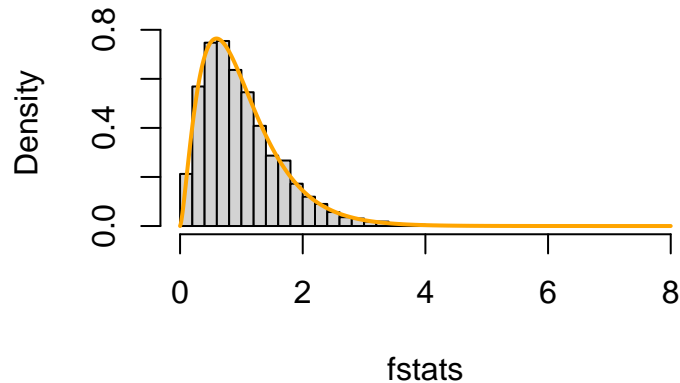# Studio 8 Simulation and computation of NHST
18.05, Spring 2022

Histogram of fstats

Overview of the studio

In this studio we will get some practice using R to simulate and run some common significance tests. As usual, we will provide a lot of hints and examples, but one of the goals of this studio is to get practice using R help and Google to figure out how to make R do what we want.

Here is a quick summary of the questions.
1. Simulation of the f-statistic: draw histogram and graph
2. z-test. Code by hand a z-test
3. Chi-square test for independence
4. ANOVA. using aov()

R introduced in this studio

See the sample code for examples.
F distribution: functions `df, pf,` etc.
Significance tests: `t.test, chisq.test, aov`

Download the zip file

- You should have downloaded the file `mit18_05_s22_studio8.zip` from our MITx site.

- Unzip it in your 18.05 studio folder.

- You should see the following R files

  `mit18_05_s22_studio8.r`
  `mit18_05_s22_studio8-samplecode.r`
  `mit18_05_s22_studio8-test.r`

  and the following other files

  `mit18_05_s22_studio8-test-answers.html`
  `mit18_05_s22_studio8-onesampledata.txt`
  `mit18_05_s22_studio8Problem3_test.tbl`

Prepping R Studio

- In R studio, open `mit18_05_s22_studio8-samplecode.r` and `mit18_05_s22_studio8.r`

- Using the Session menu, set the working directory to source file location. (This is a good habit to develop!)

- Answer the questions in the detailed instructions just below. Your answers should be put in `mit18_05_s22_studio8.r`

- Solution code will be posted tomorrow at 10 pm

Detailed instructions for the studio

• Go through **mit18_05_s22_studio8-samplecode.r** as a tutorial.

Problem 1

Problem 1. Here you will finish the code for the function
```
studio8_problem_1 = function(n, m, mu, sigma, n_trials)
```

The arguments to this function are:
  `n` = number of groups
  `m` = number of data points in each group (same for all groups)
  `mu` = common mean: same for each group
  `sigma` = common standard deviation (same for each group)
  `n_trials` = number of trials to run in the simulation

Here the goal is to do a simulation and show graphically that the f-statistic does indeed follow an f-distribution.

One simulation will consist of:
(i) Use the function parameters to draw m samples for each of n groups of data from a Norm(mu, sigma^2) distribution.
(ii) Compute the f-statistic of the data.

Do this n_trials times and make a density histogram of the f-statistics. Overlay a graph of the pdf of the f-distribution with the correct degrees of freedom.

Problem 2

Problem 2. Here you will finish the code for the function

```
studio8_problem_2 = function(data, mu0, known_sigma, alpha)
```

The arguments to this function are:
    `data` = list of sample values
    `mu0` = hypothesized mean in the null hypothesis
    `known_sigma` = known value of the standard deviation
    `alpha` = significance for the test

In this problem, you are given some data drawn from a normal distribution with a known standard deviation. The null hypothesis is that the mean of the normal distribution is `mu0`. The alternative hypothesis is that the mean is not `mu0`.

Your job is to run a z-test on the data using the given significance. Print out the mean of the data, z-statistic, p-value, and your conclusion on whether or not to reject the null hypothesis.

Problem 3

Problem 3. Here you will finish the code for the function

```
studio8_problem_3 = function(data_table_file, alpha)
```

The arguments to this function are:
    `data_table_file` = data file containing a two-by-two contingency table of married and college data.
    `alpha` = significance for the test.

In this problem, you will run a chi-square test for independence. The data is a two-by-two table (called a contingency table) from a sample of women showing the counts of their number of marriages and their college education. (When you see the table printed out, it should be clear what is in it.)

The data for testing your code was taken from the book by Rice, example on p. 489. For grading, we will make up more data.

Let the null hypothesis be that the number of marriages is independent of the level of education. Run a chi-square test for independence to test this hypothesis. You may do this by 'hand' or by using a built-in test in R.

Print out the test statistic and the p-value.

Using significance level alpha, print out whether or not you reject H_0.

Remember, you need to write code that can deal with different values of alpha and different p-values.

Problem 4

Problem 4. Here you will finish the code for the function

```
studio8_problem_4 = function(T1, T2, T3, alpha)
```

The arguments to this function are:

`T1, T2, T3` = 3 sets of data. All the same length
`alpha` = significance of the test.

The data sets T1, T2, T3 each represent the (simulated) levels of pain associated with patients receiving one of three treatments.

In this problem you should use the `aov()` function to run a one-way ANOVA with null hypothesis that the mean levels of pain for all 3 treatments are the same.

To help, we supply code to build the data into a data.frame. We then print it out so you can see how it looks.

You should use the `aov()` function to run the test.

Use the `summary()` command get a summary of the results of `aov()`. Then extract the p-value from the summary.

Print out the summary and the p-value. Then, using significance level alpha, print out your conclusion on whether or not to reject H0

Testing your code

For each problem, we ran the problem function with certain parameters. You can see the function call and the output in `mit18_05_s22_studio8-test-answers.html`. If you call the same function with the same parameters, you should get the same results as in `mit18_05_s22_studio8-test-answers.html` – if there is randomness involved the answers should be close but not identical.

For your convenience, the file `mit18_05_s22_studio8-test.r` contains all the function calls used to make `mit18_05_s22_studio8-test-answers.html`.

Before uploading your code

1. Make sure all your code is in `mit18_05_s22_studio8.r`. Also make sure it is all inside the functions for the problems.

2. Clean the environment and plots window.

3. Source the file.

4. Call each of the problem functions with the same parameters as the test file `mit18_05_s22_studio8-test-answers.html`.

5. Make sure it runs without error and outputs just the answers asked for in the questions.

6. Compare the output to the answers given in `mit18_05_s22_studio8-test-answers.html`.

Upload your code

Use the upload link on our MITx site to upload your code for grading.

Leave the file name as `mit18_05_s22_studio8.r`. (The upload script will automatically add your name and a timestamp to the file.)

You can upload more than once. We will grade the last file you upload.

Due date

Due date: The goal is to upload your work by the end of class.

If you need extra time, you can upload your work any time before 10 PM ET the day after the studio.

Solutions uploaded: Solution code will be posted on MITx at 10 PM the day after the studio.

MIT OpenCourseWare

https://ocw.mit.edu

18.05 Introduction to Probability and Statistics

Spring 2022