one way ANOVA: Use to 3 or more independent groups (categorical) on their means (dependent continuous variable)

	equal group sample sizes	**some definitions**	**unequal group sizes**
Null hypothesis	H_o: $\mu_1=\mu_2=\mu_3=$ ………$=\mu k$ H_a: $\mu_1\neq \mu_2\neq \mu_3\neq \mu_4$		same
Assumptions	• normality (dependent variable for each group is normally distributed) (use SPSS to check for normality – look at the curve – can transform the variable if it is severely skewed – p. 318) • independence (the observations are independent) – the research design should take care of this; if not use Repeated Measure ANOVA • homogeneity of variance use Levene's test in SPSS *(p should be greater than 0.05 if there is homogeneity of variance)* If sample sizes are equal or close, this can be ignored. If the population is asymmetric or similar in shape (skewness) and the largest variance is no more than 4x the smallest, the variance is probably valid. p. 316. If this is violated, use the Welch procedure p. 317.	other ways to deal with the lack of homogeneity of variance are: *use multiple comparison procedures – Games-Howell, Dunnett's C • use statistical tests that make fewer assumptions – like Kruskal-Wallis One Way ANOVA SPSS: Analyze, Compare Means, One Way ANOVA.	same
Set the criteria for rejecting the null	$\alpha = 0.05$ two degrees of freedom df_1=K-1 and df_2=N-K Then look up on Table 1 p. 672 in Howell to get the F_{cv} *The underlying distribution is the F distribution.*	N = subjects K = groups df_1 = between (numerator) df_2 = within (denominator)	same
Compute the test statistics	F = (between group variance)/ within group variance or F = (δ2 between)/ (δ2 within) or $\frac{SS_{between} / K-1}{SS_{within} / N-K}$ $SS_b = n\,\Sigma\,(X_j - X..)^2$ $SS_{total} = \Sigma(X_{ij} - X..)^2$ $SS_w = SS_{total} - SS_b$ Then take those scores – SS_b, SS_w and SS_{total}, and do this: $MS_{bet} = SS_{bet}/df_1$ $MS_{within} = SS_{within}/df_2$ Finally do this: $F=MS_{between}/MS_{within}$ Then arrange in a table: Source SS df MS F Between # # # # Within # # # Total # #	n = number of subjects Xj = an individual score – you have to do the formula for each score –the Σ means do that set for every score and add the results together Xbarj = group mean X.. = the whole group mean / grand mean / mean of all the scores $SS_{treat} = SS_{between} = SS_b$ $SS_{error} = SS_{within} = SS_w$ MS = mean square	The SS_{bet} has to take into consideration the different group sizes. So instead of the n times everything, the n's are multiplied by their groups. $SS_{bet} = \Sigma\, n_j(x_j - x..)^2$ nj is the group sample size
Make a decision	If F is greater than the critical F, reject the null. If p is smaller than 0.05, reject the null.	You can look at the means to see what the difference is, but if they are close, other analyses need to be done to see what the difference is between the groups.	
Effect Size	eta squared (η^2) or $\eta^2 = SS_{bet}/SS_{total}$ omega squared (ω^2) = $\frac{SS_{bet} - (K-1)MS_w}{SS_{total} + MS_w}$	*eta squared = x% of the variance can be explained by the differences in the treatment given to the groups *omega squared is considered better for balanced designs / equal group size but SPSS doesn't give it to you	0.01 – small 0.06 – medium 0.14 - large
Basic ANOVA Model	$Xij = \mu + \tau j + \varepsilon ij$ $Xij = \mu + (\mu j - \mu) + \varepsilon ij$	Xij - *i*th score in group *j* μ - grand mean (mean for all subjects) τj = (μj - μ) - effect of belonging to group *j* εij - random error associate with the individual score Xij	fixed models – independent variable is fixed. random model – treatments are

<table>
<tr><td></td><td></td><td></td><td>randomly chosen</td></tr>
<tr><td><u>Type 1 errors</u>
PC: error rate per comparison
FW: family wise error rate</td><td colspan="3">A Priori Multiple Comparisons
Determine the pairs of groups to compare prior to the study. Do these after the one-way ANOVA is significant and you want to find out how it is significant.</td></tr>
<tr><td></td><td>multiple t-tests</td><td colspan="2">linear contrasts</td></tr>
<tr><td>Concepts & assumptions</td><td>*homogeneity of variance assumed; equal sample size
-use MS_w with df_w

*homogeneity of variance NOT assumed, equal sample size
-use 2(n-1) df

*homogeneity of variance NOT assumed, unequal sample size
-use individual sample variance with df Welch Satterthwaite (Howell 202)</td><td colspan="2" rowspan="5">use this when you want to compare one group with a set of groups, or one set of groups with another set of groups

Linear combination:
$L = a_1x_1 + a_2x_2 + a_3x_3 \ldots\ldots\ldots + a_kx_k = \Sigma a_j\, x_j$

Linear contrast:
$\Sigma a_j = 0$

Then develop the linear contrast.
(Split the comparisons across so that they add up to 0. i.e. two groups compared would be -1 and 1; 5 groups compared in a set of 2 and a set of 3 would be .5, .5, -.33, -.33, -.33; etc.)

Then compute Σa_j
for each contrast. i.e. Sum the contrast squares for each row in the linear contrast table.

Then compute Psi. $\psi = \Sigma (a_j)(x_j)$
i.e. sum the contrast times the mean for each row.

Then compute ums of squares for each contrast:
$SScontrast = \frac{n\psi^2}{\Sigma aj^2}$
Do that for each row.

Develop the ANOVA table.

Between groups SS df MS F Sig
contrast a # # # # #
contrast b # # # # #
contrast c # # # # #

Within groups # # # # #
Total # #</td></tr>
<tr><td>null hypothesis</td><td>H_o: $\mu_{fw} = \mu_{2d}$
H_A: $\mu_{fw} < \mu_{2d}$
these are named by the trials</td></tr>
<tr><td>set the criteria for rejecting null</td><td>see assumptions for setting the df

and/or get it from SPSS (MS_w)

t_{cv} comes from Table on p. 682?</td></tr>
<tr><td>compute the test statistics</td><td>$t = \frac{xi - xj}{\sqrt{(2MSw/n)}}$
(for homogeneity & equal sample size)</td></tr>
<tr><td>make a decision</td><td>if the absolute value of t is greater than t_{cv}, reject the null</td></tr>
<tr><td>SPSS</td><td></td><td colspan="2">Analyze, compare means, one way ANOVA, contrast</td></tr>
</table>

Multiple Comparison Procedures

Type 1 errors	**Post hoc**				**Trend analysis**
PC: error rate per comparison FW: family wise error rate	Conduct procedures to determine which pairs are significantly different (after doing ANOVA). The problem is that “by making comparison after data is collected, the focus will be on large discrepancies and not based on probability.” –Bill Auxier family wise error: $\alpha_E = c\alpha$ where c = number of comparisons and α = a priori level of significance i.e. divide the .05 across the number of comparisons or use one of these others				Determine trend relationship between continuous independent and dependent variables
	Tukey	Student Newman Keuls SNK	unequal sample size Tukey/Kramer	Other procedures: If you’ve met homogeneity of variance, you can choose Tukey, SNK, Bonferroni, or Ryan (REGWQ). If you cannot assume equal population variances, you could choose Games-Howell. Remember: use Levene’s test in SPSS *(p should be greater than 0.05 if there is homogeneity of variance)*	
Concepts & assumptions		this is a less conservative method r = number of steps between ordered means *(i.e. put them in order from highest to lowest)*			You’re looking to see if there is a significant linear, quadratic, cubic or higher order relationship between the independent and dependent variables. Is the F ratio for the trend statistically significant.
null hypothesis	Ho: $\mu i = \mu j$ for $i \neq j$	same	same		SPSS:
set the criteria for rejecting null	$\alpha = 0.05$ using the $df = df_w$ *Uses Studentized Range (Q)* Howell p. 679 r = number of group means	$\alpha = 0.05$ a different r for each set of group means to compare, 4 if there’s 4, 3 if there’s 3, etc. $df = df_w$ Then look up the cv on Howell p. 679.	same You can also get the df_w from SPSS.		a. Analyze, Compare Means, One Way Anova b. Factor is the independent variable c. Click option, descriptive, homogeneity and means plot. d. click Contrast e. check polynomial to the 4^{th} degree
compute the test statistics	$Q_r = \frac{x_i - x_j}{\sqrt{(MS_w/n)}}$ First make a chart like the top of page 3 in lesson 2.2. Order the means, then subtract each one from each other in order to keep organized until you’ve subtracted them all. Use the absolute values. Then get MS_w from SPSS. Calculate this: $\sqrt{(MS_w/n)}$ to get the bottom of the formula. Then finish the formula by dividing each score in the chart by the answer you got from $\sqrt{(MS_w/n)}$.	Make the computation table the same as Tukey’s procedure. It’s different in that it’s more powerful to detect differences because of the extra dfs.	$Q = \frac{x_i - x_j}{\sqrt{[(MS_w/n_i) + (MS_w/n_j)]/2}}$ If you do it this way, you have to use each group comparison n (n_i and n_j) and do it multiple times. So some authors have suggested this way - and SPSS does it this way – and it’s a lot easier this way: $Q_r = \frac{x_i - x_j}{\sqrt{(MS_w/n_h)}}$ k $nh = \frac{k}{1/n_1 + 1/n_2 + \ldots\ldots 1/n_j}$ k is the number of groups n_h is the harmonic mean		
make a decision	If the Q is greater than Q_{cv}, then the result is significant.				

Simple Repeated Measures ANOVA: Use with dependent groups (same subjects with multiple treatments). An extension of the dependent samples t-test.

		Definitions	**SPSS notes**
Null hypothesis	Ho: $\mu 1 = \mu 2 = \mu 3$ Ha: $\mu i \neq \mu j$		
Assumptions	i. the sample is randomly selected from the population. ii. the dependent variable is normally distributed. iii. the **population** variance for treatments or occasions are equal iv. the **population** correlation coefficients between pairs of test occasion scores are equal.	iii & iv are *compound symmetry* *sphericity* = the equality of variance of the differences between the level of treatments of the repeated measures factor (present when there is compound symmetry)	SPSS Analyze, General Linear Model, Repeated Measure Then in the "within subject name" type the name of the independennt variable. Then type the number of treatments/observations in the Number of Levels. Click Add, Define. Move the variables over to the Within Subjects box. Click Options, move the independent variable over, check compare main effects, choose bonferroni. Check Descriptives & estimates of effect size. Click continue. Click Plots, put the independent variable in the horizontal axis, click Add. Click Continue, OK. When making the ANOVA table, get the BetweenTreatments row from the top row of the Tests of Within-Subjects Effects. Get the BetweenSubjects row from the bottom row (Error) of the Tests of Between Subjects Effects. Get the Error row from the bottom row (Error) from the Tests of WithinSubjects Effects.
Set the criteria for rejecting the null	$\alpha = 0.05$ *F distribution is the underlying distribution* $df_{bt} = K-1$ $df_{error} = (n-1)(K-1)$ Look up in Howell p. 672.	K is groups n is the number of subjects	
Compute the test statistics	We need: $F = (MS_{between\ treatments})/(MS_{error})$ To get there, we first make a chart with each subject down the left, the observations in the columns, and the total means for rows and columns in the far right (subject means) and bottom (group means) columns. Then we figure each of these: $SS_{total} = \sum (X_{ij} - \bar{X}_{..})^2$ (i.e. subtract the total mean from each score, square each one, then add them all together) $SS_{bs} = k\sum (\bar{X}_s - \bar{X}_{..})^2$ (i.e. subtract the total mean from each subject mean (far right column), square each one, add them all together). $SS_{bt} = n\sum (\bar{X}_j - \bar{X}_{..})^2$ (i.e. subtract the total mean from each column mean, square them, and add them together). $SS_{error} = SS_{total} - SS_{bs} - SS_{bt}$ Then make the Repeated Measures ANOVA Table In the rows, divide SS by df to get MS. $F = (MS_{between\ treatments})/(MS_{error})$ Source / SS / df / MS / F / F_{cv} Between Subjects / # / # Between Treatments / # / # / MStreatments / # / # Error / # / # / MSerror Total / #	*testing sphericity* Option 1: Variance-Covariane Matrix. In SPSS, run Analyze, Correlate, Bivariate. Choose Options, check "cross product deviation & covariances". At a glance, if the covariance rows seem fairly similar, then you can assume sphericity. Option 2: Mauchly's Test of Sphericity *p* values greater than or equal to 0.05 are required to assume sphericity. Run this in SPSS and look for the p value. (See instructions to the right, SPSS will automatically included it.) If it's violated, then use the F ratio with the Greenhouse-Geiser or Huynh-Feldt adjustment (These are also automatically given to you in SPSS).	
Make a decision	If F is greater than the critical F, reject the null. If p is smaller than 0.05, reject the null.		
Effect Size	partial eta squared or partial $\eta^2 = (SS_{between\ treatment})/(SS_{between\ Treatment} + SS_{error})$		0.01 – small 0.06 – medium 0.14 - large
Simple Repeated Measures Model	Structural model is: $Xij = \mu + \pi i + \tau j + eij$ Can be reduced mathematically to: SStotal =SSbs + SSbt +SSerror	Where Xij = score of person *i* in group *j* μ = the grand mean πi = constant associated with Person *i* τj = constant associated with Group *j* eij = error associated with Person *i* in Group *j*	*degrees of freedom* SStotal = N-1 SSbs = n-1 SSbt = K-1 SSerror = (n-1)(K-1)

Factorial ANOVA (Two-way, three-way, etc.) Comparing means (continuous dependent variable) for two or three or four independent variables (categorical).

		Definitions	
Null hypotheses	$H_{o1}: \mu_{a1} = \mu_{a2}$ (Factor A) $H_{o2}: \mu_{b1} = \mu_{b2} = \mu_{a3}$ (Factor B) $H_{o3}: \alpha\beta_{ij} = 0$ (There is no interaction between Factors A and B)	The first two hypotheses deal with main effects, the third with interaction effects.	
Assumptions	The samples are independent, random samples from defined population The scores on the dependent variable are normally distributed. The population variances in all cells of the factorial design are equal. For this, use Levene's test in SPSS. *(p should be greater than 0.05 if there is homogeneity of variance)*		**Configuration in Factorial ANOVA** 2-way ANOVA A x B 3-way ANOVA A x B x C 4-way ANOVA A x B x C x D A, B, C, D represent different independent variables each with values of 2 or greater. In the case of a 3 x 4 Two-way ANOVA, the first independent variable has *three* levels, treatments, or conditions, while the second independent variable has *four* conditions. In the case of a 3 x 4 x 2 Three-way ANOVA, the first independent variable has *three* conditions, the second IV *four* conditions, and the third IV *two* conditions. In factorial ANOVA, it's important to determine which independent variable is primary and which is (are) secondary.
Set the criteria for rejecting the null	$\alpha = 0.05$ *F distribution is the underlying distribution* $df_A = R-1$ $df_B = C-1$ $df_{AB} = (R-1)(C-1)$ $df_{error} = RC(n-1)$ $df_{total} = N-1$ Then use df_A df_B and df_{AB} as the numerator and df_{error} as the denominator to figure the three critical values. Look up in Howell p. 672.	R = number of levels for Factor A C = number of levels for Factor B N = total sample size n = sample size in each cell	
Compute the test statistics	First figure all the group means. Each of the groups. For each set for Factor A (cell mean). For each set for Factor B (cell mean). And the overall group mean. Factor B are the columns and Factor A are the rows in your chart. (See lecture notes p. 6) $SS_{total} = \sum (X_{ijk} - \bar{X}_{...})^2$ (i.e. subtract the total mean from each score, square each one, then add them all together) $SS_A = nc\sum (\bar{X}_{i.} - \bar{X}_{...})^2$ (subtract the total mean from each mean for Factor A (row means), square each one, then add them together and multiply by n and c) $SS_B = nr\sum (\bar{X}_{j.} - \bar{X}_{...})^2$ (subtract the total mean from each mean for factor B (column means), square each one, add then together and multiply by r) $SS_{cells} = n\sum (\bar{X}_{ij} - \bar{X}_{...})^2$ (subtract the total mean from each cell mean, square each one, add them together and multiply by n) $SS_{AB} = SS_{cells} - SS_A - SS_B$ $SS_{error} = SS_{total} - SS_{cells}$ Now construct the ANOVA table. In the rows, divide SS by df to get MS. $F = (MS_{between\ treatments})/(MS_{error})$ Source / SS / df / MS / F / F_{cv} Factor A / # / # / # / # / # Factor B / # / # / # / # / # Interaction (A*B) / # / # / # / # / # Error / # / # / # Total / # / #	c = number of conditions in Factor B or number of columns n = number of subjects in each cell r = number of conditions in Factor A or number of rows cell means = mean for each group that is defined by one condition for Factor A and one condition for Factor B A = Factor A B = Factor B AB = interaction between Factor A and B μ – grand mean α_i – effect of being in Group (Factor) A β_j – effect of being in Group (Factor) B $\alpha\beta_{ij}$ – Interaction of Factor A and Factor B ε_{ijk} – error associated with observation X_{ijk} X_{ijk} – Any observation	
Make a decision	Compare F and $F_{cv.}$ If F is greater than F_{cv}, then reject the null. If p is smaller than 0.05, reject the null.		Two Way Structural Model $X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ Can be reduced to: $SS_{total} = SS_A + SS_B + SS_{AB} + SS_{error}$
Effect Size	eta squared $\eta^2 = SS_B/SS_{total}$ to figure effect size for Factor B. Use this principle to figure the other effect sizes. It tells you what percent of the variance can be explained by that factor. 0.01 – small; 0.06 – medium; 0.14 - large		

<table>
<tr><th colspan="3">Interaction Effects for Factorial ANOVA To find out which interactions are significant after doing factorial analysis.</th><th>Factorial Analysis: Unequal Sample Sizes</th></tr>
<tr><td></td><td></td><td>SPSS</td><td>(use the harmonic mean; easiest in SPSS)</td></tr>
<tr><td>Set the criteria for rejecting the null</td><td>$\alpha = 0.05$
Use the same critical values for Factor A and B as used in the two-way ANOVA.</td><td></td><td rowspan="3">To run Factorial Analysis, run General Linear Models, Univariate.
SPSS will create weighted means to deal with the unequal sample sizes.

Look for homogeneity of variance as usual. In Tests of Between-Subjects Effects, look for significant main effect and interaction effects.

Then run the syntax to the left to examine the Test of Simple Effects. Look at the pairwise comparison tables to find significant differences.

To run three way analysis of variance, use SPSS. General Linear Models, Univariate.

To get the Test of Simple Effects for three way, use this syntax: (File, New, Syntax)

UNIANOVA
DependentVariable BY FactorA FactorB FactorC
/METHOD = SSTYPE(3)
/intercept=include
/emmeans = tables (FactorA)
/emmeans = tables (FactorB)
/emmeans = tables (maritalstatus)
/emmeans = tables (FactorA * FactorB * FactorC) compare (FactorA) adj(bonferroni)
/emmeans = tables (FactorA * FactorB * FactorC) compare (FactorB) adj(bonferroni)
/emmeans = tables (FactorA * FactorB * FactorC) compare (FactorC) adj(bonferroni)
/print = descriptive etasq homogeneity
/criteria = alpha(.05)
/design = FactorA FactorB FactorC FactorA * FactorB FactorA * FactorC FactorB * FactorC
FactorA * FactorB * FactorC.</td></tr>
<tr><td>Compute the test statistics</td><td>Compute Sums of Squares for each group (A) at each level of Factor (B).

Make a chart with the means for each cell (Factor B is columns; factor A is rows).

For each factor B at each level of group Factor A (n = number of subjects in each cell):

$SS_B = n\sum(\bar{X}_{i.} - \bar{X}_{...})^2$ (subtract that row's mean for from each mean in the row, square each one, then add them together and multiply by n)

For each group Factor A at each level of Factor B

$SS_A = n\sum(\bar{X}_{j.} - \bar{X}_{...})^2$ (subtract that column's mean for from each mean in the row, square each one, then add them together and multiply by n)

Now construct the ANOVA table.
(a) divide sums of squares by respective degrees of freedom to obtain mean squares
(b) divide mean squares by MSerror (found in the -2-way ANOVA table) to obtain the F ratio;
(c) compare obtained F ratio to Critical F
(d) if statistically significant, conduct post hoc multiple comparison procedures.

In the rows, divide SS by df to get MS. Divide MS by MSerror to get F.
<table><tr><th>Source</th><th>SS</th><th>df</th><th>MS</th><th>F</th></tr><tr><td>Factor B</td><td></td><td></td><td></td><td></td></tr><tr><td>Condition</td><td>#</td><td>#</td><td>#</td><td>#</td></tr><tr><td>Condition</td><td>#</td><td>#</td><td>#</td><td>#</td></tr><tr><td>Factor A</td><td></td><td></td><td></td><td></td></tr><tr><td>Condition</td><td>#</td><td>#</td><td>#</td><td>#</td></tr><tr><td>Condition</td><td>#</td><td>#</td><td>#</td><td>#</td></tr><tr><td>Condition</td><td>#</td><td>#</td><td>#</td><td>#</td></tr><tr><td>Error</td><td>#</td><td>#</td><td>#</td><td></td></tr></table></td><td>After entering or opening the data, go to:

File, New, Syntax

Copy and paste this syntax. Red stays the same. Change the variable names to match your data's variables.

UNIANOVA
DependentVariable BY FactorA FactorB
/METHOD = SSTYPE(3)
/intercept=include
/emmeans = tables (FactorA)
/emmeans = tables (FactorB)
/emmeans = tables (FactorA*FactorB) compare (FactorA) adj(bonferroni)
/emmeans = tables (FactorA * FactorB) compare (FactorB) adj(bonferroni)
/print = descriptive etasq homogeneity
/criteria = alpha(.05)
/design = FactorA FactorB
FactorA * FactorB.

Then choose Run, All.
Look particularly at the pairwise comparison tables to analyze the simple main effects.</td></tr>
<tr><td>Make a decision</td><td>Look at F and compare to F_{cv}. If F is greater than F_{cv}, reject the null. Next, run it in SPSS to determine which pairs of groups are significantly different.</td><td></td></tr>
<tr><td>Effect Size</td><td>$\eta^2 = (SS_{effect})/(SS_{effect} + SS_{error})$</td><td></td><td></td></tr>
</table>

<table>
<tr><td colspan="5">Multiple Regression To develop a "good" prediction equation, to determine the explained variance, and to test the significance of the individual predictors</td></tr>
<tr><td>Steps</td><td></td><td>Definitions</td><td>Test statistical significance of individual predictors</td><td>SPSS</td></tr>
<tr>
<td>Develop Regression Model…</td>
<td>…using method of least squares or so that $\sum(Y-Y')^2$ is at a minimum (the square of differences between actual points and the regression line).

$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots\ldots + b_kX_k$

First, in SPSS, get the means, standard deviations, and inter-correlations.
Analyze, Correlate, Bivariate, Option, Means&SD
This will give you $\bar{Y}$, $\bar{X}$ (all of them), and the r's.

Next, compute standardized regression coefficients.
$\beta_1 = \frac{r_{y1} - (r_{y2})(r_{12})}{1 - r^2_{12}}$
$\beta_2 = \frac{r_{y2} - (r_{y1})(r_{12})}{1 - r^2_{12}}$

Then compute unstandardized regression coefficients:
$b_1 = \beta_1(S_y/S_1)$
$b_2 = \beta_2(S_y/S_2)$

Then compute a, regression constant:
$a = \bar{Y} - \sum b_i\bar{X}_i$

Construct regression equation:
$Y = a + b_1X_1 + b_2X_2$</td>
<td rowspan="3">This set for the left column.
k = number of predictors
a = regression constant
b = unstandardized regression coefficients
β = standardized regression coefficient

$\bar{Y}$ = Mean of Y, criterion variable
$\bar{X}$ = Mean of predictor variables
DV = Dependent variable
IV = Independent variables

– standard deviation of the DV
S_1 – standard deviation of 1st IV
S_2 – standard deviation of 2nd IV

r_{y1} – correlation btw DV and 1st IV
r_{y2} – correlation btw DV & 2nd IV
r_{12} – correlation btw the two IVs

This set for the right column.
s_{bi} – standard error of the b

SS_{xi} = sums of squares for each predictor = $(n-1)S^2_{xi}$

SS_Y = sums of squares of Y (criterion) = $(n-1)S^2_Y$

$R^2_{1..23..k}$ = square of the multiple correlation between the first predictor variable and the remaining predictor variables

R = multiple correlation
k = number of predictors

Rsquared = explained variance

Standard error of the regression coefficient (Sb) is the variability (standard deviation) of the statistic over repeated sampling.</td>
<td rowspan="2">Test the significance of b of each predictor variable
a. H_o: β=0
b. α=0.05; underlying t distribution, df=n-k-1 p.682
c. compute test statistics

Calculate sums of squares for each predictor variable:
SS_{xi} = sums of squares for each predictor = $(n-1)S^2_{xi}$
SS_Y = sums of squares of Y (criterion) = $(n-1)S^2_Y$

Next calculate the standard error of the estimate:
$S^2_{y.12\ldots k} = \sqrt{\frac{SS_Y(1-R^2)}{n-k-1}}$

Compute s_{b1} and s_{b2}
$s_{b1} = \sqrt{\frac{S^2_{y.12}}{SS_{x1}(1 - R^2_{1.2})}}$
$\mathbf{s_{b2}} = \sqrt{\frac{S^2_{y.12}}{SS_{x2}(1 - R^2_{1.2})}}$

Now we can compute t:
$t_1 = b_1/s_{b1}$
$t_2 = b_2/s_{b2}$

d. If t is greater than t_{cv}, then reject the null.</td>
<td rowspan="3">Analyze, Regression, Linear
The criterion/dependent variable goes in the dependent variable box.
The predictor variable(s) go in the independent variables box.
Click Statistic, Estimates, Confidence Interval, Model Fit, Descriptives.

Note the means & SDs.

Note the correlation coefficients and which ones are significant.

Look at the Model Summary table to get R and the standard error of the estimate.

Look at the ANOVA table for the test of whether R is significantly different than 0.

Look at the coefficients table for the pieces of the regression equation – in the B column under Unstandardized coefficients. Note also the standard error in that column.

Look at the t & .Sig columns for the variable that is statistically significant for the prediction.

Look at the correlation columns.
Partial correlation is the correlation between two variables X and Y when the influence a third (X_2) or more variables (X_3, X_4, X_5) are removed from both X and Y.

Part or semipartial correlation is the correlation between two variables X and Y when the influence of a third (X_2) or more (X_3, X_4, X_5) variables are removed from EITHER X OR Y, BUT NOT BOTH.</td>
</tr>
<tr>
<td rowspan="2">Evaluate how "good" the model is</td>
<td>1. Is R significantly different than 0?
$R = \sqrt{(\beta_1 r_{y1} + \beta_2 r_{y2} + \ldots + \beta_k r_{yk})}$
Fill this in with the numbers from above. Then use hypothesis testing procedures to test for statistical significance:
a. H_o:R=0
b. α=0; underlying distribution is the F ratio; degrees of freedom: k and n-k-1, Howell p. 672
c. compute test statistic
$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$
d. if F is greater than F_{cv}, reject the null</td>
</tr>
<tr>
<td>2. Determine explained variance (R^2)
Just square R from above. Use it to convert to % explained variance. In social sciences 70+% explained variance is very good.</td>
<td>Interesting tidbit: Mathematically, ANOVA and multiple regression are both general linear models.</td>
</tr>
</table>

<table>
<tr><th colspan="5">Regression Diagnostics and Model Building</th></tr>
<tr><th></th><th></th><th colspan="3">Building the Regression Model</th></tr>
<tr><th>Regression Diagnostics</th><th>Definitions</th><th>Standard / Direct</th><th>Stepwise / Statisical</th><th>Sequential / Hierarchical</th></tr>
<tr>
<td rowspan="2">1. Inspect the normality of the variables.
In SPSS, do Analyze, Descriptive Stats, Descriptives. Click Options. Check Min, Max, Skewness, Kurtosis. Look at skewness (0 is perfectly normal, negative is negatively skewed, positive is positively skewed. Hopefully it's .5 or less.

2. Inspect the linearity of the variables.
In SPSS, do Analyze, Correlate, Bivariate. With ideal data, the predictors are all highly correlation with the dependent variable, but they are weakly correlated with each other. Also check the scatter plots. In SPSS do Graphs, Scatter, Matrix. Look for linear trends.

3. Check Distance, Leverage and Influence to examine outliers that could severely influence the regression model.
In SPSS, do Analyze, Regression, Linear. Add the dependent and independent variables. Click Save. Check Unstandardized for predicted & residuals, check studentized, Cook's & leverage values. For distance, look at the residuals row. Large residuals indicate large errors of prediction. For leverage, first calculate 3(k+1)/N. Then look in the Centered Leverage Value row to see if there are any greater than that number. For influence, look at the row for Cook's D. Values greater than 1 should be examined.

"If there are problems, follow this:
1. Make sure the 'outliers' are indeed valid outliers. If not, correct them to valid values, if you know what the 'correct' values are.
2. If the 'outliers' are valid values, run separate analysis with and without those outliers and see if the analyses are drastically different. If not, proceed by including them.
3. Or delete outliers from your analysis. Fitting a regression model is typically done on the basis of 'normal' data (fitting it to the most common data). Outliers are 'uncommon' and can be ethically deleted." (quote from Dr. Kijai)</td>
<td rowspan="2">Collinearity: When the predictor variables are intercorrelated with each other.

Distance: identifies potential outliers in the criterion (dependent) variable.

Leverage: examines potential outliers in the predictor (independent) variables. According to Howell, possible leverage values range from a low of 1/N to a high of 1.0. Values that exceed 3(k+1)/N should be examined.

Influence: a combination of distance and leverage which identify unusually influential cases.

Cook's D: a measure of influence.

Backward elimination: Enter all the predictor variables, then eliminate individual predictors if they do not significantly contribute to the regression model.
Forward selection: Enter one predictor at time starting with the predictor that has the highest zero order correlation with the criterion. Keep the variables in the model no matter what.
Stepwise regression:
Similar to forward, but, as each variable is entered, the contributions of previously entered variables are re-evaluated and could be dropped from subsequent model if their contributions are deemed 'not significant'.</td>
<td rowspan="2">Analyze, Regression, Linear. Leave the method as "Enter". Put the independent and dependent variables in.

Look at the three tables:
Model summary – get the R and the Rsquare for explained variance.

Look at the ANOVA table to see if R is significantly different from zero.

Look at the coefficients table to get the full regression equation. Write it out using the B column under unstandardized coeffificients.

Dependent variable = constant + (each variable multiplied by B from the SPSS table).

Double check collinearity in the coefficients table. Tolerance should be greater than .10 and VIF should be smaller than 10.</td>
<td rowspan="2">Three types within this category: backward elimination, stepwise regression, forward selection. See definitions.

Backward Elimination: Analyze, Regression, Linear. Change method to "backward". Look at the notes below the model summary to see which variables are in each step. Look at the Change Statistics to see if the changes are significantly different than 0.

Forward Selection. Analyze, Regression, Linear. Change method to "forward". SPSS will stop automatically. The last model is the good one.

Stepwise regression. Analyze, Regression, Linear. Change method to "stepwise". It's probably similar or identical to forward.</td>
<td>Analyze, Regression, Linear. This time the researcher says the order the variables are added. Leave method as Enter. Enter the variables in the order you want them analyzed.</td>
</tr>
<tr>
<td>Cross Validation
Run the Direct Method Regression on data set A using the significant predictors.

Then use that equation to run on data set B.
Transform, Compute, and put in the regression equation with the variables & constant etc.

Then check the correlation between the original dependent/criterion variable and the predicted variable (the new data column made with Transform, Compute).
Hopefully the correlation will be close to the R from data set A.</td>
</tr>
</table>

Logistic Regression (Use to predict categorical variables when the predictors are a mix of continuous and categorical, or if the continuous variables aren’t normally distributed.)		
Steps	**Definitions**	**Univariate Analysis**
Regression equation: $\text{Log} \frac{Y'}{(1-Y')} = a + b_1X_1 + b_2X_2 \ldots\ldots$ 1. develop regression equation 2. determine if the model is statistically significant 3. determine explained variance 4. determine significance of individual predictors SPSS Analyze, Regression, Binary Logistic Put dependent variable in. Put independent/predictor variables in the covariate box Options, check Hosmer goodness of fit, iteration history, CI for Ex. In the variables in the equation box look for B, SE, Wald, Sig, Exp(B). If sig is less than 0.05 then that variable is considered statistically significant. Exp(B) is the odds ratio. Use it to say the odds of the event happening. The Omnibus Tests of Model Coefficients table tells you if the model significantly predicts the dependent variable. The model summary table gives you the explained variance. Use the Nagelkerke one. Look at the Sig. column in the Variables in the Equation table to find which variables are significant predictors. Then use those to run a restricted model with just those variables. Look at the Hosmer and Lemeshow Test table to see if there is linearity between the log odds of the criterion/dependent variable and the predictor/independent variables. There should be. This is an important assumption of logistic regression. Sig should be greater than .05.	Odds – the chance of being in that group Odds ratio – the ratio of two odds a - constant b - regression coefficients X - predictors (which may be categorical or continuous) Y’ – predicted probabilities of the event Wald statistic – the square of B/SE. Tests whether the regression coefficient is significantly different from zero.	Univariate analysis can give a better picture of the relationship between the criterion/dependent and predictor/independent variables. Run ChiSquare to examine the relationship between work status and the presence of children. Read the Likelihood row in SPSS. Use the Group Statistics and T-Test to see if the differences are significant.