Chapter 3
Oneway ANOVA

1. Review of the tests we have covered so far

- One sample with interval scale DV

    - One sample z-test
      Used to compare a sample mean to a hypothesized value when the population is normally distributed with a <u>known variance.</u>

    - One sample t-test
      Used to compare a sample mean to a hypothesized value when the population is normally distributed (or large) with <u>unknown variance.</u>

- Two-independent samples with interval scale DV

    - Two independent samples t-test
      Used to compare the difference of two sample means to a hypothesized value (usually zero) when both populations are normally distributed with <u>unknown but equal variances</u>.

    - Welch's two independent samples t-test
      Used to compare the difference of two sample means to a hypothesized value (usually zero) when both populations are normally distributed with <u>unknown variances that may or may not be equal</u>.

- Two-independent samples tests ordinal DV

    - Mann-Whitney U test
      A non-parametric test used to measure the separation between two sets of sample scores (using the rank of the observations). Can also be used in place of the two independent samples t-test when the data do not satisfy t-test assumptions.

- Two (or more) nominal variables

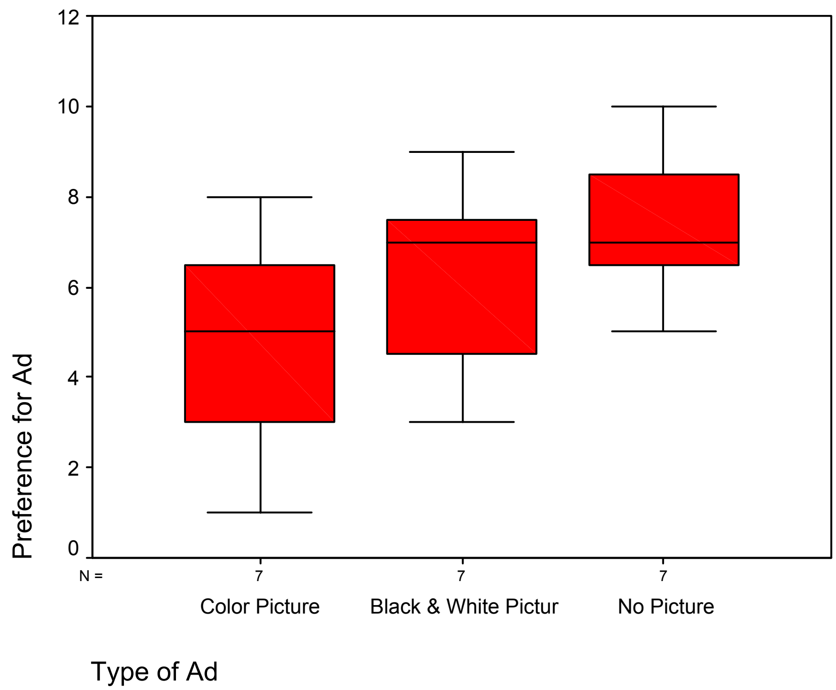
    - Pearson Chi-square test of independence
      A non-parametric test used to test the independence of (or association between) two or more variables.

2. What is an Analysis of Variance (ANOVA)?

Because sometimes, two groups are just not enough . . .

- An Advertising Example: What makes an advertisement more memorable? Three conditions:
  - Color Picture Ad
  - Black and White Picture Ad
  - No Picture Ad

  - DV was preference for the ad on an 11 point scale



ANOVA

Preference for Ad

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25.810	2	12.905	2.765	.090
Within Groups	84.000	18	4.667		
Total	109.810	20			

3. Terminology in ANOVA/Experimental Design
   - Overview of Experimental Design

   - Terminology
     - Factor = Independent variable
     - Level = Different amounts/aspects of the IV
     - Cell = A specific combination of levels of the IVs

   - A one-way ANOVA is a design with only one factor

		Factor A		
Level 1	Level 2	Level 3	Level 4	Level 5
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}
x_{51}	x_{52}	x_{53}	x_{54}	
	x_{62}			
$\bar{X}_{\bullet 1}$	$\bar{X}_{\bullet 2}$	$\bar{X}_{\bullet 3}$	$\bar{X}_{\bullet 4}$	$\bar{X}_{\bullet 5}$
n_1	n_2	n_3	n_4	n_5

x_{ij} i = indicator for subject within level j
j = indicator for level of factor A

Note that the null hypothesis is now a bit less intuitive:

H_0 : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
H_1 : Not all $\mu_i s$ are equal

The alternative hypothesis is NOT $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$
The null and alternative hypotheses must be:
- mutually exclusive
- exhaustive

The overall test of this null hypothesis is referred to as the <u>omnibus F-test</u>.

- A two way ANOVA has two factors. It is usually specified as an A*B design
  - A = the number of levels of the first factor
  - B = the number of levels of the second factor

x_{ijk} i = indicator for subject within level jk
j = indicator for level of factor A
k = indicator for level of factor B

  - Example of a 4x3 design

	Factor A				
	Level A1	Level A2	Level A3	Level A4	
Level B1	$\bar{X}_{\cdot 11}$	$\bar{X}_{\cdot 21}$	$\bar{X}_{\cdot 31}$	$\bar{X}_{\cdot 41}$	$\bar{X}_{\cdot\cdot 1}$
Level B2	$\bar{X}_{\cdot 12}$	$\bar{X}_{\cdot 22}$	$\bar{X}_{\cdot 32}$	$\bar{X}_{\cdot 42}$	$\bar{X}_{\cdot\cdot 2}$
Level B3	$\bar{X}_{\cdot 13}$	$\bar{X}_{\cdot 23}$	$\bar{X}_{\cdot 33}$	$\bar{X}_{\cdot 43}$	$\bar{X}_{\cdot\cdot 3}$
	$\bar{X}_{\cdot 1\cdot}$	$\bar{X}_{\cdot 2\cdot}$	$\bar{X}_{\cdot 3\cdot}$	$\bar{X}_{\cdot 4\cdot}$	$\bar{X}_{\cdot\cdot\cdot}$

  - Let's take a closer look at cell 23

x_{123} x_{223} x_{323} . . . x_{n23}
$\bar{X}_{\cdot 23}$

  - And now there are multiple effects to test

The effect of Factor A $H_0: \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$
The effect of Factor B $H_0: \mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3}$
The effect of the combination of Factor A and Factor B

To keep things simple, we will stick to the one-way ANOVA design for as long as possible!

4. Understanding the F-distribution

- Let's take a step back and examine the sampling distribution of s^2

  - We'll start by making no assumptions.

$$\begin{aligned} x_i - \mu &= x_i - \mu + (\overline{X} - \overline{X}) && \text{Add and subtract } \overline{X} \\ &= x_i - \overline{X} + \overline{X} - \mu && \text{Re-arrange terms} \\ &= (x_i - \overline{X}) + (\overline{X} - \mu) \end{aligned}$$

  - Now square both sides of the equation:
    [And remember from high school algebra: $(a+b)^2 = a^2 + b^2 + 2ab$]

$$\begin{aligned} (x_i - \mu)^2 &= [(x_i - \overline{X}) + (\overline{X} - \mu)]^2 \\ &= (x_i - \overline{X})^2 + (\overline{X} - \mu)^2 + 2(x_i - \overline{X})(\overline{X} - \mu) \end{aligned}$$

  - This equation is true for each of the n observations in the sample. Next, let's add all n equations and simplify:

$$\begin{aligned} \sum_{i=1}^{n}(x_i - \mu)^2 &= \sum_{i=1}^{n}[(x_i - \overline{X})^2 + (\overline{X} - \mu)^2 + 2(x_i - \overline{X})(\overline{X} - \mu)] \\ &= \sum_{i=1}^{n}(x_i - \overline{X})^2 + \sum_{i=1}^{n}(\overline{X} - \mu)^2 + \sum_{i=1}^{n}2(x_i - \overline{X})(\overline{X} - \mu) \end{aligned}$$

  - Note that 2 and $(\overline{X} - \mu)$ are constants with respect to summation over i. Constants can be moved outside of the summation

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \overline{X})^2 + (\overline{X} - \mu)^2\sum_{i=1}^{n}1 + 2(\overline{X} - \mu)\sum_{i=1}^{n}(x_i - \overline{X})$$

  - We can use two facts to simplify this equation:

$$\sum_{i=1}^{n}1 = n$$

$$\sum_{i=1}^{n}(x_i - \overline{X}) = 0$$

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \overline{X})^2 + n(\overline{X} - \mu)^2 + 0$$

- Next, let's divide both sides of the equation by σ^2

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{\sigma^2} + \frac{n(\overline{X}-\mu)^2}{\sigma^2}$$

- And then rearrange the terms

$$\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\overline{X})^2 + \left(\frac{(\overline{X}-\mu)}{\sigma/\sqrt{n}}\right)^2 \qquad \text{(eq. 3-1)}$$

- Up to this point, we have made no assumptions about *X*. To make additional progress, we now have to make a few assumptions
  - *X* is normally distributed. That is $X \sim N(\mu,\sigma)$
  - Each x_i in the sample is independently sampled

- First, let's consider the left side of eq. 3-1: $\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$

  $\frac{x_i-\mu}{\sigma}$ is the familiar form of a z-score

  Hence $\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$ is the sum of n squared z-scores

  - From our review of the Chi-square distribution we know that
    - One squared z-score has a chi-square distribution with 1df
    - The sum of *N* squared z-scores have a chi-square distribution with *N* degrees of freedom

  - Now we can say the left hand side of eq 3-1 has a Chi-square distribution

$$\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}(z_i)^2 \sim \chi_n^2$$

- Next, let's consider $\left(\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right)^2$
  - We know that the sampling distribution of the mean for data sampled from a normal distribution is also normally distributed:
  $$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
  - Hence, $\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}$ is also a z-score
  - $\left(\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right)^2$ is a single squared z-score. But squared z-scores follow a chi-square distribution. So we know that $\left(\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$

- Putting the pieces together, we can rewrite eq. 3-1 as

$$\chi_n^2 \sim \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{X})^2 + \chi_1^2$$

$$\chi_n^2 - \chi_1^2 \sim \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{X})^2$$

- Because of the additivity of <u>independent</u> chi-squared variables, this equation simplifies to:

$$\chi_{n-1}^2 \sim \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{X})^2$$

- Now let's divide both sides of the equation by n-1

$$\frac{\chi_{n-1}^2}{n-1} \sim \frac{1}{\sigma^2}\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}$$

- We recognize $\hat{\sigma}^2$ and substitute it into the equation

$$\frac{\chi^2_{n-1}}{n-1} \sim \frac{1}{\sigma^2}\hat{\sigma}^2$$

- Rearranging, we finally obtain:

$$\hat{\sigma}^2 \sim \frac{\sigma^2 \chi^2_{n-1}}{n-1}$$

- In other words, with the assumptions of normality and independence, $\hat{\sigma}^2$ has a chi-squared distribution. (But notice, σ^2 must be known!)

- Sampling Distribution of the Variance
  - Assumption: X is drawn from a normally distributed population:

$$X \sim N(\mu_X, \sigma_X)$$

Then for a sample of size n:

$$\hat{\sigma}^2 \sim \frac{\sigma_x^2 \chi^2_{n-1}}{n-1}$$

  - Facts about the Chi-square distribution:

$$E\left(\chi^2_n\right) = n$$
$$Var\left(\chi^2_n\right) = 2n$$

  - We can use these facts to check if $\hat{\sigma}^2$ is an unbiased and consistent estimator of the population variance.

    - What is the expected value of $\hat{\sigma}^2$?

$$\begin{aligned} E(\hat{\sigma}^2) &= E\left(\frac{\sigma^2 \chi^2_{n-1}}{n-1}\right) \\ &= \frac{\sigma^2}{n-1} E\left(\chi^2_{n-1}\right) \\ &= \frac{\sigma^2}{n-1}(n-1) \\ &= \sigma^2 \end{aligned}$$

$\hat{\sigma}^2$ is an unbiased estimator of σ^2

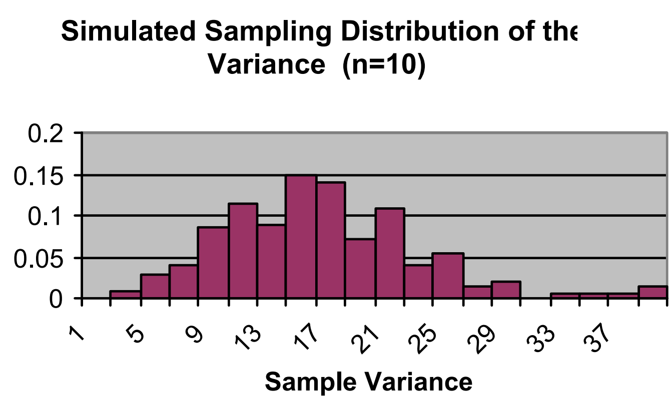
- What is the variance of the sampling distribution of $\hat{\sigma}^2$?

$$Var(\hat{\sigma}^2) = Var\left(\frac{\sigma^2 \chi^2_{n-1}}{n-1}\right)$$
$$= \left(\frac{\sigma^2}{n-1}\right)^2 Var\left(\chi^2_{n-1}\right)$$
$$= \frac{\sigma^4}{(n-1)^2} 2(n-1)$$
$$= \frac{2\sigma^4}{(n-1)} \qquad \hat{\sigma}^2 \text{ is a consistent estimator of } \sigma^2$$

- Example #1:
  Suppose we have a sample n=10 from $X \sim N(0,4)$ $\left[\sigma^2 = 16\right]$

$$\hat{\sigma}^2 \sim \frac{\sigma_x^2 \chi^2_{n-1}}{n-1} = \frac{16\chi^2_9}{9} = 1.778 * \chi^2_9$$

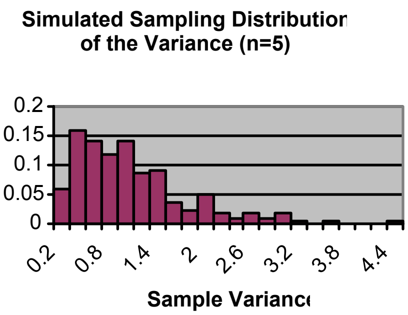
$$E(\hat{\sigma}^2) = \sigma^2 = 16 \qquad Var(\hat{\sigma}^2) = \frac{2\sigma^4}{(n-1)} = \frac{2*256}{9} = 56.889$$



- Example #2:
  Suppose we have a sample n=5 from $X \sim N(0,1)$ $[\sigma^2 = 1]$

$$\hat{\sigma}^2 \sim \frac{\sigma_x^2 \chi_{n-1}^2}{n-1} = \frac{1\chi_4^2}{4} = .25 * \chi_4^2$$

$$E(\hat{\sigma}^2) = \sigma^2 = 1 \qquad Var(\hat{\sigma}^2) = \frac{2\sigma^4}{(n-1)} = \frac{2*1}{4} = .50$$



In this simulated distribution:

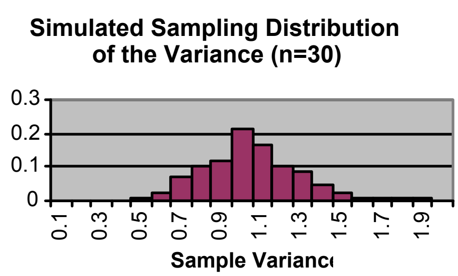
$\mu = .988$

$\sigma^2 = .538$

- Example #3:
  Suppose we have a sample n=30 from $X \sim N(0,1)$ $[\sigma^2 = 1]$

$$\hat{\sigma}^2 \sim \frac{\sigma_x^2 \chi_{n-1}^2}{n-1} = \frac{1\chi_{29}^2}{29} = \frac{\chi_{29}^2}{29}$$

$$E(\hat{\sigma}^2) = \sigma^2 = 1 \qquad Var(\hat{\sigma}^2) = \frac{2\sigma^4}{(n-1)} = \frac{2*1}{29} = .0690$$



In this simulated distribution:

$\mu = .9995$

$\sigma^2 = .0581$

- Confidence intervals to make inference about σ^2
  - As an interesting aside, we can use the equation for the sampling distribution of the variance to construct confidence intervals around the population variance (and to can make inferences about σ^2) <u>if</u> X is normally distributed with known variance. Let's construct a 90% confidence interval around σ^2:

    Lower bound: $\frac{\sigma^2 \chi^2_{n-1}(.05)}{n-1}$ Upper bound: $\frac{\sigma^2 \chi^2_{n-1}(.95)}{n-1}$

    Where $\chi^2_{n-1}(p)$ is the critical chi-square value with n-1 dfs and with p area to the left of the critical value (NOTE: EXCEL uses area to the right of the critical value).

  - <u>Example #1</u>: Suppose we draw a random sample of size 13 (n = 13) from a normally distributed population with $\sigma^2 = 34$. Then for a 90% CI around σ^2:

    Lower bound: $\frac{34 * \chi^2_{12}(.05)}{12} = \frac{34 * 5.23}{12} = 14.81$

    Upper bound: $\frac{34 * \chi^2_{12}(.95)}{12} = \frac{34 * 21.03}{12} = 59.57$

    - We are 90% confident that if the sample is drawn from the known population, the interval $(14.81, 59.57)$ will cover the true value of σ^2

    - Because the chi-squared distribution is not symmetric, this confidence interval will not be symmetric

      |34 – 14.81| = 19.19 |34 – 59.57| = 25.57

  - <u>Example #2</u>: Suppose we draw a random sample of size 30 (n = 30) from a normally distributed population with $\sigma^2 = 4$. Then to obtain a 95% CI around σ^2:

    Lower bound: $\frac{4 * \chi^2_{29}(.025)}{29} = \frac{4 * 16.05}{29} = 2.21$

    Upper bound: $\frac{4 * \chi^2_{29}(.975)}{29} = \frac{4 * 45.72}{29} = 6.31$

    - Thus under the null hypothesis, we are 95% confident that interval $(2.21, 6.31)$ will cover the true value of σ^2

- Remember, our goal was to derive the F-distribution, so let's keep going!
  - We just derived that $\hat{\sigma}^2 \sim \frac{\sigma^2 \chi^2_{n-1}}{n-1}$ when x is sampled independently from a normal distribution. If we were to draw two independent samples, from two different normally distributed populations, then we would have:

$$\hat{\sigma}_1^2 \sim \frac{\sigma_1^2 \chi^2_{n_1-1}}{n_1-1} \quad \text{and} \quad \hat{\sigma}_2^2 \sim \frac{\sigma_2^2 \chi^2_{n_2-1}}{n_2-1}$$

  - We could take the ratio of these two variance estimates

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim \frac{\sigma_1^2 \chi^2_{n_1-1} / n_1-1}{\sigma_2^2 \chi^2_{n_2-1} / n_2-1}$$

  - We'll make one additional assumption to simplify this equation. Let's <u>assume that the variances of the two populations are equal</u>.

$$\sigma_1^2 = \sigma_2^2$$

    Then

$$\frac{s_1^2}{s_2^2} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim \frac{\chi^2_{n_1-1} / n_1-1}{\chi^2_{n_2-1} / n_2-1}$$

    - This ratio of chi-squared distributions divided by their degrees of freedom is the F distribution with parameters n_1-1 and n_2-1

$$F(n_1-1, n_2-1) = \frac{\chi^2_{n_1-1} / n_1-1}{\chi^2_{n_2-1} / n_2-1}$$

$$\text{or more generally:}$$

$$F(\nu_1, \nu_2) = \frac{\chi^2_{\nu_1} / \nu_1}{\chi^2_{\nu_2} / \nu_2}$$

- The F-distribution is a ratio of two chi-squared distributions (divided by their degrees of freedom)

$$F(\nu_1, \nu_2) = \frac{\chi^2_{\nu 1} / \nu_1}{\chi^2_{\nu 2} / \nu_2}$$

- The F-distribution can be used for testing whether two unknown population variances are equal by taking the ratio of two sample variances from these populations

- Thus, one use of a F-ratio is to test the equality of two population variances (Note this test is always one-tailed).

$$H_0: \sigma^2_{Min} \geq \sigma^2_{Max}$$
$$H_1: \sigma^2_{Min} < \sigma^2_{Max}$$

$$F(df_{larger}, df_{smaller}) = \frac{s^2_{larger}}{s^2_{smaller}} \qquad df_{larger} = n_{larger} - 1 \qquad df_{smaller} = n_{smaller} - 1$$

  - Example: Do the <u>variances</u> of IQ scores for schizophrenics and manic-depressives differ? A sample of 31 schizophrenics was found to have $s_s^2 = 148$ and a sample of 101 manic-depressives was found to have $s_{md}^2 = 302$.

    - We must assume that IQ scores are <u>normally distributed</u> in both populations, and that observations were <u>independently sampled</u> from each of the populations

$$F = \frac{s^2_{larger}}{s^2_{smaller}} = \frac{s^2_{md}}{s^2_s} = \frac{302}{148} = 2.04$$

$$\begin{aligned} df &= (n_{md} - 1, n_s - 1) \\ &= (101 - 1, 31 - 1) \\ &= (100, 30) \end{aligned}$$

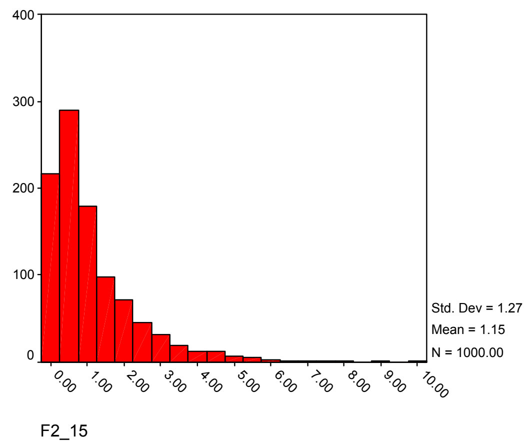
    - Looking up the associated p-value, we find for $F(100,30), p = .0139$
      Reject H_0 at $\alpha = .05$

- A more common use of the F is to test the equality of two or more population means (and we'll see why this is an appropriate use of the F-test in just a bit!)

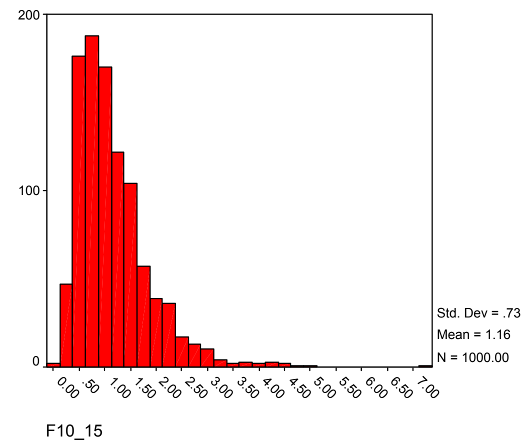
- Let's recap the assumptions we made while deriving the F-test:
  - Both populations are <u>normally distributed.</u>
  - Both populations have a <u>common variance.</u>
  - Both <u>samples were drawn independently</u> from each other.
  - Within each sample, the <u>observations were sampled randomly and independently</u> of each other.

- Fun facts about the F-distribution
  - It is positively skewed.
  - The mode is always less than 1.
  - If the numerator and the denominator degrees of freedom are both large, then the F-distribution approximates a normal distribution.
  - The distribution of this ratio was discovered by R.A. Fisher in 1924 and was given the label F in his honor.

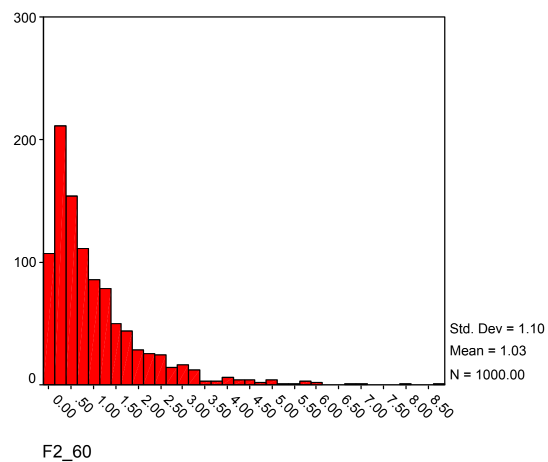
$F(2,15)$



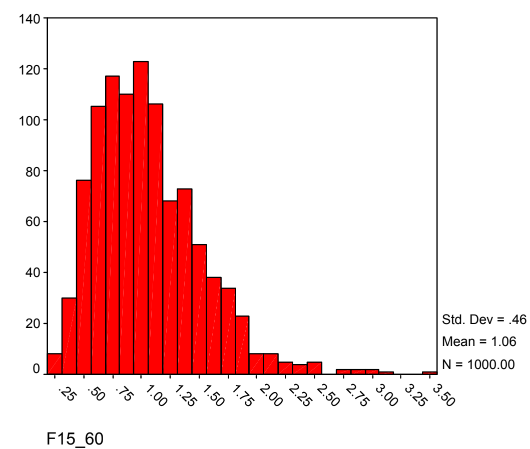
$F(10,15)$



$F(2,60)$



$F(15,60)$



5. Three ways to understand ANOVA
    - A generalization of the t-test
    - The structural model approach
    - The variance partitioning approach

ANOVA as a generalization of the t-test

- Let's consider the familiar case of a two group analysis
- Example: Suppose that two brands of soda are being compared. Participants are randomly assigned to try one soda. They rate five questions about that soda on 5-point scales, with higher numbers indicating greater liking.

Soda P	Soda C
15	13
18	15
12	11
15	17
17	12
14	13
19	
21	
16	

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	16.33333	13.5
Variance	7.5	4.7
Observations	9	6
Pooled Variance	6.423077	
Hypothesized Mean Difference	0	
df	13	
t Stat	2.121179	
P(T<=t) two-tail	0.053705	
t Critical two-tail	2.160368	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	28.9	1	28.9	4.499401	0.053705	4.667186
Within Groups	83.5	13	6.423077			
Total	112.4	14				

Using an independent samples t-test, we find: $t(13)=2.12, p = .054$

Using an ANOVA, we find: $F(1,13)=4.50, p = .054$

- Note that the p-values of these two tests are identical and that the value of the t-test squared equals the value of the F-test

  $$(2.12)^2 = 4.50$$

  - In general, the following relationship holds between the t and F tests:

    $$t(a) = b, p = c \Leftrightarrow F(1,a) = b^2, p = c$$

- In other words, ANOVA can be thought of as a generalization of the independent samples t-test.
  - For the case of two-groups, ANOVA is identical to the independent samples t-test.
  - For the case of more than two groups, ANOVA is a generalization of the same analysis.

- In this general case, the null hypothesis is:

  $$H_0 : \mu_j = \mu \text{ for all } j$$

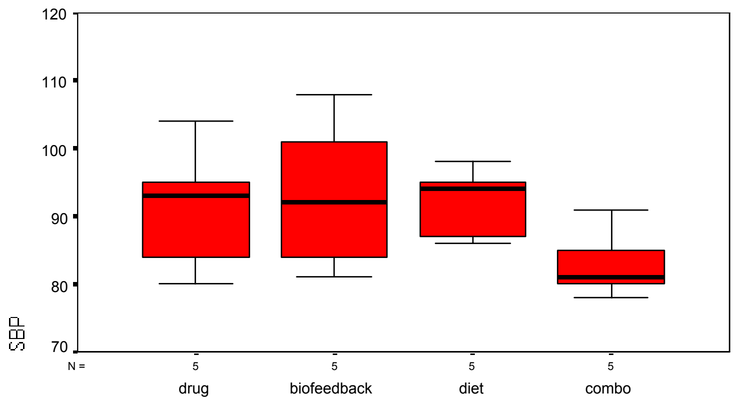
The structural model approach to ANOVA

- It is easiest to consider this approach with an example. Let's consider a program to lower blood pressure. Participants are assigned to one of four groups: drug therapy, biofeedback, diet, and a combination treatment. Below are systolic blood pressures (SBP) after two weeks of treatment. The research hypothesis is that a combination of treatments will be more effective than each of the individual treatments.

	Condition		
Drugs	Biofeedback	Diet	Combo
84	81	98	91
95	84	95	78
93	92	86	85
104	101	87	80
80	108	94	81

- Always remember to look at the data first!

```
EXAMINE VARIABLES=sbp BY group
  /PLOT BOXPLOT.
```



Descriptives

Systolic Blood Pressure

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Drug	5	91.2000	9.4710	4.2356	79.4402	102.9598
Biofeedback	5	93.2000	11.3446	5.0735	79.1138	107.2862
Diet	5	92.0000	5.2440	2.3452	85.4887	98.5113
Combo	5	83.0000	5.1478	2.3022	76.6081	89.3919
Total	20	89.8500	8.6223	1.9280	85.8146	93.8854

- The grand mean is $\bar{Y} = 89.85$. This is an unbiased estimate of the true population mean, μ

- Suppose that we did not know which group a participant was in, but we still wanted to estimate his/her SBP. Our best guess at this person's SBP would be the grand mean, $\overline{Y}$ and any deviation from the mean in that person's score would be unexplained, or error in our simple prediction.

$$Y_{ij} = \mu + \varepsilon_{ij}$$
$$Y_{ij} = 89.85 + \varepsilon_{ij}$$

ε_{ij} denotes the unexplained part of the score associate with the i^{th} person in the j^{th} group (or the error in the model).

  - This is the simplest model we can develop to explain SBP scores. It is sometimes called the <u>reduced model</u> (because it does not contain all the information we have about the participants; we have not included group into the model).

- We can improve our prediction of SBP scores if we know the treatment condition of the participant by using the group mean.

$$\begin{aligned} Y_{i1} &= \mu_1 + \varepsilon_{i1} & \qquad Y_{i1} &= 91.20 + e_{i1} \\ Y_{i2} &= \mu_2 + \varepsilon_{i2} & \qquad Y_{i2} &= 93.20 + e_{i2} \\ Y_{i3} &= \mu_3 + \varepsilon_{i3} & \qquad Y_{i3} &= 92.00 + e_{i3} \\ Y_{i4} &= \mu_4 + \varepsilon_{i4} & \qquad Y_{i4} &= 83.00 + e_{i4} \end{aligned}$$

  - We can rewrite the effect of each group as a deviation from the grand mean

$$\overline{Y}_1 = \hat{\mu} + \hat{\alpha}_1$$
$$91.20 = 89.85 + \hat{\alpha}_1$$
$$\hat{\alpha}_1 = 1.35$$

  - Across all groups, SBP scores average 89.85. But for those people in the drug therapy group, SBP scores are, on average, 1.35 units higher than the overall mean.

    - $\hat{\alpha}_1$ is interpreted as the specific effect of the drug therapy on SPB scores, relative to the other conditions

- For a one-way ANOVA, the group effects are easy to calculate:

$$\hat{\alpha}_j = \bar{Y}_{.j} - \bar{Y}_{..}$$

  - There is a constraint placed on the $\hat{\alpha}_j$s such that they must sum to zero:

$$\sum \hat{\alpha}_j = 0$$

  - Thus, we can write a participant's score as a grand mean, a treatment effect, and participant-specific error:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

μ	The overall mean of the scores
α_j	The effect of being in level j
ε_{ij}	The unexplained part of the score

  - This model is sometimes called the <u>full model</u> (because it takes into account all the information in the design).

- From a structural model perspective, we can state the null hypothesis as the lack of group effects:

$$H_0 : \alpha_j = 0 \text{ for all } j$$

- Using the structural model, we can decompose each observation into its component parts:

μ	The overall mean of the scores	89.85

α_j The effect of being in level *j*

$$\hat{\alpha}_1 = \overline{Y}_1. - \overline{Y}.. = 91.2 - 89.85 = 1.35$$
$$\hat{\alpha}_2 = \overline{Y}_2. - \overline{Y}.. = 93.2 - 89.85 = 3.35$$
$$\hat{\alpha}_3 = \overline{Y}_3. - \overline{Y}.. = 92.0 - 89.85 = 2.15$$
$$\hat{\alpha}_4 = \overline{Y}_4. - \overline{Y}.. = 83.0 - 89.85 = -6.85$$

Note that $\sum_j \alpha_i = 0$

ε_{ij} The unexplained part of the score

$$e_{11} = y_{11} - \overline{Y}._1 = 84 - 91.2 = -7.2$$
$$e_{21} = y_{21} - \overline{Y}._1 = 95 - 91.2 = 3.8$$
$$e_{31} = y_{31} - \overline{Y}._1 = 93 - 91.2 = 1.8$$
$$e_{41} = y_{41} - \overline{Y}._1 = 104 - 91.2 = 12.8$$
$$e_{51} = y_{51} - \overline{Y}._1 = 80 - 91.2 = -11.2$$

Note that $\sum_i e_{i1} = 0$

$$e_{12} = y_{12} - \overline{Y}._2 = 81 - 93.2 = -12.2$$
$$e_{22} = y_{22} - \overline{Y}._2 = 84 - 93.2 = -9.2$$
$$e_{32} = y_{32} - \overline{Y}._2 = 92 - 93.2 = -1.2$$
$$e_{42} = y_{42} - \overline{Y}._2 = 101 - 93.2 = 7.8$$
$$e_{52} = y_{52} - \overline{Y}._2 = 108 - 93.2 = 14.8$$

Note that $\sum_i e_{i2} = 0$

$$e_{13} = y_{13} - \overline{Y}._3 = 98 - 92.0 = 6.0$$
$$e_{23} = y_{23} - \overline{Y}._3 = 95 - 92.0 = 3.0$$
$$e_{33} = y_{33} - \overline{Y}._3 = 86 - 92.0 = -6.0$$
$$e_{43} = y_{43} - \overline{Y}._3 = 87 - 92.0 = -5.0$$
$$e_{53} = y_{53} - \overline{Y}._3 = 94 - 92.0 = 2.0$$

Note that $\sum_i e_{i3} = 0$

$$e_{14} = y_{14} - \overline{Y}._4 = 91 - 83.0 = 8.0$$
$$e_{24} = y_{24} - \overline{Y}._4 = 78 - 83.0 = -5.0$$
$$e_{34} = y_{34} - \overline{Y}._4 = 85 - 83.0 = 2.0$$
$$e_{44} = y_{44} - \overline{Y}._4 = 80 - 83.0 = -3.0$$
$$e_{54} = y_{54} - \overline{Y}._4 = 81 - 83.0 = -2.0$$

Note that $\sum_i e_{i4} = 0$

Note that $\sum_j \sum_i e_{ij} = 0$

- We have decomposed each participant's score into a grand mean, a treatment effect, and participant-specific error:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

$$y_{11} = 89.85 + 1.35 - 7.20$$
$$y_{54} = 89.85 - 6.85 - 2.00$$

- An Aside: This method of estimating the population parameters is referred to as the "method of least squares." Let's see why!

  - From the structural model approach, we obtained
$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

  - We can rewrite this equation:
$$\varepsilon_{ij} = Y_{ij} - \mu - \alpha_j$$
$$\varepsilon_{ij} = Y_{ij} - (\mu + \alpha_j)$$
$$\varepsilon_{ij} = Y_{ij} - \mu_j$$

    - Where ε_{ij} is the residual / error / unexplained part of the DV
      Y_{ij} is the observed score
      μ_j is the expected score

    - Residual = observed score – expected score

  - A desirable property of any fitted model is to have the smallest possible errors. But how do we define "smallest average error"?

    - Averaging the actual residuals would not work; the positive and negative residuals would cancel each other.

    - Taking the average absolute value of the residuals is possible, but messy mathematically.

    - It can be shown that if we define the residual to be $(Y_{ij} - model)$, then $(Y_{ij} - model)^2$ are the squared residuals. When the cell mean, $\bar{Y}_{\cdot j}$, is used as the model, the squared residuals are smaller than they would be if any other model were to be used. Thus, the cell mean is called the least squares estimate of μ_j

The variance partitioning approach

- Let's start by examining the formula for the estimate of the variance

$$Var(Y) = \frac{\sum (Y_i - \bar{Y})^2}{N-1}$$

$$Var(Y) = \frac{SS}{df}$$

- From the two-independent samples t-test we can see s^2_{pooled} fits nicely into this formula:

$$s^2_{pooled} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$s^2_{pooled} = \frac{SS_1 + SS_2}{(n_1 + n_2 - 2)}$$

$$s^2_{pooled} = \frac{SS \text{ from group means}}{df}$$

- In a more general case, to obtain an estimate of the variance, one takes the sum of squared deviations from each group mean (the numerator), and divides by the total number of subjects minus the number of groups (the denominator)

$$s^2_{general} = \frac{SS_1 + SS_2 + \ldots + SS_a}{(n_1 + n_2 + \ldots n_a - a)}$$

- The idea behind variance partitioning is to divide the total variability in the sample into two parts:
  - Variance due to the factor (the IV)
  - Variance that is unexplained (not due to the factor)

- Let's start by examining the formula for the estimate of the variance

$$Var(Y) = \frac{\sum (Y_i - \bar{Y})^2}{N-1}$$

$$Var(Y) = \frac{SS}{df}$$

- Now, let's try to partition the total variability in the data into two parts we can interpret.

$$\text{SST} = \sum_j^a \sum_i^n (y_{ij} - \bar{y}_{..})^2$$

$$\sum_j^a \sum_i^n (y_{ij} - \bar{y}_{\cdot j} + \bar{y}_{\cdot j} - \bar{y}_{..})^2$$

$$\sum_j^a \sum_i^n [(\bar{y}_{\cdot j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{\cdot j})]^2$$

$$\sum_j^a \sum_i^n (\bar{y}_{\cdot j} - \bar{y}_{..})^2 + \sum_j^a \sum_i^n (y_{ij} - \bar{y}_{\cdot j})^2 + \sum_j^a \sum_i^n 2(\bar{y}_{\cdot j} - \bar{y}_{..})(y_{ij} - \bar{y}_{\cdot j})$$

  - As we have done before, we can pull out the constants

$$\sum_j^a (\bar{y}_{\cdot j} - \bar{y}_{..})^2 \sum_i^n 1 + \sum_j^a \sum_i^n (y_{ij} - \bar{y}_{\cdot j})^2 + 2\sum_j^a (\bar{y}_{\cdot j} - \bar{y}_{..}) \sum_i^n (y_{ij} - \bar{y}_{\cdot j})$$

  - We note that

$$\sum_j^a (\bar{y}_{\cdot j} - \bar{y}_{..}) = 0 \qquad \text{and} \qquad \sum_i^n 1 = n$$

$$\text{SST} = n\sum_j^a (\bar{y}_{\cdot j} - \bar{y}_{..})^2 + \sum_j^a \sum_i^n (y_{ij} - \bar{y}_{\cdot j})^2$$

$$\text{SST} = \text{SSB} + \text{SSW}$$

$$\text{SST} = n\sum_{j}^{a}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 + \sum_{j}^{a}\sum_{i}^{n}(y_{ij} - \bar{y}_{\cdot j})^2$$

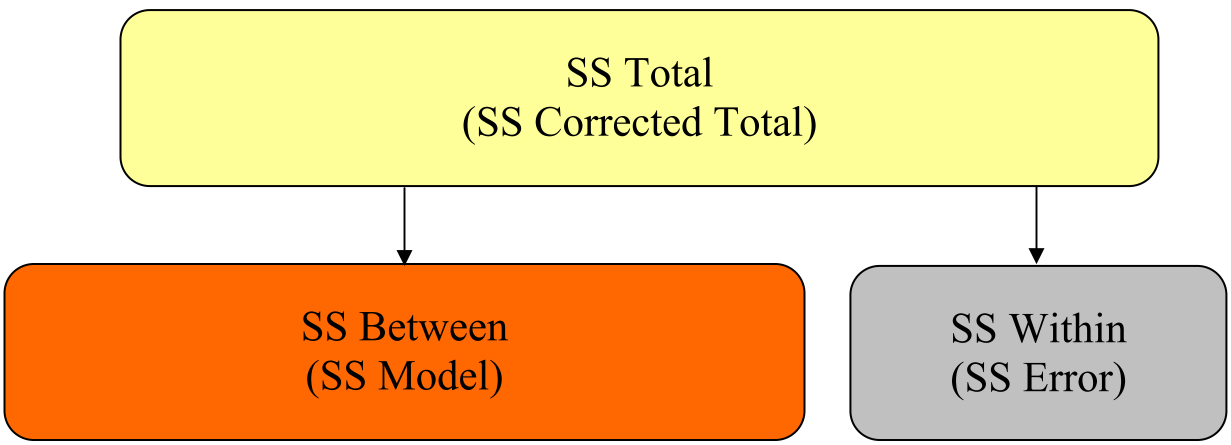
$$\text{SST} = \text{SSB} + \text{SSW}$$

$n\sum_{j}^{a}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$ is the sum of squares between groups (SSB)

SSB is the variation of all the group means around the grand mean (intergroup variability)

$\sum_{j}^{a}\sum_{i}^{n}(y_{ij} - \bar{y}_{\cdot j})^2$ is the sum of squares within groups (SSW)

SSW is the variation of all the individual observations around their group mean, the generalization of s^2_{pooled} (intragroup variability)



- Thus far, we have computed sums of squares, not variances. Remember that an estimate of a variance is given by

$$Var(Y) = \frac{SS}{df}$$

  - Let's construct a variance estimate from each of the components we derived:

$$\hat{\sigma}^2_{Between} = \frac{SSB}{df_b}, \text{ where } df_b = \text{number of groups} - 1$$

$$\hat{\sigma}^2_{Between} = \frac{SSB}{a-1}$$

$$\hat{\sigma}^2_{Within} = \frac{SSW}{df_w}, \text{ where } df_w = \text{(number of groups)*(n for each group minus 1)}$$

$$\hat{\sigma}^2_{Within} = \frac{SSW}{a(n-1)} = \frac{SSW}{N-a}$$

  - In ANOVA terminology, an estimate of the variance is called the <u>mean square</u>

$$MSB = \frac{SSB}{a-1} \qquad MSW = \frac{SSW}{a(n-1)}$$

  - An F-test is obtained by taking the ratio of two variances. Let's create and F-test with our two variance components

$$F(df_b, df_w) = \frac{MSB}{MSW} = \frac{SSB/_{a-1}}{SSW/_{a(n-1)}}$$

    How should we interpret this test?

- The denominator, MSW, is a measure of variability within each group around its mean. In other words, MSW tells us how much of the total variability is due to *error*.

$$SSW = \sum_{j}^{a}\sum_{i}^{n}(y_{ij} - \bar{y}_{\cdot j})^2$$

- The numerator, MSB, is a measure of variability in the group means. In other words, MSB tells us how much of the total variability is due to *differences in group means*.

$$SSB = n\sum_{j}^{a}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$$

- Suppose the null hypothesis is true and the group means are actually identical to each other in the population. In this case,
    - We will observe some differences in the group means, but these differences will be entirely due to error.
    - MSW and MSB will be different estimators of the same quantity: error.
    - The F-test should be near 1, because the numerator and denominator are estimating the same quantity.

- Suppose the alternative hypothesis is true and the group means are actually different from each other in the population. In this case,
    - We will observe some differences in the group means and these differences will be due to error + true differences in the population group means.
    - MSW and MSB will be estimating of the different quantities.
    - The F-test should be greater than 1, because the numerator is estimating the same quantity as the denominator PLUS the true group effects.

- From a variance partitioning perspective, we can state the null hypothesis that *MSB* and *MSW* are estimates of the same error (because there are no group effects for MSB to detect):

$$H_0 : MSB = MSW$$

Relating the structural model and the variance partitioning approaches

- From the structural model approach, we obtained

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

We defined ε_{ij} as the <u>residual</u>:

$$\varepsilon_{ij} = Y_{ij} - \mu_j$$

- It can be shown that $Var(\varepsilon_{ij}) = MSW$
    - Intuitively, this relationship should make a lot is sense. The residuals are due to error and MSW is a measure of the variance of random error.
    - A rigorous proof is tedious, and complicated, so let's settle for an intuitive proof and an example [the curious can tune into Kirk (1995) pp. 91-92]

$$\begin{aligned}
Var(\varepsilon_{ij}) &= \frac{\sum_{j=1}^{a}\sum_{i=1}^{n}(\varepsilon_{ij} - \bar{\varepsilon})^2}{df} \\
&= \frac{\sum_{j=1}^{a}\sum_{i=1}^{n}(\varepsilon_{ij} - 0)^2}{df} \\
&= \frac{\sum_{j=1}^{a}\sum_{i=1}^{n}(Y_{ij} - \bar{Y}_{\cdot j})^2}{df} \\
&= \frac{SSW}{df} \\
&= \frac{SSW}{a(n-1)} = MSW
\end{aligned}$$

- Let's return to our blood pressure example. Here is the ANOVA on the raw scores for reference

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	322.95	3	107.65	1.58	0.23	3.24
Within Groups	1089.6	16	68.1			
Total	1412.55	19				

- Let's create a new data set of residuals by subtracting the group mean from each observed score

  If (group=1) sbp_e = sbp – 91.2.
  If (group=2) sbp_e = sbp – 93.2.
  If (group=3) sbp_e = sbp – 92.0.
  If (group=4) sbp_e = sbp – 83.0.

DrugE	BiofeedbackE	DietE	ComboE
-7.2	-12.2	6	8
3.8	-9.2	3	-5
1.8	-1.2	-6	2
12.8	7.8	-5	-3
-11.2	14.8	2	-2

- Descriptive statistics on the residuals

Residuals	
Mean	0.00
Standard Error	1.69
Median	0.30
Standard Deviation	7.57
Sample Variance	57.35
Minimum	-12.20
Maximum	14.80
Sum	0.00
Count	20.00

  - From the ANOVA table, we found $MSW = 68.1$
  - From the analysis of the residuals, we found $Var(\varepsilon_{ij}) = 57.35$

- But for the analysis of residuals, the variance was computed using N-1 in the denominator; we need the denominator to be N-r. So, we have to multiply the $Var(\varepsilon_{ij})$ by a correction factor:

$$Var(\varepsilon_{ij}) * \frac{N-1}{N-4} = 57.35 * \frac{19}{16} = 68.10 = MSW$$

TaDa!

- Just for fun, let's run an ANOVA on the residuals. What do you think will happen?

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.00	3	0.00	0.00	1.00	3.24
Within Groups	1089.6	16	68.1			
Total	1089.6	19				

- Why is the *MSB* zero?

* To recap, we have just demonstrated that the variance of the residuals of the ANOVA model is equal to the *MSW*. Because the residuals are often thought of as errors, the *MSW* is sometimes called the Mean Square Error (*MSE*). In other words, the *MSW* is an unbiased estimate of the true error variance of the model, σ_ε^2

$$E(MSW) = \sigma_\varepsilon^2$$

* Let's return to the second variance component we derived: MSB
  - We can also derive the expected value of the MSB. However, again this is tedious and rather unenlightening, so we'll skip to the main result:

$$E(MSB) = \sigma_\varepsilon^2 + \frac{\sum n_i \alpha_i^2}{a-1}$$

$\frac{\sum n_i \alpha_i^2}{a-1}$ is a measure of the variability of the treatment effects

- **If the null hypothesis is true**:
  - Under the null hypothesis, all treatment effects are zero, $\alpha_j = 0$ for all j, and the MSB also estimates σ_ε^2

$$E(MSB) = \sigma_\varepsilon^2$$
$$F = \frac{MSB}{MSE} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2} = 1$$

  - Putting MSB and MSW together, the F-statistic with $df_1 = a$-1 and $df_2 = N$-a is a ratio of variances estimating the same quantity.

- **If the null hypothesis is false:**
  - Under the alternative hypothesis, at least one of the treatment effects is not zero, and

$$E(MSB) = \sigma_\varepsilon^2 + \frac{\sum n_i \alpha_i^2}{a-1}$$
$$F = \frac{MSB}{MSW} = \frac{\sigma_\varepsilon^2 + \frac{\sum n_i \alpha_i^2}{a-1}}{\sigma_\varepsilon^2} > 1$$

  - When the null hypothesis is true, the *MSB/MSW* ratio has an F-distribution with $df_1 = a$-1 and $df_2 = N$-a

- To recap:
  - When the null hypothesis is true, *MSB* has a chi-squared distribution and the F-test, the ratio of two independent chi-squared variables, follows an F-distribution.
  - When the null hypothesis is false, *MSB* no longer follows a chi-squared distribution. The F-test now follows a non-central F-distribution.

- A quick return to the t-test
  - I claim that s^2_{pooled} from the t-test is <u>exactly</u> equal to MSW from an ANOVA on two groups.
  - Why?

- Let's start with the formula for s^2_{pooled} from the two-independent samples t-test:

$$s^2_{pooled} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$s^2_{pooled} = \frac{SS_1 + SS_2}{(n_1 + n_2 - 2)}$$

$$s^2_{pooled} = \frac{SS \text{ within each group}}{df}$$

  - Thus, when there are two groups, $s^2_{pooled} = MSW$

6. A recap of the F-test
   - We have now found two uses of the F-test
   - We can use the F-test to examine if the population variances of two independent samples are equal.

$$H_0: \sigma^2_{Min} \geq \sigma^2_{Max}$$
$$H_1: \sigma^2_{Min} < \sigma^2_{Max}$$
$$F(df_{larger}, df_{smaller}) = \frac{s^2_{larger}}{s^2_{smaller}} \qquad df_{larger} = n_{larger} - 1 \qquad df_{smaller} = n_{smaller} - 1$$

   - We can also use the F-test to examine if the population means of two or more groups are equal

$$H_0: \quad \mu_1 = \mu_2 = ... = \mu_a$$
$$H_1: \quad \text{Not all } \mu_i s \text{ are equal}$$

$$F(df_b, df_w) = \frac{MSB}{MSW} = \frac{SSB / a-1}{SSW / a(n-1)} \qquad df_b = a - 1 \qquad df_w = a(n-1)$$

   - These tests examine different hypotheses, are set-up differently, and result in very different conclusions! Be sure not to confuse these two different uses of the F-test!!!

7. Completing the one-way ANOVA table

- The results of a one-way analysis of variance are almost always displayed in a table similar to the following:

ANOVA

Source of Variation	SS	df	MS	F	P-value
Between Groups	SSB	a-1	SSB/df_b	MSB/MSW	
Within Groups	SSW	N-a	SSW/df_w		
Total	SST=SSB+SSW	N-1			

- The p-value listed is the p-value associated with $F(a\text{-}1, N\text{-}a)$. Although you should know how to determine if an observed F is significant based on tabled F-values, in general, it is better to look up exact p-values with a computer program.

- This table provides a nice summary of
  - The decomposition of *SST* into *SSB* and *SSW*
  - The decomposition of degrees of freedom
  - The pooled error variance (*MSW*)

- Here is the ANOVA table from the blood pressure example.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	322.95	3	107.65	1.58	0.23	3.24
Within Groups	1089.6	16	68.1			
Total	1412.55	19				

  - What is the null hypothesis? Should it be retained or rejected?

- Be able to fill in a partially completed ANOVA table (assuming equal *n*)

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	64.2	3				
Within Groups	52.3					
Total		39				

8. Confidence Intervals in ANOVA
   - Remember a confidence interval is determined by:

     $$Estimate \pm (Critical\ Value \;\; * \;\; Std\ Error)$$

   - For the one-sample t-test, we obtained the following formula:

     $$\bar{x}_{obs} \pm \left(t_{crit} * \frac{s}{\sqrt{n}} \right)$$

   - We need three parts:
     - The Estimate
     - The Standard Error
     - The Critical Value

   - Let's determine the parts of the confidence interval for ANOVA
     - The Estimate

       This is the easy part: we can use the cell mean

     - The Standard Error

       Our estimate of the variance is *MSW*

       For *n*, we should use n_j, the cell size for the mean of interest

       $$\sqrt{\frac{MSW}{n_j}}$$

     - The Critical Value

       We are computing an interval around a single mean, so we can use the critical value of the t-distribution or of the F-distribution:

       We can use a t-statistic: $t_{crit}(df_w)$

       Or we can use an F-statistic: $\sqrt{F_{crit}(1, df_W)}$

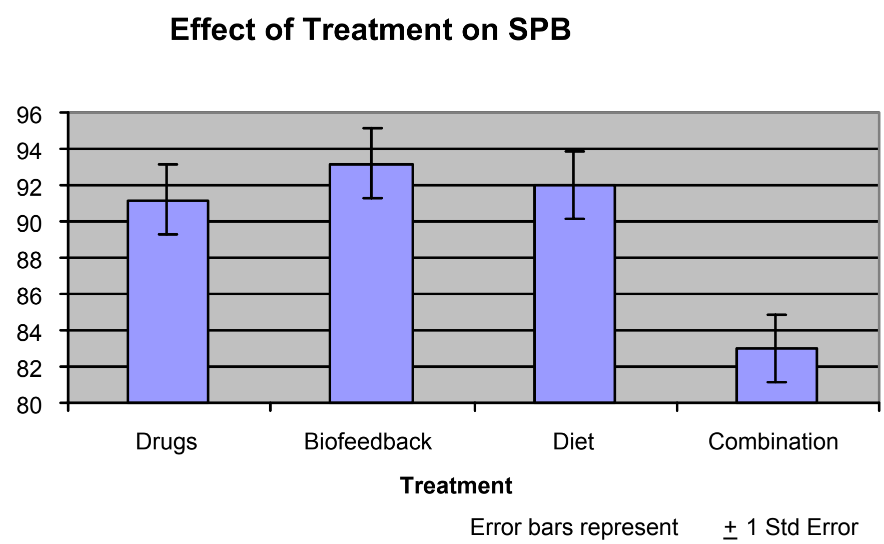
     - Putting it all together:

       $$\bar{x}_{.j} \pm \left(t_{crit}(df_W) * \sqrt{\frac{MSW}{n_j}} \right)$$

- Returning to the blood pressure example, to construct a 95% CI:
  - For the drug group $\bar{X}_1 = 91.20, n_1 = 5, MSW = 68.10$
  - Find t_{crit}: from EXCEL we find that $t_{.05/2}(16) = 2.12$

$$91.20 \pm \left(2.12 * \sqrt{\frac{68.10}{5}}\right) \text{ or } 91.20 \pm (2.12 * 3.69)$$

(83.38, 99.03)



9. Effect sizes in ANOVA

- Extending g: δ
  - Remember that for the two groups situation

$$g = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{pooled}}$$

  - In the ANOVA case, we have MSW instead of σ^2_{pooled}

$$g^{\sim} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{MSW}}$$

  - One possibility is to report a d value for each pair of means, but this is cumbersome. A more workable approach is to report the effect size of the range of group means

$$\delta = \frac{\overline{X}_{MAX} - \overline{X}_{MIN}}{\sqrt{MSW}}$$

  - To interpret δ:

δ = .25	small effect size
δ = .75	medium effect size
δ = 1.25	large effect size

- Cohen's f
  - Another approach is to calculate the standard deviation of the group means, $\hat{\sigma}_m$, and compare it to the within-group standard deviation.

$$\hat{\sigma}_m = \sqrt{\frac{\sum_{j=1}^{r}(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})^2}{a}}$$

  - Then a standardized effect size measure would be given by

$$f = \frac{\hat{\sigma}_m}{\hat{\sigma}_e} = \frac{\hat{\sigma}_m}{\sqrt{MSW}}$$

  - Interpretation of f (for the multi-group case) is the same as the interpretation of d (in the two group case):

f = .10	small effect size
f = .25	medium effect size
f = .40	large effect size

- Eta Squared (η^2) aka R^2
  - A measure of the proportion of the variance accounted for by the effect

$$\eta^2 = \frac{SSBetween}{SSBetween + SSWithin} = \frac{SSBetween}{SSTotal}$$

  - Compared to the reduced model (a model with the grand mean only), what percent of the variance is explained by the factor?
  - It is equivalent to the correlation between the predicted scores (from the full model) means and the observed scores
  - η^2 is based on the specific sample from which it was calculated and can not be generalized to the population (It overestimates the true variance accounted for). Because η^2 is biased, and there are better measures available, you should not report η^2. Unfortunately, this is the only measure of effect size that SPSS computes.

- Omega Squared(ω^2)
  - An alternative measure of the proportion of variance that corrects the positive bias of η^2

$$\hat{\omega}^2 = \frac{SSBetween - (a-1)MSWithin}{SSTotal + MSWithin}$$

  - Interpretation of ω^2:

$\omega^2 = .01$	small effect size
$\omega^2 = .06$	medium effect size
$\omega^2 = .15$	large effect size

- Summary:
  - f is best measure to use to capture effect size
  - ω^2 is the best measure to use to discuss the percentage of variance accounted for
  - However, with more then two groups, it is very difficult to interpret these values. What is the effect? Why is the effect large?
  - Many wise people have suggested that effect sizes should not be reported for omnibus tests.
  - ?? Why can't we report an r effect size for ANOVA??

- Let's examine each of these measures with our blood pressure example:

$$\delta = \frac{\overline{X}_{MAX} - \overline{X}_{MIN}}{\sqrt{MSW}} = \frac{93.2 - 83}{\sqrt{68.1}} = 1.23$$

$$\sigma_m = \sqrt{\frac{\sum_{j=1}^{r} (\mu._j - \mu..)^2}{a}} = \sqrt{\frac{(91.2 - 89.85)^2 + (93.2 - 89.85)^2 + (92 - 89.85)^2 + (83 - 89.85)^2}{4}} = 4.02$$

$$f = \frac{\hat{\sigma}_m}{\sqrt{MSW}} = \frac{4.02}{\sqrt{68.1}} = .49$$

$$\eta^2 = \frac{322.95}{322.95 + 1089.60} = .228$$

$$\omega^2 = \frac{SSBetween - (a-1)MSWithin}{SSTotal + MSWithin} = \frac{322.95 - (3)68.1}{1412.55 + 68.1} = .08$$

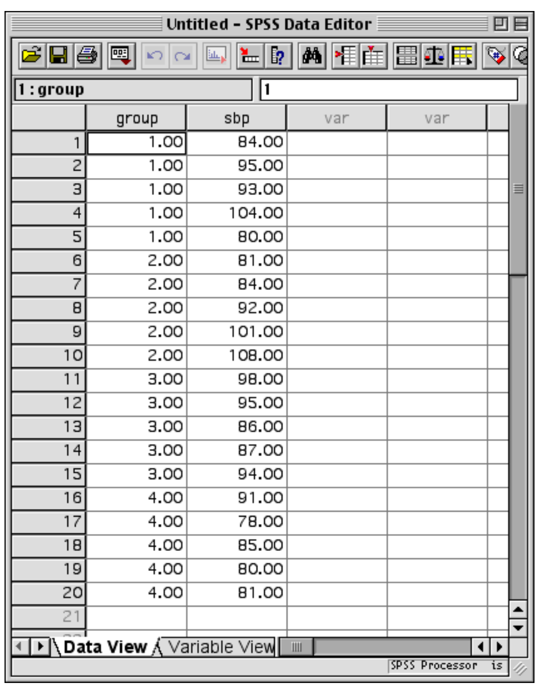
10.ANOVA in SPSS

- Data entry
  - Data must be entered with the condition in one column and the dependant variable in a second column.

```
DATA LIST FREE
  /group sbp.
BEGIN DATA.
1 84
1 95
.
.
4 80
4 81
END DATA.

VARIABLE LABELS
  group 'Experimental Condition'
  sbp 'Systolic Blood Pressure'.

VALUE LABELS
  group 1 'Drug'
          2 'Biofeedback'
          3 'Diet'
          4 'Combo'.
```



	group	sbp	var	var
1	1.00	84.00		
2	1.00	95.00		
3	1.00	93.00		
4	1.00	104.00		
5	1.00	80.00		
6	2.00	81.00		
7	2.00	84.00		
8	2.00	92.00		
9	2.00	101.00		
10	2.00	108.00		
11	3.00	98.00		
12	3.00	95.00		
13	3.00	86.00		
14	3.00	87.00		
15	3.00	94.00		
16	4.00	91.00		
17	4.00	78.00		
18	4.00	85.00		
19	4.00	80.00		
20	4.00	81.00		
21				

- Using ONEWAY

```
ONEWAY sbp BY group
 /stat=descriptives.
```

Descriptives

Systolic Blood Pressure

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Drug	5	91.2000	9.47101	4.23556	79.4402	102.9598	80.00	104.00
Biofeedback	5	93.2000	11.34460	5.07346	79.1138	107.2862	81.00	108.00
Diet	5	92.0000	5.24404	2.34521	85.4887	98.5113	86.00	98.00
Combo	5	83.0000	5.14782	2.30217	76.6081	89.3919	78.00	91.00
Total	20	89.8500	8.62234	1.92801	85.8146	93.8854	78.00	108.00

ANOVA

Systolic Blood Pressure

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	322.950	3	107.650	1.581	.233
Within Groups	1089.600	16	68.100		
Total	1412.550	19			

$$F(3,16) = 1.58, p = .23$$

- With ONEWAY SPSS gives us a confidence interval for $\bar{X}_1 = (79.44, 102.95)$
- We previously computed this CI to be $\bar{X}_1 = (83.38, 99.03)$. Why the difference from what we previously computed?

- Suppose you had a sample of n=5 with $\bar{X}_1$ = 91.20 and s^2 = 89.68. How would you construct a 95% CI?

    - Find t_{crit}: from EXCEL we find that $t_{.05/2}(4) = 2.78$

$$91.20 \pm \left(2.78 * \sqrt{\frac{89.68}{5}}\right) \text{ or } 91.20 \pm (2.78 * 4.235)$$

(79.42, 102.97)

- We did not use s^2 for the group; we used MSW as our estimate of the variance.
  - But to conduct an ANOVA, we MUST assume that the variances of all the groups are equal.
  - If this assumption is satisfied, then MSW is an estimate of the variance of group 1 and it is based on a larger sample than s_1^2. By the Law of Large numbers, MSW should be a better estimate of the variance than s_1^2

- SPSS ONEWAY, however, does NOT use a pooled variance estimate (using information from all the groups) to construct the confidence intervals. Instead, it uses only the information from the particular experimental group to construct the confidence interval.

- In other words, when you run an ANOVA, and the assumptions of an ANOVA are upheld, then you SHOULD NOT use the confidence intervals computed by SPSS ONEWAY (but you can use GLM/UNIANOVA).

- Using UNIANOVA/GLM

```
UNIANOVA sbp  BY group
  /EMMEANS = TABLES(group).
```

Tests of Between-Subjects Effects

Dependent Variable: Systolic Blood Pressure

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	322.950[a]	3	107.650	1.581	.233	.229
Intercept	161460.450	1	161460.450	2370.932	.000	.993
GROUP	322.950	3	107.650	1.581	.233	.229
Error	1089.600	16	68.100			
Total	162873.000	20				
Corrected Total	1412.550	19				

a. R Squared = .229 (Adjusted R Squared = .084)

Experimental Condition

Dependent Variable: Systolic Blood Pressure

Experimental Condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Drug	91.200	3.691	83.376	99.024
Biofeedback	93.200	3.691	85.376	101.024
Diet	92.000	3.691	84.176	99.824
Combo	83.000	3.691	75.176	90.824

  - Note that these Confidence Intervals are correct (assuming that the homogeneity of variances assumption is satisfied). You CAN use UNIANOVA to obtain valid CIs.
  - To obtain these confidence intervals, you must ask for a table of expected means; descriptives does not output confidence intervals .

- Oneway ANOVA in EXCEL is not recommended
  - o EXCEL can be useful in some circumstances, but it is very limited in its capabilities

- All procedures give identical results for the omnibus F-test.

$$F(3,16) = 1.58, p = .23$$

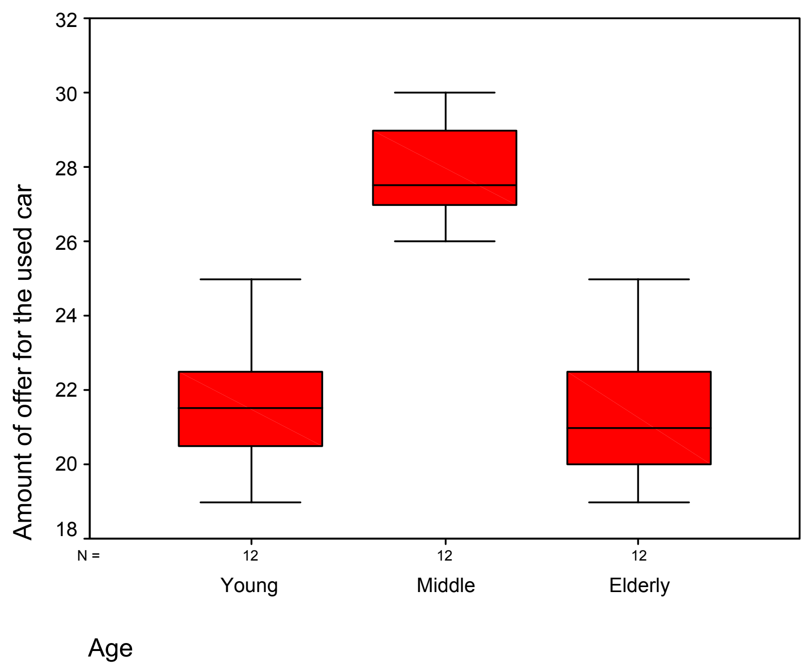
11. Two more examples

- Example #1: Age and car-buying (Neter, 1996, Ex 16.13)
  An organization wanted to investigate the effect of the car owner's age on the amount of cash offered to buy a used car. All "car owners" took a medium priced six year old car to 36 randomly selected dealers in the region. Here is a list of the offers received (in the hundreds):

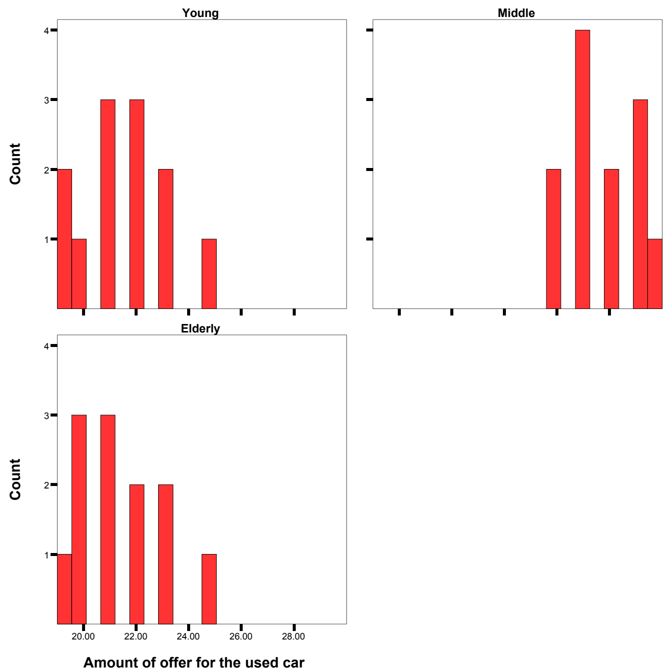
Young		Middle		Elderly	
23	20	28	27	23	21
25	23	27	30	20	20
21	19	27	38	25	19
22	22	29	27	21	20
21	19	26	26	22	22
22	21	29	29	23	21

- Step 1: Exploratory Data Analysis: Look at the data

  - Side-by-Side Boxplots



- Side-by-Side Histrograms
  (Use Interactive Graphs in SPSS)



- Look at the descriptive statistics
  EXAMINE VARIABLES=price BY age.

Descriptives

	Age		Statistic	Std. Error
Amount of offer for the used car	Young	Mean	21.5000	.50000
		Median	21.5000	
		Variance	3.000	
		Std. Deviation	1.73205	
		Interquartile Range	2.5000	
		Skewness	.315	.637
		Kurtosis	.234	1.232
	Middle	Mean	27.7500	.37183
		Median	27.5000	
		Variance	1.659	
		Std. Deviation	1.28806	
		Interquartile Range	2.0000	
		Skewness	.249	.637
		Kurtosis	-1.000	1.232
	Elderly	Mean	21.4167	.48396
		Median	21.0000	
		Variance	2.811	
		Std. Deviation	1.67649	
		Interquartile Range	2.7500	
		Skewness	.729	.637
		Kurtosis	.434	1.232


```
EXAMINE VARIABLES=price.
```

Descriptives

		Statistic	Std. Error
Amount of offer for the used car	Mean	23.5556	.56265
	Median	22.5000	
	Variance	11.397	
	Std. Deviation	3.37592	
	Interquartile Range	6.0000	
	Skewness	.427	.393
	Kurtosis	-1.186	.768

- Step 2: Set-up the Hypothesis Test and Conduct the Analysis

$H_0 : \mu_1 = \mu_2 = \mu_3$
H_1 : Not all μ_i's are equal

Use α=.05

```
ONEWAY price BY age
  /STAT=DESC.
```

Descriptives

Amount of offer for the used car

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Young	12	21.5000	1.73205	.50000	20.3995	22.6005
Middle	12	27.7500	1.28806	.37183	26.9316	28.5684
Elderly	12	21.4167	1.67649	.48396	20.3515	22.4819
Total	36	23.5556	3.37592	.56265	22.4133	24.6978

ANOVA

Amount of offer for the used car

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	316.722	2	158.361	63.601	.000
Within Groups	82.167	33	2.490		
Total	398.889	35			

$$F(2,33) = 63.60, p < .01$$

- Reject null hypothesis & conclude that not all age groups received the same cash offers for the car.

- Step 3: Calculate Confidence Intervals and Effect Sizes

  - Confidence Intervals:

    Use SPSS:

```
UNIANOVA price  BY age
  /EMMEANS = TABLES(age).
```

Age

Dependent Variable: Amount of offer for the used car

Age	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Young	21.500	.456	20.573	22.427
Middle	27.750	.456	26.823	28.677
Elderly	21.417	.456	20.490	22.343

Or calculate them by hand:

$$\bar{x}_{\cdot j} \pm \left(t_{crit}(df_W) * \sqrt{\frac{MSW}{n_j}} \right)$$

$t_{crit}(33) = 2.0345$ for $\alpha = .05$, two-tailed
$MSW = 2.490$
$n_j = 12$ for all j

Youth: $21.5 \pm \left(2.0345 * \sqrt{\frac{2.490}{12}} \right)$ (20.57, 22.43)

Middle: $27.75 \pm \left(2.0345 * \sqrt{\frac{2.490}{12}} \right)$ (26.82, 28.68)

Elderly: $21.4167 \pm \left(2.0345 * \sqrt{\frac{2.490}{12}} \right)$ (20.49, 22.34)

- Effect Sizes:

$$\delta = \frac{\overline{X}_{MAX} - \overline{X}_{MIN}}{\sqrt{MSW}} = \frac{27.72 - 21.42}{\sqrt{2.490}} = 3.99$$

$$\sigma_m = \sqrt{\frac{\sum_{j=1}^{r}(\mu._j - \mu.)^2}{a}} = \sqrt{\frac{(21.5 - 23.56)^2 + (27.75 - 23.56)^2 + (21.42 - 23.56)^2}{3}} = 2.965$$

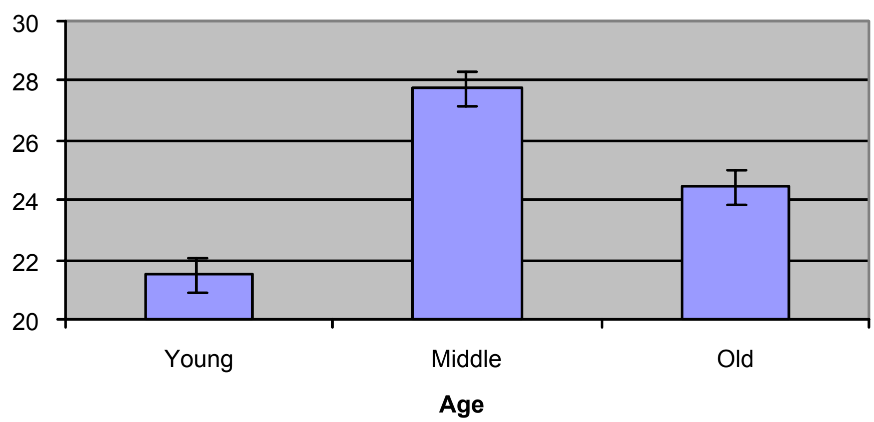
$$f = \frac{\sigma_m}{\sqrt{MSW}} = \frac{2.965}{\sqrt{2.490}} = 1.88$$

$$\eta^2 = \frac{316.722}{398.889} = .794$$

$$\omega^2 = \frac{SSBetween - (a-1)MSWithin}{SSTotal + MSWithin} = \frac{316.722 - (2)2.49}{398.889 + 2.49} = .777$$

$F(2,33) = 63.60$, $p < .01$, $f = 1.88$

Age Discimination in Used Car Prices



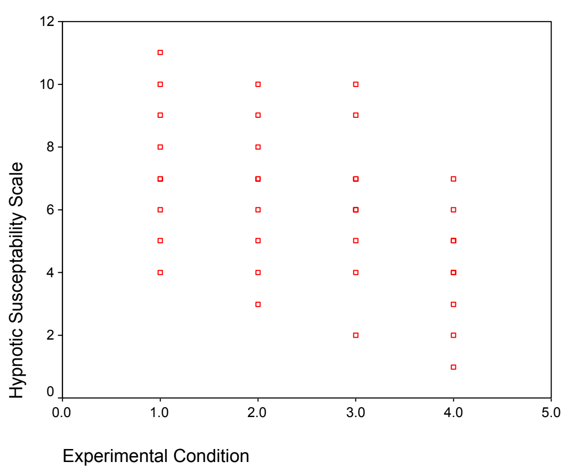
Error bars represent + 1 Std Error

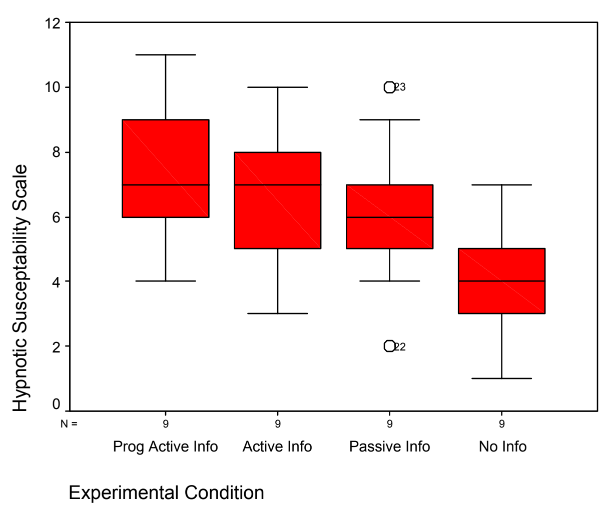
- Example #2: Susceptibility to Hypnosis (Kirk, 1995, Exercise 5.4)
  A researcher wanted to design instructions to maximize hypnotic susceptibility. Thirty-six hypnotically naive participants were randomly assigned to one of four groups:
  - Group 1: Programmed active information
  - Group 2: Active information
  - Group 3: Passive information
  - Group 4: Control group: no information

  All participants then completed the Stanford Hypnotic Susceptibility Scale. The following data were observed:

Group 1	Group 2	Group 3	Group 4
4	10	4	4
7	6	6	2
5	3	5	5
6	4	2	7
10	7	10	5
11	8	9	1
9	5	7	3
7	9	6	6
8	7	7	4

  - Step 1: Exploratory Data Analysis: Looking at the data
    - Scatterplot of Group by DV and Side-by-Side Boxplots



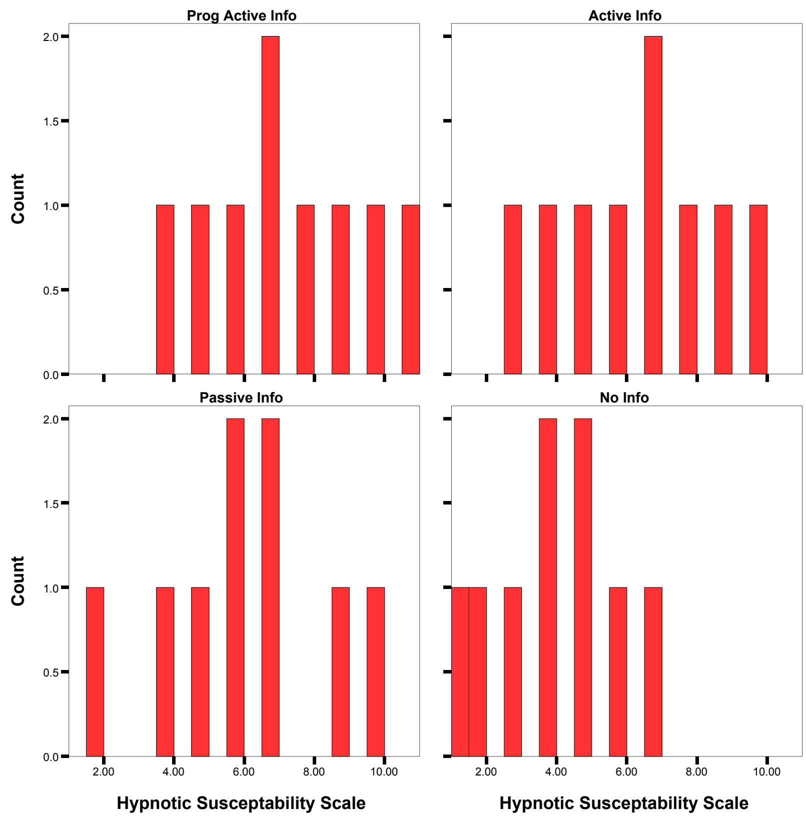


- Examine Descriptive Statistics

Descriptives

	Experimental Condition		Statistic	Std. Error
Hypnotic Susceptability Scale	Prog Active Info	Mean	7.4444	.76578
		Median	7.0000	
		Variance	5.278	
		Std. Deviation	2.29734	
		Interquartile Range	4.0000	
		Skewness	.092	.717
		Kurtosis	-.828	1.400
	Active Info	Mean	6.5556	.76578
		Median	7.0000	
		Variance	5.278	
		Std. Deviation	2.29734	
		Interquartile Range	4.0000	
		Skewness	-.092	.717
		Kurtosis	-.828	1.400
	Passive Info	Mean	6.2222	.81271
		Median	6.0000	
		Variance	5.944	
		Std. Deviation	2.43812	
		Interquartile Range	3.5000	
		Skewness	-.131	.717
		Kurtosis	.040	1.400
	No Info	Mean	4.1111	.63343
		Median	4.0000	
		Variance	3.611	
		Std. Deviation	1.90029	
		Interquartile Range	3.0000	
		Skewness	-.202	.717
		Kurtosis	-.490	1.400

- Side-by-Side Histograms



- Step 2: Set-up the Hypothesis Test and Conduct the Analysis

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1 : Not all $\mu_i 's$ are equal

Use α=.05


```
ONEWAY shsc BY group
 /STAT=DESC.
```

ANOVA

Hypnotic Susceptability Scale

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53.861	3	17.954	3.571	.025
Within Groups	160.889	32	5.028		
Total	214.750	35			

$$F(3,32) = 3.57, p = .025$$

- Step 3: Calculate Confidence Intervals and Effect Sizes
  - Confidence Intervals:

$$\bar{x}_{\cdot j} \pm \left(t_{crit}(df_W) * \sqrt{\frac{MSW}{n_j}} \right)$$

$t_{crit}(32)$ = 2.0369 for $\alpha = .05$, two-tailed
MSW = 5.028
n_j = 9 for all j

PAI: $7.444 \pm \left(2.0369 * \sqrt{\frac{5.028}{9}} \right)$ (5.92, 8.97)

Active Info: $6.556 \pm \left(2.0369 * \sqrt{\frac{5.028}{9}} \right)$ (5.03, 8.08)

Passive Info: $6.222 \pm \left(2.0369 * \sqrt{\frac{5.028}{9}} \right)$ (4.70, 7.75)

Control: $4.111 \pm \left(2.0369 * \sqrt{\frac{5.028}{9}} \right)$ (2.59, 5.63)

Experimental Condition

Dependent Variable: Hypnotic Susceptability Scale

Experimental Condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Prog Active Info	7.444	.747	5.922	8.967
Active Info	6.556	.747	5.033	8.078
Passive Info	6.222	.747	4.700	7.745
No Info	4.111	.747	2.589	5.634

- Effect Sizes:

$$\delta = \frac{\overline{X}_{MAX} - \overline{X}_{MIN}}{\sqrt{MSW}} = \frac{7.444 - 4.111}{\sqrt{5.028}} = 1.48$$

$$\sigma_m = \sqrt{\frac{\sum_{j=1}^{r}(\mu_{.j} - \mu_{..})^2}{a}}$$

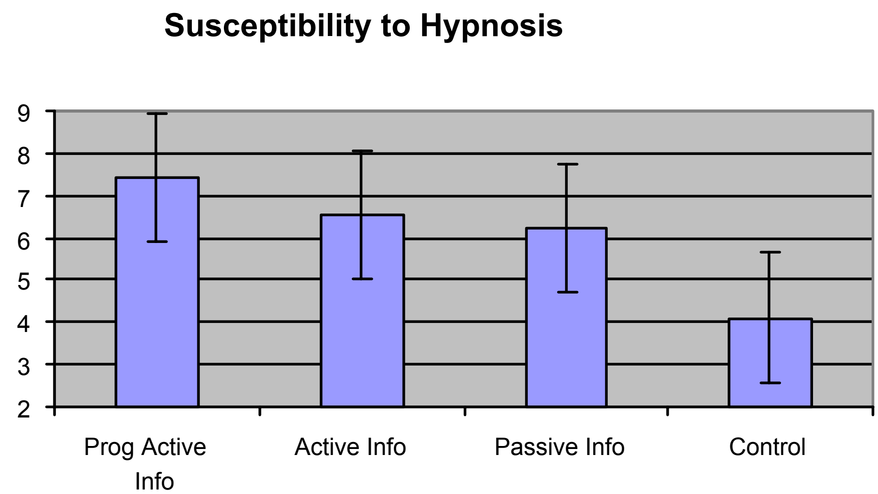
$$= \sqrt{\frac{(7.44 - 6.08)^2 + (6.56 - 6.08)^2 + (6.22 - 6.08)^2 + (4.11 - 6.08)^2}{4}} = 1.223$$

$$f = \frac{\sigma_m}{\sqrt{MSW}} = \frac{1.223}{\sqrt{5.028}} = .55$$

$$\eta^2 = \frac{53.861}{214.750} = .251$$

$$\omega^2 = \frac{SSBetween - (a-1)MSWithin}{SSTotal + MSWithin} = \frac{53.861 - (3)5.028}{214.75 + 5.028} = .176$$

$F(3,32) = 3.57$, $p = .025$, $f = .55$



Error Bars Represent 95% Confidence Intervals

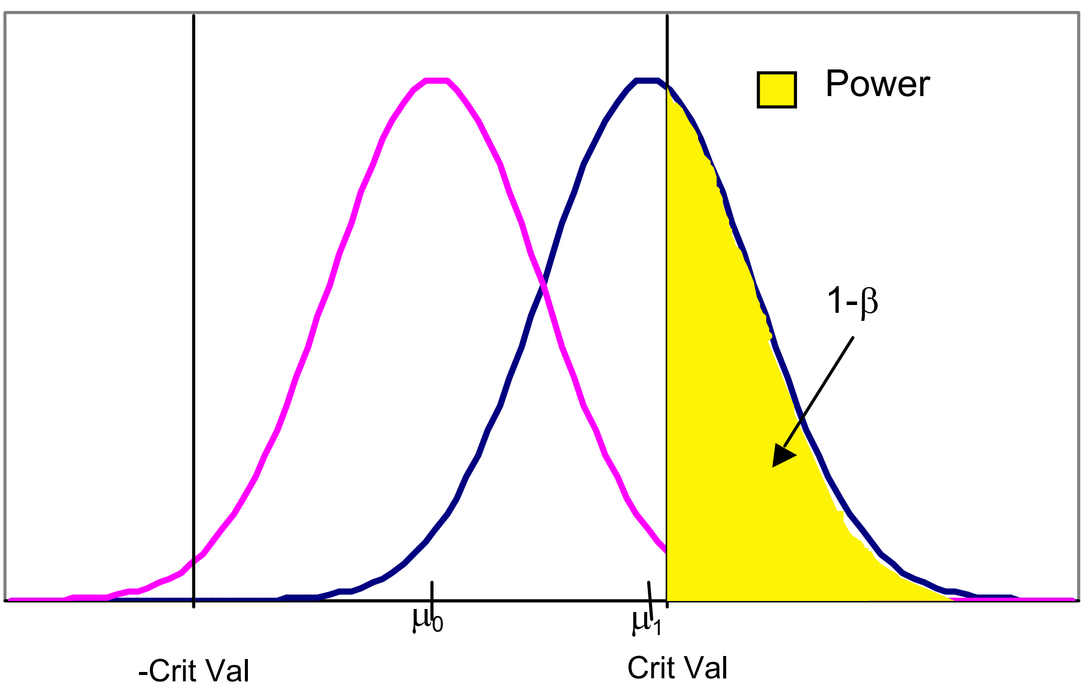
12.Power in ANOVA

- Three approaches to power
  - Post-hoc power
  - Power analysis to determine sample size for a t-test
  - Power analysis to determine sample size in ANOVA

- A quick review of statistical power

State of the world	Decision: Reject Null Hypothesis	Decision: Fail to Reject Null Hypothesis
Null Hypothesis TRUE	Type 1 Error Probability = α	Correct Decision Probability = $1 - \alpha$
Null Hypothesis FALSE	Correct Decision Probability = $1 - \beta$ (POWER)	Type 2 Error Probability = β

  - Power is the ability to detect a difference that is present
  - Because power concerns a difference that is present, all power calculations take place under the alternative hypothesis



- Post-hoc power
  - After you conduct a statistical test, you can compute its power.
  - In practice, this is a relatively pointless exercise:
    - If you reach statistical significance, then you must have had a large enough sample to detect the effect, and your power will be high.
    - If you did not reach statistical significance, then you must not have had a large enough sample to detect the effect, and your power will be low
    - Thus, post-hoc power will correlate strongly with the significance level
    - Unfortunately, this type of power is what SPSS gives you when you ask for power.

  - Calculating post-hoc post lays the framework for the concepts we will use for a power analysis to plan sample size, so it is enlightening to see how power is calculated after the data is collected.
    - See Appendix A for details on how to calculate post-hoc power in cases where you have are one or two independent samples.

- A more common use of a power analysis is to plan the sample size before running a study so that the study will have sufficient power.

- Power analysis to determine sample size for a two-sample t-test
  - A common power benchmark is $1-\beta=.80$ so that 80% of the time an effect is present our study will detect it.

  - Cohen (1998) has provided tables that simplify the process of determining the necessary sample size for a test. To use his tables, you need to specify
    - α The Type I error rate
    - Whether you will conduct a one-tailed or two tailed test
    - d Cohen's effect size

  - We generally set $\alpha_2 = 0.05$, so that quantity is known in advance
  - Thus, we need to estimate d in order to determine the sample size for $1-\beta=.80$.

- To determine sample size:
  - Set $\alpha_2 = 0.05$ (usually)
  - Estimate the effect size (*d*) you will find in your study
  - Use charts and tables to look up the sample size necessary to obtain the desired power – usually $1-\beta = .80$
  - CAUTION: Cohen's tables provide the necessary sample size *per group* to achieve the desired power. For a two-sample t-test, you must double the tabled value.

- Example #1: Suppose we want to find a small effect (d = .2) with $\alpha_2 = 0.05$ and $1-\beta = .80$
  - Yikes! We need 393 per group for a total sample of N = 786

- Example #2: Suppose we want to find a small effect (d = .2), but we are willing to settle for 60% power. ($\alpha = 0.05$ and $1-\beta = .60$)
  - We still need 246 per group for a total sample of N = 492

- Example #3: Suppose we want to find a medium effect (d = .5) with $\alpha = 0.05$ and $1-\beta = .80$
  - Now we *only* need 64 per group for a total sample of N = 128

- Power analysis to determine sample size for a oneway ANOVA

  - In this case, we need to estimate/know the following information:
    - α The Type I error rate
    - u The degrees of freedom in the numerator of the F-test
      a-1
    - f The standardized effect size

    - Again, Cohen's tables provide the necessary sample size *per group* to achieve the desired power. For an ANOVA design, you must multiply the tabled value by the number of groups.

  - Example #1: Suppose we want to find a small effect (f = .10) with $\alpha = 0.05$ for a two-group design (u = 1) with 80% power.
    - We find that n = 393. Thus, we require a sample of N = 786

- Example #2: Suppose we want to find a medium effect (f = .25) with $\alpha = 0.05$ for a five-group design (u = 4) with 80% power.
  - We find that n = 39. Thus, we require a sample of N = 195

- Example #2: Suppose we want to find a medium effect (f = .25) with $\alpha = 0.05$ for a five-group design (u = 4), but we are willing to settle for 50% power.
  - Now, we find that n = 21. Thus, we require a sample of N = 105. (But with this sample, we would only have a 50% chance of detecting a medium sized effect!)

- For different types of analyses, a power analysis to determine sample size boils down to the same procedure: You must estimate some sort of effect size and then use this information to determine the necessary sample size for a given level of power.

- Words of caution about power analyses:
  - Be realistic about your assumptions! Small changes in your assumptions can result in large changes in the necessary sample
  - All these calculations assume equal n per cell. In general it is unwise to plan a study with unequal n. However, if you know in advance that the sample sizes will be unequal, you should take that information into account in your power analysis
  - These power analyses only determine the sample size to detect an omnibus effect. If you do not care about the omnibus effect, then different procedures should be used.

- I am NOT a fan of power analyses.
  - If the estimates for a power analysis are off, the estimated sample size will be off, sometimes dramatically.
  - Frequently people attempt to conduct a power analyses when they have too little information. In these cases, a pilot study should be conducted.
  - Pilot studies can be beneficial in determining necessary sample size, and in refining the manipulation/materials to decrease error and increase the effect size.

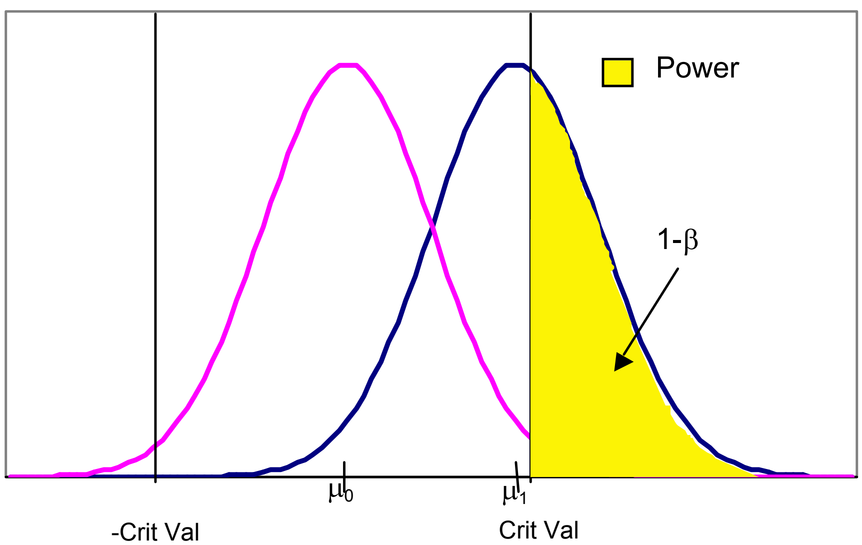
- Power analyses are required for all grant proposals, so you need to be familiar with them.

Appendix

A. Post-Hoc Power Analyses

- A Post-hoc power analysis is the process of determining the power of a statistical test after you have collected the data and conducted the test.

- To calculate the power of a test, we need three pieces of information:
    - The raw effect size
    - The standard error (the variability of the sampling distribution)
    - The alpha level (which we use to determine the location of the critical value, determined under the null hypothesis, in the alternative hypothesis distribution



- The three step procedure to calculating post-hoc power:
  - Step #1: Under the H_0 curve, find the critical value, $\bar{X}_{crit}$, in the units of the original scale, associated with critical value of the test statistic.
  - Step #2: Switch to the H_1 curve, and identify the z-score associated $\bar{X}_{crit}$.
  - Step #3: Under the H_1 curve, find the p-value associated with the z-score and determine the power of the test.

- Example #1: one-sample t-test (TV viewing example, p. 2-30)
  - $\bar{X}_{obs} = 22.53 \qquad \mu_0 = 21.22$
  - $s = 2.85$
  - $n = 50 \quad df = 49$
  - For two-tailed test with $\alpha = .05, t_{crit} = \pm 2.00957$
  - Step #1: Find $\bar{X}_{crit}$ (the value of $\bar{X}$ exactly associated with t_{crit}) under the null hypothesis curve

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$\pm 2.00957 = \frac{\bar{X} - 21.22}{\frac{2.84}{\sqrt{50}}} \qquad \bar{X}_{crit} = 20.378 \text{ and } \bar{X}_{crit} = 22.027$$

  - Step #2: Find the z-score associated with the appropriate $\bar{X}_{crit}$ under the alternative hypothesis curve
    - What do we know about the H_1 curve
      - We know its mean: $\bar{X}_{obs} = 22.53$
      - We know its standard deviation: $\frac{s}{\sqrt{n}} = .4016$
      - We know its approximate shape: it approximates a normal curve
      - $H_1 \sim N\left(\bar{X}_{obs}, \frac{s}{\sqrt{n}}\right)$ or $N(22.53,.4016)$

- Using this information, we can find the z-score associated with appropriate $\overline{X}_{crit}$ for a $N(22.53,.4016)$ curve:

$$z = \frac{\overline{X}_{crit} - \overline{X}_{obs}}{\frac{s}{\sqrt{n}}} = \frac{22.027 - 22.53}{\frac{2.84}{\sqrt{50}}} = -1.25$$

  - Step #3: Find the p-value associated with the z-score and determine the power of the test

    For z = -1.2523, p = .1052

- Interpret p (draw curves!)

  In this case, estimated power = 1-.1052 = 0.8947

- For example #1, when we calculate post-hoc power in this manner, we are only approximating the power because we have made a key simplification:
  - We have assumed that the alternative hypothesis distribution is normally distributed (we used z-tables to determine the area under its curve).
  - Actually, the alternative hypothesis distribution follows a non-central t-distribution. The non-central t-distribution has two parameters:

    ν = degrees of freedom

    δ = non-centrality parameter

  We can use SPSS UNIVANOVA to calculate the exact observed power.

- Example #2 (revisited)
  - First, let's subtract the null hypothesis from the data, so that the null hypothesis becomes $H_0 : \mu = 0$

```
compute t_hour = hours - 21.22.
```


- Now, we can use UNIANOVA to test if the mean differs from zero (that is, to conduct a one-sample t-test examining differences from zero), and we can ask for the observed power:

```
UNIANOVA t_hour
  /PRINT = OPOWER.
```

Tests of Between-Subjects Effects

Dependent Variable: T_HOUR

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	85.439	1	85.439	10.529	.002	10.529	.889
Error	397.629	49	8.115				
Total	483.068	50					
Corrected Total	397.629	49					

a. Computed using alpha = .050

- Note that our estimate of the power was close
  Estimated power = 0.895
  Actual power = 0.889

- SPSS also prints the non-centrality parameter for the non-central F-distribution
  $\hat{\phi} = 10.529$

  Note: ϕ is the non-centrality parameter of the F-distribution
  δ is the non-centrality parameter of the t-distribution

- Understanding the non-centrality parameter

$$\delta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{2}{n}}\sigma} = d\sqrt{\frac{n}{2}} \qquad \phi = \frac{\sqrt{\frac{\sum(\mu_i - u)^2}{a}}}{\frac{\sigma}{\sqrt{n}}} = \frac{\sigma_m}{\frac{\sigma}{\sqrt{n}}} = f\sqrt{n}$$

- The non-centrality parameter depends on
  - The difference between the population means
  - The population variance
  - The sample size