R Quiz Instructions
18.05, Spring 2022

Overview

The quiz format is identical to that of the R studios and practice quiz.

- READ THE INSTRUCTIONS PDF!
- Read the problems carefully
- Save your work frequently
- You may use any resources except another person
- (This includes, paper, books, code on your computer, code on the internet ...)
- If code is given with a question it is meant to be used. Don't just ignore it.
- Remember to use print or cat statements to print the values asked for
- Before uploading the code: clear your environment and source the entire file (choose source from the code menu)
- Make sure that it runs without error and outputs just the answers asked for in the questions.
- Due Friday 5/6 by 5 PM.

Download the zip file

- You should have downloaded the file `mit18_05_s22_RQuiz.zip` from our MITx site.
- Unzip it in your 18.05 studio folder.
- You should see the following R files
  `mit18_05_s22_RQuiz.r`
  `mit18_05_s22_RQuiz-test.r`
  `mit18_05_s22_RQuiz-samplecode.r`

  and the following other files

  `mit18_05_s22_RQuiz-test-answers.html`
  `mit18_05_s22_RQuiz_data_prob4_test.txt`

General instructions

- Using the Session menu, set the working directory to source file location.
- Answer the questions in the detailed instructions just below. Your answers should be put in `mit18_05_s22_RQuiz.r`

- For each question, you will complete the code for the corresponding function.
- As usual, use the function's arguments and any other code given in the function.
- Do not print out things that are not asked for.

Note: The file `mit18_05_s22_RQuiz-samplecode.r` contains only a few lines of code, but they may be useful to you.

Detailed instructions for this quiz

0. Clean your space

Problem 1 (20 points)

This problem will ask you to do several different short tasks. The parts are not related.

Problem 1a (5 points) Graphing

Here you will finish the code for the function:

`rquiz_problem_1a(mu, sigma, w_shape, w_scale, a, b)`

Arguments:

`mu` = mean of the normal pdf to plot
`sigma` = standard deviation of the normal pdf to plot
`w_shape` = shape parameter for Weibull pdf
`w_scale` = scale parameter for Weibull pdf
`a, b` = endpoints of the range of x for the plot

On the same plot for x between a and b put graphs of the pdfs of:
(i) Norm(mu, sigma)
(ii) Weibull distribution with shape and scale parameters `w_shape` and `w_scale`.

The graphs should be in different colors

Every plot should be a line graph (type='l')

Problem 1b (5 points) Combination and factorials.

Here you will finish the code for the function: `rquiz_problem_1b(n, k, m)`

Arguments:

`n` = see instructions below
`k` = see instructions below
`m` = see instructions below

This function should compute and print the following values:

(i) n choose k
(ii) m factorial
(iii) log of n choose k (for this use the function lchoose to avoid overflow)

Problem 1c (10 points) Bayesian success.

Here you will finish the code for the function:

`rquiz_problem_1c(theta_values, num_patients, num_cured)`

Arguments:

`theta_values` = List of possible values of θ.
`num_patients` = The number of patients in the trial.
`num_cured` = The number of successes in the trial.

A treatment with unknown probability θ of success is tried on `num_patients` patients yielding `num_cured` successes. The possible values of the unknown θ are given in `theta_values`. That is, we only entertain a finite number of hypotheses for the value of θ.

Your code should use the data and do each of the following:

(i) Compute the maximum likelihood estimate (MLE) for θ. (HINT: the function which.max might be useful.) Print out the MLE

(ii) Suppose there is a flat prior, i.e. each of the possible values of θ is equally likely. Find the prior predictive probability that a single patient will be cured. Print out the predictive probability. (This part does not use the experimental data.)

(iii) Use the data and a flat prior to do a Bayesian update to find the posterior probability for θ. Print out the posterior.

Problem 2 (20 points)

This problem is on making histograms.

Problem 2a (10 points) Here you will finish the code for the function:

```
rquiz_problem_2a = function(n_draws, k, bin_width)
```

Arguments:

`n_draws` = Number of sample points in the histogram
`k` = Number of degrees of freedom for the chi-square distribution
`bin_width` = Bin width for histogram

Your code should simulate `n_draws` draws from a chi-square distribution with `k` degrees of freedom. Use the results to plot a density histogram. Use a bin width of `bin_width`. Finally, plot the $\chi^2(k)$ probability density function on top of it.

Problem 2b (10 points) Here you will finish the code for the function:

```
rquiz_problem_2b = function(n_trials, n_draws, k, bin_width)
```

Arguments:

`n_trials` = Number of trials
`n_draws` = Number of sample points in each trial
`k` = Number of degrees of freedom for the chi-square distribution
`bin_width` = Bin width for histogram

This problem will illustrate the central limit theorem as follows.

- One trial will consist of drawing a sample of size `n_draws` from a $\chi^2(k)$ distribution.
- Simulate `n_trials` trials.
- For each trial, compute the standardized mean. It should help to know that the $\chi^2(k)$ distribution has mean k and **variance** $2k$.

- Plot a density histogram of the `n_trials` standardized means. Use bin width `bin_width`.
- Add a graph of the standard normal pdf to the histogram.

Problem 3 (10 points)

Here you will finish the code for the function

`rquiz_problem_3 = function(our_data, alpha)`

Arguments:

`our_data` = data from some experiment.
`alpha` = Significance level for the Shapiro-Wilk test

You have collected data and before running a t-test, you want to check if the data comes from a normal distribution. To do this, you run a Shapiro-Wilk test for normality.

Run the Shapiro-Wilk test. Then, print out the null hypothesis, p-value and whether or not to reject the null hypothesis.

You need to use code to find and print the p-value and decide whether to reject or not. You shouldn't just read the p-value off the screen and enter that value. That is, your code should work correctly even if we change the data or significance level for the test.

Problem 4 (Extra credit: 5 points)

This problem is for extra credit if you have time.

Make sure you have set the WORKING DIRECTORY to the source file location, so R will find the file. The lists are data from independent random normal trials.

Here you will finish the code for the function

`rquiz_problem_4 = function(data_file_name, alpha)`

Arguments:

`data_file_name` = data file
`alpha` = Significance level for t-test

The given code extracts two lists, `x` and `y` from the data file.

Run a two sample t-test with unequal variances to test if x and y are drawn from distributions with the same mean

Print out the p-value and whether or not to reject the null hypothesis at significance level `alpha`

You need to use code to find and print the p-value and decide whether to reject or not. You shouldn't just read the p-value off the screen and enter that value. That is, your code should work correctly even if we change the data file or significance level for the test.

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_RQuiz-test-answers.html`. If you call the same function with the same parameters, you should get the same results as in

`mit18_05_s22_RQuiz-test-answers.html` – if there is randomness involved the answers should be close but not identical.

For your convenience, the file `mit18_05_s22_RQuiz-test.r` contains all the function calls used to make `mit18_05_s22_RQuiz-test-answers.html`.

Before uploading your code

1. Make sure all your code is in `mit18_05_s22_RQuiz.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_RQuiz-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_RQuiz-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_RQuiz.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: Friday 5/6 at 5 pm

18.05 Introduction to Probability and Statistics
Spring 2022