

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

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):	<pre>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)</pre>
-------------------	--

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

1. Code (-5):	<pre>frq(gss\$fairearnnew, out = "v")</pre>																								
2. Table:	how fair is what r earn on the job (x) <numeric> <table><tr><th>val</th><th>label</th><th>frq</th><th>raw.prc</th><th>valid.prc</th><th>cum.prc</th></tr><tr><td>1</td><td>less than they deserve</td><td>844</td><td>23.81</td><td>44.23</td><td>44.23</td></tr><tr><td>2</td><td>about as much and more than they deserve</td><td>1064</td><td>30.02</td><td>55.77</td><td>100.00</td></tr><tr><td>NA</td><td>NA</td><td>1636</td><td>46.16</td><td>NA</td><td>NA</td></tr></table> <i>total N=3544 · valid N=1908 · $\bar{x}$=1.56 · σ=0.50</i>	val	label	frq	raw.prc	valid.prc	cum.prc	1	less than they deserve	844	23.81	44.23	44.23	2	about as much and more than they deserve	1064	30.02	55.77	100.00	NA	NA	1636	46.16	NA	NA
val	label	frq	raw.prc	valid.prc	cum.prc																				
1	less than they deserve	844	23.81	44.23	44.23																				
2	about as much and more than they deserve	1064	30.02	55.77	100.00																				
NA	NA	1636	46.16	NA	NA																				
3. Interpretation:	<i>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.</i>																								

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):	<pre>sjt.xtab(gss\$born, gss\$fairearnnew, show.row.prc = TRUE)</pre>
----------------------	---

2. Table:

was r born in this country	how fair is what r earn on the job		Total
	less than they deserve	about as much and more than they deserve	
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 %

$\chi^2=0.033 \cdot df=1 \cdot \phi=0.006 \cdot p=0.855$

3. Interpretation:

Respondents' immigrant status has NO effect on perceived pay fairness since the p value is HIGHER than 0.05. We can conclude respondents who were born in this country and who were not born in this country have similar perception of pay fairness.

2)

2a) First, create a descriptive table for “**wwwhr**”. Paste **(5 points)** and interpret the table **(5 points)**.

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

4. Response set	<i>no response sets for continuous variables</i>															
5. Code (-5):	<i>descr(gss\$wwwahr, out = "v", show = "short")</i>															
6. Table:	<table><tr><th colspan="5">Basic descriptive statistics</th></tr><tr><th>Variable</th><th>N</th><th>Missings (%)</th><th>Mean</th><th>SD</th></tr><tr><td>dd</td><td>2239</td><td>36.82</td><td>15.51</td><td>19.48</td></tr></table>	Basic descriptive statistics					Variable	N	Missings (%)	Mean	SD	dd	2239	36.82	15.51	19.48
Basic descriptive statistics																
Variable	N	Missings (%)	Mean	SD												
dd	2239	36.82	15.51	19.48												
7. Interpretation:	<i>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.</i>															

2b) Second, recode “**wwwahr**” into three categories. The new variable name will be “**wwwahrnew**”. **(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]

Code (-5):	<pre>gss\$wwwahrnew <- rec(gss\$wwwahr, rec = "0:7=1 [low level of internet use]; 8:35=2 [middle level of internet use]; 36:200=3 [high level of internet use]", append = FALSE)</pre>
-------------------	---

2c) Third, create a frequency table for “**wwwahrnew**”. Paste **(5 points)** and interpret the table **(10 points)**.

1. Code (-5):	<code>frq(gss\$wwwahrnew, out = "v")</code>
----------------------	---

2. Table:

www hours per week (x) <numeric>

val	label	freq	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=3544 · valid N=2239 · $\bar{x}=1.66$ · $\sigma=0.68$

3. Interpretation:

The internet use in hours variable shows that 45.65% of the respondents has a low level of internet use; 42.61% of the respondents has a middle level of internet use; 11.75% of the respondents has a high level of internet use.

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

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

6. Table:

race of respondent (x) <numeric>					
val	label	freq	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

total N=3544 · valid N=3491 · $\bar{x}=1.40$ · $\sigma=0.69$

7. Interpretation:

The respondents' race variable shows that 72.01% of the respondents are white; 16.18% of the respondents are black, and 11.80% of the respondents are other race.

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)**.

1. Code (-5):

sjt.xtab(gss\$wwwhrnew, gss\$race, show.row.prc = TRUE)

2. Table:

www hours per week	race of respondent			Total
	white	black	other	
low level of internet use	692 69 %	187 18.6 %	124 12.4 %	1003 100 %
middle level of internet use	733 77.6 %	118 12.5 %	94 9.9 %	945 100 %
high level of internet use	181 69.9 %	49 18.9 %	29 11.2 %	259 100 %
Total	1606 72.8 %	354 16 %	247 11.2 %	2207 100 %

$\chi^2=20.804 \cdot df=4 \cdot \text{Cramer's } V=0.069 \cdot p=0.000$

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.