SEX DISCRIMINATION PROBLEM

7. Checking the Assumptions Underlying Inferential Methods

7.1 What the methods of inference should be used?

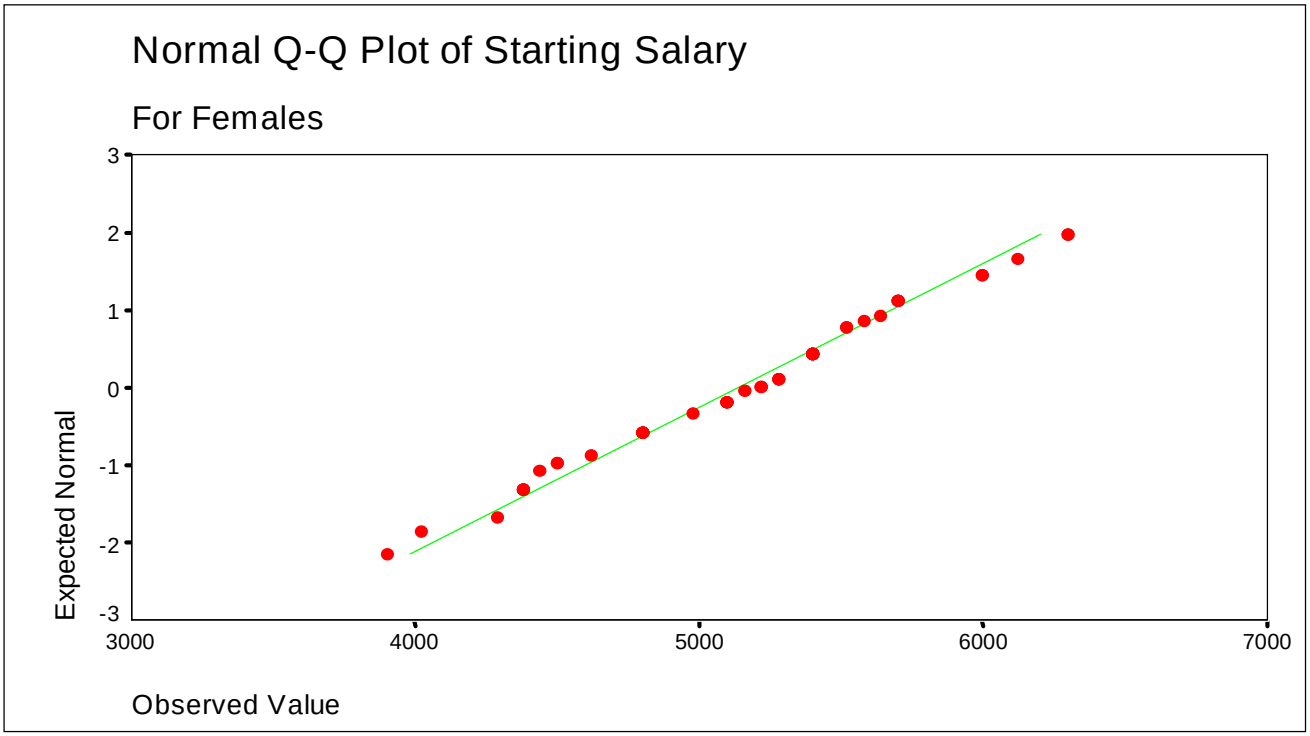
With the meaning of the populations of interest specified in Section 3.2, we can consider the males and females in our case study as two random samples from the corresponding two populations. As the population standard deviations are unknown, two-sample t-test should be used to determine whether there are significant differences between the population means. The Independent-Samples T Test procedure available in SPSS compares the means of one variable (beginning salary) for two groups of cases (males, females). Ideally, for this test, the subjects should be randomly assigned to two groups, so that any difference in response is due to treatment and not to other factors. This is not the case if you compare average starting salary for males and females. A person is not randomly assigned to be a male or female.

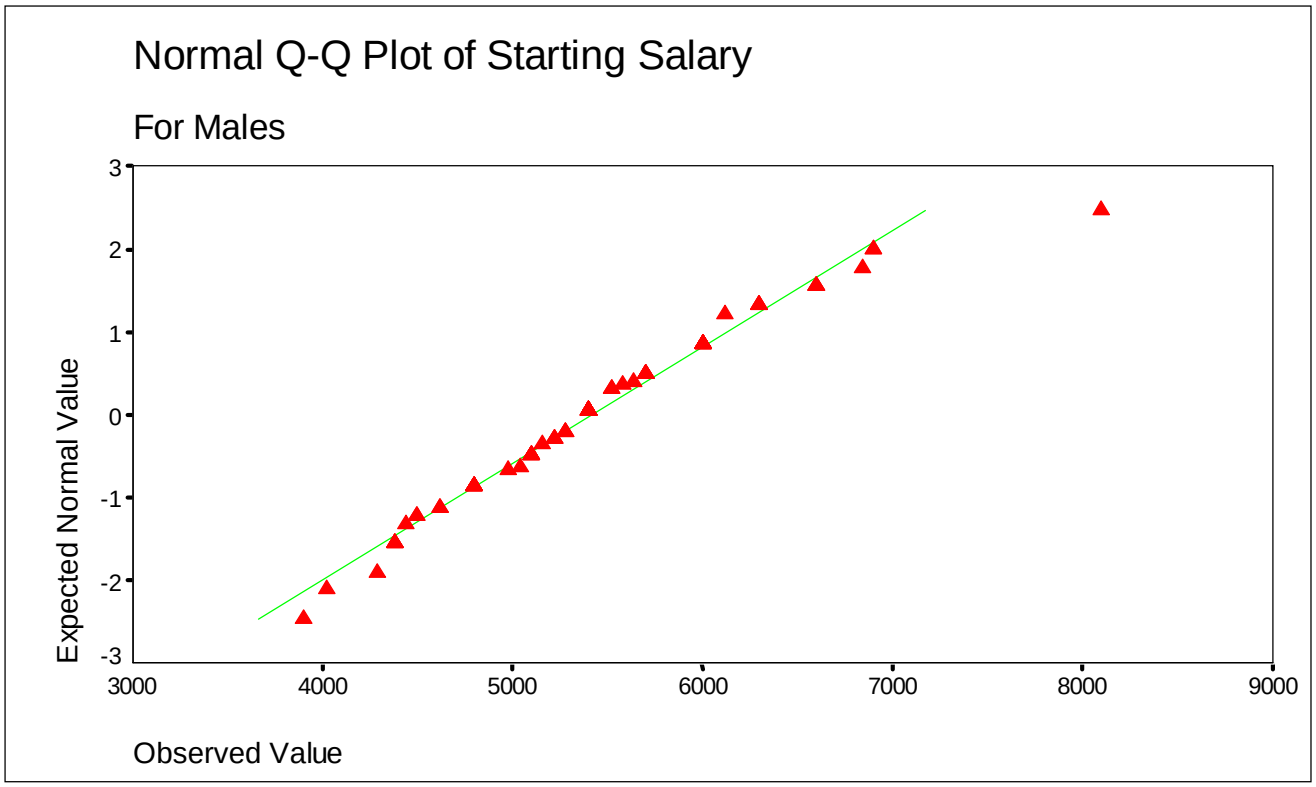
7.2 What assumptions should be satisfied and how to verify them?

The assumptions of the Independent-Samples T Test procedure are that both samples are random samples from their respective populations, the two samples are independent of one another, and both populations are normal.

In order to determine whether or not a variable is normally distributed, you can use one of the two available procedures in SPSS: *Normal Q-Q* Plot or *Normal P-P* Plot. The *Normal Q-Q* plot plots the quantiles of a variable's distribution against the quantiles of the normal distribution. If the data come from a normal distribution, the plot should resemble a straight line.

The SPSS Instructions to obtain the normal probability plot for each gender are given in Section 13.





Both plots do not indicate any significant departure from a straight line. Thus, there is no reason to suspect that the assumption of normality is violated.

The *Normal P-P* plot plots the cumulative proportions of a variable's distribution against the cumulative proportions of the normal distribution. Similarly, if the sample is from a normal distribution, points will cluster around a straight line.