Quantitative Methods II
Assignment 1

August 28, 2011

Solutions

Problem: High Students [[10]]

Samples of size ten ($n = 10$) were taken from the population of males and the population of females who took a Statistics course in the summer of 2010. Heights of the samples were measured to determine if males were no shorter than females. If we specify μ_M as the average height for males who took Statistics in the summer of 2010 and μ_F the average height of females who took Statistics in the summer of 2010, then our null and alternative hypotheses are

$$H_0 : \mu_m \geq \mu_f$$

$$H_A : \mu_m < \mu_f$$

We are supplied with the information that the two sub-populations have heights that *are* Normally distributed. Figure 1 supplies the boxplots of the data for each sample. These boxplots can be used to determine if the Normality assumption is reasonable. These boxplots do not indicate outliers, nor do they indicate that the data is severely skewed. From this, it appears reasonable to assume the measurements are Normally distributed in each group.

With that, we can now use the t-test to test the null hypothesis that the average height for males is not less than the average height for females in Statistics classes in the summer of 2010. Let us note that the boxplots in Figure 1 suggest that the variance of the heights in the femal population is larger than the variance in the heights of the male population.

As such, we will need to use Welch's Two-Sample t-test, as it does not require the two groups have equal variances.

The results of the t-test indicates that the null hypothesis is supported by the data. Thus, we can reasonably conclude at the $\alpha = 0.05$ level that males who took Statistics in the summer of 2010 do not have smaller heights than similar females ($t = -1.510; \nu = 16; p = 0.9247$).

⋆⋆⋆

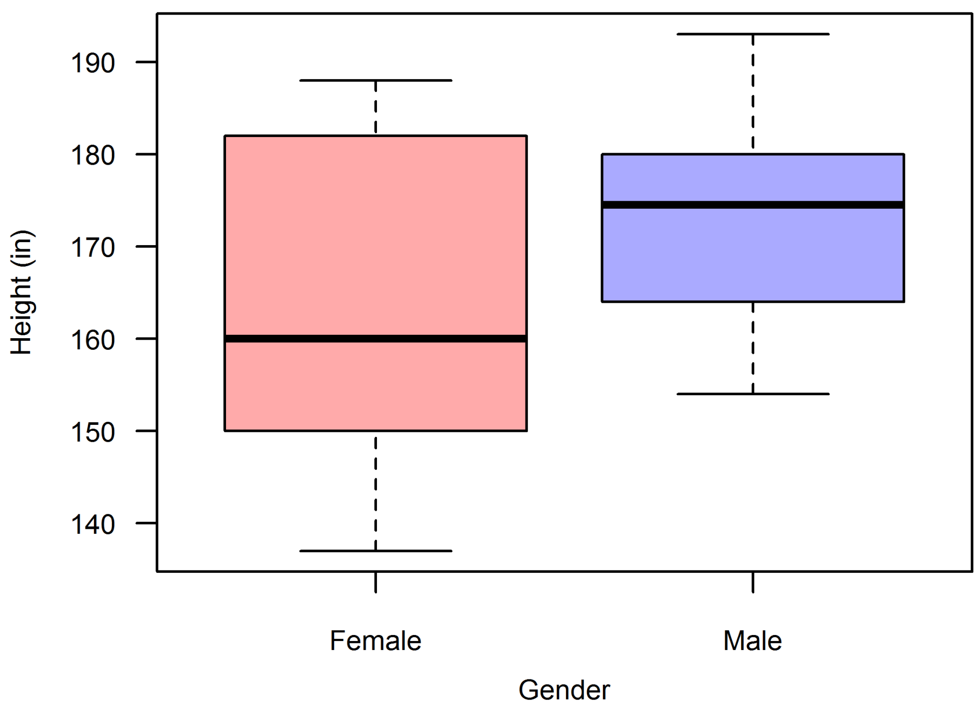

Figure 1. *A boxplot of the* `studentHeight` *data file. Note that the assumption of Normality seems reasonable for this data. Remember that each group has size* $n = 10$, *as such, deviations from Normality are difficult to detect.*

Problem: Higher Students? ⟦15⟧

Samples of size ten ($n = 10$) were taken from the population of males and the population of females who took a Statistics course in the summer of 2010. Heights of the samples were measured to determine if males were no shorter than females. If we specify μ_M as the average height for males who took Statistics in the summer of 2010 and μ_F the average height of females who took Statistics in the summer of 2010, then our null and alternative hypotheses are

$$H_0 : \mu_m \geq \mu_f$$

$$H_A : \mu_m < \mu_f$$

The previous analysis explicitly made the assumption of Normality. This analysis does not make that assumption. As it has a higher power than non-parametric tests, we would like to use the t-test to test the null hypothesis. The primary assumption of the t-test is that the measurements in each group are Normally distributed. To determine if this assumption is reasonable, we can use graphical methods or numerical methods. The two graphical methods we will employ are the boxplot and the histogram; the numerical method, the Shapiro-Wilk test of Normality.

As previously discussed, the boxplots in Figure 1 suggest that such an assumption is not unreasonable. Figure 2 provides histograms for the two samples. If the data comes from a Normal population, we would expect to see histograms that are symmetric and 'mound-shaped.' Neither histogram gives us this picture — although the male group is closer to our expectations. Thus, the assumption of Normality is called into question here.

We must note, however, that the sample size is $n = 10$ for each group. Small sample sizes have great variability and even Normally distributed samples may not appear as such. Thus, while the Normality assumption is called into question, we cannot conclude definitely that it is unreasonable.

The final method we can use to determine the reasonableness of the Normality assumption is a numerical test, such as the Shapiro-Wilk test. Performing this test suggests that the assumption of Normality is not unreasonable for either the males ($W = 0.974; p = 0.923$) or the females ($W = 0.929; p = 0.437$).

As neither graphical test conclusively discounted the assumption of Normality and as the numeric test concurred, we can conclude that the assumption of Normality for this data is reasonable. As such, we can use Welch's Two-Sample t-test to test the null hypothesis.

The results of the test indicate that we cannot reject the null hypothesis at the $\alpha = 0.05$ level. Thus, we can conclude that it is reasonable that males who took Statistics in the

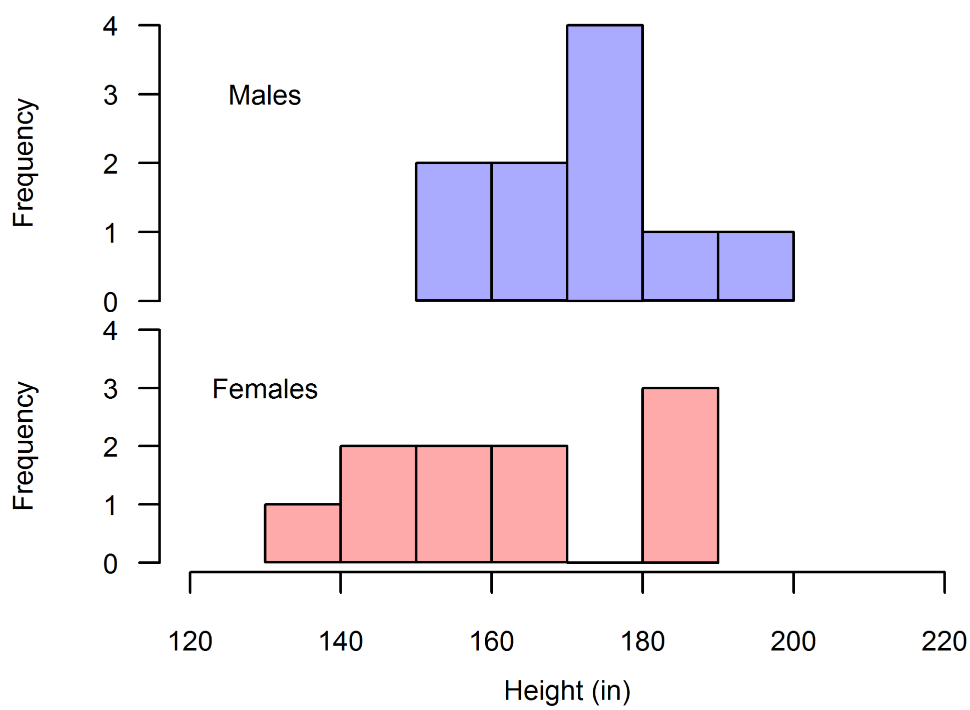

Figure 2. *Two histograms of the* `studentHeight` *data file — one for each group. Note that the assumption of Normality does not seem too unreasonable for this data. Remember that each group has size $n = 10$, as such, deviations from Normality are difficult to detect.*

summer of 2010 are not shorter on average than females who took Statistics in the summer of 2010 ($t = -1.510; \nu = 16; p = 0.925$).