	One sample case for the mean			**One sample case for the correlation**			
	Z-Test: Known Population Variance (σ2) (uncommon)	**T-Test: Unknown Population Variance (σ2)**	**SPSS**	**Known population correlation** (uncommon)	**Unknown population correlation**		**SPSS**
State the null hypothesis	H_o: μ=100 (or something)	Ho : μ=162 (or something) Ha : μ≠162 (non directional)	*Go to Analyze, Compare Means, One Sample T-Test*	Ho : ρ=a *(is the correlation the same as the population?)* Ho : ρ=0.35 Ha : ρ≠0.35	Ho : ρ=0 Ha : ρ≠0 *Is the correlation significantly different from 0?*	*alternative method* Ho : ρ=0 Ha : ρ≠0	*Go to Anaylze, Correlate, Bivariate*
Set the criteria for rejecting the null	(α) =0.05 Critical value = 1.96 for a two tailed test or 1.645 for a one tailed test. *See also Table 8.1 p. 189* *Based on normal distribution*	(α) =0.05 Critical value = Look at Table C3 Use n-1 for df *Based on Student's t distribution*	(α) =0.05	(α) =0.05 Critical value = 1.96 for a two tailed test or 1.645 for a one tailed test. *See Table 8.1 p189* *Based on normal distribution*	(α) =0.05 Critical value = Look it up on Table C3 Use n-2 for df *Based on Student's t distribution*	(α) =0.05 Critical value = Look it up on Table C3 Use n-2 for df	(α) =0.05
Compute the test statistics	***Z-test for the one sample mean*** $Z = \frac{x - \mu}{\sigma m}$ $\sigma m = \frac{\sigma}{\sqrt{n}}$	***t-test for the one sample mean*** $t = \frac{x - \mu}{Sm}$ $Sm = \frac{S}{\sqrt{n}}$	Look for the p value in the sig. (2 tailed) column.	$Z = \frac{Zr - Z\rho}{Szr}$ Use Table C6 to convert r to Fisher's Z (zr) $Szr = \sqrt{(1/(n-3))}$	$t = r\frac{\sqrt{(n-2)}}{(1-r^2)}$		Look for the p value in the sig. (2 tailed) row. (It's the % error you make if you reject the null.
Construct a confidence interval	(CI) = Statistic (X) +/- (critical value x standard error of the statistic) 95% if we set (α) =0.05	CI = X +/- (CV) (Sm)		CI = Zr +/- (CV) (Szr)	CI = Zr +/- (CV) (Szr) Use Table C6 to convert r to Zr and then when done to convert both interval endpoints back to r		
Make a decision	If the Z score is less than CV, retain the null. If the Z score is greater than the CV, reject the null.	If the t score is less than CV, retain the null. If the t score is greater than the CV, reject the null.	Compare p value to (α) =0.05. If it is greater, reject the null.	If the Z score is less than CV, retain the null. If the Z score is greater than the CV, reject the null.	If the t score is less than CV, retain the null. If the t score is greater than the CV, reject the null. If the hypothesized correlation is outside the CI_{95}, reject the null.	If ρ is greater than the CV, it is significantly different than 0. If ρ is less than the CV, it is not significantly different than 0.	Compare p value to (α) =0.05. If it is 0.05 or less, reject the null.
Effect Size		$d = \frac{\text{mean difference}}{\text{standard deviation}}$ 0.2 small 0.5 medium 0.8 large					
Notes	x = sample mean μ = population / hypothesized mean σm = Standard error of the mean	df = degrees of freedom d= effect size S = standard deviation					

	Two sample case for the mean						
	FIRST *Test homogeneity of variance*	t-test for independent samples homogeneity of variance assumed	SPSS	t-test for independent samples homogeneity of variance NOT assumed	SPSS	t-test for dependent samples (or correlated samples) examples: pretest/ posttest; husbands / wives; twins, etc. n = number of paired observations	SPSS
Assump-tions	***Independent samples***: scores in one group are not influenced by scores in other group ***Homogeneity of variance***: variances in the population are equal. $\sigma 1^2 = \sigma 2^2$	1st test if we meet the homogeneity of variance assumption (see column 1)	*Go to Analyze, Compare Means, Independent Samples T-Test Test variable is the continuous variable; grouping variable is the two groups.* Look under Levine's test for Sig. If p>0.05 or α, assume homogeneity. Read the row that matches the homogeneity results (top row for homogeneity).	1st test if we meet the homogeneity of variance assumption (see column 1) *df is different* *standard error is different*	*Analyze, Compare Means, Independent Samples T-Test* Look under Levine's test for Sig. If p>0.05 or α, assume homogeneity. Read the row that matches the homogeneity results (bottom row for no homogeneity)		*Go to Analyze, Compare Means, Paired-Samples T Test* *Add the two variables.*
Hypo-thesis	Ho: $\sigma_1^2 = \sigma_2^2$ (desired outcome)	Ho: $\mu 1 = \mu 2$ or Ho: $\mu 1 - \mu 2 = 0$ Ha : $\mu 1 \neq \mu 2$ or Ha: $\mu 1 - \mu 2 \neq 0$		H0: $\mu 1 = \mu 2$ Ha: $\mu 1 \neq \mu 2$		H0: $\mu 1 - \mu 2 = d = 0$ Ha: $\mu 1 \neq \mu 2$	
Criteria for rejecting null	(α) =0.05 Fcv two dfs: n1– 1 and n2 -1 *Based on F distribution (Table C5)*	(α) =0.05 or 0.01 Look at Table C3 df = n1+n2 -2 *Based on t-distribution*		df= $\frac{[(S_1^2/n1)+(S_2^2/n2)]^2}{[(S_1^2/n_1)^2/(n_1-1)]+[(S_2^2/n_2)^2/(n_2-1)]}$ Look at Table C3 *Based on t-distribution*		(α) =0.05 / 0.01 df= n-1 Table C3 *Based on t-distribution*	
Test Statistic	Ratio between the variances Put the larger number in the numerator F= $\frac{S_1^2 \text{ (largest)}}{S_2^2 \text{ (smallest)}}$ Remember that S^2 is the variance.	find: $s^2 = \frac{[\Sigma X_1^2 - (\Sigma X_1)^2 / n_1] + [\Sigma X_2^2 - (\Sigma X_2)^2 / n_2]}{n_1 + n_2 - 2}$ then find: $S_{m1-m2} = \sqrt{[s^2 (1/n_1 + 1/n_2)]}$ then find: $t = \frac{(M_1 - M_2)}{S_{m1-m2}}$		find: $S_{m1-m2}=\sqrt{[(S_1^2/ n_1)+ (S_2^2/ n_2)]}$ then find $t = \frac{(M1 - M2)}{S_{m1-m2}}$		Find standard error $S_{md}= S_d / \sqrt{n}$ S_d=standard deviation of difference M_d=Mean of difference between pairs $t=M_d / S_{md}$	
CI	in SPSS, look at the sig for the F/Levene's test. If p<0.05, reject null & assume no h. of pop.var.	(CI) = mean difference between groups $(M_1 - M_2)$ +/- tcv x S_{m1-m2}	Look at the CI lower & upper.	(CI) = mean difference between groups $(M_1 - M_2)$ +/- tcv x S_{m1-m2}	Look at the confidence interval lower & upper.	CI = $M_d \pm$ tcv x S_{md}	
Effect Size (d)	remember: 0.2 small 0.5 medium 0.8 large	$\frac{M_1 - M_2}{s}$ where $s = \sqrt{s^2}$ or $d = t\sqrt{[(n_1+n_2) / n_1 n_2]}$		$\frac{M_1 - M_2}{s}$ or $d = t\sqrt{[(n_1+n_2) / n_1 n_2]}$		M_d / S_d or $t/\sqrt{n}$	
decision	If F is equal to or greater than Fcv, reject the null. (no homogeneity of variance)	If t is less than tcv, retain the null. If the hypothesized value (0) is within the interval, retain the null.	Look at Sig. (2-tailed). If p> α, retain the null. If the CI contains 0, retain null.	If t is less tcv, retain the null. If the hypothesized value (0) is within the interval, retain the null.	Look at Sig. (2-tailed). If p> α, retain the null. If the CI contains 0, retain the null.	If t is greater than tcv, reject the null. If CI does not contain 0, reject the null.	Look at Sig. (2-tailed). If p is greater than α, retain the null. If the CI contains 0, retain the null.
Notes	Independent variable (the two groups) Dependent variable – the continuous variable that the groups vary on (in theory)	m1= group 1 mean; m2 = group 2; mean; m1-m2= absolute value of mean difference between groups S_{m1-m2}= standard error of the difference between the group means s^2 = pooled est. of pop variance	(α) =0.05 or 0.01	remember $s = \sqrt{s^2}$	(α) =0.05 or 0.01	Make a column to find the difference. Find the mean (M_d) and the std deviation (S_d)	(α) =0.05 or 0.01

Linear Regression Steps

x is the independent variable; y is the dependent
Develop the regression equation using the ***method of least squares***:

1st calculate b.

$$b = \frac{n\Sigma XY - \Sigma X\Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

To get all these sums, go to SPSS, use Compute to create a variable column for XY; create another for X^2 ; then go to Analyze, Descriptive Statistics, Descriptives.

or for b use this:
$b = r\,(S_y/S_x)$
(S is the standard deviation)

Then calculate a.

$$a = \frac{\Sigma Y - b\Sigma X}{n}$$

Then plug the numbers into the regression equation:
$Y' = a + bX$

How good is it?
Compute the square of the correlation between the criterion & predictor variables. Converted to percent, this indicates x percent of the criterion scores can be predicted by the predictor variable. (This is also called variance.)

Is it significant?
Is the correlation (r) statistically significant?
Check p compared to 0.05 (should be less)

Predicting scores
Plug X into the regression equation.
$Y' = a + bX$
Can do in SPSS or by hand.

Errors of prediction / standard error of estimate
Higher the correlation, the smaller the prediction/residual errors. To find the standard error of the estimate / the standard deviation of the distribution of errors of prediction:
$S_{y.x} = S_y[\sqrt{(1-r^2)}][\sqrt{(n-1)/(n-2)}]$ (6.11)
$S_{y.x} = S_y\sqrt{(1-r^2)}$ (6.12 for large samples, 30+)

Reminders:
a = Y intercept = regression constant
b = slope = regression coeffecient

Conditional Distributions

Assumptions:
The conditional distribution is normal.
The mean of the distribution is the predicted score Y'
The standard deviations of all the conditional distributions are equal (homoscedasticity).
The standard deviation of the conditional distributions is the standard error of the estimate ($S_{y.x}$)

Probability of getting the predicted score or higher
$Z = (Y - Y')/ S_{y.x}$

Y – Criterion Variable
Y' – the predicted score for each fixed value of X; or the mean of the distribution of Y for each value of X.
Sy.x – Standard error of estimate or standard deviation of each conditional distribution

Put the z score in the Normal Distribution Calculator or (Table C1)
http://davidmlane.com/hyperstat/z_table.html
Click radio box for ***above*** Z
Read the shaded area and convert to % probability of getting that score.

Probability of getting another score or higher
Same as above, put the "another score" in Y
Click radio box for ***above*** Z
You're predicting y given x

Confidence Intervals
$CI = Y' \pm (t_{cv})S_{Y'}$

Y' = predicted score
tcv – critical value for t for df n-2 (Table C3)

$$S_{Y'} = S_{y.x}\sqrt{\left\{1 + (1/n) + \left[\frac{(X - Xbar)^2}{(n-1)S^2_X}\right]\right\}}$$

Look back at the all the numbers you got in SPSS in step 1 under Linear Regression to fill this in.

Test Significance

In simple linear regression (only one predictor variable), if the correlation is significant, the regression coefficient is significant.

Hypothesis
H_o: β=0 (β is the population regression coefficient)

Set Criterion
(α) =0.05 or 0.01
Look at Table C3
df = n-2
Based on t-distribution

Test Statistics
$t = (b - \beta)/S_b$
remember β=0

$Sb = S_{y.x} / \sqrt{((n-1)S^2_X)}$
Get S^2_X from the SPSS sums.
$S_{y.x}$ See errors of prediction, left.

Make Decision
If t is greater than t_{cv}, the null must be rejected.

----
x= independent / predictor
y= dependent / criterion

SPSS *(Thanks Cheryl Kisunzu)*

Descriptives (means & std devs):
To get all the sums you need, go to SPSS, use Compute to create a variable column for XY; create another for X^2 ; then go to Analyze, Descriptive Statistics, Descriptives.

Correlation
Anaylze, Correlate, Bivariate

Linear Regression
Analyze, Regresssion, Linear then click the dependent variable to move it into the dependent variable box; followed by clicking the independent variable and moving it into the independent variable box then click OK

Scatterplot with regression line
Graphs, Interactive, Scatterplot, check regression under Fit tab.

How good is it?
Look at the Model Summary. R Square tells you how "good" it is; percent of variance in the criterion that's explained by the predictor.

Chi-Square Tests of Frequencies

	Goodness-of-fit (one-sample case)	SPSS	Test of independence *(or test of homogeneity or association)*	SPSS	Test of Two-Dependent Samples or McNemar Test	SPSS
Assumptions	*one categorical/nominal variable *used to determine if there is relative agreement between observed frequencies and expected frequencies *expected frequencies come from the research scenario	You may have to reconstruct the data (enter the number of scores for each category). Analyze, nonparametric tests, chi-square. Asymp. Sig is the p value for the chi-square. If the p value is greater than our criterion of α =0.05, then we retain the null.	*two nominal variables	You may have to reconstruct the data (enter the number of scores for each category). Analyze, Descriptives, CrossTabs, under the Cells button check Observed, Expected & standardized residuals; under the Stats button check Chi-Square and Phi/Cramer's V. Then run it. If the p value is greater than our criterion of α =0.05, then we retain the null.	Tests for significance of change. The non-parametric equivalent of the paired sample t-test. The frequencies are of a category instead of measurement along a continuous scale.	You may have to reconstruct the data (enter the number of scores for each category). Analyze, Descriptives, CrossTabs, under the Cells button check Observed, Expected & standardized residuals; under the Stats button check Chi-Square and Phi/Cramer's V. Then run it. If the p value is greater than our criterion of α =0.05, then we retain the null.
Hypothesis	Null: There is no difference between the categories. Or the categories fit an expected frequency. (stated in words, not symbols)		Null: There is no relationship between the two variables or there is no difference among the categories.		Null: An equal number of people change their mind from before to after.	
Criteria for rejecting null	$\alpha = 0.05$ df=C-1 (C=number of categories) *If there are more than 30 categories, use formula 21.2:* $z = \sqrt{2}\,\chi^{2} - \sqrt{2df-1}$ $\chi^2_{(cv)}$ Look up in Table C4 *based on Chi-Square distribution*		$\alpha = 0.05$ df = (R-1)(C-1) R = # of rows C = # of columns $\chi^2_{(cv)}$ Look up in Table C4 *based on Chi-Square distribution*		$\alpha = 0.05$ df = (R-1)(C-1) $\chi^2_{(cv)}$ Look up in Table C4 *based on Chi-Square distribution*	
Test Statistic	$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$ *(it should make a long formula for each instance – i.e.* $(1-2)^2/2 + (2-3)^2/3 + \ldots$*)*		$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$ Expected frequencies are computed: $E = \frac{f_r \times f_c}{N}$ *It's ok to get these in SPSS.*		$\chi^2 = (A-D)^2/(A+D)$ A & D are the cells representing changes (i.e. in opinion). It doesn't matter which is A or D.	
Decision	If the observed χ2 is less than the critical value, retain the null.		If the observed χ2 is less than the critical value, retain the null.		If the observed χ2 is less than the critical value, retain the null.	
Effect Size	$ES = \chi^2 / (N)(K-1)$ K= number of categories 0.10 - Small 0.30 - Moderate 0.50 - Larg		Cramer's V is used as an index ES = √[(χ2/N)/((the smaller of R or C) -1)] R = # of rows C = # of columns		Effect size is the difference between the proportions of people who changed their opinion. Proportion is number who changed / N. Calculate for both. Then subtract the smaller from the larger. -------------------------- $R = \frac{O-E}{\sqrt{E}}$	
Standardized residual	$R = \frac{O-E}{\sqrt{E}}$ This determines the statistical significance. If the residual is greater than 2, it is a major contributor to the significant χ^2 value.		If the null is not rejected, there is no need to calculate standardized residuals. $R = \frac{O-E}{\sqrt{E}}$			
Notes	O = Observed frequency E = Expected frequency					