The page isn't wide enough for all the useful columns. Find them on the separate pages here.

Page 1: Notation, dist'n, and conditions to use that dist'n.

Pages 2 and 3: Formulas for CI and HT and how to find the precise sampling dist'n (degrees of freedom)

Page 4: Formulas for the desired sample size. Formulas for the SE, standard error of the test statistic.

Page 5: Degrees of freedom for two-sample t-procedures.

Type of inference	**Parameter (or question for HT)**	**Sample statistic and theoretical sampling dist'n of test statistic**	**Conditions needed to use theoretical sampling dist'n of test statistic**
Inference on one mean	μ(mu): μ population mean	$\bar{X}$ *t* dist'n	Dist'n normal or CLT applies, meaning $n \geq 30$ approximately
Inference on one proportion	p	$\hat{p}$ normal dist'n	$np \geq 10$ AND $n(1-p) \geq 10$ If np is quite samll, we don't have symmetric distribution
Inference on two means	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$ *t* dist'n	In EACH group, Dist'n normal or CLT applies, meaning $n \geq 30$ approximately.
Inference on two proportions	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$ normal dist'n	In EACH group: $np \geq 10$ AND $n(1-p) \geq 10$
Inference on the mean of differences from matched pairs study	μ_d	$\bar{X}_d$ *t* dist'n	Dist'n of differences normal OR CLT applies, meaning $n \geq 30$ approximately, where *n* is number of pairs.
Inference on the correlation coefficient	ρ (called "rho")	*r* *t* dist'n	See regression model conditions. Generally linear pattern (rather than a different pattern,) same variance across x-values, residuals are independent and have normal dist'n.
Inference on the Slope Coefficient in a regression model	β (called "beta")	*b* *t* dist'n	See regression model conditions. Generally linear pattern (rather than a different pattern,) same variance across x-values, residuals are independent and have normal dist'n.
Test of goodness of fit	Do the data fit a particular specified dist'n?	χ_p^2 (chi-squared, *p* degrees of freedom)	Each expected count is at least 5.
Test of association of two categorical variables	Are the two categorical variables associated?	χ_p^2 (chi-squared, *p* degrees of freedom)	Each expected count is at least 5.
Analysis of Variance (ANOVA) for difference of means	Is there a difference in the means of two or more groups?	$F_{p,q}$ (*F* statistic, *p* and *q* degrees of freedom)	In EACH group, Dist'n normal OR CLT applies, meaning $n \geq 30$ or so. Variability is similar in all groups.
Analysis of Variance for regression	Is at least one variable in the model useful in predicting the response variable?	$F_{p,q}$ (*F* statistic, *p* and *q* degrees of freedom)	For EACH explanatory variable, same conditions as linear model for single explanatory variable. Check with residual analysis, including plots. To start, plot the residuals vs the fitted values.

Type of inference	**Sample statistic**	**Confidence Interval formula** SE is "standard error. Generally, use software to obtain this. See last page for formulas.	**Theoretical dist'n of Test statistic**	**Degrees of freedom**
Inference on one mean	$\bar{X}$	$\bar{X} - t^* \cdot SE \le \mu \le \bar{X} + t^* \cdot SE$	*t*-dist'n	*df* = $n-1$
Inference on one proportion	$\hat{p}$	$\hat{p} - z^* \cdot SE \le p \le \hat{p} + z^* \cdot SE$	Normal dist'n	Not relevant
Inference on two means	$\bar{X}_1 - \bar{X}_2$	$\left(\bar{X}_1 - \bar{X}_2\right) - t^* \cdot SE \le \mu_1 - \mu_2$ $\le \left(\bar{X}_1 - \bar{X}_2\right) + t^* \cdot SE$	*t* -dist'n	*df* = Smaller of $n_1 - 1$ and $n_2 - 1$ or Satterthwaite approximation
Inference on two proportions	$\hat{p}_1 - \hat{p}_2$	$\left(\hat{p}_1 - \hat{p}_2\right) - z^* \cdot SE \le p_1 - p_2$ $\le \left(\hat{p}_1 - \hat{p}_2\right) + z^* \cdot SE$	Normal dist'n	Not relevant
Inference on the mean of differences from matched pairs study	$\bar{X}_d$	$\bar{X}_d - t^* \cdot SE \le \mu_d \le \bar{X}_d + t^* \cdot SE$	*t* -dist'n	*df* = $n-1$ where *n* is the number of pairs
Inference on the correlation coefficient	*r*	$r - t^* \cdot SE \le \rho \le r + t^* \cdot SE$	*t* -dist'n	*df* = $n-2$
Inference on the Slope Coefficient in a regression model	*b*	$b - t^* \cdot SE \le \beta \le b + t^* \cdot SE$	*t* -dist'n	*df* = $n-2$ for simple regression
Test of goodness of fit	χ^2_{df}	This investigation is a test. No parameters are estimated with this procedure.	χ^2_p (chi-squared) dist'n	*df* = Number of categories minus 1
Test of association of two categorical variables	χ^2_{df}	This investigation is a test. No parameters are estimated with this procedure.	χ^2_p (chi-squared) dist'n	*r* = number of rows *c* = number of columns *df* = $(r-1)(c-1)$
Analysis of Variance (ANOVA) for difference of means	$F_{p,q}$	This investigation is a test. No parameters are estimated with this procedure.	F dist'n	*k* = number of groups *n* = total sample sizes *df: p = k – 1,* *q= n – k*
Analysis of Variance for regression	$F_{p,q}$	This investigation is a test. No parameters are estimated with this procedure.	F dist'n	*k* = number of explanatory variables *n* = sample size *df: p= k ,* *q= n – k – 1*

Type of inference	**Sample statistic**	**Test statistic** SE is "standard error. Generally, use software to obtain this. See next page for formulas.	**Theoretical dist'n of Test statistic**	**Degrees of freedom**
Inference on one mean	$\bar{X}$	$t = \dfrac{\bar{X} - \mu_0}{SE}$	*t*-dist'n	*df* = $n-1$
Inference on one proportion	$\hat{p}$	$z = \dfrac{\hat{p} - p_0}{SE}$	Normal dist'n	Not relevant
Inference on two means	$\bar{X}_1 - \bar{X}_2$	$t = \dfrac{\left(\bar{X}_1 - \bar{X}_2\right) - 0}{SE}$	*t* -dist'n	*df* = Smaller of $n_1 - 1$ and $n_2 - 1$ or Satterthwaite approximation (page 5 here)
Inference on two proportions	$\hat{p}_1 - \hat{p}_2$	$z = \dfrac{\left(\hat{p}_1 - \hat{p}_2\right) - 0}{SE}$	Normal dist'n	Not relevant
Inference on the mean of differences from matched pairs study	$\bar{X}_d$	$t = \dfrac{\bar{X}_d - \mu_0}{SE}$	*t* -dist'n	*df* = $n-1$ where *n* is the number of pairs
Inference on the correlation coefficient	*r*	$t = \dfrac{r - \rho_0}{SE}$	*t* -dist'n	*df* = $n-2$
Inference on the Slope Coefficient in a regression model	*b*	$t = \dfrac{b - \beta_0}{SE}$	*t* -dist'n	*df* = $n-2$ for simple regression
Test of goodness of fit	χ^2_{df}	$\chi^2 = \sum \dfrac{(Observed\ -\ Expected)^2}{Expected}$	χ^2_p (chi-squared) dist'n	*df* = Number of categories minus 1
Test of association of two categorical variables	χ^2_{df}	$\chi^2 = \sum \dfrac{(Observed\ -\ Expected)^2}{Expected}$	χ^2_p (chi-squared) dist'n	*r* = number of rows *c* = number of columns *df* = $(r-1)(c-1)$
Analysis of Variance (ANOVA) for difference of means	$F_{p,q}$	$F = \dfrac{\text{Mean Square Error Between Groups}}{\text{Mean Square Error Within Groups}}$	F dist'n	*k* = number of groups *n* = total sample sizes *df: p = k – 1,* *q= n – k*
Analysis of Variance for regression	$F_{p,q}$	$F = \dfrac{\text{Mean Square Error Between Groups}}{\text{Mean Square Error Within Groups}}$	F dist'n	*k* = number of explanatory variables *n* = sample size *df: p= k ,* *q= n – k – 1*

Sample size for estimating one proportion	**Sample size for estimating one mean**
$n=\left(\frac{z^*}{ME}\right)^2\cdot\tilde{p}(1-\tilde{p})$, where ME is the chosen margin of error and we use $\tilde{p}=0.5$ or some other value of $\tilde{p}$ if available.	$n=\left(\frac{z^*\cdot\tilde{\sigma}}{ME}\right)^2$, Where ME is the chosen margin of error and $\tilde{\sigma}$ is an estimate of the population standard deviation.

Type of question and type of inference	Standard Error formula for CI	Standard Error formula for HT
One mean	$SE=\frac{s}{\sqrt{n}}$	$SE=\frac{s}{\sqrt{n}}$
One proportion	$SE=\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$SE=\sqrt{\frac{p_0(1-p_0)}{n}}$ Where p_0 is the value in the null hypothesis
Two means	$SE=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	$SE=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
Two proportions	$SE=\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions from the two separate samples	$SE=\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1}+\frac{\bar{p}(1-\bar{p})}{n_2}}$ For testing whether the pop'n proportions are equal. Here $\bar{p}$ is the "pooled" proportion. $\bar{p}=\frac{\text{sum of counts from both samples}}{\text{sum of trials from both samples}}$
Mean of differences from matched pairs data	$SE=\frac{s_d}{\sqrt{n_d}}$ where the subscripts refer to using the differences	$SE=\frac{s_d}{\sqrt{n_d}}$ where the subscripts refer to using the differences
Correlation coefficient	$SE=\sqrt{\frac{1-r^2}{n-2}}$	$SE=\sqrt{\frac{1-r^2}{n-2}}$
Slope coefficient	Obtain with technology	Obtain with technology
Test of goodness of fit	Not relevant	Not relevant
Test of association of two categorical variables	Not relevant	Not relevant
Analysis of Variance (ANOVA) for difference of means	Not relevant	Not relevant
Analysis of Variance for regression	Not relevant	Not relevant

Standard Error (degrees of freedom) for two-sample t-procedures:

In two-sample t procedures, in order to show that the test statistic has an exact t-dist'n, we must have that the two population variances are equal. In that case, $df = n_1 + n_2 - 2$

If it is not appropriate to assume that the two population variances are equal, then a "conservative" approach (does not overstate our confidence in our answers) is to use the smaller of $n_1 - 1$ and $n_2 - 1$.

An adjustment can be made to the degrees of freedom to take into account how different the variances are and how different the sample sizes are.

Most statistical software will use this Satterthwaite approximation as the degrees of freedom for two-sample t procedures. (It is derived by a modification of the method of moments method of estimation.)

$$\mathrm{df} = \frac{\left(\dfrac{s_1^{\,2}}{n_1} + \dfrac{s_2^{\,2}}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^{\,2}}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^{\,2}}{n_2}\right)^2}.$$

Don't do this "by hand."

It is included here because, as you use software, you will see degrees of freedom that do not fit the "simple" method given in this handout and in many applied statistics texts.