Name of Test	Person Responsible, Lecture Notes #	Situation to use in	Assumptions and limitations	Formula to do by hand	R code	SAS code	Examples (ones we've seen before in class, or if you find a relevant one)
Assessing normality (graphically)	Audrey, 3 (pg 28)	To see if normal distribution is a reasonable fit for data - will draw a QQ plot. A straight line means normality assumption is reasonable			qqnorm(galton)	proc univariate; qqplot galton;	
One Sample T-test	Audrey, 6 (pg 4)	Test null hypothesis on mean or variance of sample. Can test difference in paired samples (like twins... or before/after measurements)	-Observations are independent. -Small sample, and sample is from normal distribution	pg 4	t.test(x)	proc ttest; var x;	
One Sample Z-test	Audrey, 6 (pg 11)	Same as One Sample T-test (but on large sample)	-Observations are independent. -Large sample, and sample does not need to be normally distributed (because of CLT)	pg 14			
Sign Test (small sample)	Audrey, 6 (pg 19)	Same as One Sample T-test (but on small, non-normal (ie binomial) sample) -Test for consistent differences between pairs of samples	-Observations are independent -Small sample -Non-normal	pg 19 - 25	pg 26	proc univariate; var x;	
Sign Test (large sample)	Audrey, 6 (pg 30)	Same as One Sample T-test (but on large, non-normal (ie binomial) sample) -Test for consistent differences between pairs of samples	-Large sample (n>=40).. pg 31 -May need to do Binomial continuity correction (pg 34)				
Wilcoxon Signed Rank Test	Audrey, 6 (pg 38)	Similar to One Sample T-test (but on paired non-normal samples)	-Each pair is independent -Data is paired and come from same population	pg 38-55 (hopefully will not need to do this by hand)	wilcox.test(x)		
Two Sample T-test	Haley, 7 (pg 5-11)	Comparing sub-samples from a single cross-sectional sample Comparing samples from two different populations Comparing subjects randomly allocated to different interventions from a single sample ***note: helpful summary table for two-sample test statistics on pg. 29 of this set of slides	-The two groups are independent random samples from two normal distributions with equal variances	pg 5, example on pg 9	pg 10	proc ttest; class y; var x; (pg 11)	pg 9-11
Homogeneity of Variance	Haley, 7 (pg 12-19)	Testing whether the variances of two random variables are equivalent	-The two groups are independent samples from normal distributions	pg 15, example on pg 17	pg 19	same as two-sample ttest (pg 18 shows relevant output)	pg 17-19 (continuation of birthweight example introduced on page 6)
Large Sample Approximation	Haley, 7 (pg 22-24)	Testing whether the means of two groups are equivalent, when the variances are unknown and not equivalent	-Large independent samples -Homogeneity of variance and normality assumptions no longer needed	pg 23, example on pg 24			pg 24
Welch-Satterthwaite Approximation	Haley, 7 (pg 25-28)	Same as large sample approximation (different assumptions)	-Normality -Small independent samples -Unequal variances	pg 25, example on pg 27	pg 28		pg 27-28
Wilcoxon (Mann-Whitney) Rank Sum Test	Haley, 7 (pg 31-46)	Testing whether the samples come from the same distribution (non-parametric alternative to the two-sample t-test) ***Default non-parametric test (pg 72)	-Independent samples	pg 31 and 33, example on pg 34	pg 46 and pg 58	proc npar1way wilcoxon correct=no; class y; var x; (pg. 45)	pg 34-38 pg 41-46
Wilcoxon Rank Sum Exact Test	Haley, 7 (pg 47-50)	Same as above	-Independent samples	similar to above, but add pg 47, example on pg 48	pg 49	proc npar1way wilcoxon; class y; var x; (pg 50)	pg 47-50
Mann-Whitney Test	Haley, 7 (pg 51-53)	Equivalent to Wilcoxon	-Independent samples	pg 51, example on pg 52			pg 52-23

Name of Test	Person Responsible, Lecture Notes #	Situation to use in	Assumptions and limitations	Formula to do by hand	R code	SAS code	Examples (ones we've seen before in class, or if you find a relevant one)
Permutation Test	Haley, 7 (pg 60-65)	Testing whether the samples come from populations with the same distribution ***Asymptotically equivalent to t-test, thus the most powerful asymptotically under normality, but computationally intensive and sensitive to outliers (pg 72)	-Random assignment to groups	pg 60, example on pg 61			pg 61-64
Kolmogorov-Smirnov Test	Haley, 7 (pg 66-71)	Testing whether the samples come from populations with the same distribution) ***Employ if trying to detect difference in distribution other than location shift (pg 72)	-Independent samples	pg 66-67, example on pg 67-69	pg 70	proc npar1way; class y; var x; exact ks; (pg 70-71)	pg 67-70
Poisson Confidence Intervals	Joyce, 11 (pg 11-15)	Finding either an exact or approximate confidence interval for λ	-For normal approximation, λ either >= 100 or >= 30 (see page 12)	pg 11-15	exact: poisson.test(x) where x is observed count		
Homogeneity of λ Test (Poisson)	Joyce, 11 (pg 16-22)	H0: $X_1,...,X_n \sim$ Poisson(λ) Observed counts may have larger variance than expected under Poisson model if assumption of homogeneity of the λs is not satisfied -> Test homogeneity of λs	-$X_1,...,X_n$ are Poisson random variables	pg 17-18, chi squared goodness of fit test	chisq.test(x) where x is vector of observed counts		
Linear Regression Confidence Intervals and Hypothesis Tests	Joyce, 12 (pg 23-35)	Get a confidence interval for β ($Y = \alpha + \beta X + \varepsilon$)	Assumptions for Linear Regression (HILE Gauss): -linearity: $Y_i = \alpha + \beta X_i + \varepsilon_i$ -Xs are fixed constants -homogeneity of variance: ε_i are iid ~N(0,σ^2) pg 25, if σ^2 is known, use s^2 and t distribution	pg 24-25,27	pg 28	pg 29	pg 23-35
Linear Regression Prediction Intervals	Joyce, 12 (pg 36-43)	Get a prediction interval for a new or future observation (Yx), given X = x (Yx_hat = α_hat + β_hat*x)	ε is normally distributed, future Yx for predicition is independent from the sample	pg 39	pg 42	pg 43	pg 36-43
Linear Regression - 2 Sample t-test	Joyce, 13 (pg 11-20)	X = 1 if in group 1 0 if in group 2 test if mean outcome is equal across groups, or test if β = 0 (β_hat = Y1_bar - Y2_bar) advantage of regression: t-tets can compare 2 groups, but cannot adjust for confounders	same assumptions for linear regression as above	pg 12-16		proc reg (pg 20)	pg 17-20
Linear Regression Diagnostics	Joyce, 13 (pg 21-40)	test assumptions and validity of linear regression model				pg 26, 28, 34	pg 23-40
Multiple Linear Regression	Joyce, 14, (pg 3-25)	$Y = \beta X + \varepsilon$ where Y, β, and ε are vectors and X is a matrix Reasons to use multiple linear regression: -determine best set of covariates to predict outcome -adjust for confounders -investigate interaction effect -use categorical predictor with more than 2 categories	$Y = \beta X + \varepsilon$ where Y, β, and ε are vectors and X is a matrix Assumptions: -linearity: each X variable is linearly associated with Y -the values of each X variable are fixed constants -ε_i iid ~ N(0, σ^2)	pg 7-8	pg 15	proc reg (pg 10, 13, 17-18, 20, 22) proc glm (pg 23)	

Name of Test	**Person Responsible, Lecture Notes #**	**Situation to use in**	**Assumptions and limitations**	**Formula to do by hand**	**R code**	**SAS code**	**Examples (ones we've seen before in class, or if you find a relevant one)**
Correlation Test (and Fisher's Transformation)	Joyce, 14, (pg 26-80)	H0: ρ = 0 vs. H1: ρ not = 0 or H0: ρ1 = ρ2 = ... = ρk vs. H: at least one inequality for k independent samples (Fisher's)		pg 27, 31, 33-34	cor.test() pg 43 (Fisher's)	proc corr (pg 36, 43)	throughout pg 26-80
ANOVA	Joyce, 15 (all pages)	Test hypotheses about mean of more than 2 groups Categorical predictor variables H0: μ1 = μ2 = ... = μk (means for k groups are the same) H1: at least one inequality	assume Yij ~ N(μi, σ^2) GLM assumptions (HILE Gauss)	pg 6-8	pg 11, 27	pg 11, 24-26	pg 9-11, 22-27
Log-Rank Test	Haley, 24 (pg 26-38)	Test whether two survival functions are different, in the presence of right censoring (without censoring, you would use a rank test - e.g. Wilcoxon rank sum test)		pg 29-30, example on pg 34 (data given on 31)	pg 35	pg 36-38 (two different ways to do this in SAS)	pg 31-38