Statistics Review

I. Types of Variables
   
   A. Qualitative vs. Quantitative variables
      
      Qualitative variables: Not inherently numerical
      Examples: Gender, race, sexual preference
      Nominal scale of measurement (name only)

      Quantitative variables: Have numerical properties
      Examples: Height, Weight, Response Time,
      Number of Words recalled, I.Q., Score on a test
I. Types of Variables
(continued)

B. Quantitative Variables have different scales of measurement:
- Ordinal (1st, 2nd, 3rd, ...)
- Interval (equal differences imply equal distances)
- Ratio (distance from zero is meaningful)

C. Coding variables in SPSS

Statistics Review
- I. Types of variables
- II. Descriptive Techniques
- III. Inferential Techniques
II. Descriptive Techniques

A. Frequency Distributions

B. Measures of Central Tendency

C. Measures of Variability

A. Frequency Distributions

Graphical representation of a set of observations:

Y = The number (frequency) of each of the scores

X = categories of observations (scores)

Frequency Distributions

Quantitative Example

<table>
<thead>
<tr>
<th>X</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>
Frequency Distributions come in different shapes.

- Symmetrical
- Positive Skew
- Negative Skew

Frequency Distributions

- How to do a frequency distribution in SPSS
II. Descriptive Techniques

A. Frequency Distributions
B. Measures of Central Tendency
C. Measures of Variability

B. Measures of Central Tendency

Where do the observation generally fall on the x-axis?

Three measures:
- Mode – most frequent
- Median – the middle of the distribution
- Mean – the arithmetic average

\[ M = \frac{\sum x}{n} \]

Measures of Central Tendency

<table>
<thead>
<tr>
<th>X</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Quantitative Example

<table>
<thead>
<tr>
<th>X</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.00</td>
<td>2</td>
</tr>
<tr>
<td>9.00</td>
<td>10.00</td>
</tr>
<tr>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td>9.00</td>
<td>8.00</td>
</tr>
<tr>
<td>8.00</td>
<td>8.00</td>
</tr>
<tr>
<td>8.00</td>
<td>8.00</td>
</tr>
<tr>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td>5.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Mode = 9
Median = 8
Mean = (\sum x)/n
= 196/25
= 7.84
Conclusion on Measures of central tendency

Mean is the most common and useful, but it may be a poor measure with highly skewed distributions.

II. Descriptive Techniques

A. Frequency Distributions
B. Measures of Central Tendency
C. Measures of Variability
C. Measures of Variability

How spread out or variable are the set of scores on a quantitative scale?

Three most common measures:
- SS
- Variance
- Standard Deviation

Measures of Variability

SS (sum of the squared deviations)

<table>
<thead>
<tr>
<th>x</th>
<th>x-M</th>
<th>(x-M)^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8-3</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>1-3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3-3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0-3</td>
<td>9</td>
</tr>
</tbody>
</table>

\[
\Sigma = 12 \\
M = 12/4 = 3.0
\]

Measures of Variability

Variance (\(\sigma^2\))

Population Variance is the average of the squared deviations from the mean:

\[
\sigma^2 = \frac{\sum(x-M)^2}{N} = \frac{SS}{N} \quad (N = \text{population size})
\]

Example: \(35/4 = 9.5\)

\(\sigma^2 = \text{sample variance} = \frac{SS}{n-1}\)

\(= 38/3 = 12.67\)
Measures of Variability

Standard Deviation ($\sigma$) for a population

Square root of variance

$$\sigma = \sqrt{\frac{\sum(x-M)^2}{N}}$$

$$= \sqrt{\frac{SS}{N}}$$

Example: $\sqrt{9.5} = 3.08$

$s$ = sample standard deviation = $\sqrt{SS/(n-1)}$

Descriptive Statistics

- SPSS Descriptive Statistics.

Statistics Review

I. Types of variables
II. Descriptive Techniques
   - Central tendency
   - Variability
III. Inferential Techniques
III. Inferential Techniques

A. Rational & purpose of inferential Statistics

B. Sampling Distributions

C. Student’s t test

D. Simply 1-way ANOVA

---

III. Inferential Techniques

A. Purpose of inferential Statistics

Inferential statistics enable us to reach conclusions (make inferences) about a population based on sample data.

Addresses the issue of reliability

If I collected another sample, will I get the same results?

---

III. Inferential Techniques

B. Inferential statistics use the mean and variance of a sample to infer (guess) the mean of a population.

Sample Mean

Dependent Variable (y)

Sample Variance

Population mean is somewhere in here

This guess is subject to sampling error, and thus comes with a confidence interval, e.g., 95%. The width of the interval is determined by sample size and variance.

---
III. Inferential Techniques

C. Non-numerical example:

<table>
<thead>
<tr>
<th>Dependent Variable (X)</th>
<th>Sample Mean Group 1</th>
<th>Sample Mean Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% Confidence Interval for the population mean of Group 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% Confidence Interval for the population mean of Group 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion: We are more than 95% sure that groups 1 and 2 are from different populations. We would say there is statistically significant (p < .05) difference between the groups.

III. Inferential Techniques

A. Purpose of inferential Statistics
B. Sampling Distributions
C. Student's t test

Sampling Distributions

How do we guess the population mean based on a sample?
How do we know the accuracy (confidence limits) of that guess?

The answers to both of these questions are in the central limits theorem.
### The Central Limits Theorem

1. The distribution of sample means has a mean equal to the population mean.

\[ \mu_M = \mu \]

\( \mu_M \) is called the expected value of the mean.

---

2. The distribution of the sample means approaches a normal distribution.

Because the distribution is shaped like a normal distribution, we can use tables of the normal to estimate confidence intervals.
The Central Limits Theorem

3. The standard deviation of the distribution of sample means is equal to the standard deviation of the population divided by the square root of n, (n = sample size).

\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \]

\( \sigma_{\bar{X}} \) is called the standard error of mean.

With a sample, this would be \( S_{\bar{X}} \).

Example:
Sample, n = 100, \( \mu = 10, s = 2 \)

\[ \mu_{\bar{X}} = \mu = 10 \]

\[ S_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2 \]
Interval Estimation

We should guess that the population mean is 10.
We can be 95% confident that the population mean is between:

\[
\text{confidence interval} = M \pm S_m \times t
\]

(use t not Z, because we estimated standard deviation with s)

\[
M = 10, S_m = .2, t = 2.00
\]

\[
10 + .2(2.00) = 10 + .4 = 10.4
\]

\[
10 - .2(2.00) = 10 - .4 = 9.6
\]
What if we have two groups?

Example:
Sample 1: \( n_1 = 100, \mu_1 = 10, s_1 = 2 \)
Sample 2: \( n_2 = 100, \mu_2 = 12, s_2 = 2 \)

Two possibilities

Population
\( \mu, \sigma \)

Sample 1: \( \mu_1, s_1 \)
Sample 2: \( \mu_2, s_2 \)

\( \mu_1 = \mu \)
\( \mu_2 = \mu \)
But: \( \mu_1 \neq \mu_2 \)
This is called the null hypothesis:
\( H_0: \) no treatment effect

Alternatively

Population
\( \mu_1, \sigma \)

Sample 1: \( \mu_1, s_1 \)
Sample 2: \( \mu_2, s_2 \)

\( \mu_1 = \mu \)
\( \mu_2 = \mu \)
But: \( \mu_1 \neq \mu_2 \)
This is called the alternative hypothesis:
\( H_1: \) there is a treatment effect
With two samples

Doing the math as before we get the following estimates:

Group 1 = 10 + .2(2.00) = 10 + .4 = 10.40
And 10 - .2(2.00) = 10 - .4 = 9.60

Group 2 = 12 + .2(2.00) = 12 + .4 = 12.40
And 12 - .2(2.00) = 12 - .4 = 11.60

\[ t \text{ test for two groups} \]

We would conclude that we are 95% confident that the Group 1 mean is less than Group 2 mean.

Rather than go through all the steps of constructing confidence intervals, we can construct a test to see if the two means are the same or different.

We test two hypotheses:

\( H_0 \): the means are from the same population (null hypothesis)
\( H_1 \): no treatment effect, \( \mu_1 = \mu_2 \)

Versus

\( H_0 \): there is treatment effect,

a) \( \mu_1 < \mu_2 \) (one-tailed test)
b) \( \mu_1 = \mu_2 \) (treatment lowers scores – one tailed test)
c) \( \mu_1 > \mu_2 \) (treatment raises scores – one tailed test)
With two samples: 

\[ s_{M-M} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

Where \( s^2 \) is the pooled (average) variance

\( s^2 \) is average of \( s_1^2 \) and \( s_2^2 \), or the average of \( \sigma_1^2 \) and \( \sigma_2^2 \), which is \( \sigma^2 \)

\[ s_{M-M} = \sqrt{\left( \frac{4}{100} + \frac{4}{100} \right)} = .2828 \]

\[ t = \frac{(M_1 - M_2)}{s_{M-M}} \]

\[ t = \frac{(12 - 10)}{.28} = 7.14 \]

\( df \) for this test are \( (n_1 - 1) + (n_2 - 1) = 198 \)

Thus \( t_{critical} = \pm 1.98 \)

\( t_{obs} = 7.14 \)
t-test

$t_{\text{obs}}$ is past the $t_{\text{crit}}$, therefore we reject $H_0$.

In a sentence:
The mean for Group 1 ($M = 10$) was significantly less than the mean for Group 2 ($M = 12$), $t(198) = 7.14, p < .05$.

---

t-test

Independent samples $t$ in SPSS.

---

Analysis of Variance
ANOVA

What if there are more than two groups?
Analysis of Variance
ANOVA

- Use M₁, M₂, and M₃ to calculate between groups variance (SSₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑₑᵉ
- Use variance within the groups to construct within groups variance (SS_within; error variance)
- Calculate \( F(k-1, N-1) = \frac{SS_{between}}{SS_{within}} \)