Hypothesis Testing 2

Lisa Gandolfi

Instructions

```
library(DescTools)
library(Hmisc)
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:DescTools':
##
##     %nin%, Label, Mean, Quantile

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(gplots)
```

```
## Registered S3 method overwritten by 'gplots':
##   method         from
##   reorder.factor DescTools

##
## Attaching package: 'gplots'

## The following object is masked from 'package:DescTools':
##
##     reorder.factor

## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(descr)
library(effectsize)

load("data/countries2.rda")

##Independent:fem_finance: % of women with account at financial institution
#or with mobile money-service provider
```

```r
##Dependent: gender_inequality: Gender Inequality Index:
#inequality in reproductive health, empowerment and
#the labor market

#Transform the independent variable
countries2$fem_finance1 <- cut2(countries2$fem_finance, g=4)
summary(countries2$fem_finance1)
```

```
## [ 1.67, 30.0) [30.03, 53.6) [53.59, 81.3) [81.34,100.0]          NA's
##            39            39            39            39            39
```

```r
#Assign names to the levels
levels(countries2$fem_finance1) <-c("Low", "Limited", "Medium", "High")
```

1. State the null and the alternative hypotheses.

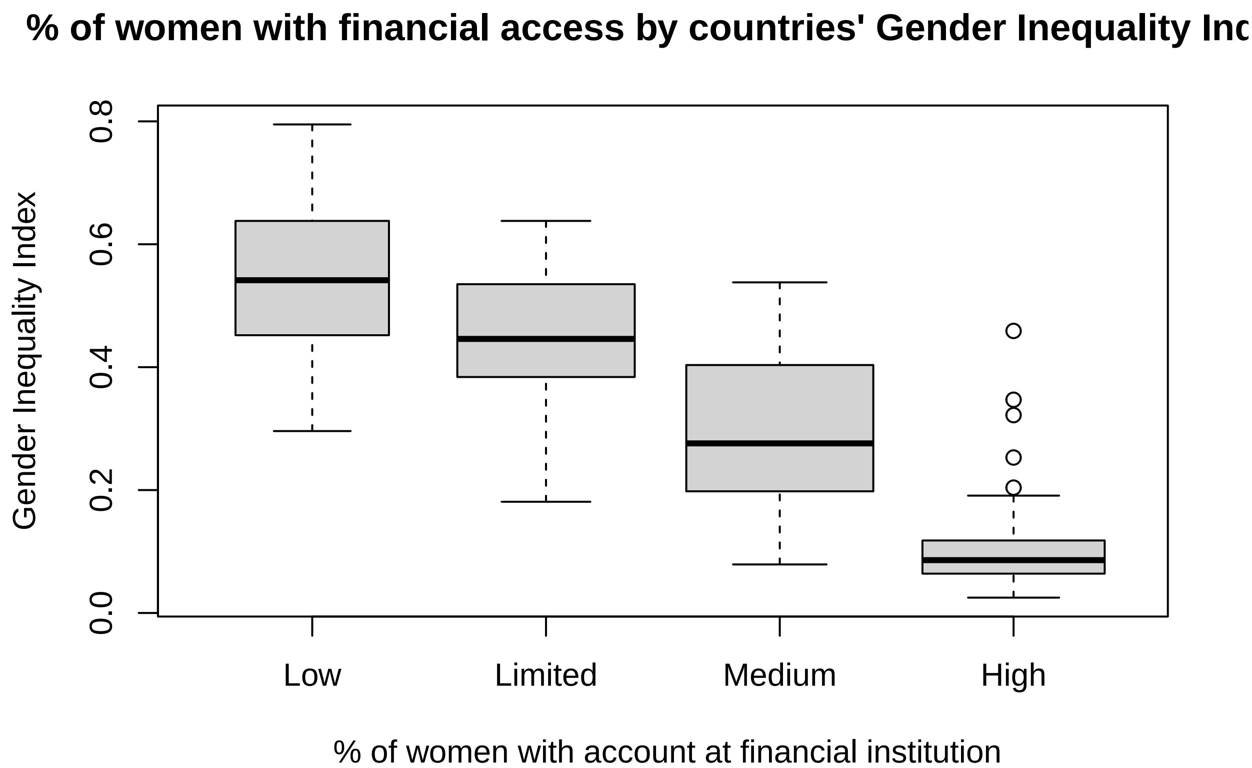
```r
##Ho:m1 = m2 = m3 = m4
##Ha:the mean level of percentage of females with financial access will vary across
#different levels of the Gender Inequality Index
```

I am hypothesizing that the mean across different levels of female access to financial institutions will vary as the Gender Inequality Index varies across countries. This is because countries that have low access for females in financial institutions might also have a high degree of prejudice or societal norms that prevent women from being regarded equally to men. This will lead to a high degree of difference among the Gender Inequality Index.

2. Using `compmeans` describe what the relationship looks like.

```r
compmeans(countries2$gender_inequality, countries2$fem_finance1, xlab=" % of women with account at finan
```

```
## Warning in compmeans(countries2$gender_inequality, countries2$fem_finance1, :
## 49 rows with missing values dropped
```




```
## Mean value of "countries2$gender_inequality" according to
## "countries2$fem_finance1"
##              Mean   N  Std. Dev.
## Low      0.5405000  32 0.11476568
## Limited  0.4412703  37 0.11510132
## Medium   0.2940000  39 0.12873473
## High     0.1166053  38 0.09334983
## Total    0.3391781 146 0.19444211
```

From the plot, we can see that there is a relationship between gender inequality and the percentage of women who have access to financial institutions. In particular, countries with a very low percentage of women with access to finance ("Low"), are also the countries with a high gender inequality index, nearing a median of 0.6. Countries with a limited access to financial institutions for women have a slightly smaller median gender inequality index, around 0.4. Countries with "Medium" access have a gender inequality index slightly bigger than 0.2, and those with "High" access have the smallest gender inequality index, nearing zero. This confirms our hypothesis that the two variables will have different relationships across levels of financial access for women. This also tells us that countries with "Low", "Limited" and "medium" access all seem to have similar standard deviations, however countries with "Low" access seem to have smaller standard deviations, but also appear to have a few extreme outliers that have a "High" access to finance for females and yet still have significant gender inequality indexes.

3. Test for statistical significance using ANOVA. You should also conduct the appropriate test of the strength of the relationship.

```r
fem_fin_ineq<-aov(countries2$gender_inequality~countries2$fem_finance1)
summary(fem_fin_ineq)
```

```
##                          Df Sum Sq Mean Sq F value Pr(>F)
## countries2$fem_finance1   3  3.645  1.2149   93.89 <2e-16 ***
## Residuals               142  1.837  0.0129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 49 observations deleted due to missingness
```

```r
#Test strenght
eta_squared(fem_fin_ineq)
```

```
## For one-way between subjects designs, partial eta squared is equivalent
##   to eta squared. Returning eta squared.

## # Effect Size for ANOVA
##
## Parameter               | Eta2 |       95% CI
## ---------------------------------------------
## countries2$fem_finance1 | 0.66 | [0.59, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

The sum of squares between's degrees of freedom for this test is 3, while the sum of squares within's degrees of freedom is 142. The mean square between is 1.2149, while the mean squared error is 0.0129. The f-stat for this test is 93.89 and the p-value is <2e-16. Since our p-value is very close to zero, and is < than the alpha value of 0.05 which is commonly used for hypothesis testing. As a result, we can reject the null hypothesis and we can conclude that differences in the Gender Inequality Index are related to differences in the percentage of women who have access to financial institutions.

In order to determine the strenght of this relationship, I utilized the Eta-squared command. The results from running the command tell us that differences in Gender Inequality Index explain about 66% of the variation in female access to financial institutions across nations. This represents quite a strong relationship, as exemplified by Table 11.2 on page 271 of the textbook. As a result, we can conclude that our hypothesis test is reliable and we can uphold our previous conclusion.