DAY 3 ACTIVITIES

Activity 1 (Slide 13):

The following atrazine concentrations were measured before and after application (June, September) at 24 shallow groundwater sites. Note that <0.01 means that the concentration was below the detection limit of 0.01. Conduct a sign test to see if the median difference is 0, or if concentrations are higher after application. (An empty column has been provided for you to enter whether the difference is positive or negative.) You will need to use *Distributions → Discrete → Binomial → Tail probs*

June atrazine	Sept atrazine	Diff. pos or neg?	June atrazine	Sept atrazine	Diff. pos or neg?
0.38	2.66		<0.01	0.02	
0.04	0.63		<0.01	<0.01	
<0.01	0.59		<0.01	0.5	
0.03	0.05		<0.01	0.03	
0.03	0.84		0.02	0.09	
0.05	0.58		0.03	0.06	
0.02	0.02		0.02	0.03	
<0.01	0.01		0.02	0.01	
<0.01	<0.01		0.05	0.1	
<0.01	<0.01		0.03	0.25	
0.11	0.09		0.05	0.03	
0.09	0.31		<0.01	88.36	

S+ = _____, n = ____

Activity 2 (Slide 19):

1. Generate 2 random samples of n = 15 from a lognormal distribution with $\mu = 0$ and $\sigma = 0.5$:
Distributions → Continuous distributions → Lognormal distribution →Sample from lognormal
Enter mean = 0, std dev = 0.5, (rows) = 15, (columns) = 2, and uncheck the "sample mean" box. (The box is set up to put the samples into rows, but we are putting samples in columns.)
2. Go to "View data set" to make sure you have 2 columns of 15 numbers each.
3. Make a new variable by adding 0.5 to each of the observations in the 2nd sample:
Data → Manage variables... → Compute new variable, Call it "shift"; expression is obs2+.5
4. Create a histogram for each sample to see skewed shape. (*Graphs → Histogram*). Don't save them.
5. Stack the two variables called "obs1" and "shift" into a single column:
Data → Active data set → Stack variables in active data set [Choose obs1 and shift, no other changes.]
6. Create a boxplot of the stacked data, by groups (*Graphs →Boxplot, plot by groups;* use defaults shown.) You should see the "obs1" boxplot at least slightly lower than the "shift" boxplot.
7. Do an independent samples t-test to compare the two variables (*Statistics →Means →Indepen...*)
8. Do a Wilcoxon test to compare the two variables (*Statistics →Nonpar → Two-sample Wilcoxon*)
9. Compare the results from Steps 7 and 8. Which had a smaller *p*-value? Which is more appropriate?

Activity 3 (Slide 17, Afternoon):

Go to the following website: http://onlinestatbook.com/stat_sim/reg_by_eye/index.html
Try to figure out where to draw the line that minimizes MSE (and maximizes *r*). Play with it several times with new data.

Activity 4 (Slide 31, Afternoon):

Go to the following website: http://illuminations.nctm.org/LessonDetail.aspx?ID=L455

Add a clump of points in the lower left, then notice what happens when you add an outlier in the upper right. Continue to play with adding outliers in different locations, to see how they affect the regression line and other values.