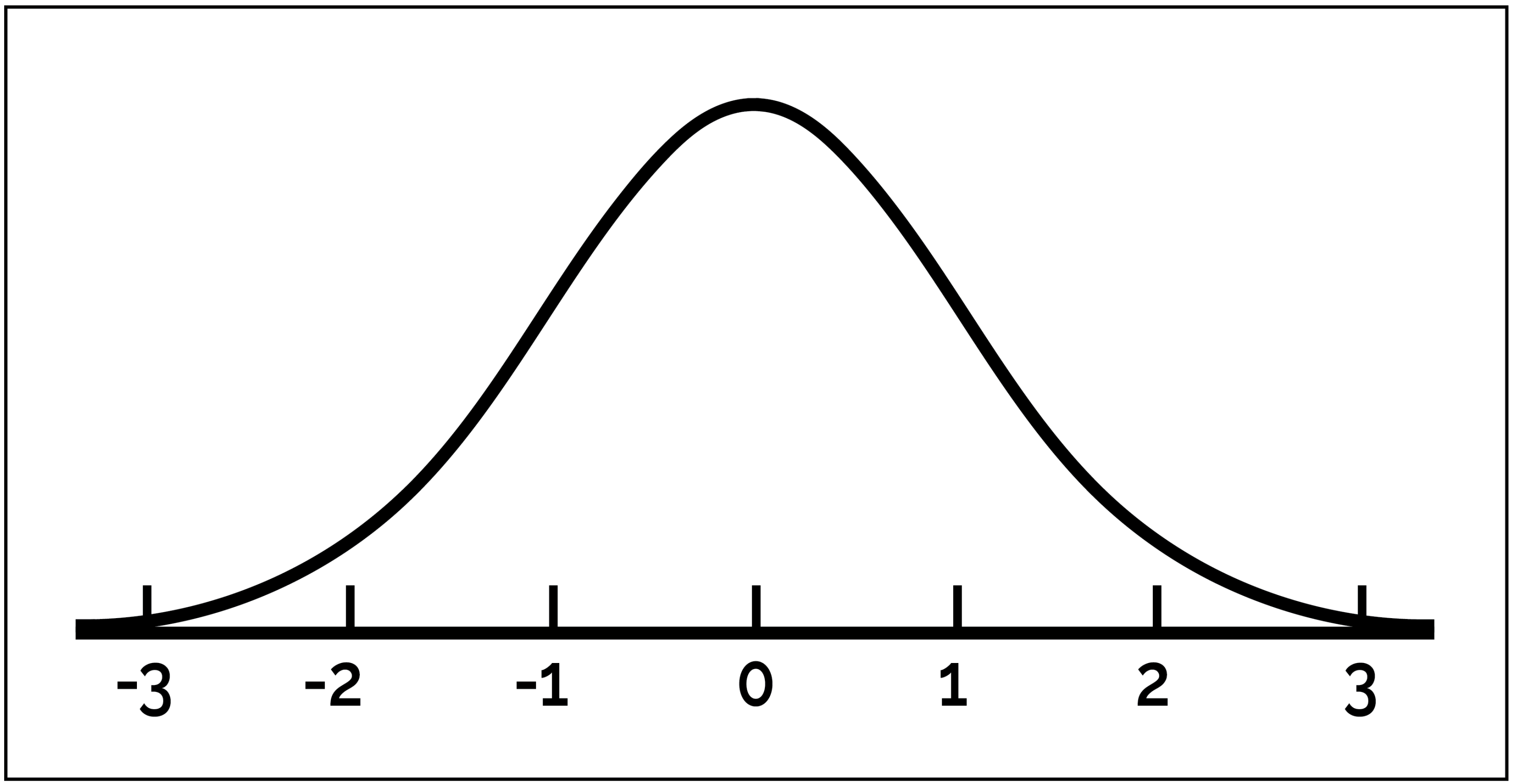
# Hypothesis Testing – Two Categorical Variables

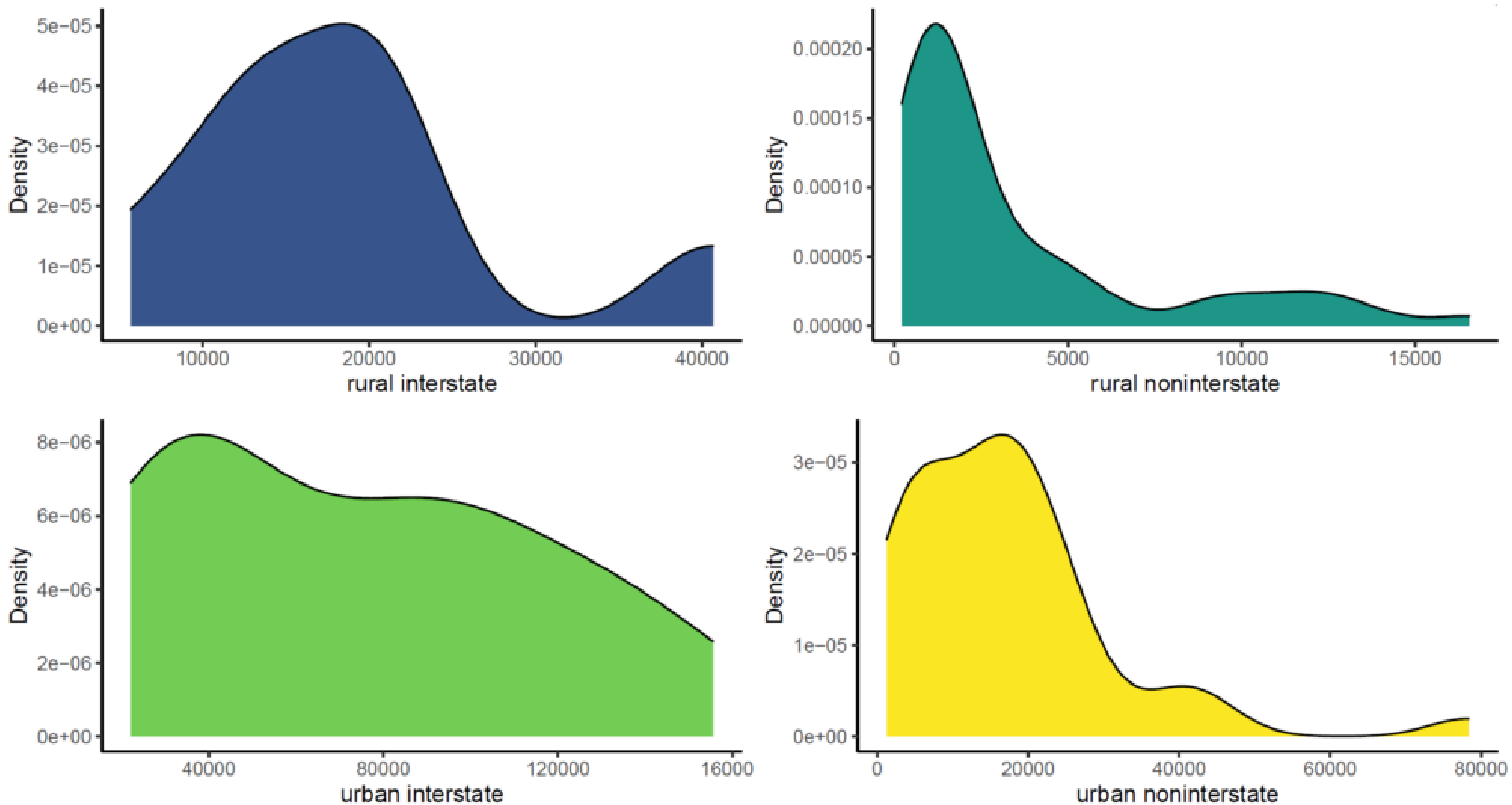
# Parametric Statistics

These are statistics that are used when the sampling distribution is **normal distributed**.



# Non-Parametric Statistics

These are statistics that are used when the sampling distribution is **not normal distributed**.

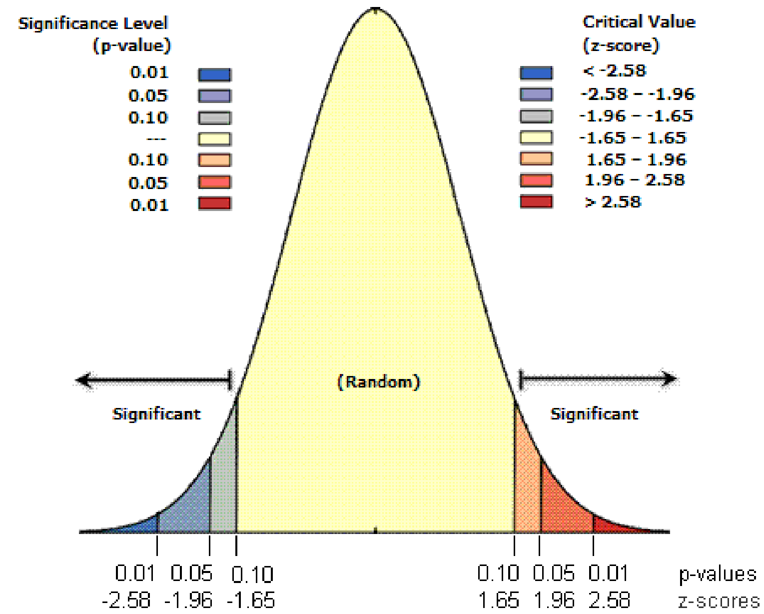


# Statistical Significance

This is a claim that a set of observations are not due to random chance.

Used in hypothesis testing to help decide whether to reject the null hypothesis.

Leads to a lot of problems within research when relied on.



# Chi Square

This is used when the IV and DV are both categorical and the data is not normally distributed (i.e., non-parametric).

$$x^2 = \frac{\Sigma(O - E)^2}{E}$$

O = Observed frequency

E = Expected frequency

$$E = \frac{(\text{row total})(\text{column total})}{(\text{sample size})}$$

# Chi-Square Example

Is gender and education level dependent at a 5% level of significance?

In other words, in the data below, is there a relationship between the gender of an individual and the level of education that they have obtained?

| | High School | Bachelors | Masters | Total |
|---|---|---|---|---|
| Female | 60 | 54 | 46 | 160 |
| Male | 40 | 44 | 53 | 137 |
| Total | 100 | 98 | 99 | 297 |

# Chi-Square Example

Let's first state our null and alternative hypothesis:

$H_0$ Gender and education level are independent, there is no relationship between gender and education level.

$H_A$ Gender and education level are dependent, there is a relationship between gender and education level.

# Chi-Square Example

$$E = \frac{\text{(row total)(column total)}}{\text{(sample size)}}$$

| | High School | Bachelors | Masters | Total |
|---|---|---|---|---|
| Female | $\frac{(160)(100)}{297}$ | $\frac{(160)(98)}{297}$ | $\frac{(160)(99)}{297}$ | 160 |
| Male | $\frac{(137)(100)}{297}$ | $\frac{(137)(98)}{297}$ | $\frac{(137)(99)}{297}$ | 137 |
| Total | 100 | 98 | 99 | 297 |

# Chi-Square Example

Observed

| | High School | Bachelors | Masters |
|---|---|---|---|
| Female | 60 | 54 | 46 |
| Male | 40 | 44 | 53 |

Expected

| | High School | Bachelors | Masters |
|---|---|---|---|
| Female | 53.87 | 52.68 | 53.45 |
| Male | 46.13 | 44.32 | 46.55 |

# Chi-Square Example

$$\chi^2 = \left( \frac{(60 - 53.87)^2}{53.87} \right) + \left( \frac{(54 - 52.68)^2}{52.68} \right) + \left( \frac{(46 - 53.45)^2}{53.45} \right) +$$

$$\chi^2 \approx 3.45$$

# Degrees of Freedom

The degrees of freedom for the chi-square are calculated using the following formula:

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

r = row number

c = column number

# Chi-Square Example

Now, we need to compare this value to the critical value from the chi-square distribution at a 5% significance level.

$$(2 - 1) \times (3 - 1) = 2$$

2 degrees of freedom

If the observed chi-square test statistic is greater than the critical value, the null hypothesis can be rejected.

If the observed chi-square test statistic is less than the critical value, we fail to reject the null hypothesis.

# Chi-Square Example

Now lets take our df value of 12 and find our critical value with our 5% (0.05) significance level and locate the critical value in a chi-square distribution chart (p. 375).

The critical value is 5.99

| | P | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | 0.995 | 0.975 | 0.2 | 0.1 | 0.05 | 0.025 | 0.02 | 0.01 | 0.005 | 0.002 | 0.001 |
| 1 | .0004 | .00016 | 1.642 | 2.706 | 3.841 | 5.024 | 5.412 | 6.635 | 7.879 | 9.55 | 10.828 |
| 2 | 0.01 | 0.0506 | 3.219 | 4.605 | 5.991 | 7.378 | 7.824 | 9.21 | 10.597 | 12.429 | 13.816 |
| 3 | 0.0717 | 0.216 | 4.642 | 6.251 | 7.815 | 9.348 | 9.837 | 11.345 | 12.838 | 14.796 | 16.266 |
| 4 | 0.207 | 0.484 | 5.989 | 7.779 | 9.488 | 11.143 | 11.668 | 13.277 | 14.86 | 16.924 | 18.467 |
| 5 | 0.412 | 0.831 | 7.289 | 9.236 | 11.07 | 12.833 | 13.388 | 15.086 | 16.75 | 18.907 | 20.515 |
| 6 | 0.676 | 1.237 | 8.558 | 10.645 | 12.592 | 14.449 | 15.033 | 16.812 | 18.548 | 20.791 | 22.458 |
| 7 | 0.989 | 1.69 | 9.803 | 12.017 | 14.067 | 16.013 | 16.622 | 18.475 | 20.278 | 22.601 | 24.322 |
| 8 | 1.344 | 2.18 | 11.03 | 13.362 | 15.507 | 17.535 | 18.168 | 20.09 | 21.955 | 24.352 | 26.124 |
| 9 | 1.735 | 2.7 | 12.242 | 14.684 | 16.919 | 19.023 | 19.679 | 21.666 | 23.589 | 26.056 | 27.877 |
| 10 | 2.156 | 3.247 | 13.442 | 15.987 | 18.307 | 20.483 | 21.161 | 23.209 | 25.188 | 27.722 | 29.588 |
| 11 | 2.603 | 3.816 | 14.631 | 17.275 | 19.675 | 21.92 | 22.618 | 24.725 | 26.757 | 29.354 | 31.264 |
| 12 | 3.074 | 4.404 | 15.812 | 18.549 | 21.026 | 23.337 | 24.054 | 26.217 | 28.3 | 30.957 | 32.909 |
| 13 | 3.565 | 5.009 | 16.985 | 19.812 | 22.362 | 24.736 | 25.472 | 27.688 | 29.819 | 32.535 | 34.528 |
| 14 | 4.075 | 5.629 | 18.151 | 21.064 | 23.685 | 26.119 | 26.873 | 29.141 | 31.319 | 34.091 | 36.123 |
| 15 | 4.601 | 6.262 | 19.311 | 22.307 | 24.996 | 27.488 | 28.259 | 30.578 | 32.801 | 35.628 | 37.697 |
| 16 | 5.142 | 6.908 | 20.465 | 23.542 | 26.296 | 28.845 | 29.633 | 32 | 34.267 | 37.146 | 39.252 |
| 17 | 5.697 | 7.564 | 21.615 | 24.769 | 27.587 | 30.191 | 30.995 | 33.409 | 35.718 | 38.648 | 40.79 |
| 18 | 6.265 | 8.231 | 22.76 | 25.989 | 28.869 | 31.526 | 32.346 | 34.805 | 37.156 | 40.136 | 42.312 |
| 19 | 6.844 | 8.907 | 23.9 | 27.204 | 30.144 | 32.852 | 33.687 | 36.191 | 38.582 | 41.61 | 43.82 |
| 20 | 7.434 | 9.591 | 25.038 | 28.412 | 31.41 | 34.17 | 35.02 | 37.566 | 39.997 | 43.072 | 45.315 |

# Chi-Square Example

$\chi^2$ = 3.45

The critical value = 5.99

$3.483 < 5.99$, we fail reject the null hypothesis.

Therefore, at a 5% significance level, we do not have enough evidence to conclude that there is a relationship between gender and education level.