

Statistical Methods and Data Analysis I

Lecture 17: Using F-Test.

Oleg Goldshmidt

`oleg.goldshmidt@post.idc.ac.il`

Arison School of Business
Interdisciplinary Center (IDC)
Herzliya, Israel

May 21, 2019

Customer Service Example: Scenario

Scenario: mobile phone service centre

A manager of a mobile communications company's service centre notices that some days his representatives are very busy, and on other days — quite idle. He mentions this to a colleague, the manager of another such centre. She says, however, that she has never noticed such a problem, her representatives' workload is quite stable day to day.

Customer Service Example: Scenario

Scenario: mobile phone service centre

A manager of a mobile communications company's service centre notices that some days his representatives are very busy, and on other days — quite idle. He mentions this to a colleague, the manager of another such centre. She says, however, that she has never noticed such a problem, her representatives' workload is quite stable day to day.

They call a data analyst in for help...

Customer Service Example: Scenario

Scenario: mobile phone service centre

A manager of a mobile communications company's service centre notices that some days his representatives are very busy, and on other days — quite idle. He mentions this to a colleague, the manager of another such centre. She says, however, that she has never noticed such a problem, her representatives' workload is quite stable day to day.

They call a data analyst in for help...

- The data analyst asks for historical data about the number of customers served by both branches on a daily basis. He suspects that the number of customers in the first branch varies a lot more than in the second one.

Customer Service Example: Scenario

Scenario: mobile phone service centre

A manager of a mobile communications company's service centre notices that some days his representatives are very busy, and on other days — quite idle. He mentions this to a colleague, the manager of another such centre. She says, however, that she has never noticed such a problem, her representatives' workload is quite stable day to day.

They call a data analyst in for help...

- The data analyst asks for historical data about the number of customers served by both branches on a daily basis. He suspects that the number of customers in the first branch varies a lot more than in the second one.
- The data are:

```
> X <- c(156, 278, 134, 202, 236, 198, 187, 199, 143, 165, 223)
> Y <- c(345, 332, 309, 367, 388, 312, 355, 363, 381)
```

Customer Service Example: Scenario

Scenario: mobile phone service centre

A manager of a mobile communications company's service centre notices that some days his representatives are very busy, and on other days — quite idle. He mentions this to a colleague, the manager of another such centre. She says, however, that she has never noticed such a problem, her representatives' workload is quite stable day to day.

They call a data analyst in for help...

- The data analyst asks for historical data about the number of customers served by both branches on a daily basis. He suspects that the number of customers in the first branch varies a lot more than in the second one.
- The data are:

```
> X <- c(156, 278, 134, 202, 236, 198, 187, 199, 143, 165, 223)
> Y <- c(345, 332, 309, 367, 388, 312, 355, 363, 381)
```
- Degrees of freedom: 10 and 8.

Customer Service Example: Using F Distribution



Customer Service Example: Using F Distribution

- The data analyst formulates his $H_0 : \sigma_X^2 = \sigma_Y^2$ and $H_A : \sigma_X^2 \neq \sigma_Y^2$...

Customer Service Example: Using F Distribution

- The data analyst formulates his $H_0 : \sigma_X^2 = \sigma_Y^2$ and $H_A : \sigma_X^2 \neq \sigma_Y^2$...
- ... chooses the desired confidence level (95%)...

Customer Service Example: Using F Distribution

- The data analyst formulates his $H_0 : \sigma_X^2 = \sigma_Y^2$ and $H_A : \sigma_X^2 \neq \sigma_Y^2$...
- ... chooses the desired confidence level (95%)...
- ... and computes the variances, and the “F-score”:

```
> var(X)
[1] 1828.564
> var(Y)
[1] 795.1944
> var(X)/var(Y)
[1] 2.299518
```

Customer Service Example: Using F Distribution

- The data analyst formulates his $H_0 : \sigma_X^2 = \sigma_Y^2$ and $H_A : \sigma_X^2 \neq \sigma_Y^2$...
- ... chooses the desired confidence level (95%)...
- ... and computes the variances, and the “F-score”:

```
> var(X)
[1] 1828.564
```

```
> var(Y)
[1] 795.1944
```

```
> var(X)/var(Y)
[1] 2.299518
```

- The p -value can be computed in 2 ways:

```
> 1-pf(2.3, 10, 8)
[1] 0.1252775
```

```
> pf(2.3, 10, 8, lower.tail=FALSE)
[1] 0.1252775
```

Customer Service Example: Using F Distribution

- The data analyst formulates his $H_0 : \sigma_X^2 = \sigma_Y^2$ and $H_A : \sigma_X^2 \neq \sigma_Y^2$...
- ... chooses the desired confidence level (95%)...
- ... and computes the variances, and the “F-score”:

```
> var(X)
[1] 1828.564
> var(Y)
[1] 795.1944
> var(X)/var(Y)
[1] 2.299518
```

- The p -value can be computed in 2 ways:

```
> 1-pf(2.3, 10, 8)
[1] 0.1252775
> pf(2.3, 10, 8, lower.tail=FALSE)
[1] 0.1252775
```

- The 97.5% quantile (not 95% because we look only at the right tail) is

```
> qf(0.975, 10, 8)
[1] 4.295127
```

Customer Service Example: Using F-Test

- To double-check himself the data analyst runs R's F-test:

```
> var.test(X,Y,alternative="greater",conf.level=0.95)
```

```
F test to compare two variances
```

```
F = 2.2995, num df = 10, denom df = 8, p-value = 0.1253  
alternative hypothesis: true ratio of variances is greater than 1  
sample estimates:  
ratio of variances  
      2.299518
```

- The conclusion is, no, the variances in daily number of customers are not significantly different, even though they *look* different.
- The difference in standard deviations is only ~ 1.5
- This may well be random...