Formula Sheet

Sample Statistics:

Sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$

Linear Transformations/Combinations:

Mean: $E(aX+b) = aE(X)+b, E\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_iE(X_i)$

Variance: $V(aX+b) = a^2V(X), V\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jcov(X_i, X_j)$

Two-Sample Confidence Intervals

Two sided C.I. (means):

$\bar{x} - \bar{y} \pm t_{1-\alpha/2,v}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$ where $v = \frac{\left(\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}\right)^2}{\frac{(s_x^2/n_x)^2}{n_x-1}+\frac{(s_y^2/n_y)^2}{n_y-1}}$.

Test statistic: $T.S. = \frac{\bar{x}-\bar{y}-\Delta_0}{\sqrt{s_x^2/n_x+s_y^2/n_y}} \overset{H_0}{\sim} t_v$.

If we assume $\sigma_x^2 = \sigma_y^2$, replace s_x^2 and s_y^2 with $s_p^2 = \left(\frac{n_x-1}{n_x+n_y-2}\right)s_x^2 + \left(\frac{n_y-1}{n_x+n_y-2}\right)s_y^2$ and $v = n_x + n_y - 2$.

Two sided C.I. (proportions):

$\widetilde{p}_x - \widetilde{p}_y \pm z_{1-\alpha/2}\sqrt{\frac{\widetilde{p}_x(1-\widetilde{p}_x)}{\widetilde{n}_x} + \frac{\widetilde{p}_y(1-\widetilde{p}_y)}{\widetilde{n}_y}}$ where $\widetilde{n}_x = n_x + 2$ and $\widetilde{p}_x = \frac{x+1}{\widetilde{n}_x}$.

Two sided C.I. for ratio of variances $\frac{\sigma_x^2}{\sigma_y^2}$: $\left(\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{1-\alpha/2,n_x-1,n_y-1}}, \frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{\alpha/2,n_x-1,n_y-1}}\right)$

Test statistic: $T.S. = \frac{\frac{s_x^2}{s_y^2}}{\Delta_0} \overset{H_0}{\sim} F_{n_x-1,n_y-1}$.

Wilcoxon signed-rank test:
1. calculating the differences $d_i = (x_i - y_i) - \Delta_0$
2. proceed from step 2 in the Wilcoxon signed-rank one population setting.

Contingency tables:
Expected value: $E_{ij} = \frac{n_{i+}n_{+j}}{n}$
Test statistic: $T.S. = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-E_{ij})^2}{E_{ij}} \overset{H_0}{\sim} \mathcal{X}^2_{(r-1)(c-1)}$

Regression:

Model: $Y_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_p x_{p,i} + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
Estimation (for simple regression only):
$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{(\sum_{i=1}^{n} x_i y_i)-n\bar{x}\bar{y}}{(\sum_{i=1}^{n} x_i^2)-n\bar{x}^2} = r\frac{s_y}{s_x}$
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$.
Sum of Squares: note $SST = SSE + SSR$.
$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$, with $n-1$ degrees of freedom
$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, with $n-p-1$ degrees of freedom
$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, with p degrees of freedom
Mean Squares: $MS = SS/df$.
Conditional variance: $s^2 = MSE$.

Model evaluation tools:
$R^2 = SSR/SST = 1 - SSE/SST$
$R^2_{adj} = R^2 + (1-R^2)\frac{p}{n-p-1} = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$
$AIC = 2(p+1) + n \cdot log\left(\frac{SSE}{n}\right)$

Inference in Regression

Individual β's:
C.I.: $\hat{\beta}_j \pm t_{1-\alpha/2,n-p-1}s_{\beta_j}$
Test: $T.S. = \frac{\hat{\beta}_j-\beta_{j0}}{s_{\beta_j}} \overset{H_0}{\sim} t_{n-p-1}$ where, for simple regression only, $s_{\beta_1} = s/\sqrt{\sum_{j=1}^{n}(x_j-\bar{x})^2}$.
Otherwise, read it from the standard error output from R.
C.I. on mean response (for simple regression only): $\hat{y} \pm t_{1-\alpha/2,n-2}\left(s\sqrt{\frac{1}{n} + \frac{(x_{obs}-\bar{x})^2}{\sum_{j=1}^{n}(x_j-\bar{x})^2}}\right)$
Prediction Interval on response (for simple regression only): $\hat{y} \pm t_{1-\alpha/2,n-2}\left(s\sqrt{1 + \frac{1}{n} + \frac{(x_{obs}-\bar{x})^2}{\sum_{j=1}^{n}(x_j-\bar{x})^2}}\right)$

Nested F test:

$$T.S. = \frac{\frac{SSE_{red}-SSE_{full}}{dfE_{red}-dfE_{full}}}{\frac{SSE_{full}}{dfEfull}} \overset{H_0}{\sim} F_{dfE_{red}-dfE_{full},dfE_{full}}$$