# **SAMPLE** RStudio lab assignment: chi-square



[How to submit an assignment]

[Code templates in R]

[Interpretation templates]

[Variables in GSS]

## **Assignment instructions**

In this lab assignment, you are asked to complete three questions.

Paste the specific codes you used for each question. When I run the codes you pasted below, it should generate the tables/analyses on my end.

The video below shows how to complete this assignment with different variables from start to finish. Watch the video and read the instructions together. You need to attend the lecture or watch the lecture video to understand this assignment.

[SAMPLE ASSIGNMENT VIDEO]

1)

1a) First, create a frequency table for "*fairearn*." Paste (5 points) and interpret the table (5 points).

1. Variable name (-1 point):	fairearn							
2. What it measures? (-1 point):	perceived pay fairness							
3. Full wording of the question (-1 point):	How fair is what you earn on your job in comparison to others doing the same type of work you do?							
4. Response set (-1 point):	(1: much less than they deserve; 2: somewhat less than they deserve; 3: about as much as they deserve; 4: somewhat more than they deserve; 5: much more than they deserve)							
5. Code (-5):	frq(gss\$fairearn, out = "v")							
6. Table:	how fair is what r earn on the job (x) <numeric>           val         label         frq         raw.prc         valid.prc         cum.prc           1         much less than you deserve         234         6.60         12.26         12.26           2         somewhat less than you deserve         610         17.21         31.97         44.23           3         about as much as you deserve         909         25.65         47.64         91.88           4         somewhat more than you deserve         127         3.58         6.66         98.53           5         much more than you deserve         28         0.79         1.47         100.00           NA         NA         1636         46.16         NA         NA           total N=3544 · valid N=1908 · <math>\bar{x}</math>=2.53 · σ=0.85</numeric>							
7. Interpretation:	The perceived pay fairness variable shows that 12.26% of the respondents think they earn much less than they deserve; 31.97% of the respondents think they somewhat less than they deserve; 47.64% of the respondents think they earn about as much as they deserve; 6.66% of the respondents think they earn somewhat more than they deserve; 1.47% of the respondents think they earn much more than they deserve.							

- 1b) Second, recode "fairearn" into two categories. The new variable name will be "fairearnnew". (10 points).
  - 1 and 2 will be 1 [less than they deserve]
  - 3 and 4 and 5 will be 2 [about as much and more than they deserve]

# Code (-5): gss\$fairearnnew <- rec(gss\$fairearn, rec = "1,2=1 [less than they deserve]; 3,4,5=2 [about as much and more than they deserve]", append = FALSE)

1c) Third, create a frequency table for "fairearnnew" Paste (5 points) and interpret the table (5 points).

1. Code (-5):	frq(gss\$fairearnnew, out = "v")
2. Table:	how fair is what r earn on the job (x) <numeric>vallabelfrqraw.prcvalid.prccum.prc1less than they deserve84423.8144.2344.232about as much and more than they deserve106430.0255.77100.00NANA163646.16NANAtotal N=3544 · valid N=1908 · <math>\bar{x}</math>=1.56 · <math>\sigma</math>=0.50</numeric>
3. Interpretation:	The perceived pay fairness variable shows that 44.23% of the respondents think they earn less than they deserve; 55.77% of the respondents think they earn about as much and more than they deserve.

1d) Finally, run chi-square analysis for "born" (independent variable) and "fairearnnew" (dependent variable). Paste (5 points) and interpret the chi-square table (10 points).

1. Code (-5):	sjt.xtab(gss\$born, gss\$fairearnnew, show.row.prc = TRUE)

2. Table:	how fair is what r earn on the job					
	was r born in this country	less than they deserve	about as much and more than they deserve	Total		
	yes	735 44.1 %	932 55.9 %	1667 100 %		
	no	107 45 %	131 55 %	238 100 %		
	Total	842 44.2 %	1063 55.8 %	1905 100 %		
		$\chi^2 = 0.033$	3 · df=1 · φ=0.006 ·	p=0.855		
3. Interpretation:	perceived pay fai 0.05. We can cor	rness since th clude respon- were not born	s has NO effect on the p value is HIGH dents who were bo to in this country ha	rn in this		

# 2)

# 2a) First, create a descriptive table for "wwwhr". Paste (5 points) and interpret the table (5 points).

1. Variable name (-1 point):	wwwhr
2. What it measures? (-1 point):	internet use in hours
3. Full wording of the question (-1 point):	Not counting email, how many hours per week do you use the Web?

4. Response set	no response sets for continuous variables						
5. Code (-5):	descr(gss\$wwwhr, out = "v", show = "short")						
6. Table:	Basic descriptive statistics						
	Variable N Missings (%) Mean SD						
	dd 2239 36.82 15.51 19.48						
7. Interpretation:	The internet use in hours variable shows that the average hours of the respondents using the web is 15.51, with standard deviation 19.48.						

2b) Second, recode "wwwhr" into three categories. The new variable name will be "wwwhrnew". (10 points).

- 0 to 7 will be 1 [low level of internet use]
- 8 to 35 will be 2 [middle level of internet use]
- 36 to 200 will be 3 [high level of internet use]

#### 

2c) Third, create a frequency table for "wwwhrnew". Paste (5 points) and interpret the table (10 points).

1. Code (-5):	frq(gss\$wwwhrnew, out = "v")

2. Table:	www hours per week (x) < numeric>						
	val	label	frq	raw.prc	valid.prc	cum.prc	
	1	low level of internet use	1022	28.84	45.65	45.65	
	2	middle level of internet use	954	26.92	42.61	88.25	
	3	high level of internet use	263	7.42	11.75	100.00	
	NA	NA	1305	36.82	NA	NA	
		total N=	3344 · 1	valid N=2.	$239 \cdot \bar{x} = 1.6$	δ · σ=0.68	
3. Interpretation:	the re	nternet use in hours va espondents has a low le espondents has a midd % of the respondents l	evel c	of internation	et use; 4 ernet use	2.61% of e;	

2d) Now, create a frequency table for "*race*". Paste (5 points) and interpret the table (5 points).

1. Variable name (-1 point):	race
2. What it measures? (-1 point):	respondents' race
3. Full wording of the question (-1 point):	What's your race?
4. Response set (-1 point):	(1: white; 2: black; 3: other)
5. Code (-5):	frq(gss\$race, out = "v")

6. Table:	race of respondent (x) < numeric>						
	val	label	frq	raw.prc	valid.prc	cum.prc	
	1	white	2514	70.94	72.01	72.01	
	2	black	565	15.94	16.18	88.20	
	3	other	412	11.63	11.80	100.00	
	NA	NA	53	1.50	NA	NA	
		otal N=.	3544 · 1	alid N=3	$491 \cdot \bar{x} = 1.4$	$0 \cdot \sigma = 0.69$	
7. Interpretation:	respond	lents are	white;	16.18% of	nows that 72 the respon ents are oth		

2e) Finally, run chi-square analysis for "wwwhrnew" (independent variable) and "race" (dependent variable). Paste (5 points) and interpret the chi-square table (10 points).

sjt.xtab(gss\$wwwhrnew, gss\$race, show.row.prc = TRU						
	race	of respon	ıdent			
www hours per week	white	black	other	Total		
low level of	692	187	124	1003		
internet use	69 %	18.6 %	12.4 %	100 %		
middle level of	733	118	94	945		
internet use	77.6 %	12.5 %	9.9 %	100 %		
high level of	181	49	29	259		
internet use	69.9 %	18.9 %	11.2 %	100 %		
T-4-1	1606	354	247	2207		
1 otai	72.8 %	16 %	11.2 %	100 %		
	low level of internet use middle level of internet use high level of	race www hours per week  low level of 692 internet use 69 %  middle level of 733 internet use 77.6 %  high level of 181 internet use 69.9 %  Total 1606	race of response           www hours per week         white         black           low level of internet use         692         187           internet use         69%         18.6%           middle level of internet use         77.6%         12.5%           high level of internet use         181         49           internet use         69.9%         18.9%           Total         1606         354	race of respondent           www hours per week         white         black         other           low level of internet use         692         187         124           internet use         69%         18.6%         12.4%           middle level of internet use         77.6%         12.5%         9.9%           high level of internet use         181         49         29           internet use         69.9%         18.9%         11.2%           Total         1606         354         247		

### 3. Interpretation:

Respondents' race has an effect on internet use in hours since the p value is LESS than 0.05. We can conclude that white respondents, black respondents, and other race respondents 40-59 age group have substantially different level of internet use.