

STA130H1S – Winter 2020

Week 4 Practice Problems - Sample Answers

N. Moon and L. Bolton

Instructions

How do I hand in these problems for the January 30 deadline?

Your complete .Rmd file that you create for *QUESTIONS 1 & 2* of these practice problems and the resulting pdf (i.e., the one you ‘Knit to PDF’ from your .Rmd file) must be uploaded into a Quercus assignment (link: <https://q.utoronto.ca/courses/138992/assignments/284428>) by 11:59PM, on January 30. Late problem sets or problems submitted another way (e.g., by email) are *not* accepted.

What should I bring to tutorial on January 31?

R output (e.g., output and explanations) for *QUESTIONS 1 & 2* only. You can either bring a hard copy or bring your laptop with the output.

Tutorial Grading

Tutorial grades will be assigned according to the following marking scheme.

	Mark
Completion of required problems (due on Quercus the day before your tutorial)	1
Attendance for the entire tutorial	1
In-class exercises	4
Total	6

Practice Problems

Question 1

A 1920s tea party

British statistician Ronald Fisher was at a tea party in the 1920s. One of the other guests was algae scientist Dr Muriel Bristol, who refused a cup of tea from Fisher because he put milk in BEFORE pouring the tea in. Bristol was convinced she could taste the difference, and much preferred the taste of tea where the milk was poured in afterwards. Fisher didn't think that there could be a difference and proposed they tested this.

The test was set up as follows: 8 cups of tea were made, 4 with milk in first and 4 with tea in first.

The result: Bristol correctly identified whether the tea or milk was poured first for all 8 cups.

Fisher, being a good Statistician, wondered if this happened just by chance (Bristol was just guessing, 50/50), or whether it seemed more likely that Bristol was not guessing.

Your turn

Consider running an experiment like this with your fellow students in STA130. You get a random sample of 50 STA130 students to each taste one British-style cup of tea and tell you whether they think the milk or tea was poured first. 31 students correctly state which was poured first. Go through the steps to test whether students are just guessing or not.

- What are appropriate null and alternative hypothesis to test the claim? Make sure you define the parameter in context.

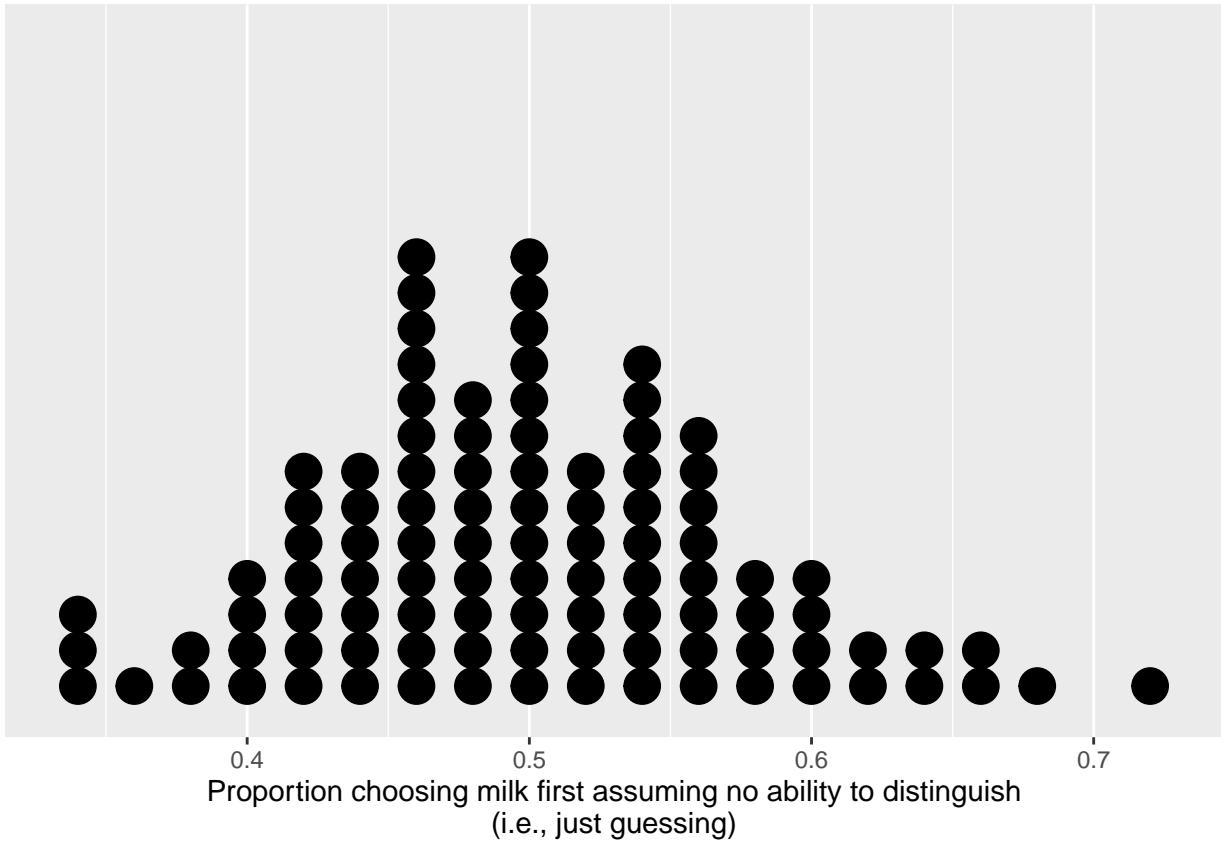
$$H_0 : p = 0.5$$

$$H_A : p \neq 0.5$$

where p is the proportion of STA130 students who can taste a difference between tea with the milk or tea poured in first.

- Assume you conduct a test of significance using simulation and get the following estimated sampling distribution of the test statistic assuming the null hypothesis is true. For simplicity, this distribution shows the results of only 100 simulations. There are 100 dots on the plot, one for each simulation. (In practice, 100 simulations is not sufficient to obtain a good estimate of the sampling distribution.)

```
ggplot(sim, aes(p_correct)) +  
  geom_dotplot() +  
  labs(x="Proportion choosing milk first assuming no ability to distinguish\n(i.e., just guessing)") +  
  scale_y_continuous(NULL, breaks = NULL) # get rid of strange y-axis label
```



- (i) What does each single dot in the plot represent?

Each dot represents the result for one simulation (i.e., the proportion of people who guess ‘milk first’ in one simulated sample of 50 people who are just guessing). For each simulation, 50 values are randomly generated that are equally likely to be correct or incorrect, since the null hypothesis is that people are equally likely to be right or wrong in identifying whether the milk or tea was poured first (this would be the case if they were really just guessing). The dots are placed at the proportion of the 50 guesses that are ‘milk first’ in the simulation. These values are generated under the assumption that the null hypothesis is true.

- (ii) Based on this plot, what is your estimate of the p-value?

```
# how far is the test statistic from null value of p?
test_stat <- 31/n_observations
distance <- abs(test_stat-0.5)
# to estimate the p-value you need to find the proportion of
# simulated values at least as far away from null value as the test statistic is
est_pval <- mean(simulated_stats >= 0.5 + distance) + mean(simulated_stats <= 0.5 - distance)
est_pval

## [1] 0.14
```

Our test statistic is $31/50 = 0.62$. To estimate the p-value based on this simulation, we need to find the proportion of dots (of the 100) that are at least as far away from the null value of p as 0.62 is (i.e., find the proportion of the 100 simulated values that are at least 0.68 (≥ 0.62), or are $0.5 - (0.62 - 0.5) = 0.38$ or less (≤ 0.38)). Based on the R code above, the estimated p-value is 0.14.

- (c) What conclusion can you make based on the p-value you calculated in part b(ii)?

Based on these results, there is no evidence against the hypothesis that STA130 students tasting the tea are just guessing whether the milk was put in first. At a 5% significance level, these results are consistent with students not being able to distinguish (correctly or incorrectly) the order in which milk has been added to British-style tea.

- (d) Suppose the analysis described in (b) is repeated but this time 1000 simulations are used to get a better estimate of the p-value, and the resulting p-value is 0.064. Do not conduct this simulation. What is an appropriate conclusion based on this p-value?

We would conclude that there is weak evidence against the hypothesis that people tasting the tea are just guessing.

You may enjoy *this article* that provides more details about the tea party and the experiment. Optional.

Question 2

Approximately 10% of the general population is left-handed. Suppose that the Department of Statistical Sciences (DoSS) is conducting a study to see if this percentage is the same among their undergraduate students (that is, all students in an undergraduate DoSS statistics program). This would help inform classroom renovations and bookings to ensure sufficient left-handed (and right-handed) seating. Suppose 500 students in statistics programmes are randomly selected and asked whether or not they are left-handed. Suppose that 61 of these 500 students respond that they are left-handed.

- (a) What are appropriate null and alternative hypothesis to test the claim? Make sure you define the parameter in context.

$$H_0 : p = 0.1$$

$$H_A : p \neq 0.1$$

where p is the proportion of all students in undergraduate DoSS programs at the University of Toronto who are left-handed.

- (b) Use the `sample()` function to simulate the number of left-handed students in a random sample of 500 DoSS students, assuming that the prevalence of left-handedness is the same among DoSS students as it is in the general population. How many left-handed students did you have in your simulated sample of 500 students? How does this simulated count compare to the results of the handedness study (i.e., that 67 of the 500 students sampled were left-handed)? How does it compare to the assumption that 10% of students are left-handed.

Note: the probabilities assigned to the values in the vector from which you're sampling using the `sample()` function are considered equal by default. For example, consider simulating flipping a coin 10 times:

```
sample(c("Head", "Tail"), size=10, replace=TRUE)

## [1] "Tail" "Head" "Tail" "Tail" "Tail" "Tail" "Head" "Tail" "Tail" "Tail"

# will do the same thing as:
sample(c("Head", "Tail"), size=10, prob=c(0.5, 0.5), replace=TRUE)

## [1] "Tail" "Tail" "Head" "Tail" "Head" "Tail" "Head" "Tail" "Head" "Tail"

# Even though the exact counts of "Head" and "Tail" differ each time you
# run this code, if you simulate enough coin flips (by increasing
# the value of 'size', you'll get approximately the same proportion
# of "Head" and "Tail" outcomes)

# To modify the code to make Tails much more likely than Heads,
# we could change the probs:
sample(c("Head", "Tail"), size=10, prob=c(0.2, 0.8), replace=TRUE)

## [1] "Tail" "Tail" "Tail" "Head" "Tail" "Tail" "Tail" "Head" "Tail" "Tail"
```

Set the random number seed to the last 2 digits of your student number (not your UTORID, your 9-10 digit student number) before carrying out your simulation. We set the seed so that the results won't change each time this code is run or knitted. If we didn't do this, your interpretations and conclusions may not be relevant to the new run of the code (or when you knit your Rmd file!).

```
set.seed(27) # assume my student number ends in 27

n<-500
simulated_sample<-sample(c("Left","Right"),size=n, prob=c(0.1, 0.9), replace=TRUE)

simulated_sample

## [1] "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [10] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Left"   "Right"
## [19] "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Left"   "Right"  "Right"
## [28] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [37] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [46] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [55] "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [64] "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [73] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Left"   "Right"
## [82] "Right"  "Right"  "Right"  "Left"   "Right"  "Left"   "Right"  "Right"  "Right"
## [91] "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Left"   "Right"
## [100] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [109] "Left"   "Left"   "Right"  "Left"   "Right"  "Right"  "Right"  "Left"   "Right"
## [118] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [127] "Left"   "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"
## [136] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Left"
## [145] "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [154] "Right"  "Right"  "Right"  "Right"  "Left"   "Left"   "Right"  "Left"   "Right"
## [163] "Right"  "Left"   "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Right"
## [172] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [181] "Right"  "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"
## [190] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [199] "Right"  "Left"   "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"
## [208] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [217] "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Left"   "Right"  "Right"
## [226] "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"
## [235] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [244] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [253] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [262] "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Left"   "Left"   "Right"
## [271] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [280] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [289] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [298] "Left"   "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [307] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [316] "Right"  "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"
## [325] "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Left"   "Right"  "Right"
## [334] "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [343] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [352] "Right"  "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [361] "Right"  "Left"   "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
## [370] "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"  "Right"
```

```

## [379] "Right" "Right" "Left" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [388] "Left" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [397] "Left" "Right" "Left" "Right" "Right" "Left" "Right" "Right" "Right" "Right"
## [406] "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Left"
## [415] "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [424] "Left" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [433] "Right" "Right" "Left" "Right" "Right" "Right" "Right" "Right" "Right" "Left"
## [442] "Right" "Right" "Right" "Left" "Right" "Left" "Right" "Right" "Right" "Right"
## [451] "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [460] "Left" "Right" "Right" "Right" "Left" "Right" "Right" "Right" "Right" "Left"
## [469] "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Left" "Right"
## [478] "Right" "Left" "Right" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [487] "Right" "Right" "Left" "Right" "Right" "Right" "Right" "Right" "Right" "Right"
## [496] "Right" "Left" "Right" "Right" "Right" "Right"

```

```

sum(simulated_sample=="Left")

```

```

## [1] 59

```

```

mean(simulated_sample=="Left")

```

```

## [1] 0.118

```

There were 59 left-handed students in my simulated sample of 500 students selected under the assumption that 10% of the population of DoSS students are left-handed. 11.8% (59/500x100) of the simulated sample of students were left-handed. This is 17 students more than the results of the study, but only 1.8% more than the assumption that 10% of students are left-handed. If we changed the random seed, or did not set one each time we used the `sample()` function, we would end up with a slightly different simulated count/percentage each time we ran this code.

- (c) Use R to estimate the sampling distribution of the test statistic under the assumption that the prevalence of left-handedness among DoSS students matches the general population. Use 1000 repetitions and set the seed to the last 2 digits of your student number. Generate the plot of this estimated sampling distribution and describe the distribution in a few sentences.

Use `labs(x="your title here")` to give the x-axis a better name than the default. You can split your axis title across different lines to make it easier to read by adding `\n` OR typing an “enter”/carriage return. See the code shown in Question 1b) for an example.

```

library(tidyverse)

set.seed(21) # if last 2 digits of student number are 21
repetitions <- 1000
simulated_stats <- rep(NA, repetitions) #initialize this vector to store
# simulated proportions of left-handed students

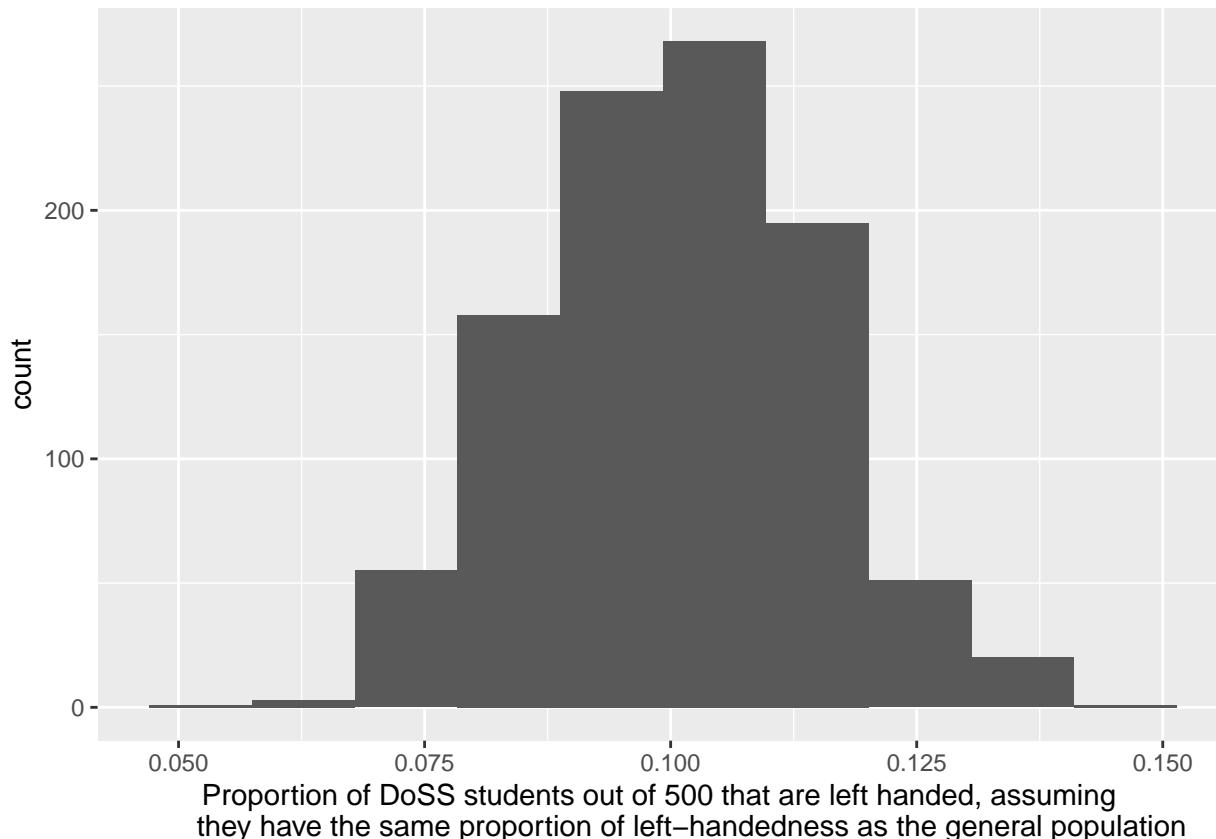
n_observations <- 500
for (i in 1:repetitions)
{
  new_sim <- sample(c("Left", "Right"), size=n_observations, prob=c(0.1, 0.9), replace=TRUE)
  sim_p <- sum(new_sim == "Left") / n_observations
}

```

```

    simulated_stats[i] <- sim_p
}
sim <- tibble(prop_Left = simulated_stats)
ggplot(sim, aes(prop_Left)) +
  geom_histogram(bins=10) +
  labs(x="Proportion of DoSS students out of 500 that are left handed, assuming
they have the same proportion of left-handedness as the general population")

```



This estimated sampling distribution of proportions of left-handed students in simulated samples of 500 students is unimodal and approximately symmetric and centred around 0.1. This is not surprising since we would expect that about 10% of students in samples of 500 students would be left-handed assuming students are just as likely to be left-handed as the general population. These simulated proportions of left-handed students range from around 0.05 to 0.15.

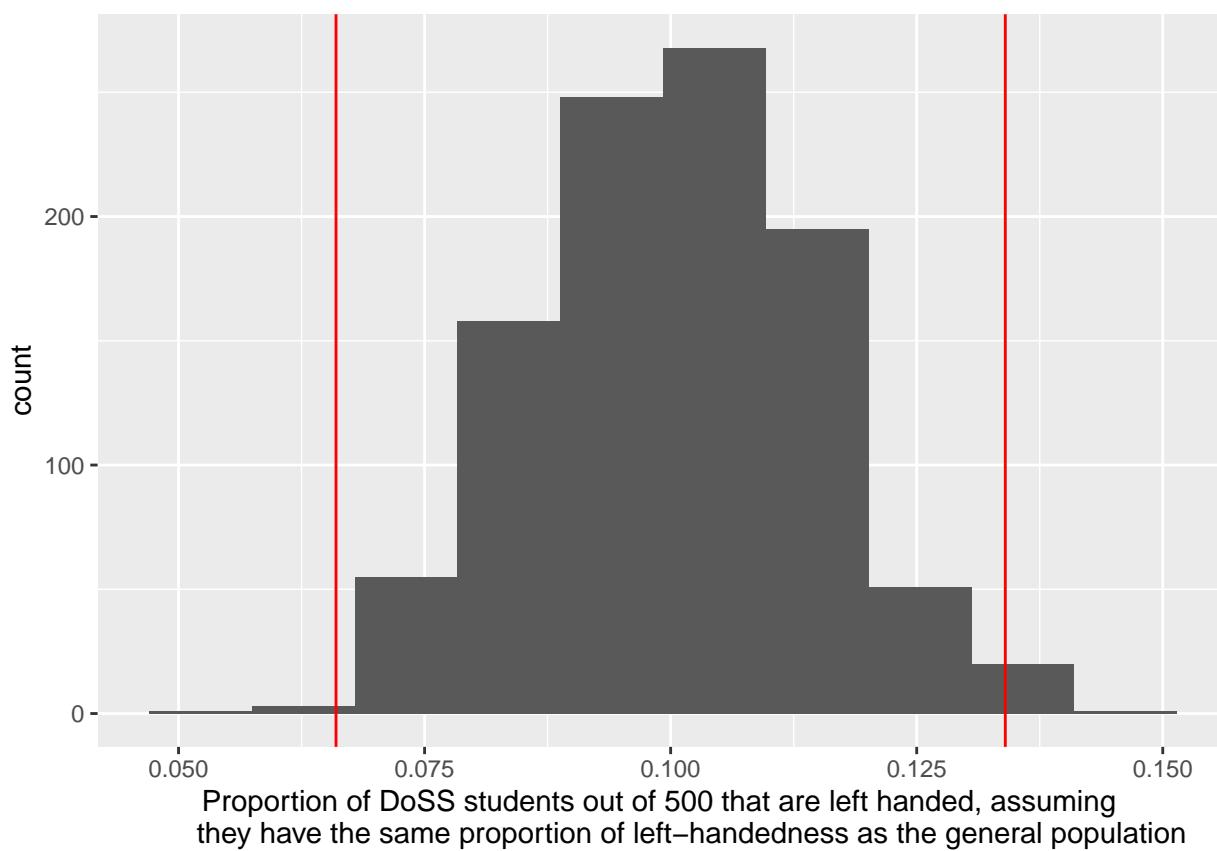
- (d) Use R to compute the p-value of this hypothesis test based on the sampling distribution that you estimated in part (c).

```

test_stat <- 67/n_observations
repetitions <- 1000

ggplot(sim, aes(prop_Left)) +
  geom_histogram(bins=10) +
  geom_vline(xintercept=test_stat, color="red") +
  geom_vline(xintercept=0.1-(test_stat-.1), color="red") +
  labs(x="Proportion of DoSS students out of 500 that are left handed, assuming
they have the same proportion of left-handedness as the general population")

```



```
sim %>%
  filter(prop_Left >= test_stat | prop_Left <= 0.1-(test_stat-0.1)) %>%
  summarise(p_value = n() / repetitions)
```

```
## # A tibble: 1 x 1
##   p_value
##   <dbl>
## 1 0.018
```

- (e) Which of the following statements is/are valid description(s) of the p-value you computed in (d):
- The probability that the proportion of DoSS students who are left-handed matches the general population.
 - The probability that the proportion of DoSS students who are left-handed does not match the general population.
 - The probability of obtaining a number of left-handed students in a sample of 500 students at least as extreme as the result in this study.
 - The probability of obtaining a number of left-handed students in a sample of 500 students at least as extreme as the result in this study, if the prevalence of left-handedness among all DoSS students matches the general population.

Only iv. is a valid interpretation. iii. is close but missing the important fact that the p-value is calculated by comparing the value of the test statistic to results generated *assuming that the null hypothesis is true*. i.

and *ii.* are incorrect because a p-value does not estimate the probability of H_0 or H_A being true. H_0 either is or isn't true; we just don't know. Rather, the p-value gives us a measure of how unusual (or not) our results look, assuming that the value of the parameter in our null hypothesis is true. When the p-value is small, the test statistic looks extreme (or unusual or surprising) under the assumption that H_0 is true, which makes us doubt the reasonableness of that assumption, and side with the alternative hypothesis. When the p-value is large, the test statistic does not look unusual (i.e., it seems consistent with H_0 since it is within the range of values you'd expect to see when H_0 is true). Therefore, large p-values don't provide evidence against H_0 .

- (f) Write a conclusion to this hypothesis test based on the p-value you computed in part (d).

The p-value I computed was 0.018. This means that there is moderate evidence against the hypothesis that the prevalence of left-handedness among students matches that of the general population (i.e., 10%).

Question 3

A Scottish woman noticed that her husband's scent changed. Six years later he was diagnosed with Parkinson's disease. His wife joined a Parkinson's charity and noticed that odour from other people. She mentioned this to researchers who decided to test her abilities. They recruited 6 people with Parkinson's disease and 6 people without the disease. Each of the recruits wore a t-shirt for a day, and the woman was asked to smell the t-shirts (in random order) and determine which shirts were worn by someone with Parkinson's disease. She was correct for 11 of the 12 t-shirts! You can read about this here.

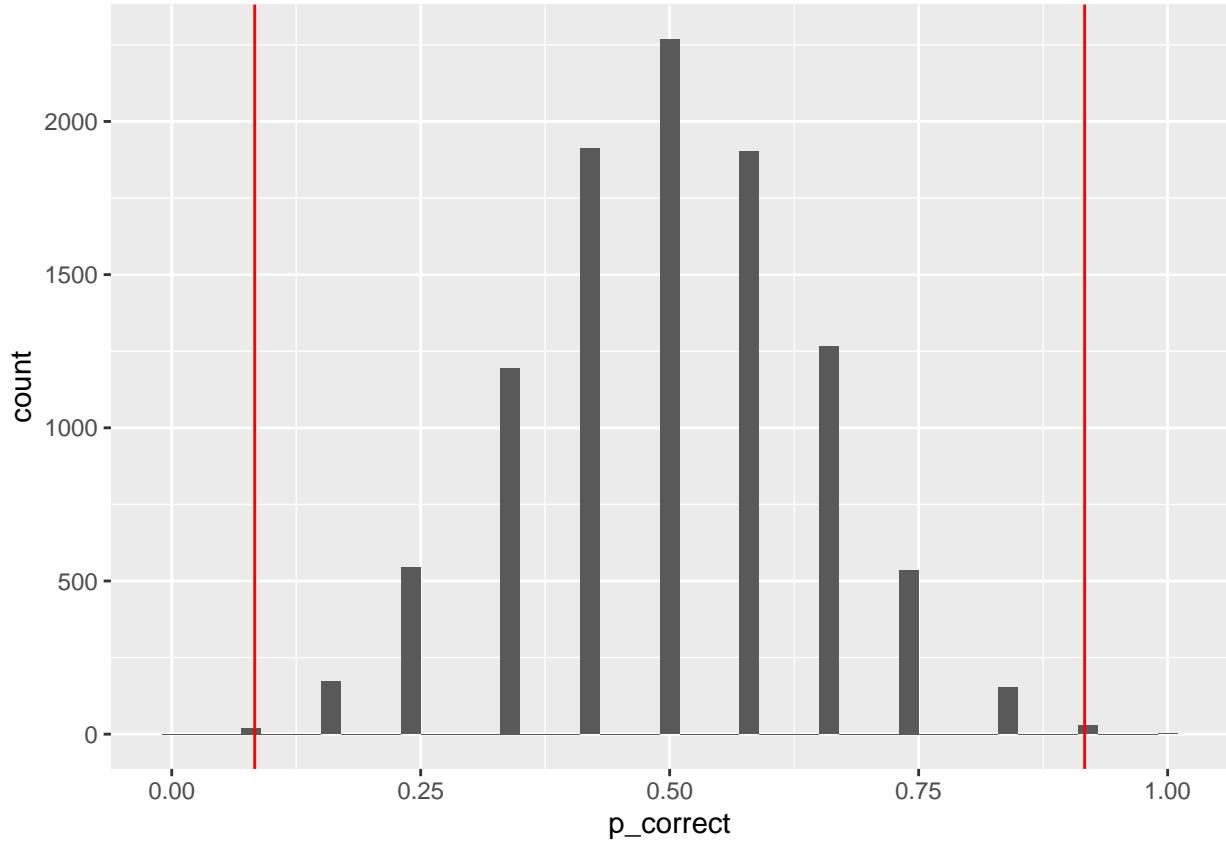
- (a) Without conducting a simulation, describe what you would expect the sampling distribution of the proportion of correct guesses about the 12 shirts to look like if someone was just guessing.

It is expected that about 50% of the guesses would be correct if someone was guessing at random. Having very large or very small percentages of correct guesses wouldn't happen as often. So, the sampling distribution of the proportion of correct guesses about the 12 shirts would likely be roughly symmetric, unimodal, and centred at around 0.5.

- (b) Carry out a test using simulation to determine if there is evidence that this woman has some ability to identify Parkinson's disease by smell, or if she was a lucky guesser.

Set the random number seed to the last two digits of your student number before carrying out your simulation. Use 10,000 repetitions. (This simulation is similar to the code in Question 1, but with many more simulated values of the test statistic under the null hypothesis. 10,000 is a lot of repetitions - more than is likely needed - but we'll do this many repetitions this time anyways.)

```
set.seed(19) # assume the last 2 digits of my student number are 19
repetitions <- 10000
simulated_stats <- rep(NA, repetitions)
n_observations <- 12
test_stat <- 11/12
for (i in 1:repetitions)
{
  new_sim <- sample(c("correct", "incorrect"), size=n_observations, replace=TRUE)
  sim_p <- sum(new_sim == "correct") / n_observations
  simulated_stats[i] <- sim_p
}
sim <- tibble(p_correct = simulated_stats)
ggplot(sim, aes(p_correct)) +
  geom_histogram(binwidth=0.02) +
  geom_vline(xintercept=11/12, color="red") +
  geom_vline(xintercept=0.5-(11/12-.5), color="red")
```



```
sim %>%
  filter(p_correct >= 11/12 | p_correct <= 0.5-(11/12-.5)) %>%
  summarise(p_value = n() / repetitions)
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1 0.0052
```

```
sum(sim$p_correct>=11/12)
```

```
## [1] 31
```

```
sum(sim$p_correct<=11/12)
```

```
## [1] 21
```

The hypotheses being tested are

$$H_0 : p = 0.5$$

$$H_A : p \neq 0.5$$

where p is the proportion of shirts that the woman correctly identified.

The test statistic is $\hat{p} = 11/12 = 0.9167$.

We simulate 10,000 values assuming that the woman is randomly guessing, so the probability she will be correct for each of the 12 shirts is 0.5. From these 10,000 simulations, 39 give a value greater than or equal to $11/12=0.9167$ and 31 are less than or equal to $1/12=0.0833$. So the proportion of observations in our simulation that are as extreme or more extreme than what we observed is $(39+31)/10000=0.007$. This is our estimate of the p-value.

We conclude that we have strong evidence against the hypothesis that this women is just guessing whether or not the person has Parkinson's. This suggest the woman is able to identify people with Parkinson's based on their scent.

Note: A larger number of repetitions results in a more precise estimate of the p-value. So you should see less variability among the p-values you and your friends obtained this question than if you had used a smaller number of repetitions.

- (c) The woman correctly identified all 6 people who had been diagnosed with Parkinson's but incorrectly identified one of the others as having Parkinson's. Eight months later he was was diagnosed with the disease. So the woman was actually correct 12 out of 12 times. Are you able to get the p-value for the test using this new data, without running a new simulation? What would you change from your answer to (b)? What wouldn't you change?

The hypotheses and number of observations (12) would be the same so there would be no need to re-run the simulation. To get the p-value, we would now look at how many of the simulations resulted in a proportion of correct values that are 12 or 0. Nothing else would change. In this case, the estimated p-value would be 0.0007 (see R code and output below for calculation of the p-value). So, there would be very strong evidence against the null hypothesis that she was just guessing on the 12 shirts. Although we cannot conclusively say this woman is able to detect Parkinson's disease, these results do provide compelling evidence suggesting she is indeed able to detect Parkinson's disease by scent.

```
sim %>%
  filter(p_correct >= 12/12 | p_correct <= 0.5-(12/12-.5)) %>%
  summarise(p_value = n() / repetitions)
```

```
## # A tibble: 1 x 1
##   p_value
##   <dbl>
## 1 0.0004
```

```
sum(sim$p_correct>=12/12)
```

```
## [1] 3
```

```
sum(sim$p_correct<=0/12)
```

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