# Chi Square

## EDP 613

### Week 12

# A Note About The Slides

Currently the equations do not show up properly in Firefox. Other browsers such as Chrome and Safari do work.

# Independence

Two variables that have no association with each other are **statistically independent**.

# Frequencies

- **expected frequencies**
  - written $f_e$
  - is what you would *expect* in a bivariate table if two variables were statistically independent.
  - calculated by assuming the null hypothesis is true.
- **observed frequencies**
  - written $f_o$
  - is what you would *observe* in a bivariate table given what you have.
  - you know this already.

# Chi-Square Test

- Written $\chi^2$.
- Is an inferential test to find significant relationships between two variables.
- Calculated by

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

- with

$$df = (r-1)(c-1)$$

# Assumptions

Only that the data is derived from *random sampling*.

# Example

Fake scenario: From 2 political parties and 1 other group (Democrats, Republicans, and Independent), you take a poll from a sample of US citizens asking which category they support (i.e., their likely vote). You might wonder whether the people's likely votes are equally distributed between the 3 political parties (your $H_0$).

Samples

- Democrats: 587
- Republicans: 552
- Independents: 480

Total

$$587 + 552 + 480 = 1619.$$

Proportions

- Democrats: $587/1619 \approx 0.3626$
- Republicans: $552/1619 \approx 0.3410$
- Independents: $480/1619 \approx 0.2965$

Expected Frequencies

- We have three categories
- $H_0$ : voters are equally distributed across the categories
- Expected value is found by dividing the total by $1/3 \approx 0.3333$. So

  $1619 \cdot 0.3333 \approx 539.7$

| Type | Democrats | Republicans | Independents |
|---|---|---|---|
| **Observed** | 587 | 552 | 480 |
| **Expected** | 539.67 | 539.67 | 539.67 |

So

$$\chi^2 = \frac{(587 - 539.67)^2}{539.67} + \frac{(552 - 539.67)^2}{539.67} + \frac{(480 - 539.67)^2}{539.67}$$

$$\approx 11.03$$

with

$$\begin{aligned} df &= (2 - 1) \cdot (3 - 1) \\ &= 2 \end{aligned}$$

# Now let's try this in R!

```
votes = c(D=587, R=552, I=480); votes
```

```
##   D   R   I
## 587 552 480
```

# Calculate observed proportions

```
N = sum(votes); N
```

```
## [1] 1619
```

```
votes_freq = votes/N

votes_freq
```

```
##         D         R         I
## 0.3625695 0.3409512 0.2964793
```

# Determine expected frequencies

```
prop = 1/3; prop
```

```
## [1] 0.3333333
```

```
f_e = N * prop; f_e
```

```
## [1] 539.6667
```

| Type | Democrats | Republicans | Independents |
| --- | --- | --- | --- |
| **Observed** | 587 | 552 | 480 |
| **Expected** | 539.67 | 539.67 | 539.67 |

```
chisq_rs = (587 - 539.67)^2/539.67 + (552-539.67)^2/539.67 + (480 - 539.
chisq_rs
```

```
## [1] 11.0302
```

with

```
## [1] 2
```

# Significance

```
p_val = pchisq(chisq_rs, df = 2, lower.tail = FALSE)

p_val
```

```
## [1] 0.00402553
```

# I lied. This is not the easiest way!

You can actually do this in one line.

```
not_a_lie = chisq.test(votes); not_a_lie
```

```
##
##      Chi-squared test for given probabilities
##
## data:  votes
## X-squared = 11.03, df = 2, p-value = 0.004025
```

# That’s it. Take a break before our real R session!