

Homework5

Swetha Jayapathy

11/4/2020

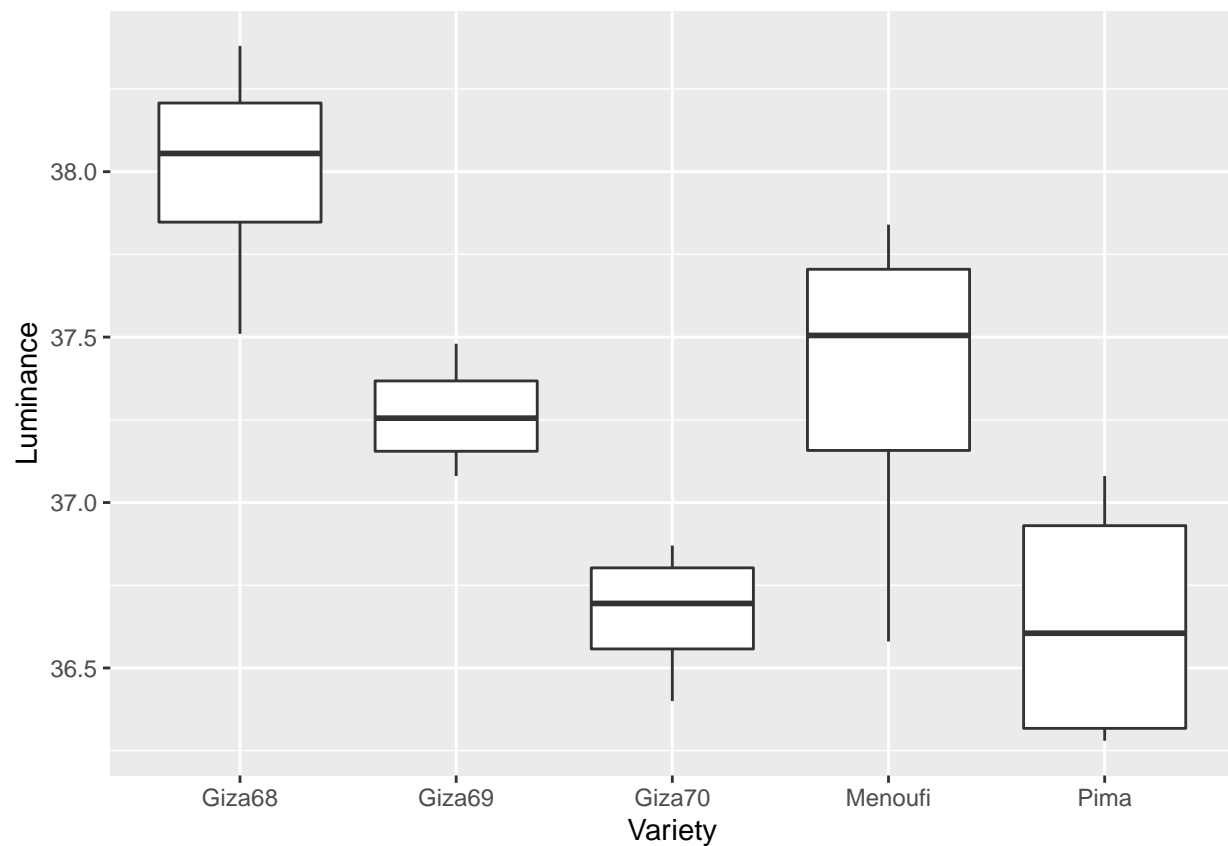
R Markdown

```
library(Sleuth3)
library(ggplot2)
```

```
HW5Dat <- read.csv("D:/D drive contents/Fall 2020/Stats 511/Class/HW/HW5/Cotton.csv")
```

Question 1a:

```
qplot(Variety, Luminance, data=HW5Dat, geom="boxplot")
```



Question 1b: Normality : Looking at the box plot, since the median line is almost at the middle of the boxes and the data does not look skewed, we could say that the Normality assumption holds good.

Equal Standard Deviation : The height of the boxes differ so much and hence the data is not equally spread. Hence we cannot say that the equal standard deviation holds for this data.

Independence: Since 4 samples of each variety of cotton are randomly selected, we could say that the independence assumption holds good.

Question 1c:

```
with(HW5Dat, unlist(lapply(split(Luminance, Variety), length)))
```

```
## Giza68 Giza69 Giza70 Menoufi Pima
##      4      4      4      4      4
```

```
with(HW5Dat, unlist(lapply(split(Luminance, Variety), sd)))
```

```
## Giza68 Giza69 Giza70 Menoufi Pima
## 0.3690528 0.1750000 0.2069622 0.5564396 0.3986958
```

We could see from the data that the sample sizes for all the varieties are equal. We could also see from the above result that the ratio of the largest of standard deviation to the smallest of standard deviation is 3.179 which is not much larger than 3. The population is also normal as from the above box plot it can be seen that the data is not that much skewed and the median is almost at the middle of the box for all the varieties. Therefore, it can be said that ANOVA F-test is appropriate for the cotton data.

Question 1d: Let $\mu_1, \mu_2, \mu_3, \mu_4$ and μ_5 be the population mean of Giza68, Giza69, Giza70, Menoufi and Pima respectively.

Null Hypothesis : $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ Alternate Hypothesis : Let μ_i denote any of the above mentioned population mean. H_A : At least one of the μ_i is different.

Question 1e:

```
HW5Dat_aov <- aov(Luminance~Variety, data=HW5Dat)
HW5Dat_aov
```

```
## Call:
## aov(formula = Luminance ~ Variety, data = HW5Dat)
##
## Terms:
##              Variety Residuals
## Sum of Squares  5.061930  2.034725
## Deg. of Freedom      4      15
##
## Residual standard error: 0.3683047
## Estimated effects may be unbalanced
```

One-way ANOVA test and Output:

```
anova(HW5Dat_aov)
```

```
## Analysis of Variance Table
##
## Response: Luminance
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Variety      4 5.0619  1.26548    9.3291 0.0005416 ***
## Residuals    15 2.0347  0.13565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 1f: Statistical Conclusion: Significant evidence that the population mean of the luminance of Giza68, Giza69, Giza70, Menoufi and Pima are all not equal. That is, at least one of the population mean is different from other ($p=0.0005416$, one way ANOVA F-test).

Question 1g: Calculate a 95% confidence interval for the difference between the population mean luminance for the Giza68 and Giza69 cotton :

```
with(HW5Dat, unlist(lapply(split(Luminance, Variety), mean)))
```

```
## Giza68 Giza69 Giza70 Menoufi Pima
## 38.0000 37.2675 36.6650 37.3575 36.6425
```

```
qt(0.975, 15)
```

```
## [1] 2.13145
```

```
(38.0000-37.2675)-2.13145*sqrt(0.13565)*sqrt(1/4 + 1/4)
```

```
## [1] 0.1774015
```

```
(38.0000-37.2675)+2.13145*sqrt(0.13565)*sqrt(1/4 + 1/4)
```

```
## [1] 1.287598
```

Thus CI ranges from 0.1774015 to 1.287598 cd/m^2 .

Question 1h: Statistical Conclusion : We estimate that the mean luminance of the Giza68 variety is 0.1774 to 1.2875 cd/m^2 more than the mean luminance of Giza69 variety (95% Confidence Interval).

Question 1i: Scope of Inference : The samples were not randomly selected from a general population. Therefore we could say that there is no causation as it is an observational study. Since the sample was randomly selected from each variety of cotton, we could say that inferences to the population can be drawn. It can be seen from the ANOVA F-test that there is a difference in the population means between the variety of cottons. For example the mean luminance of the Giza68 is 0.1774 to 1.2875 cd/m^2 more than the mean luminance of Giza69.