

Problem Set 7

Due date: 6 November

Table of contents

Question 1	1
Part A	2
Part B	3
Part C	4
Part D	5
Question 2	5

Please upload your completed assignment to the ELMs course site (under the assignments menu). Remember to include an annotated script file for all work with R and show your math for all other problems (if applicable, or necessary). Please also upload your completed assignment to the Github repository that you have shared with us. *We should be able to run your script with no errors.*

Total points: 20

Question 1

Total points: 10

I hypothesize that the older someone is, the more they like watching golf. I conduct a survey based on an SRS of adults in America. The independent variable is a four-category measure of age: 21-35, 36-50, 51-65, 66-plus. The dependent variable is a three-category measure of how much they enjoy watching golf: not at all, somewhat, very much.

Age_Range		21-35	36-50	51-65	66-plus	All
not at all	N	5	1	2	3	11
	% row	45.5	9.1	18.2	27.3	100.0
somewhat	N	6	5	3	4	18
	% row	33.3	27.8	16.7	22.2	100.0
very much	N	1	3	3	4	11
	% row	9.1	27.3	27.3	36.4	100.0
All	N	12	9	8	11	40
	% row	30.0	22.5	20.0	27.5	100.0

Part A

Points: 4

Make a cross-tab, by hand, of the relationship between these two variables based on the following (completely made up) data. Make sure that you present both raw counts and the appropriate percentages to examine the relationship.

R1: 21-35, Somewhat	R11: 51-65, somewhat	R21: 66-plus, very much	R31: 66-plus, not at all
R2: 51-65, not at all	R12: 21-35, not at all	R22: 21-35, somewhat	R32: 21-35, not at all
R3: 66-plus, very much	R13: 36-50, somewhat	R23: 36-50, very much	R33: 66-plus, very much
R4: 21-35, not at all	R14: 21-35, very much	R24: 66-plus, somewhat	R34: 21-35, somewhat
R5: 36-50, somewhat	R15: 66-plus, not at all	R25: 21-35, not at all	R35: 51-65, very much
R6: 21-35, not at all	R16: 51-65, not at all	R26: 36-50, very much	R36: 66-plus, very much
R7: 36-50, not at all	R17: 36-50, somewhat	R27: 21-35, somewhat	R37: 21-35, somewhat
R8: 66-plus, somewhat	R18: 66-plus, somewhat	R28: 66-plus, somewhat	R38: 36-50, somewhat
R9: 36-50, very much	R19: 51-65, very much	R29: 51-65, somewhat	R39: 36-50, somewhat
R10: 51-65, very much	R20: 21-35, somewhat	R30: 51-65, very much	R40: 66-plus, not at all

Ans 1A:

Part B

Points: 2

Do the data support my hypothesis? Be sure to explain the nature of the relationship (or lack thereof, if relevant).

Ans 1B:

To determine if the data support the hypothesis that “the older someone is, the more they like watching golf,” we can look at the pattern of responses across the age groups, particularly focusing on the “very much” category as an indicator of liking golf a lot.

Here is the breakdown of the “very much” category across the different age ranges:

- 21-35: 1 respondent (9.1% of the “very much” responses)
- 36-50: 3 respondents (27.3% of the “very much” responses)
- 51-65: 3 respondents (27.3% of the “very much” responses)
- 66-plus: 4 respondents (36.4% of the “very much” responses)

Here’s the distribution for the “somewhat” category:

- 21-35: 6 respondents (33.3% of the “somewhat” responses)
- 36-50: 5 respondents (27.8% of the “somewhat” responses)
- 51-65: 3 respondents (16.7% of the “somewhat” responses)
- 66-plus: 4 respondents (22.6% of the “somewhat” responses)

In the “somewhat” category, we see that the 36-50 age group has the highest number of respondents, which might suggest a spike in interest in this middle age group.

When considering both “somewhat” and “very much” categories together, we can conclude that the data generally support the hypothesis that older individuals are more likely to enjoy watching golf, but the relationship is not strictly linear as there’s a noticeable interest in the 36-50 age range.

Finally, it is also important to consider the actual number of respondents in each age group, as percentages alone can sometimes be misleading. The 66-plus age group not only has the highest percentage of “very much” responses but also has the highest number of respondents overall, which strengthens the claim that older individuals are more inclined to enjoy watching golf according to the data provided.

We can use a chi square test to really determine the variable relationships.

Part C

Points: 2

Ans C2: Calculating the chi square value by hand

For “Not at all” interest:

- 21-35: $112 = (5 - 3.3)23.3 = (1.7)23.3 = 2.893.3 \ 0.876$, $112 = 3.3(5 - 3.3)2 = 3.3(1.7)2 = 3.32.89 \ 0.876$
- 36-50: $122 = (1 - 2.475)22.475 = (1.475)22.475 = 2.1762.475 \ 0.879$, $122 = 2.475(1 - 2.475)2 = 2.475(1.475)2 = 2.4752.176 \ 0.879$
- 51-65: $132 = (2 - 2.2)22.2 = (0.2)22.2 = 0.042.2 \ 0.018$, $132 = 2.2(2 - 2.2)2 = 2.2(0.2)2 = 2.20.04 \ 0.018$
- 66-plus: $142 = (3 - 3.025)23.025 = (-0.025)23.025 = 0.0006253.025 \ 0.0002$, $142 = 3.025(3 - 3.025)2 = 3.025(-0.025)2 = 3.0250.000625 \ 0.0002$

For “Somewhat” interest:

- 21-35: $212 = (6 - 5.4)25.4 = (0.6)25.4 = 0.365.4 \ 0.067$, $212 = 5.4(6 - 5.4)2 = 5.4(0.6)2 = 5.40.36 \ 0.067$
- 36-50: $222 = (5 - 4.05)24.05 = (0.95)24.05 = 0.90254.05 \ 0.223$, $222 = 4.05(5 - 4.05)2 = 4.05(0.95)2 = 4.050.9025 \ 0.223$
- 51-65: $232 = (3 - 3.6)23.6 = (-0.6)23.6 = 0.363.6 \ 0.1$, $232 = 3.6(3 - 3.6)2 = 3.6(-0.6)2 = 3.60.36 \ 0.1$
- 66-plus: $242 = (4 - 4.95)24.95 = (-0.95)24.95 = 0.90254.95 \ 0.182$, $242 = 4.95(4 - 4.95)2 = 4.95(-0.95)2 = 4.950.9025 \ 0.182$

For “Very much” interest:

- 21-35: $312 = (1 - 3.3)23.3 = (2.3)23.3 = 5.293.3 \ 1.603$, $312 = 3.3(1 - 3.3)2 = 3.3(2.3)2 = 3.35.29 \ 1.603$
- 36-50: $322 = (3 - 2.475)22.475 = (0.525)22.475 = 0.27562.475 \ 0.111$, $322 = 2.475(3 - 2.475)2 = 2.475(0.525)2 = 2.4750.2756 \ 0.111$
- 51-65: $332 = (3 - 2.2)22.2 = (0.8)22.2 = 0.642.2 \ 0.291$, $332 = 2.2(3 - 2.2)2 = 2.2(0.8)2 = 2.20.64 \ 0.291$
- 66-plus: $342 = (4 - 3.025)23.025 = (0.975)23.025 = 0.95063.025 \ 0.314$, $342 = 3.025(4 - 3.025)2 = 3.025(0.975)2 = 3.0250.9506 \ 0.314$

Now we add all these individual chi-square values to get the total chi-square statistic:

Total $\chi^2 = 0.876 + 0.879 + 0.018 + 0.0002 + 0.067 + 0.223 + 0.1 + 0.182 + 1.603 + 0.111 + 0.291 + 0.314$

$\chi^2 = 0.876 + 0.879 + 0.018 + 0.0002 + 0.067 + 0.223 + 0.1 + 0.182 + 1.603 + 0.111 + 0.291 + 0.314$

Total $\chi^2 = 4.5644$

This is our chi-square statistic. To determine whether this is statistically significant, we would compare this to the critical value from the chi-square distribution with 6 degrees of freedom (since we have 3 degrees of freedom from interest categories minus 1,

Part D

Points: 2

Using the chi-square statistic that you computed in question 1(c), can you reject the null hypothesis of no relationship between these two variables with 95% confidence? Why, or why not?

Ans D:

Using the previously calculated chi square value:

Since our calculated χ^2 value of approximately 4.5644 is less than the critical value of 12.59, we do not reject the null hypothesis. There is not enough evidence to suggest a statistically significant relationship between age and the level of interest in watching golf at the 95% confidence level.

Question 2

Points: 10

I hypothesize that people who identify as more liberal are less likely to see gun control as an important issue than people who identify as more conservative. Using the `nes` dataset from the `poliscidata` package, create a cross-tab examining the effect of someone's self-identification as liberal or conservative, measured using the variable `libcon3` variable, and their indication of how important gun access is to them as an issue, measured using the `gun_importance3` variable.

Do the data support my hypothesis? Be sure to explain the nature of the relationship (or lack thereof, if relevant). Report a chi-square statistic (it is fine to have R generate the statistic and p-value). What does the chi-square statistic tell you about whether you can reject the null hypothesis of no relationship between these two variables with 95% confidence? Why, or why not?

Pearson's Chi-squared test

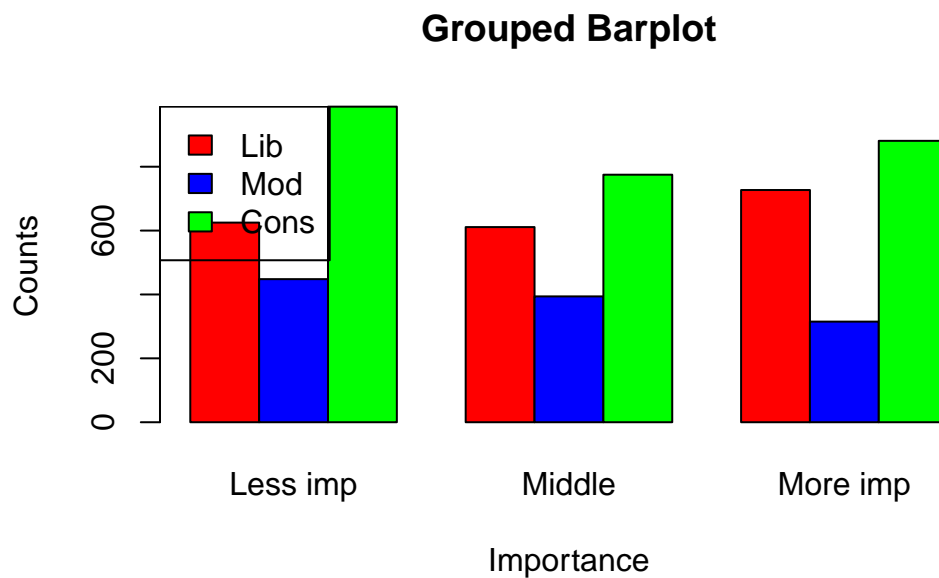
data: table

X-squared = 40.105, df = 4, p-value = 4.118e-08

Based on the p value , we can reject the null hypothesis. Note that the degree of freedom is 4.

	Less imp	Middle	More imp
Lib	625	611	727
Mod	448	394	315
Cons	988	775	881

Based on the table itself, we can see that the contingency matrix itself does not support such an assumption as there is no direct pattern between liberal and gun violence



Not much difference between values as visualized.