**Lab 1 notes - Introduction to JMP**

 **JMP documentation comes with installation**

 Menu Card, Using JMP, Basic Analyses

 **JMP online tutorial videos –** search “JMP tutorial”

 **Importing EXCEL files**

Variable names in first row. One row per observation. One worksheet

 **Designating variable type**

Continuous, ordinal or nominal

Designating the variable type helps determine what analysis will be done.

Variables are X = predictors or Y = outcomes

 **Making new variables out of existing variables**

 *Cols -> formula*

Example: BMI = 703 Weight (lbs) / Height2 (in)

 **Generating random variables**

Open new empty JMP data table (skip if using existing table)

 *File -> New -> Data Table*

Select the one column, then Rows -> Add Rows and type in number of rows to generate (for example, 100). Keep this column active. (skip if using existing table).

*Cols -> Formula -> Random -> Random Normal*(*mean*) will fill in normal random numbers with specified mean in the empty rows

 **Univariate analysis** – “**Distribution” platform**

 Set preferences (can set for any platform)

*File -> Preferences -> Platforms -> Distribution*

 Usually want to have a horizontal layout

 One way analyses: *Analyze -> Distribution*

 Can use “by” option to make strata. Carry out same analysis in each stratum

 **Normal quantile plots** (distribution platform)

Is continuous data distribution normal? Sample size n.

 Data: X1, X2, X3, … Xi … Xn

Can sort from smallest to largest.

Can compute Zi = (Xi – mean)/SD - empirical

Can get Zi as the inverse of the quantile (ie, NORM.INV) assuming normal – theory.

Quantiles = 1/n, 2/n, 3/n, … i/n, … n/n=1

If data is normal, both ways of getting Zi should be the same!

Plot of empirical Zi vs theory Zi should give a straight line.

 **Continuous data normality test** (distribution platform)

Shapiro-Wilk “W” stat is a formal test for normality. Data has a normal distribution if **W is near 1.0.** The W statistic p value tests the null hypothesis that the data is normal so the p value should be large if the data has a normal distribution.

*Analyze -> Distribution* (choose variables) *-> Normal Quantile Plot*

*Analyze -> Distribution ->* (choose variables) -> *Continuous Fit -> Fit Normal -> Goodness-of-fit*

 **Bivariate analysis** – **4 platforms**

 Analyze -> Fit Y by X



If X and Y are both continuous, get scatter plot, “bivariate” platform

If X is ordinal or nominal and Y is continuous, get “oneway” platform, means & stats by group

If X and Y are both ordinal or nominal, get “contingency” platform - frequency cross tables.

If X is continuous and Y is ordinal or nominal, get “logistic” platform. **The ROC analysis is in this platform.**