standard deviation</td>
</tr>
</tbody>
</table>
### Definitional Formulas

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample (unbiased)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS = \sum (X_i - \mu)^2$</td>
<td>$SS = \sum (X_i - \overline{X})^2$</td>
</tr>
<tr>
<td>$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$</td>
<td>$s^2 = \frac{\sum (X_i - \overline{X})^2}{N-1}$</td>
</tr>
<tr>
<td>$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N} - \mu^2}$</td>
<td>$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{N-1}}$</td>
</tr>
</tbody>
</table>

### Computational Formulas

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample (unbiased)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS = \sum X^2 - N\mu^2$ or $SS = \sum X^2 - \left(\frac{\sum X}{N}\right)^2$</td>
<td>$SS = \sum X^2 - N\overline{X}^2$ or $SS = \sum X^2 - \left(\frac{\sum X}{N}\right)^2$</td>
</tr>
<tr>
<td>$\sigma^2 = \frac{\sum X^2}{N} - \mu^2$ or $\sigma^2 = MS = \frac{SS}{N}$</td>
<td>$s^2 = \frac{\left[\sum X^2 - \left(\frac{\sum X}{N}\right)^2\right]}{N-1}$ or $s^2 = MS = \frac{SS}{N-1} = \frac{SS}{df}$</td>
</tr>
<tr>
<td>$\sigma = \sqrt{\frac{\sum X^2}{N} - \mu^2}$</td>
<td>$s = \sqrt{\left[\sum X^2 - \left(\frac{\sum X}{N}\right)^2\right]}$ or $s = \sqrt{\frac{SS}{df}}$</td>
</tr>
</tbody>
</table>
Population vs. Sample

Population the larger group to which a law applies
(ALL HU undergraduate students, ALL likely voters)
parameters (e.g. $\sigma$, $\rho$)

Sample is the smaller subset which we study
statistics (e.g. $r$, $t$, $F$, $SD$, $M$)

Randomly selected samples vs. samples of convenience

Inferential statistics are used to make an inference about the overall population from a sample that is a subset of that population.
Variables

Variables: a thing which varies (has more than one value) in the given study.

Constants: a thing which is constant (has only one value) in the given study.

Research Design: positioning of variables in a study in regard to one another (assigned roles).