

## 18.05 Problem Set 9, Spring 2022

### Problem 1. (15: 10,5 pts.)

We perform a  $t$ -test for the null hypothesis  $H_0 : \mu = 10$  at significance level  $\alpha = 0.05$  by means of a dataset consisting of  $n = 16$  elements with sample mean 11 and sample variance 4.

- (a) Should we reject the null hypothesis in favor of  $H_A : \mu \neq 10$ ?
- (b) What if we test against  $H'_A : \mu > 10$ ?

### Problem 2. (40: 10,10,5,10,5 pts.)

Jerry took a JP Licks token and asked Jon to perform a test at significance level  $\alpha = 0.05$  to investigate whether the coin is fair or biased toward tails (the side that says ‘Token’). Jon recorded the following data

THTTHTTTTTTTH

showing 3 heads and 9 tails.

Before Jon could compute the one-sided  $p$ -value for  $H_0 : \theta = 0.5$  versus  $H_A : \theta < 0.5$ , he needed to take Aviva to the playground.

- (a) Erika believes that Jon’s intention was to count the number of heads in twelve flips. Let’s call this Experiment 1. Compute the rejection region and  $p$ -value. Sketch the null-distribution and rejection region. What does Erika conclude?
- (b) Ruthi believes that Jon’s intention was to stop after the third heads and report the number of tails, e.g., in the data the third head came on flip 12 so the number of tails is 9. Let’s call this Experiment 2. Compute and sketch the corresponding null-distribution, rejection region, and  $p$ -value. What does Ruthi conclude? Hint: if a counting argument eludes you, google “negative binomial distribution”.

The R functions `dnbinom`, `pnbinom`, etc. might be helpful. For example, `dnbinom(5, 3, 0.2)` gives the probability of seeing 5 tails before the third head in a sequence of tosses of a coin with probability 0.2 of heads.

- (c) Jerry actually told Jon to count the number of heads in 100 flips (Experiment 3), so Jerry figures that Jon must have gotten bored and quit right after the 12<sup>th</sup> flip. Strictly speaking, can Jerry compute a  $p$ -value from Jon’s partial data? Why or why not?
- (d) Let’s reexamine the same data from the Bayesian perspective. What is the likelihood function in Experiment 1? What is the likelihood function in Experiment 2? How are these likelihood functions related? Given the prior  $\text{Beta}(n,m)$ , find the posterior in each case. How are they related?
- (e) Read [https://en.wikipedia.org/wiki/Likelihood\\_principle](https://en.wikipedia.org/wiki/Likelihood_principle), appreciate the volt-meter story, and summarize the main points we are getting at via the earlier parts of this problem regarding frequentist and Bayesian experiments.

### Problem 3. (10 pts.) (Chi-square for variance)

The following data comes from a normal distribution which you suspect has variance equal to 1. You want to test this against the alternative that the variance is greater than 1.

1.76, -2.28, -0.56, 1.46, 0.59, 1.26,  
-1.94, -0.79, -0.86, -1.41, 2.07, 1.30

There is a chi-square test for this. Look at

<https://www.itl.nist.gov/div898/handbook/eda/section3/eda358.htm>

and run the test with significance level 0.05. You can use R for the computations, but explain what you are doing and give the value of the test statistic, and the  $p$ -value.

**Problem 4.** (10 pts.) (**Chi-squared for categorical data**)

Jon and Jerry spent a fortune on dice and bent coins for 18.05, so they decide to submit an invoice to the math department for reimbursement. The math department suspects that their six figure expense report is made up, so they call you to test the data for fraud. You do some research and learn that accounting data should follow something called *Benford's law*. This states that the relative frequency of the first digits of each entry should have the following distribution:

First digit $k$	1	2	3	4	5	6	7	8	9
probability $p(k)$	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

You go back to the math department and tell them that the only data you need is the counts of all the first digits in their invoice. They are skeptical, but they know you have taken 18.05 and, so, must know what you are doing. They give you the following counts.

First digit $k$	1	2	3	4	5	6	7	8	9
count	7	13	12	9	9	13	11	10	16

The math department doesn't want to unjustly accuse Jon and Jerry, so they ask you to test at 0.001 significance level. Run a significance test to see how well this data fits Benford's distribution and make a recommendation to the math department.

See [https://en.wikipedia.org/wiki/Benford%27s\\_law](https://en.wikipedia.org/wiki/Benford%27s_law)

**Problem 5.** (20: 10,10 pts.) (**Two-sample F test for equal variances**)

In this problem, we want you to become comfortable using the web to learn about new tests and R commands. There are dozens of statistical software packages used by labs and industry, so being able to learn new commands is probably more important than trying to memorize them all. Look at the help file in R for `var.test`

We used R to secretly create a vector  $x$  of 20 random samples from an  $N(0, 1)$  distribution and a vector  $y$  of 20 random samples from a  $N(10, 1)$  distribution. These are listed below. With a tiny bit of editing you should be able to copy and paste these into R.

-0.802, 0.457, 0.972, 0.044, 0.318, -1.380, 0.111,  
-0.023, -0.700, -1.977, -0.497, 1.471, -1.314, -0.078,  
-0.505, 0.583, 1.363, -1.863, -2.105, 0.488

9.019, 9.852, 7.947, 9.465, 10.060, 10.508, 9.506,  
9.540, 10.218, 9.407, 11.455, 11.422, 7.698, 9.972,  
10.928, 11.577, 10.376, 8.605, 9.347, 10.715

(a) Temporarily erase from your mind the variances used in generating the data. We want to test the null hypothesis that the normal distributions which generated  $x$  and  $y$  have equal variances. Now use `var.test` to find the  $p$ -value for the test for equal variances. (You may

need to go back and look at the documentation for `var.test` to see how to do this.)

(b) You should look up the F-test for equal variances. Give the formula and compute the F statistic and  $p$ -value directly. You can use R to compute sample means, sample variances and do arithmetic as needed, but you can't use `var.test` to do all the work. You will need the R function `pf` to compute probabilities for the F distribution

Should you reject the null-hypothesis at  $\alpha = 0.05$ ?

Note: here is how you get a two-sided  $p$ -value for this test.

If the  $F$ -statistic is greater than 1 then

$$p = 2 \times \text{the right tail probability}$$

If the  $F$ -statistic is less than 1 then

$$p = 2 \times \text{the left tail probability}$$

**Problem 6.** (20: 10,10 pts.) (**One-way ANOVA**)

Read the abstract of the following paper:

<https://www.sciencedirect.com/science/article/pii/S1090513813000226>

Barnaby J. Dixson, Robert C. Brooks. *The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities*. Evolution and Human Behavior, Volume 34, Issue 3, May 2013, Pages 236-241

Note that one of the authors may have a personal bias:

[https://www.researchgate.net/profile/Barnaby\\_Dixson](https://www.researchgate.net/profile/Barnaby_Dixson)

For this problem you will need the F-test for equal means from the reading for class 19. For the purposes of the problem, we made a slight simplification to the experimental protocol and data.

(a) The table below records the mean attractiveness rating for 351 heterosexual or bisexual women who rated the attractiveness of one male face of each type from 0 (very low) to 5 (very high). So we have four groups with 351 samples per group.

facial hair state	clean	5-day	10-day	full
sample mean	1.32	1.26	1.53	1.39
sample variance	0.56	0.80	0.93	0.82

Run a one-way ANOVA (F-test) at  $\alpha = 0.01$  for equal means “by hand”. That is, compute the  $F$ -statistic and corresponding  $p$  value using the data in the table, and decide whether to reject the null hypothesis.

(b) Following up, use 3 two-sample  $t$ -tests to investigate the hypothesis that the mean for 10-day is higher than the each of the others. For each test use significance level  $\alpha = 0.01$ . (The two-sample  $t$ -test is described in the reading for class 19.)

If exactly one of these tests resulted in a rejection, would it be appropriate to conclude that we have rejected the null hypothesis in part (a) at significance level  $\alpha = 0.01$ ?

MIT OpenCourseWare

<https://ocw.mit.edu>

18.05 Introduction to Probability and Statistics

Spring 2022

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.