Chapter 10 inferential statistics and chi-square

1) What is inferential statistics?

sampling
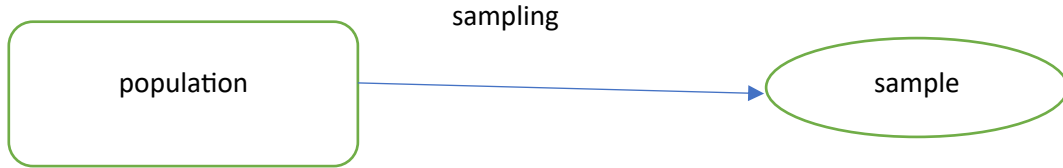
population → sample

to describe the error margins of your data, inferring from the sample back to the population.

2) Process of doing inferential statistics

X -> Y in the sample, we start with an assumption to nullify this causality relations (x-> Y) in the population. Basically, we are saying your finding in the sample is false, it would not be holding true in the population. Inferential statistics would come to assess how likely the null hypothesis is true. If the likelihood of the null hypothesis is true is low, then you can reject the null hypothesis. Otherwise, you cannot reject the null hypothesis.

The threshold in social science to reject a null hypothesis is 5%, if your P is lower than 5%, reject the null hypothesis, if P >= 5%, do not reject the null hypothesis.

P (probability): a probability that your null hypothesis is true.

	True null hypothesis	False null hypothesis
Reject (P < .05)	Type I error	OK
Not reject (P >= .05)	OK	Type II error

3) Chi-square (χ^2)

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

F_o is the observed frequency, whereas F_e is the expected frequency

$$F_e = \frac{RM \times CM}{N}$$

4) Example

	Race		
Presidential vote	white	Black	RM
Joe Biden	589 $(\frac{682 \times 1266}{1374} = 628.4)$	93 $(\frac{682 \times 108}{1374} = 53.6)$	682
Donald Trump	677 $(\frac{692 \times 1266}{1374} = 637.6)$	15 $(\frac{692 \times 108}{1374} = 54.4)$	692
CM	1266	108	N = 1,374

A) Null hypothesis: race has nothing to do with the presidential vote (independent variable has nothing to do with the dependent variable)
B) Computing the expected frequencies

$$F_e = \frac{RM \times CM}{N}$$

	Race		
Presidential vote	white	Black	RM
Joe Biden	589 $(\frac{682 \times 1266}{1374} = 628.4)$	93 $(\frac{682 \times 108}{1374} = 53.6)$	682
Donald Trump	677 $(\frac{692 \times 1266}{1374} = 637.6)$	15 $(\frac{692 \times 108}{1374} = 54.4)$	692
CM	1266	108	N = 1,374

C) Computing the chi-square

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

$$\chi^2 = \frac{(589 - 628.4)^2}{628.4} + \frac{(93 - 53.6)^2}{53.6} + \frac{(677 - 637.6)^2}{637.6} + \frac{(15 - 54.4)^2}{54.4} = 62.4$$

D) Computing df (degree of freedom)

df = (r − 1) (c − 1)   r: total number of rows, c: total number of columns

df = (2 -1)(2-1) = 1

E) Determine the p value (P (probability): a probability that your null hypothesis is true)

P < .001

F) Decision regarding the null hypothesis, type of error committed.

Reject the null hypothesis, committing type I error