Assignment 11

Your name and student ID

Today's date

- Due date: Tuesday, May 5 10:00pm.
- Late penalty: 50% late penalty if submitted within 24 hours of due date, no marks for assignments submitted thereafter.
- This assignment is marked out of 16. Marks are indicated for each question.
- Remember: autograder is meant as sanity check ONLY. It will not tell you if you have the correct answer. It will tell you if you are in the ball park of the answer so *CHECK YOUR WORK*
- Submission process: Follow the submission instructions on the final page. Make sure you do not remove any `\newpage` tags or rename this file, as this will break the submission.

Helpful hints:

- Every function you need to use was taught during lecture! So you may need to revisit the lecture code to help you along by opening the relevant files on Datahub. Alternatively, you may wish to view the code in the condensed PDFs posted on the course website. Good luck!
- Knit your file early and often to minimize knitting errors! If you copy and paste code for the slides, you are bound to get an error that is hard to diagnose. Typing out the code is the way to smooth knitting! We recommend knitting your file each time after you write a few sentences/add a new code chunk, so you can detect the source of knitting errors more easily. This will save you and the GSIs from frustration! **You must knit right before submitting**
- If your code runs off the page of the knitted PDF then you will LOSE POINTS! To avoid this, have a look at your knitted PDF and ensure all the code fits in the file. When it doesn't, go back to your .Rmd file and add spaces (new lines) using the return or enter key so that the code runs onto the next line.

Parental leave is often compensated to some degree, but the amount of compensation varies greatly. You read a research article that stated, "across people of all incomes, 47% of leave-takers received full pay during their leave, 16% received partial pay, and 37% received no pay."

After reading this, you wonder what the distribution of parental leave payment is for low income households. Suppose you conduct a survey of leave-takers within households earning less than $30,000 per year. You surveyed 225 people (selected in a random sample) and found that 51 received full pay, 33 received partial pay, and 141 received no pay.

1. [1 point] You would like to investigate whether the distribution of pay for households earning $< \$30{,}000$ is different from that of all income levels. Does this correspond to a chi-square test of independence or a chi-square test for goodness of fit?

TODO: YOUR ANSWER HERE

2. [1 point] What are the expected counts of leave-takers among households with incomes < $30,000? Assign each expected count to the appropriate variable. Make sure to remove the quotes. Round each number to 2 decimal places.

```
# put your answer here
full_pay <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
partial_pay <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
no_pay <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"



check_problem2()
```

```
## [1] "Checkpoint 1 Error: Full pay is out of right range"
## [1] "Checkpoint 2 Error: Partial pay is out of right range"
## [1] "Checkpoint 3 Error: no pay is out of right range"
##
## Problem 2
## Checkpoints Passed: 0
## Checkpoints Errored: 3
## 0% passed
## --------
## Test: FAILED
```

3. [1 point] State the null hypothesis under which the above expected counts were computed.

TODO: YOUR ANSWER HERE

4. [1 points] Compute the chi-square statistic. Round your answer to 2 decimal places.

```
# put your answer here
chi_sq_answer <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"



check_problem4()
```

```
## [1] "Checkpoint 1 Error: Answer should be numeric"
## [1] "Checkpoint 2 Error: Your answer is out of right range"
##
## Problem 4
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

5. [1 point] Uncomment which cell (i.e. term in summation) contributes the most to the statistic.

```
# UNCOMMENT THE CORRECT ANSWER

#largest_contribution <- "full pay"
#largest_contribution <- "partial pay"
#largest_contribution <- "no pay"



check_problem5()
```

```
## [1] "Checkpoint 1 Error: You should unmute one answer"
##
## Problem 5
## Checkpoints Passed: 0
## Checkpoints Errored: 1
## 0% passed
## --------
## Test: FAILED
```

6. [1 point] Compute the p-value for your test statistic. Round your answer to 2 decimal places.

```
# put your answer here
p_value <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"



check_problem6()
```

```
## [1] "Checkpoint 1 Error: Your answer is out of correct range"
## [1] "Checkpoint 2 Error: Your answer is not numeric"
##
## Problem 6
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

7. [1 point] Conclude whether you believe there is evidence against the null hypothesis in favor of the alternative hypothesis. Answer this by uncommenting the appropriate conclusion.

```
# UNCOMMENT THE CORRECT ANSWER

#conclusion <- "in favor of null"
#conclusion <- "against null"



check_problem7()
```

```
## [1] "Checkpoint 1 Error: You should unmute one answer"
##
## Problem 7
## Checkpoints Passed: 0
## Checkpoints Errored: 1
## 0% passed
## --------
## Test: FAILED
```

Human papillomavirus (HPV) is a very common STI that most sexually active persons will encounter during their lifetimes. While many people clear the virus, certain strands can lead to adverse health outcomes such as genital warts and cervical cancer.

Suppose that you selected a random sample from a population and collected these data on age and HPV status for the sample:

Age Group	HPV +	HPV -	Row total
14-19	160	492	652 (33.9%)
20-24	85	104	189 (9.8%)
25-29	48	126	174 (9.1%)
30-39	90	238	328 (17.1%)
40-49	82	242	324 (16.9%)
50-59	50	204	254 (13.2%)
Col total	515 (26.8%)	1406 (73.2%)	1921

8. [1 point] Which variable is explanatory and which is response? Uncomment the appropriate answer.

```
# UNCOMMENT THE CORRECT ANSWER

#variable_type <- c("explanatory: age group", "response: HPV status")
#variable_type <- c("explanatory: HPV status", "response: age group")



check_problem8()
```

```
## [1] "Checkpoint 1 Error: You should unmute one answer"
##
## Problem 8
## Checkpoints Passed: 0
## Checkpoints Errored: 1
## 0% passed
## --------
## Test: FAILED
```

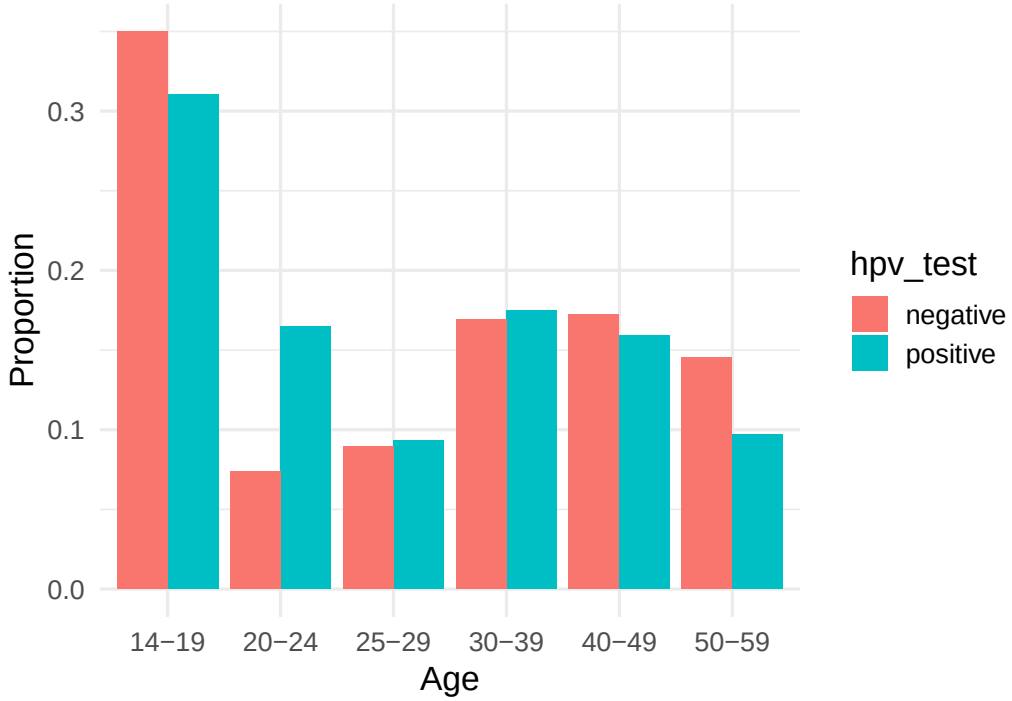
9. [2 points] Formulate null and alternative hypotheses using these data to test whether there is a relationship between age group and HPV status. State these hypotheses using the language or notation of conditional distributions.

TODO: YOUR ANSWER HERE

10. [1 point] Run the code below to examine the conditional distribution of age by HPV status. Based on this plot, which age group will contribute the most to the chi-square statistic? Explain why. (That is, can you tell based on this plot when the observed count will differ most from the expected count under the null hypothesis of no relationship between age group and HPV status?)

TODO: YOUR ANSWER HERE

```
## Warning: package 'tibble' was built under R version 3.5.2
## Warning: package 'tidyr' was built under R version 3.5.2
## Warning: package 'purrr' was built under R version 3.5.2
## Warning: package 'stringr' was built under R version 3.5.2
## Warning: package 'forcats' was built under R version 3.5.2
```



11. [2 points] Fill out the table of expected counts under the null hypothesis of no association between age group and HPV status. You don't need to show your work, but make sure you can calculate the expected counts by hand, using a calculator. Assign each appropriate cell/letter to the variable in the code. Round each number to 2 decimal places.

Expected counts:

Age Group	HPV +	HPV -
14-19	A	H
20-24	B	I
25-29	C	J
30-39	D	K
40-49	E	L
50-59	G	M

```
# put your answer here
A <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
B <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
C <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
D <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
E <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
G <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
H <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
I <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
J <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
K <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
L <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"
M <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"



check_problem11()
```

```
## [1] "Checkpoint 1 Error: A is not numeric"
## [1] "Checkpoint 2 Error: B is not numeric"
## [1] "Checkpoint 3 Error: C is not numeric"
## [1] "Checkpoint 4 Error: D is not numeric"
## [1] "Checkpoint 5 Error: E is not numeric"
## [1] "Checkpoint 6 Error: G is not numeric"
## [1] "Checkpoint 7 Error: H is not numeric"
## [1] "Checkpoint 8 Error: I is not numeric"
## [1] "Checkpoint 9 Error: J is not numeric"
## [1] "Checkpoint 10 Error: K is not numeric"
## [1] "Checkpoint 11 Error: L is not numeric"
## [1] "Checkpoint 12 Error: M is not numeric"
##
## Problem 11
## Checkpoints Passed: 0
## Checkpoints Errored: 12
## 0% passed
## --------
## Test: FAILED
```

12. [1 point] Calculate the test statistic. Round your answer to 2 decimal places.

```
# put your answer here
chi_sq_p12 <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"



check_problem12()
```

```
## [1] "Checkpoint 1 Error: Answer should be numeric"
## [1] "Checkpoint 2 Error: Your answer is out of right range"
##
## Problem 12
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

13. [1 point] Calculate the p-value for your test statistic. Round your answer to 2 decimal places.

```
# put your answer here
p_value_p13 <- "REPLACE WITH NUMBER ROUNDED TO 2 DECIMALS"



check_problem13()
```

```
## [1] "Checkpoint 1 Error: Answer should be numeric"
## [1] "Checkpoint 2 Error: Your answer is out of right range"
##
## Problem 13
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

14. [1 point] Assess whether there is evidence against the null in favor of the alternative. Answer this by uncommenting the appropriate conclusion.

```
# UNCOMMENT THE CORRECT ANSWER

#conclusion_p14 <- "in favor of null"
#conclusion_p14 <- "against null"



check_problem14()
```

```
## [1] "Checkpoint 1 Error: You should unmute one answer"
##
## Problem 14
## Checkpoints Passed: 0
## Checkpoints Errored: 1
## 0% passed
## --------
## Test: FAILED
```

Check your score

Click on the middle icon on the top right of this code chunk (with the downwards gray arrow and green bar) to run all your code in order. Then, run this chunk to check your score.

```
# Just run this chunk.
total_score()
```

```
##                      Test Points_Possible          Type
## Problem 1  NOT YET GRADED               1 free-response
## Problem 2          FAILED               1    autograded
## Problem 3  NOT YET GRADED               1 free-response
## Problem 4          FAILED               1    autograded
## Problem 5          FAILED               1    autograded
## Problem 6          FAILED               1    autograded
## Problem 7          FAILED               1    autograded
## Problem 8          FAILED               1    autograded
## Problem 9  NOT YET GRADED               2 free-response
## Problem 10 NOT YET GRADED               1 free-response
## Problem 11         FAILED               2    autograded
## Problem 12         FAILED               1    autograded
## Problem 13         FAILED               1    autograded
## Problem 14         FAILED               1    autograded
```

Submission

For assignments in this class, you'll be submitting using the **Terminal** tab in the pane below. In order for the submission to work properly, make sure that:

1. Any image files you add that are needed to knit the file are in the `src` folder and file paths are specified accordingly.
2. You **have not changed the file name** of the assignment.
3. The file is saved (the file name in the tab should be **black**, not red with an asterisk).
4. The file knits properly.

Once you have checked these items, you can proceed to submit your assignment.

1. Click on the **Terminal** tab in the pane below.
2. Copy-paste the following line of code into the terminal and press enter.

cd; cd ph142-sp20/hw/hw11; python3 turn_in.py

3. Follow the prompts to enter your Gradescope username and password. When entering your password, you won't see anything come up on the screen–don't worry! This is just for security purposes–just keep typing and hit enter.
4. If the submission is successful, you should see "Submission successful!" appear as output.
5. If the submission fails, try to diagnose the issue using the error messages–if you have problems, post on Piazza.

The late policy will be strictly enforced, **no matter the reason**, including submission issues, so be sure to submit early enough to have time to diagnose issues if problems arise.