An Analysis of the Number of Hours of Exercise (Per Week) for Male and Female Statistics Students at Longwood University

By Virginia Tam

I. Introduction

In this study, the task was to determine if there is a difference between the mean number of minutes exercised per week between all male and female statistics students at Longwood University. Drawing conclusions from the data, we have reason to believe that there is in fact a difference between the mean number of minutes exercised per week between all male and female statistics students, and that difference shows that male students exercise more than female students.

II. Data Collection

The variables studied were *Sex* and *Number of Hours of Exercise (Per Week)*. The variable Sex was our explanatory variable. As a categorical variable, the two determined levels of *Sex* were Male and Female. The population in this study is all the female and all the male students taking Math 171 and Math 301 at Longwood University. The quantitative response variable in this study was *Number of Hours of Exercise (Per Week)*.

The data set that was used for this report was comprised of data collected via survey from all sections of Dr. Lunsford's Math 171 and Math 301 classes. Given the nature of this methodology, it is reasonable to consider these data as representative samples of the populations and this study can be considered an observational study. Although this cannot technically be considered a simple random sample (SRS), we will consider it representative enough of the respective populations, and so it will be treated as an SRS.

III. Data Description

The distribution of the variable sex is such that 41 have a score of Female (Female=41) and 30 have a score of Male (Male= 30). With these values in mind, that means that 57.746% of the data set is female, and 42.254% of the data set is male.

The distribution of the variable *Number Exercise Hours per Week* for females is very right skewed. As depicted in the graph provided in Figure 1, we have a concentration of values among the fewest amount of hours with a rather long tail into the higher amounts of hours, with many gaps in between. The majority of females enrolled in Math 171 or Math 301 exercise less hours in the week, while very few females enrolled in Math 171 or Math 301 exercise more frequently. The data seems to be centered around 4-5 hours ($M = 4, \bar{x}_F = 5.341$). Note that this implies that roughly half of the females in the distribution exercise less than 4-5 hours a week. Given that s = 5.189, we can see that the majority of the data is very close to the value of the mean. Judging by the respective graph provided in Figure 3, this means the skew of the distribution is very heavy. The data range from about 0-27, so the approximate range of the data equals 27 – 0 = 27. There are two probable outliers (visible in Figure 3), with one at 27 and one at 20.

The distribution of the variable *Number of Exercise Hours per Week* for males is non-symmetrical and bimodal. As depicted in the graph provided in Figure 2, we have two concentrations of values, with a cluster of values among the fewest amount of hours and another cluster near the third quartile of the data ($Q_3 = 13$). The majority of males enrolled in Math 171 or Math 301 exercise between 3-10 hours a week, and also 17-20 hours a week. The data seems to be centered around 8-10 hours a week ($M = 8$, $\bar{x}_m = 10.867$). Note that this implies that roughly half of the males in the distribution exercise less than 8-10 hours a week. Given that $s = 7.427$, the data is spread out from the mean. Judging by the graph provided in Figure 3, the data do not seem as heavily skewed as the sample distribution for *Number Exercise Hours per Week* for females. The data range from about 1 to 25, so the approximate range of the data equals 25 – 1 = 24. There do not seem to be probable outliers in the data set.

IV. Data Analysis

In this study, we tested the following hypotheses

$$H_0: \mu_F - \mu_M = 0$$

$$H_a: \mu_F - \mu_M \neq 0$$

where μ_F was representative of the mean *Number of Exercise Hours per Week* for all female statistics students at Longwood University, and μ_M represented the mean *Number of Exercise Hours per Week* for all male statistics students at Longwood University. As depicted in Figure 3, the sample distributions of the sample means seem to be heavily right skewed, and in Figure 2 the distribution for μ_M seems to be bimodal, both indicating that the sample populations are not normally distributed. Despite this, the descriptive statistics outlined in Table 1 states that the sample sizes are likely large enough to be considered normal as independent populations ($n_F > 30$, $n_M = 30$), and as previously stated we are treating these samples as SRS's, so we technically meet the conditions to conduct the two-sample t-test. Regardless, although the distribution for μ_M only presents light skewed-ness in its boxplot (Figure 3), the histogram shows patterns of bimodal behavior which displays that μ_M may not be coming from a normal population. Likewise, the heavy skewed-ness of μ_F is so distinct (IQR = 5) that it seems unlikely for those data to have come from a normally distributed population.

Given the output in Table 2, we can see that test is significant as we can observe that p (0.00119) < 0.05 and $t(48.90) = -3.441194$. Knowing this, we have evidence that there is a difference in the mean number of minutes exercised per week between all male and female statistics students at Longwood University. We are 95% confident that $-8.701 < \mu_F - \mu_M < -2.351$. Given that $\bar{x}_M < \bar{x}_F$, this represents the mean amount of hours that male students at Longwood University spend exercising per week that is greater than that of the female students at Longwood University. Given this confidence interval we also see that our hypotheses test is supported, seeing as both the calculated difference between the means falls within the confidence interval (-8.701 < -5.436 < -2.351) and the t statistic also falls within this confidence interval (-8.701 < $t(48.90)$ = -3.441194 < -2.351).

V. Appendix

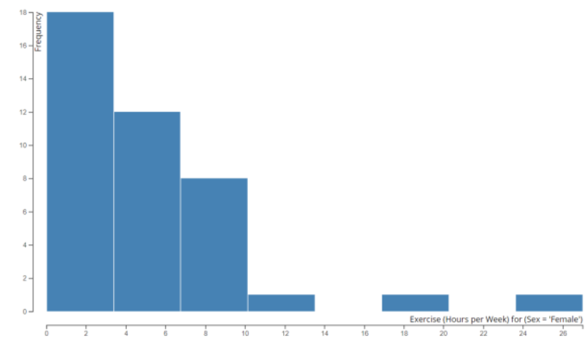

Figure 1. Histogram of Number of Hours Exercised per Week for Females

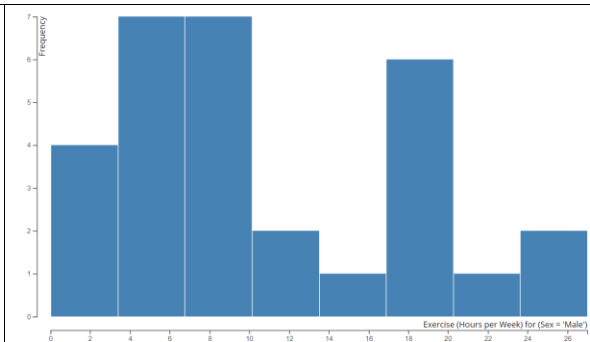

Figure 2. Histogram of Number of Hours Exercised per Week for Males

Figure 3. Box Plot Hours Exercised per Week by Sex

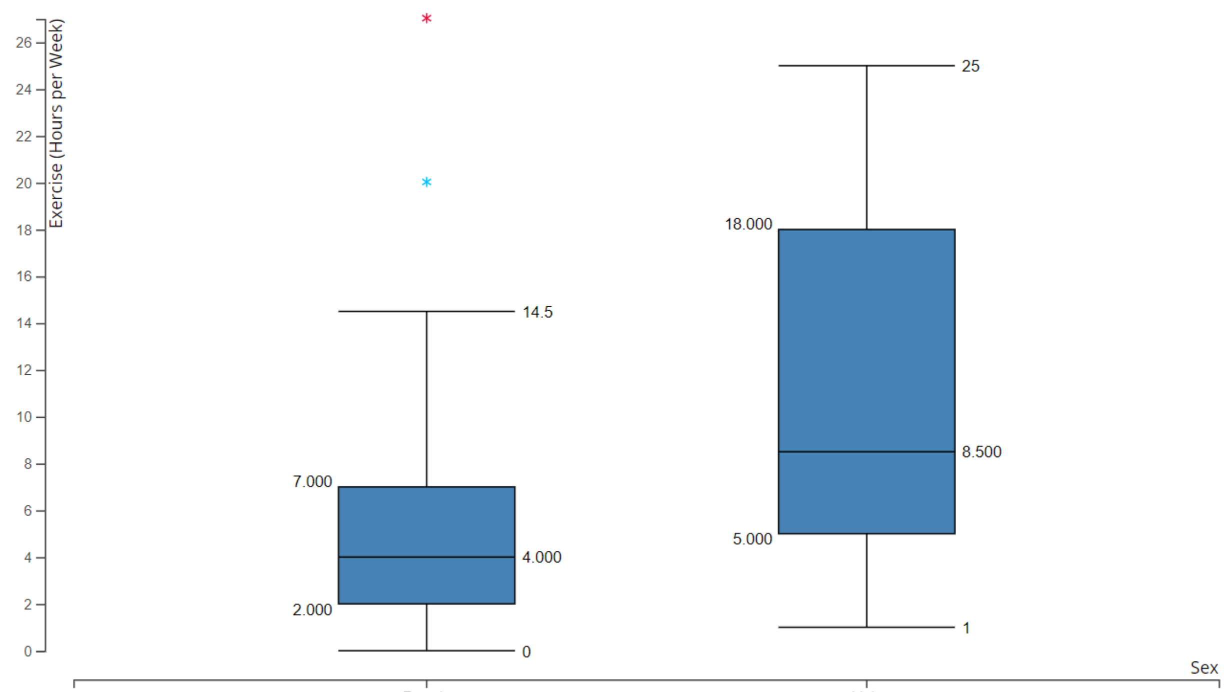

Table 1. Descriptive Statistics for Hours Exercised per Week by Sex									
Level of Sex	Measured Variable	n	Standard Deviation	Minimum Value	Q1	Median	Mean	Q3	Maximum
Female	Exercise (hours per week)	41	5.189	0	2	4	5.341	7	27
Male	Exercise (hours per week)	30	7.427	1	5	8.5	10.867	18	25

Table 2. Two-Sample t-Test Results			
Difference	-5.436	**t Ratio**	-3.441194
Std Error Diffrnce	0.545592	**DF**	48.893446
Upper CL Diffrnce	6.6611	**Prob > |t|**	0.001195
Lower CL Diffrnce	4.3906		
Confidence	0.95		