Audrey Leung November 20, 2014

1. **Task 1: Conduct a chi-square test to determine if there is an association between marital status (marital) and political orientation (politics).**

    a. **What are the null and alternative hypothesis for your test?**
    The null hypothesis is that there is no relationship between the two variables, marital status and political orientation; they are independent.
    The alternative hypothesis is that the two variables, marital status and political orientation, are not independent.

    b. **What test statistic and p-value do you get?**

```
        Pearson's Chi-squared test

data:  GSS$marital and GSS$politics
X-squared = 44.2255, df = 16, p-value = 0.0001823
```

    The test statistic is 44.2255
    The p-value is 0.0001823
    I used Pearson's product-moment correlation coefficient because the alternative hypothesis is two-sided, and I cannot predict the nature of the relationship between marital status and political orientation.

    c. **Conduct an appropriate effect size calculation for your relationship.**

```
[1] Cramer's V:
[1] 0.08756363
```

    Cramer's V is 0.08756363. I am using Cramer's V to calculate the effect size because the table has more than a 2x2 matrix of marital status x political orientation.

    d. **Evaluate your hypothesis in light of your tests of statistical and practical significance. What, if anything, can you conclude from your results?**
    For the test of statistical significance, since the p-value is < 0.05, the test is highly significant. Thus, we reject the null hypothesis and accept the alternative hypothesis. There is a relationship between marital status and political orientation. Marital status and political orientation are not independent.
    Since Cramer's V is < 0.2, the value of association between the variables is weak with low practical significance.

2. **Task 2: Conduct a correlation analysis to examine if there is an association between age when married (agewed) and hours of tv watched (tvhours).**

```
        Pearson's product-moment correlation

data:  GSS$agewed and GSS$tvhours
t = -1.0349, df = 1192, p-value = 0.3009
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.08654554  0.02681630
sample estimates:
        cor
-0.02996096
```

    a. **What are the null and alternative hypotheses for your test?**
    The Pearson's correlation coefficient measures the extent to which two metric or interval-

type variables are linearly related. In this case, there are two metric variables.
The null hypothesis is that there is no correlation between the two variables, age when married and hours of TV watched; the true correlation is equal to 0.
The alternative hypothesis is that the true correlation is not equal to 0. The age when a person is married is not correlated to the number of hours of TV watched.

b. **What test statistic and p-value do you get?**
The test statistic is -1.035.
The p-value is 0.30.

c. **Evaluate your hypothesis in light of your tests of statistical and practical significance. What, if anything, can you conclude from your results?**
For the test of statistical significance, since the p-value is > 0.05, the test is not statistically significant and we cannot reject the null hypothesis. There is no correlation between age when married and hours of TV watched. If p-value were statistically significant, then the test statistic of -1.035 would imply that the variables are negatively correlated.

3. **Task 3: Create a new binary/dummy variable, "married", that denotes whether an individual is currently married or not currently married. Conduct an independent sample t-test to evaluate the hypothesis that number of children (childs) is greater for those who are married than those who are not married.**

```
	Welch Two Sample t-test

data:  GSS$childs by GSS$married
t = -8.067, df = 1408.402, p-value = 7.641e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf -0.5521131
sample estimates:
mean in group FALSE  mean in group TRUE 
           1.482906            2.176545 
```

a. **What is the null and alternative hypotheses for your test?**
The null hypothesis is that the true difference in means is equal to 0; there is no difference in means. The number of children is the same for those who are married than those who are not.
The alternative hypothesis is that the true difference in means is greater than 0. Then number of children is greater for those who are married than those who are not.

b. **What test statistic and p-value do you get?**
The test statistic is -8.067
The p-value is 7.641e-16

c. **Conduct an appropriate effect size calculation for your relationship.**

```
> cohens_d(married_t, married_f)
[1] 0.4212165
```

Cohen's d was 0.4212165.

d. **Evaluate your hypothesis in light of your tests of statistical and practical significance.** What, if anything, can you conclude from your results?
For the p-value < 0.05, the test was highly significant, meaning that I would reject the null hypothesis and accept the alternative hypothesis. The true difference in means of the marital status and number of children is not equal to zero. Then number of children is greater for those who are married than those who are not. Since Cohen's d falls in the range from 0.3-0.5, the results have a medium effect or medium practical significance.

4. **Task 4: We want to consider just the subpopulation of 23-year olds in this sample. Conduct a Wilcox rank-sum test to determine whether your new "married" variable from Task 3 is associated with the number of children (childs) for respondents who are 23 years old.**

```
	Wilcoxon rank sum test with continuity correction

data:  age$childs by age$married
W = 19, p-value = 0.0002656
alternative hypothesis: true location shift is not equal to 0
```

a. **What is the mean of your new "married" variable among 23-year-olds (e.g., the proportion of cases in the category coded "1")?**

```
> mean(age$married, na.rm=TRUE)
[1] 0.2857143
```

The mean of the "married" variable among 23-year-olds is 0.2857.

b. **What is the null and alternative hypotheses for your test?**
The null hypothesis is that there is no difference in ranks. The number of children that a 23-year-old person has is not associated with whether they are married.
The alternative hypothesis is that there is a difference in ranks, and the number of children that a 23-year-old person has is associated with whether they are married or not.

c. **What test statistic and p-value do you get?**
The test statistic is 19.
The p-value is 0.0002656.

d. **Conduct a cohen's d effect size calculation for your relationship (you can use the** same code we used in our class R examples).

```
> cohens_d(age_married_t, age_married_f)
[1] 1.976885
```

Cohen's d effect size is 1.976885

e. **Evaluate your hypothesis in light of your tests of statistical and practical significance. What, if anything, can you conclude from your results?**
Since the p-value is <0.05, the test was highly significant, meaning that I would reject the null hypothesis and accept the alternative hypothesis. The number of children that a 23-year-old person has is associated with whether they are married or not.
Since Cohen's d is > 0.9, the practical significance is very high and the effect size is very strong. The number of children a 23-year-old person has is strongly associated with whether they are married or not.