Chi-squared test

- Used to compare two categorical variables
- Each categorical variable breaks our population into two or more subgroups
  Null hypothesis is that the two categorical variables are **independent**:
  - p-value > 0.05: Do not reject Null, i.e. the variables are independent
  - p-value <0.05: Reject null, i.e. the variables are NOT independent.
- Need to transform the data from categorical into the number of counts in each subgroup.
  - Need 80% of “expected” cell counts to be > 5
  - No cells with “expected” cell counts below 1

  NOTE:R will warn you if this is not the case.

Chi-squared example: Transform

Gender	Coffee Intake
Male	3
Male	2
Female	0
Male	1
Female	1
Male	2
Female	2

	0	1	2	3
Male	0	1	2	1
Female	2	1	1	0

Does Gender impact Coffee Intake?

Problem 10.4

Question: Are gender and coffee consumption independent?

The file LipidData contains information about gender and coffee consumption

1. State the Null Hypothesis and the alternative hypothesis
2. Construct a table of counts:

```
table(lipidDataFrame$Gender,lipidDataFrame$`Coffee intake (cups/day)`)

        0  1  2  3  4  5  7  8
female  7  5  7  2  0  2  0  1
male   35 14 12  4  2  3  1  0
```

Question: Is this data a good candidate for this test?

Problem 10.4

3. Run a chi-squared test:

```
> chisq.test(lipidDataFrame$Gender,lipidDataFrame$`Coffee intake (cups/day)`)

        Pearson's Chi-squared test

data:  lipidDataFrame$Gender and lipidDataFrame$`Coffee intake (cups/day)`
X-squared = 7.7587, df = 7, p-value = 0.3544
```

4. Conclusion?