

# Stats in Practice #11, The Chi-Square Test

*Kellin Rumsey*

*4/22/2019*

## 1. The dataset

Do males or females get killed off more often in slasher films? Does sexual activity play a role in likelihood of survival? In 2010, A. Welsh attempted to answer these questions by tracking the survival of  $n = 485$  characters from a random sample of 50 horror films. The data is given as a two-category table of observed counts where the categories are:

1. Survival or death.
2. Character features: (Combination of gender and sexual activity).

In this SIP, we will conduct a Chi-Square Test to determine if these categorical variables are independent or not.

### Load the dataset

- Open Rstudio and open a new script.
- Load the dataset by typing: `cdi = read(url('http://math.unm.edu/~knrumsey/slasher_films.rda'))`

## 2. Chi-Square test of significance

1. Look closely at the data by typing:

```
ObservedCounts
```

Reproduce this table on your writeup, and find the row/column totals.

2. Write out the hypotheses in words.
3. Assuming no relation between variables, what is the expected count of sexually active females who survive? Compare this to the observed count. Does it seem like gender and sexual activity might play a role in survival?
4. In R, we can find the table of expected counts easily by typing:

```
ExpectedCounts <- chisq.test(ObservedCounts)$expected
```

Reproduce this table on your writeup. Are the assumptions of the test met (**explain your answer**). Also check to make sure your answer in 3 matches what you get here.

5. Which groups die in slasher films *more often* than expected? Which groups die in slasher films *less often* than expected. Does this make sense to you?
6. Calculate the contribution to the  $\chi^2$  test statistic of sexually active females.
7. The complete  $\chi^2$  test statistic can be found in R by typing:

```
test_stat <- chisq.test(ObservedCounts)$statistic
```

Report the value of the test statistic.

8. What are the degrees of freedom for this test? Use the chi-square table to find a range for the p-value. Also find it “exactly” in R by typing: `pchisq(test_stat, df=?, lower.tail=FALSE)` where ? is the degrees of freedom of the test.
9. Make your conclusion in terms of the problem.