Handout #8: HW
Points: 25

Consider a study done by Timonen et al. that investigating the risk of suicide and it possible relationship with income level. This study was done in Finland because it has one of the highest death rates from suicide and treatment methods are similar across income levels (i.e. most of the population is treated similarly in public hospitals, regardless of socioeconomic status). Consider the following data from their study.

Violent* vs. Non Violent Across Occupations

Occupation	Violent Suicide	Count
Senior Position	Yes	57
Senior Position	No	30
Self Employed	Yes	33
Self Employed	No	10
Unskilled / Unemployed	Yes	137
Unskilled / Unemployed	No	51
Student	Yes	86
Student	No	12
Retired	Yes	267
Retired	No	104

*Violent Includes: Hanging, drowning, shooting, wrist cutting, jumping from a height

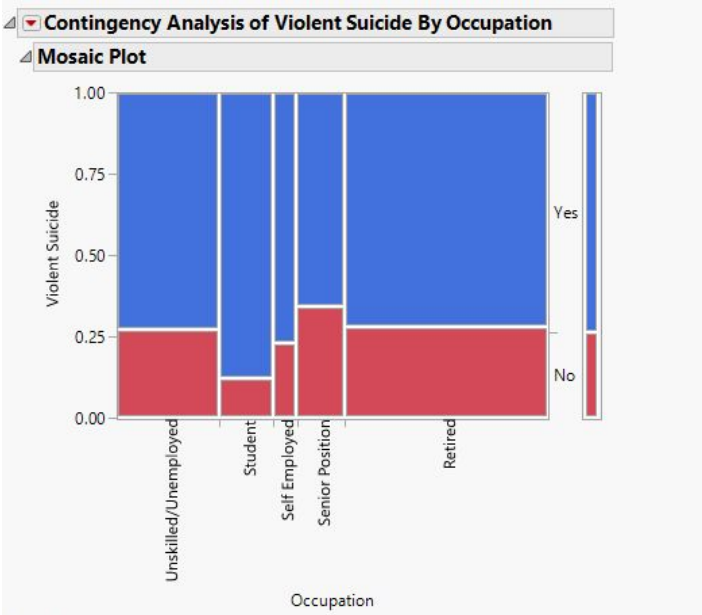
Source: Timonen, M., Viilo, K., Hakko, H., Väisänen, E., Räsänen, P., and Särkioja, T. (2001). "*Risk of suicide related to income level in mental illness: Psychiatric disorders are more severe among suicide victims of higher occupational level*." British Medical Journal, Vol. 323 (7306) : 232.

Consider the relationship between Violent/Non Violent Suicide and a person Occupation.

1. Enter this data into JMP and obtain a mosaic plot and its associated contingency table.

I included both column and row percentages because I thought they would both be useful and interesting. This data brings me back to my gerontology course where suicide of elderly people was a really big topic that no one wants to talk about. We spent a good bit of time on the topic and even compared previous occupation research as a significant variable due to income being speculated as a major influence on suicide in retired people, since retirment income is based largely on income during employment. I would be very interested to see what categories the retired individuals would've been classified as prior to retirement, to see if this could be used to further break apart into patterns (assuming there are any patterns). Eg. Is a retired senior programmer more likely to commit violent suicide over someone who was self employed prior to retirement. Obviously this would require more data.

Handout #8: HW
Points: 25



Freq: Count

Contingency Table

		Violent Suicide		
	Count	No	Yes	Total
	Col %			
	Row %			
Occupation	Unskilled/Unemployed	51	137	188
		24.64	23.62	
		27.13	72.87	
	Student	12	86	98
		5.80	14.83	
		12.24	87.76	
	Self Employed	10	33	43
		4.83	5.69	
		23.26	76.74	
	Senior Position	30	57	87
		14.49	9.83	
		34.48	65.52	
	Retired	104	267	371
		50.24	46.03	
		28.03	71.97	
	Total	207	580	787

Tests

N	DF	-LogLike	RSquare (U)
787	4	7.6777937	0.0169

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	15.356	0.0040*
Pearson	13.839	0.0078*

Describe the patterns present in the mosaic plot. For example, which Occupation has the largest proportion of violent suicides? Lowest proportion? Etc. (5 pts)

The Mosaic plot makes it obvious that students have the highest violent suicide percentage in this dataset, with those occupying senior positions having the smallest percentage. From the other direction we can see that the retired group has the most suicides overall, inline with what I mentioned above, but I am unsure if this a proportional representation or not. What I mean by that is there seems to be the fewest total suicides in the self employed category, but I'd guess that self employment is a smaller data group to begin with, so these widths wouldn't be saying much in comparison to each other unless they

are proportional and the dataset was representative. I think I need to read more on mosaic plots to make sure I'm reading that part of it correctly.

2. Conduct a Chi-Square Test of Independence to answer the following research question.

   Research Question: Is there an association between whether or not a suicide is violent in nature and one's occupation?

   You can see the relevant JMP output for this above too.

Tests

N	DF	-LogLike	RSquare (U)
787	4	7.6777937	0.0169

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	15.356	0.0040*
Pearson	13.839	0.0078*

   What is the conclusion for your test? Write this conclusion in laymen's terms. (3 pts)
   There is enough statistical evidence to say that the level of violence in a suicide depends on occupation (p = .0078). .0078 < .05

3. From the red drop-down menu on the Continency Table, select Expected. Show the math for how the Expected quantity is computed. (You only need to do this for one cell). (3 pts)

Freq: Count

Contingency Table

		Violent Suicide		
	Count Expected	No	Yes	Total
Occupation	Unskilled/Unemployed	51 49.4485	137 138.551	188
	Student	12 25.7764	86 72.2236	98
	Self Employed	10 11.31	33 31.69	43
	Senior Position	30 22.8831	57 64.1169	87
	Retired	104 97.582	267 273.418	371
	Total	207	580	787

   row_total * column_total / total_total = expected cell value
   188 * 207 / 787 = 49.4485 (top left cell non violent, unskilled/unemployed)

4. From the red drop-down menu on the Continency Table, select Cell Chi Square. Which cells contribute to the overall Chi Square value the most? Discuss. (3 pts)

Occupation	Violent Suicide		
Count Cell Chi^2	No	Yes	Total
Unskilled/Unemployed	51 0.0487	137 0.0174	188
Student	12 7.3629	86 2.6278	98
Self Employed	10 0.1517	33 0.0542	43
Senior Position	30 2.2134	57 0.7900	87
Retired	104 0.4221	267 0.1507	371
Total	207	580	787

Non-violent student suicides seems to have the largest cell chi squared value giving that cell the largest contribution. If you look at the expected value in the table from the last question compared to the actual count, this is no surprise because they are so different. We can also see that the violent student deaths come in as the second biggest chi contribution. This also doesn't surprise me since the student suicides were the first thing to stand out to me, as I mentioned above. What did surprise me is that the non-violent student suicides have a greater statisitcal impact than the violent ones. It's really just the flipside of the same coin, and since the non-violents are smaller, making differences more proportionally significant, it makes sense, but not what I would've guessed from the get go.

5. There is likely to be a larger difference in age for Occupation = Student compared the other Occupations provided here. Rerun the statistical test without Occupation = Student. Does removing this group change the conclusion of the research question? Discuss. (4 pts)

Taking students, the row containing the two largest cell chi squared values, out of the dataset intantly makes the data look more uniform. I won't bother pasting the mosaic here cause it's exactly what one would expect, though I can provide it if requested. It drastically changes the chi squared result too. Without the student group considered, we do not have enough statistical evidence to say that the level of violence in a suicide depends on occupation (p = .5090). We can see this in how the self employed non-violent group is dropped below the proportional average by almost as much as the senior position non-violent group is above that same line. They essentially cancel each other out, leaving our p value really close to .5.

Research Question: Is there an association between whether or not a suicide is violent in nature and one's occupation?

Consider another dataset provided by Timonen et al. Conduct a Chi-Square Test of Independence on this data.

Reason for Admission into Treatment
Across Occupations

Occupation	Admission Reason	Count
Senior Position	Psychosis	15
Senior Position	Drug/Alcohol	5
Self Employed	Psychosis	3
Self Employed	Drug/Alcohol	4
Unskilled / Unemployed	Psychosis	33
Unskilled / Unemployed	Drug/Alcohol	20
Student	Psychosis	7
Student	Drug/Alcohol	5
Retired	Psychosis	119
Retired	Drug/Alcohol	41

6. Write out the appropriate research question for your analysis. (3 pts)

   Research Question: Is there an association between whether a patient's admission into treatment is drug/alcohol or psychosis related and one's occupation?

7. Write a conclusion for your research question, in laymen's language. (4 pts)
   There is not enough statistical evidence to say that there is an association between whether a patient's admission into treatment is drug/alcohol or psychosis related and one's occupation ($p = .1671$). Some of these categories have very small amounts of data though which makes me personally more suspect of the results. More data is always a good thing, but where do we draw the line between enough and not enough data?