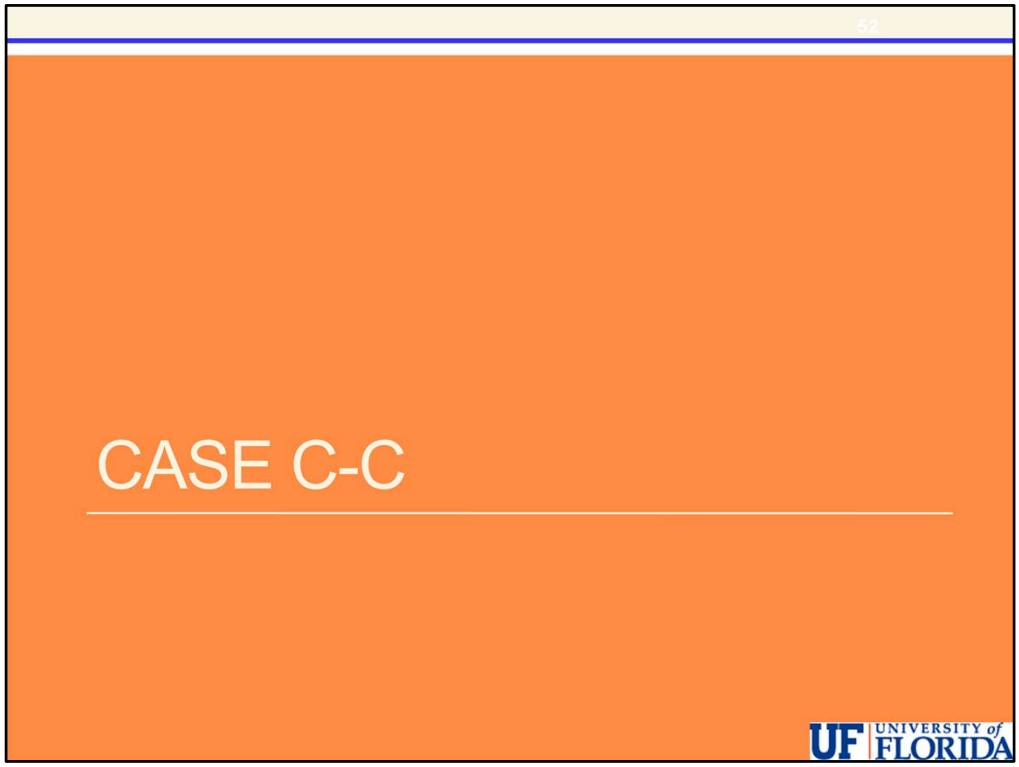


Original Data: http://bolt.mph.ufl.edu/2012/12/23/learn-by-doing-case-c-c-software/

When we discussed exploratory data analysis for Case C-C, we used a dataset based on a 1999 study at the University of Pennsylvania and Children's Hospital of Philadelphia, in which parents were surveyed about the lighting conditions under which their children slept between birth and age 2 (lamp, night-light, or no light) and whether or not their children developed nearsightedness (myopia). The purpose of the study was to explore the effect of a young child's nighttime exposure to light on later nearsightedness.

Notice this is an observational study which does not control for any other possible lurking variables.

Data: Nightlight

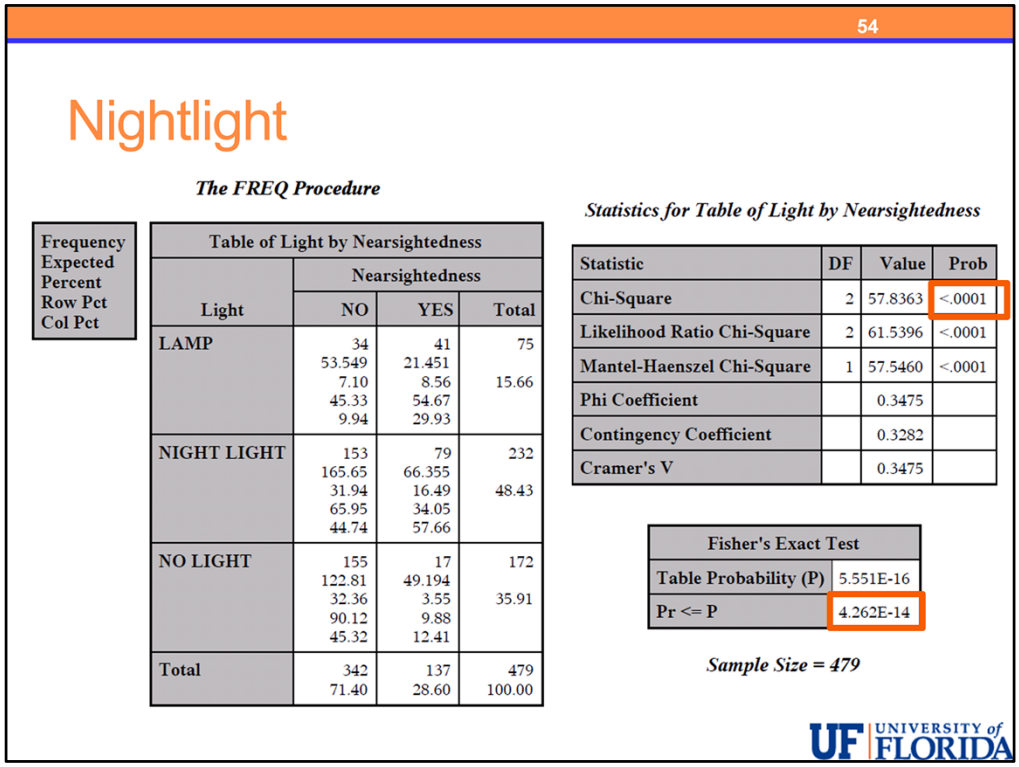
Obs	Anylight	Light	Nearsightedness
1	NO	NO LIGHT	NO
2	YES	NIGHT LIGHT	NO
3	YES	LAMP	NO
4	NO	NO LIGHT	NO
5	NO	NO LIGHT	NO
6	YES	NIGHT LIGHT	YES
7	YES	LAMP	NO
8	YES	NIGHT LIGHT	YES
9	NO	NO LIGHT	NO
10	YES	NIGHT LIGHT	YES

UF UNIVERSITY of FLORIDA

Here is a few lines of the data.

Notice the variable values are not coded.

We have added a new variable called Anylight which is NO for children with no light and YES for children with a lamp or night light.



To investigate the association between type of light and nearsightedness, using the original three level light variable, we can conduct a chi-squared test or fisher's exact test.

The null hypothesis is that there is no relationship between the type of light and future nearsightedness in other words, that type of light and future nearsightedness are independent.

The alternative hypothesis is that there IS a relationship between the type of light and future nearsightedness in other words, that type of light and future nearsightedness are dependent.

In SAS, the values in each cell are in the following order – specified in the "legend" in the upper left corner of the table. Frequency, Expected Count, Overall Percent, Row Percent, Column Percent

Using the row percentages, our contingency table shows that among children with no light, 9.88% developed nearsightedness, among children with a nightlight, 34.05% developed nearsightedness and among children with a lamp, 54.67% developed nearsightedness.

Without using any inferential statistics, this difference seems extreme. And, in fact, the p-value of both the chi-square test (given as < 0.0001) and Fisher's exact test (which gives a tiny probability of 4.3x10^(-14)) show an extremely highly significant result.

Thus we can reject the null hypothesis.

We conclude that there is enough evidence of an association between the type of light at night and the future development of nearsightedness in the population. Type of light used at night and development of nearsightedness are dependent.

Nightlight

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	57.836[a]	2	.000	.000
Likelihood Ratio	61.540	2	.000	.000
Fisher's Exact Test	61.016			.000
N of Valid Cases	479			

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 21.45.

			Nearsightedness		
			NO	YES	Total
Light	LAMP	Count	34	41	75
		Expected Count	53.5	21.5	75.0
		% within Light	45.3%	54.7%	100.0%
		% within Nearsightedness	9.9%	29.9%	15.7%
		% of Total	7.1%	8.6%	15.7%
	NIGHT LIGHT	Count	153	79	232
		Expected Count	165.6	66.4	232.0
		% within Light	65.9%	34.1%	100.0%
		% within Nearsightedness	44.7%	57.7%	48.4%
		% of Total	31.9%	16.5%	48.4%
	NO LIGHT	Count	155	17	172
		Expected Count	122.8	49.2	172.0
		% within Light	90.1%	9.9%	100.0%
		% within Nearsightedness	45.3%	12.4%	35.9%
		% of Total	32.4%	3.5%	35.9%
Total		Count	342	137	479
		Expected Count	342.0	137.0	479.0
		% within Light	71.4%	28.6%	100.0%
		% within Nearsightedness	100.0%	100.0%	100.0%
		% of Total	71.4%	28.6%	100.0%

UF UNIVERSITY of FLORIDA

The SPSS output gives exactly the same information. The only difference is the order that the cell values are presented.

In SAS the values were Frequency, Expected Count, Overall Percent, Row Percent, Column Percent.

In SPSS they are given as Count – which is the frequency, expected count, then % within light which is the ROW percent, then % within nearsightedness which is the column percent, with the overall percent being provided last.

Understanding the output provided by your software is important now and most definitely in practice.

The p-value of the appropriate chi-square test and Fisher's exact test are outlined in the table and are reported to be 0.000 which doesn't mean the p-value is exactly equal to zero but it is zero rounded to three decimal places.

Again, our conclusion is that there is a highly statistically significant association between type of light and nearsightedness.

Nightlight

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	46.041[a]	1	.000	.000	.000
Continuity Correction[b]	44.622	1	.000		
Likelihood Ratio	51.605	1	.000	.000	.000
Fisher's Exact Test				.000	.000
N of Valid Cases	479				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 49.19.

b. Computed only for a 2x2 table

			Nearsightedness		
			NO	YES	Total
anylight	NO	Count	155	17	172
		Expected Count	122.8	49.2	172.0
		% within anylight	90.1%	9.9%	100.0%
		% within Nearsightedness	45.3%	12.4%	35.9%
		% of Total	32.4%	3.5%	35.9%
	YES	Count	187	120	307
		Expected Count	219.2	87.8	307.0
		% within anylight	60.9%	39.1%	100.0%
		% within Nearsightedness	54.7%	87.6%	64.1%
		% of Total	39.0%	25.1%	64.1%
Total		Count	342	137	479
		Expected Count	342.0	137.0	479.0
		% within anylight	71.4%	28.6%	100.0%
		% within Nearsightedness	100.0%	100.0%	100.0%
		% of Total	71.4%	28.6%	100.0%

UF UNIVERSITY of FLORIDA

To investigate the association between the variable anylight and nearsightedness we can conduct a chi-squared test with a continuity correction or fisher's exact test.

The null hypothesis is that there is no relationship between whether or not the child slept with any light and future nearsightedness in other words, exposure to light during sleep and future nearsightedness are independent.

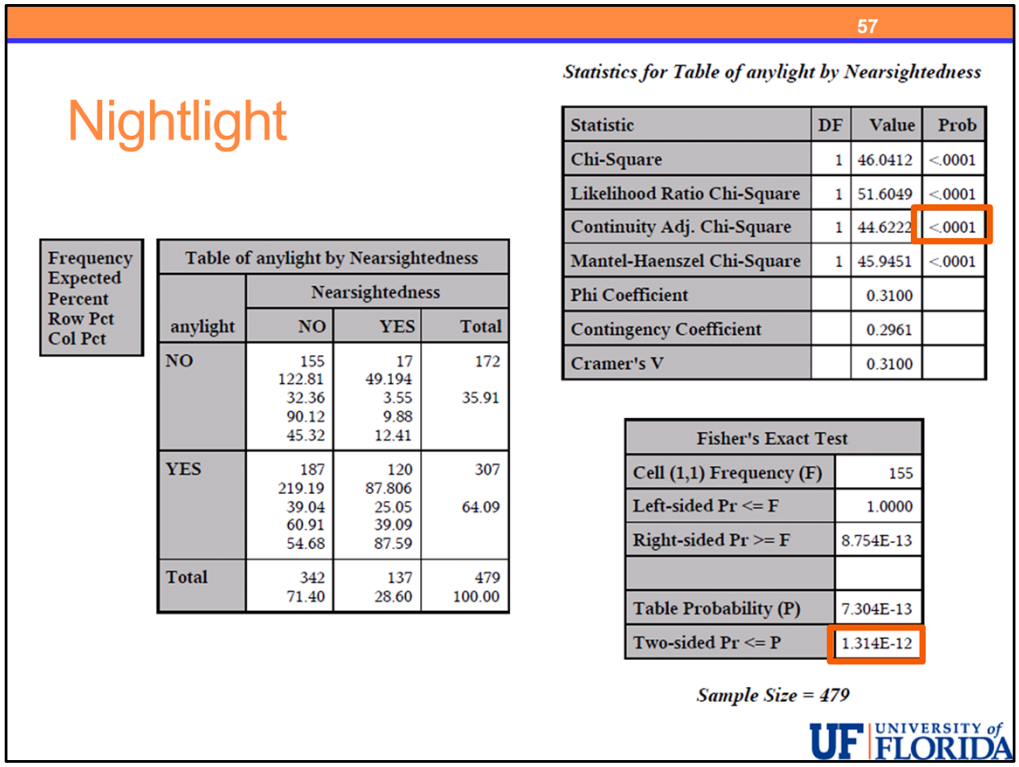
The alternative hypothesis is that there IS a relationship between whether or not the child slept with any light and future nearsightedness in other words, exposure to light during sleep and future nearsightedness are dependent.

Using the row percentages, our contingency table shows that among children with no light, 9.88% developed nearsightedness whereas among children with a nightlight or lamp, 39.09% developed nearsightedness.

In SPSS, the p-value of both the continuity adjusted chi-square test and Fisher's exact test are given as 0.000 giving an extremely highly significant result.

Thus we can reject the null hypothesis.

We conclude that there is enough evidence of an association between whether or not the child slept with any light and the future development of nearsightedness in the population. Exposure to light during sleep and future nearsightedness are dependent.



The only difference between the SAS output and SPSS output is in the reporting of the p-values.

In SAS, the p-value of the continuity adjusted chi-square test is given as < 0.0001 and for Fisher's exact test it is given as 1.3x10^(-12).

Both of these are extremely small and so we would again reject the null hypothesis.