JMP assignment 2

The “PSA.xls” dataset contains prostate specific antigen (PSA - ng/ml) values on normals and those with enlarged prostate disease (“disease”).

1. Determine if the data in each group follows a Gaussian distribution better on the original scale or a log scale. When considering the log scale, be sure to indicate whether you are using log base e or log base 10.

2. Using the best scale, report summary statistics including the mean, SD and range for each group. Is there overlap between the two distributions?

3. Consider a threshold that is halfway between the two means. Using this threshold, compute the empirical sensitivity, specificity and accuracy. Hint: Create a variable that is equal to ‘1” if an individual is above the threshold and is equal to “0” otherwise. Then make a cross tabulation.

As an alternative way to find the threshold, carry out an ROC analysis (under “Fit Y by X” where Y=group) and find the threshold that maximizes the sum of sensitivity and specificity. This alternative is optional.

4. Compute the Gaussian theoretical sensitivity and specificity using the threshold and the means and SDs reported above. Compare to the results from #3 above. Hint: Compute the Z scores and use the JMP function “NormalDistribution(z)” or the EXCEL function “=normsdist()”.

JMP hints:
Make new variable column
   Cols -> New Column

Take log of old variable
   Cols -> formula   -> transcendental -> log

Analyse -> Distribution
   Choose “normal quantile plot” under red triangle

If – else logic
   Cols -> formula -> conditional -> if
Need comparison