Chi-square Test

- C→C

The chi-square test

We want to know if the differences in sample proportions in a contingency table (i.e. two categorical variables) are likely to have occurred just by chance because of the random sampling.

We use the **chi-square (χ^2) test of independence** to assess the null hypothesis of no relationship between the two categorical variables of a two-way table.

Suppose we want to study the relationship between video game usage and later school attention issues among preschoolers.

Video Games	Attention Problems		
	No	Yes	Total
No	1000	500	1500
Yes	250	250	500
Total	1250	750	2000

What descriptive measure could I use to describe whether video game usage is associated with later attention problems?

Suppose we want to study the relationship between video game usage and later school attention issues among preschoolers.

Video Games	Attention Problems		
	No	Yes	Total
No	1000	500	1500
Yes	250	250	500
Total	1250	750	2000

It is always most interesting to ***condition on the explanatory variable*** of interest so that you can investigate whether knowing a particular value of an explanatory variable, helps us predict our response variable.

Conditional Row Percents:

- Among those who were not exposed to video games before age 2, 33% (500 out of 1500) of them have attention problems in pre-school.
- Among those who were exposed to video games before age 2, 50% (250 out of 500) of them have attention problems in pre-school.

H0: There is no relationship between video game exposure and attention problems

HA: There is a relationship between video game exposure and attention problems.

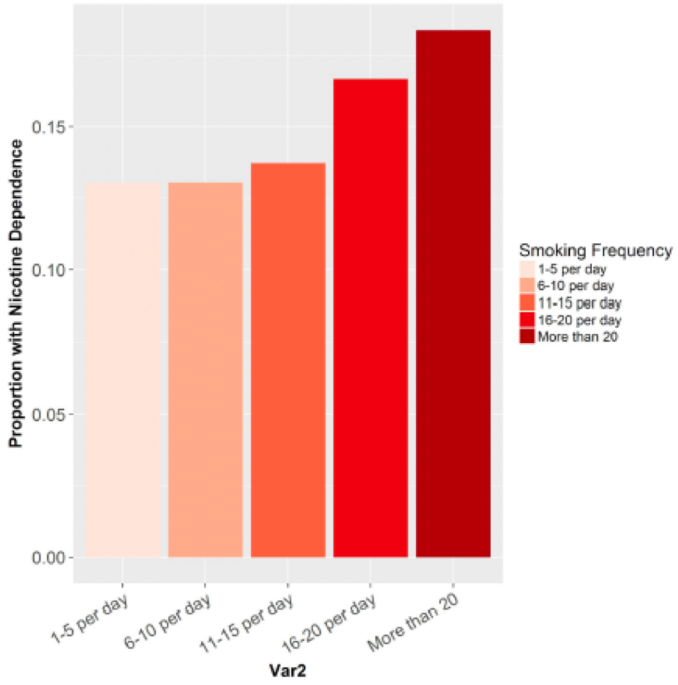
```
        Pearson's Chi-squared test with Yates' continuity correction

data:  vg and ap
X-squared = 43.736, df = 1, p-value = 3.758e-11
```

There is a significant association between video game exposure and attention problems ($X^2=43.7$, p-value<0.001).

Those exposed to video games before age 2 have a significantly higher likelihood of attention problems compared to those who were not exposed to video games before age 2.

Suppose we want to understand the relationship between nicotine use and nicotine dependence:

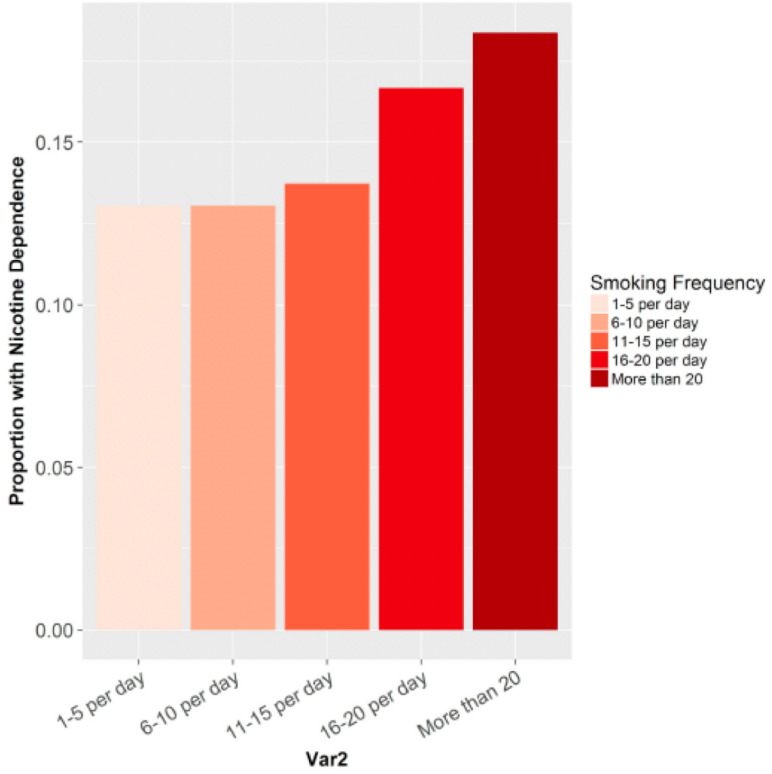



```
        Pearson's Chi-squared test

data:  new_data$cigarette_use and new_data$nic_dependence
X-squared = 22.038, df = 4, p-value = 0.0001969
```

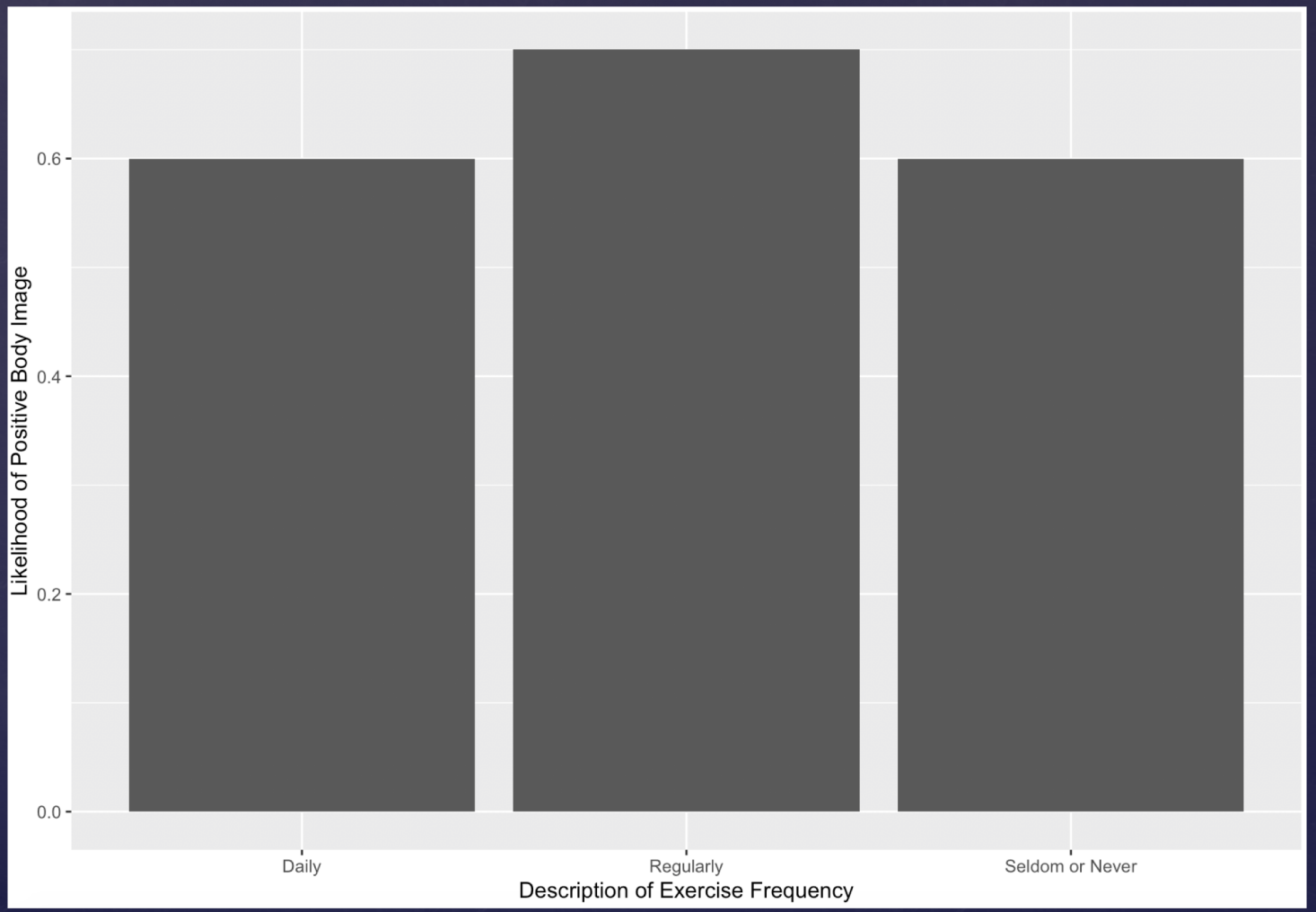
Post-Hoc Analysis

- Just as before, we can look at all pairwise comparisons:



comparison	raw.p	adj.p
1-5 per day vs. 6-10 per day	1.0000	1.0000
1-5 per day vs. 11-15 per day	0.6334	1.0000
1-5 per day vs. 16-20 per day	0.0149	0.1488
1-5 per day vs. More than 20	0.0004	0.0038
6-10 per day vs. 11-15 per day	0.6334	1.0000
6-10 per day vs. 16-20 per day	0.0149	0.1488
6-10 per day vs. More than 20	0.0004	0.0038
11-15 per day vs. 16-20 per day	0.0435	0.4351
11-15 per day vs. More than 20	0.0018	0.0184
16-20 per day vs. More than 20	0.2855	1.0000

Suppose we want to understand the relationship between Exercise Frequency and Positive Body Image . We obtain the following bivariate bar chart to describe the Relationship.



There is not enough evidence to suggest that there is a relationship between exercising and body image (X2=0.28, p-value=0.8663).

```
        Pearson's Chi-squared test

data:  ExerciseFrequency and BodyImage
X-squared = 0.28708, df = 2, p-value = 0.8663
```