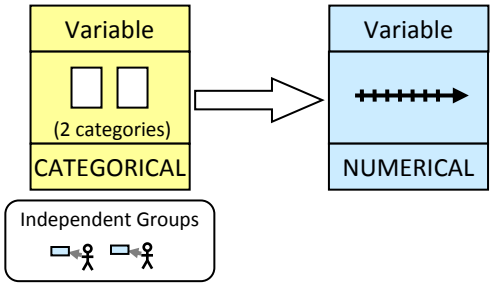
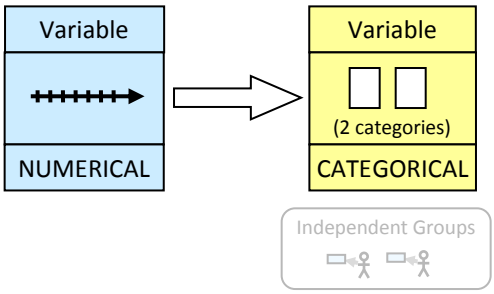
Statistical questions for statistical methods

Unpaired (two-sample) t-test

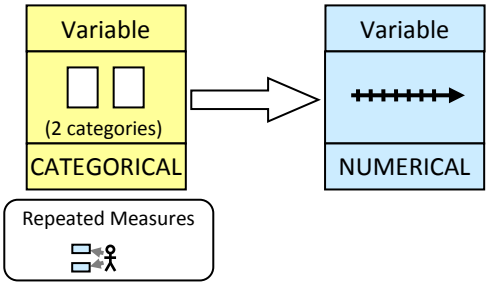


DECIDE: "Does the numerical outcome have a relationship with the categorical explanatory variable?"
"Is the mean of the outcome the same under each category?"



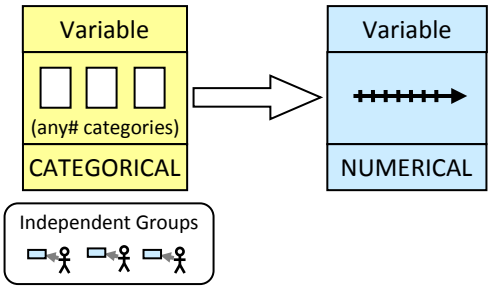
DECIDE: "Does the categorical outcome have a relationship with the numerical explanatory variable?"
"Is the mean of the explanatory variable the same under each category of the outcome?"

Paired t-test



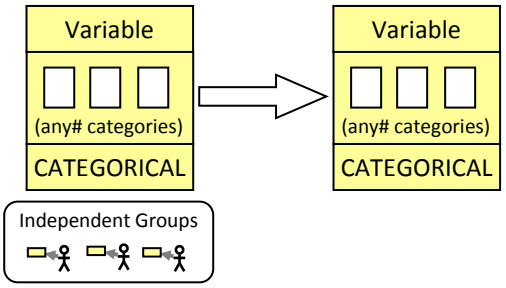
DECIDE: "Does the numerical outcome have a relationship with the categorical explanatory variable?"
"When you compare the outcome between categories for each subject, is the mean difference zero?"

ANOVA (one way)



DECIDE: "Does the numerical outcome have a relationship with the categorical explanatory variable?"
"Is the mean of the outcome the same under every category?"

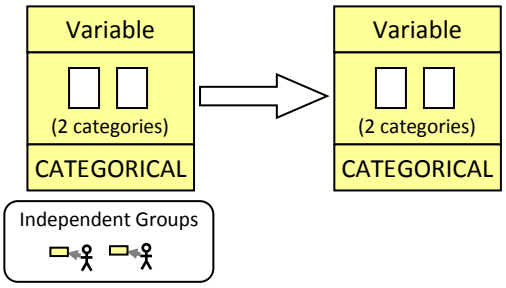
Chi-squared test (for association/independence)



DECIDE: "Does the categorical outcome have a relationship with the categorical explanatory variable?"
"Are the chances of the different categories in the outcome the same under every category of the explanatory variable?"

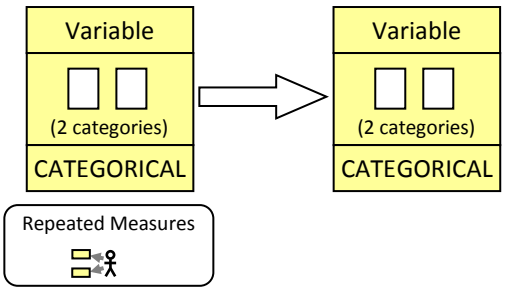
Statistical questions for statistical methods

Odds ratio



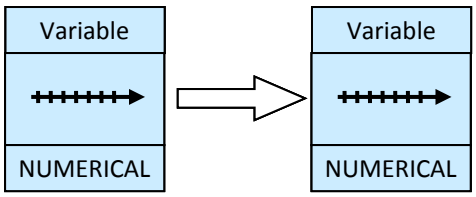
DESCRIBE: Compares the chances of one category of the outcome across the two explanatory categories *in your particular sample*. (If equally likely in both categories it will come out to 1.)

McNemar's test



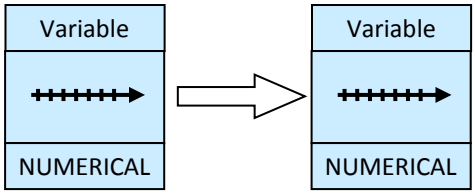
DECIDE: "Does the categorical outcome have a relationship with the categorical explanatory variable?"
"For each possible subject, are the chances of the two categories in the outcome the same under both categories of the explanatory variable?"

Linear regression (simple)



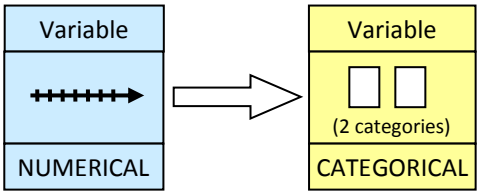
DECIDE: "Does the numerical outcome have a relationship with the numerical explanatory variable?"
PREDICT: "How do you calculate the outcome most accurately if you know the value of the explanatory variable?"

Correlation



DESCRIBE: Describes the strength of the linear relationship between the two variables *in your particular sample.* (If the relationship is perfect positive it will come out to 1, if it's perfect negative it will come out to -1, if no relationship it will come out to 0.)

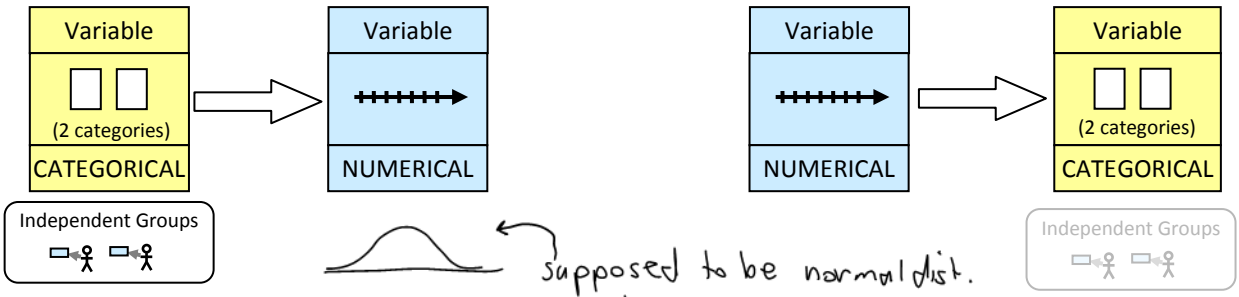
Logistic regression

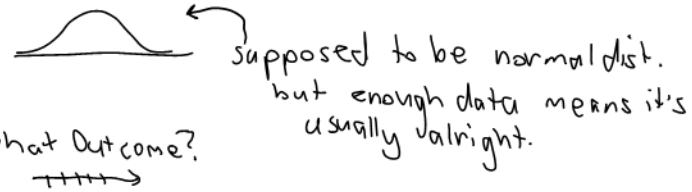


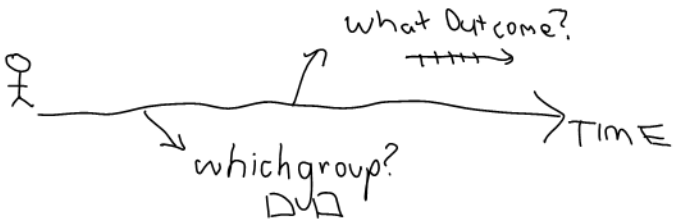
DECIDE: "Does the categorical outcome have a relationship with the explanatory variable?"
PREDICT: "How do you calculate the chances of the outcome most accurately if you know the value of the explanatory variable?"

Details about statistical methods

Unpaired (two-sample) t-test





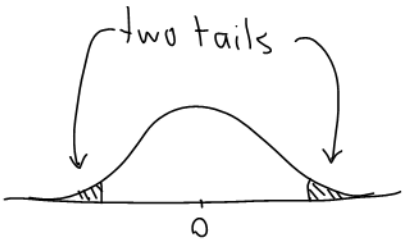


	chilli	temp
1	Y	38
2	Y	37
3	Y	38
4	Y	36
5	N	35
6	N	36
7	N	34
8	N	36

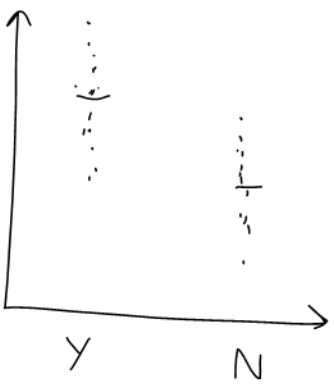
mean $\bar{x}_1$ & s.d. s_1

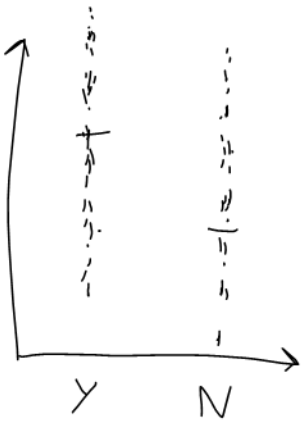
mean $\bar{x}_2$ & s.d. s_2

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2^2}}}$$



P-value

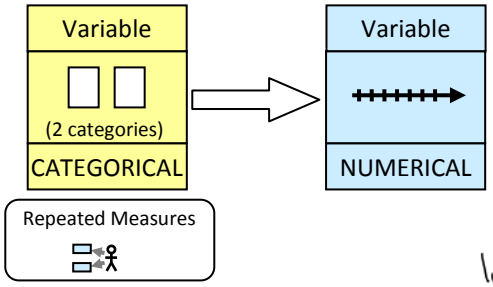


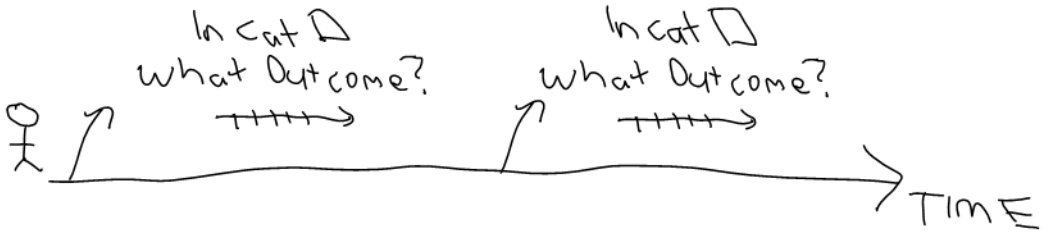


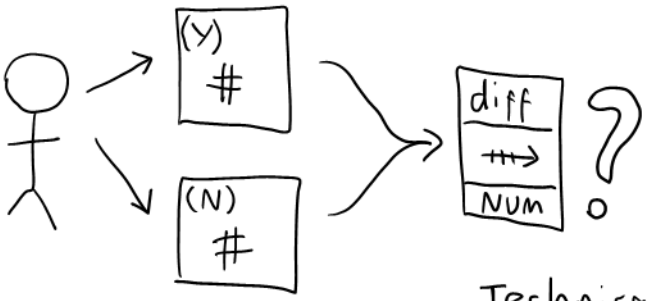
bigger spread makes it harder to see the difference

Details about statistical methods

Paired t-test







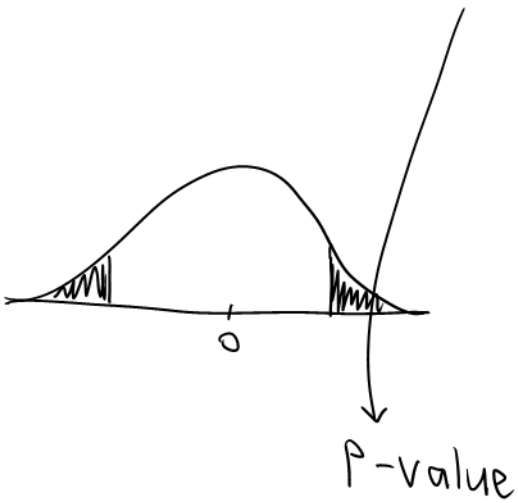
Technically it's a one-sample t-test

	temp (chilli Y)	temp (chilli N)	diff
1	#	#	#
2	#	#	#
3	#	#	#
4	#	#	#
5	#	#	#
6	#	#	#
7	#	#	#

mean $\bar{d}$

st.dev s_d

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$

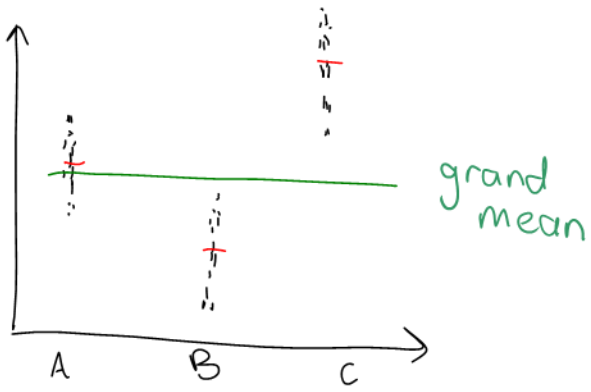


Details about statistical methods

ANOVA (one way) Analysis of Variance

Variable	→	Variable
(any# categories)		
CATEGORICAL		NUMERICAL

Independent Groups



	sum of squared differences SS	degrees of freedom df	mean sum of squares MSS	F	P-value
groups					
error					
total					

No relationship
⇒ Expected answer is 0

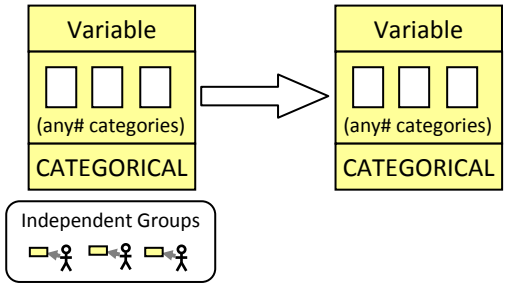
P-value

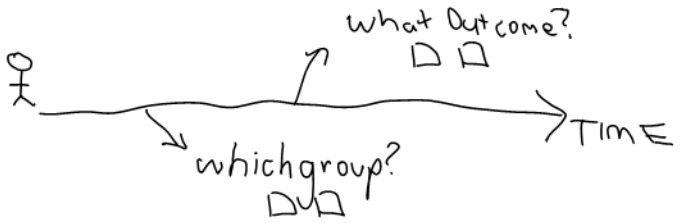
ANOVA says "there is a difference somewhere"

POST-HOC t-tests
tell you which category is different to which

Details about statistical methods

Chi-squared test (for association/independence)





"two-way table"
"contingency table"
"cross-tabulation"

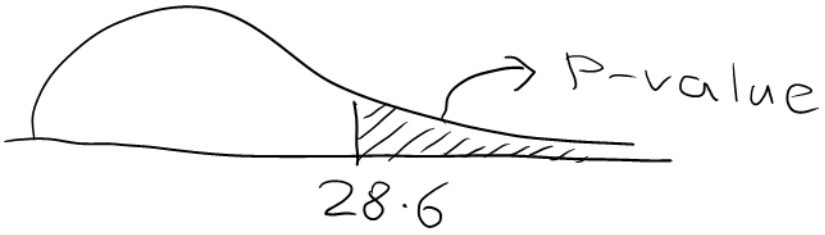
	chilli	feel hot
1	Y	Y
2	Y	Y
3	Y	Y
4	Y	N
5	N	N
6	N	Y
⋮	⋮	⋮

		chilli Y	chilli N	
feel hot	Y	13	4	17
	N	5	14	19
		18	18	

$$\chi^2 = \sum \frac{(obs-exp)^2}{exp}$$

$$= 28.6$$

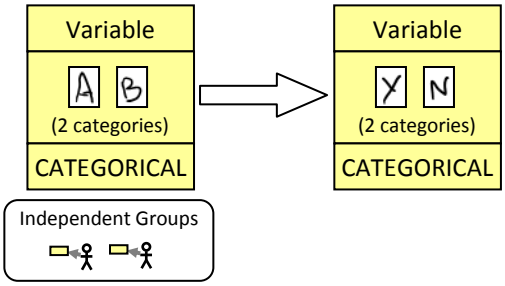
NO REL $\Rightarrow$ 0 is most likely answer



(can also do Fisher's exact test)

Details about statistical methods

Odds ratio



DESCRIBE

Strength of relationship

odds ratio of B to A = $\frac{\text{odds of picking someone Y in B}}{\text{odds of picking someone Y in A}}$

Comes out to answer 1 if the same odds.

	A	B
Y	20	40
N	30	10
	50	50

odds for Y in cat B = $\frac{40}{10}$

odds for Y in cat A = $\frac{20}{30}$

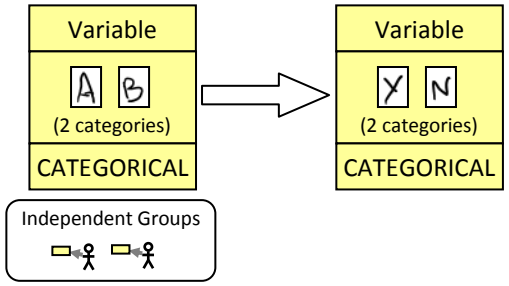
odds ratio of B to A = $\frac{\left(\frac{40}{10}\right)}{\left(\frac{20}{30}\right)} = \frac{40}{10} \times \frac{30}{20} = 6$

NOTE: If O.R. is 1.8 then you can say "A person in cat B is 1.8 times as likely to have Y than a person in cat A."

OR "A person in cat B is 80% more likely to have Y than a person in cat A."

Details about statistical methods

~~Odds ratio~~ RISK RATIO



DESCRIBE

Strength of relationship

$$\text{risk ratio of B to A} = \frac{\text{risk of person in cat B getting Y}}{\text{risk of person in cat A getting Y}}$$

Cat A: 50 all NO ⟶ 20 have YES

Cat B: 50 all NO ⟶ 40 have YES

$$\text{risk of Y in cat A} = \frac{20}{50}$$

$$\text{risk of Y in cat B} = \frac{40}{50}$$

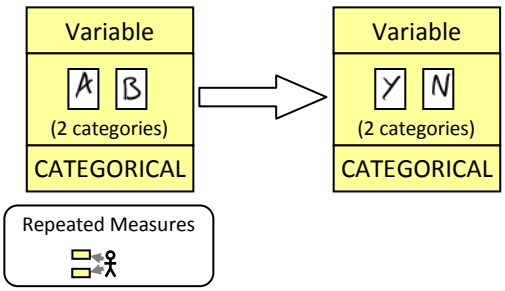
$$\text{risk ratio of B to A} = \frac{\left(\frac{40}{50}\right)}{\left(\frac{20}{50}\right)} = 2$$

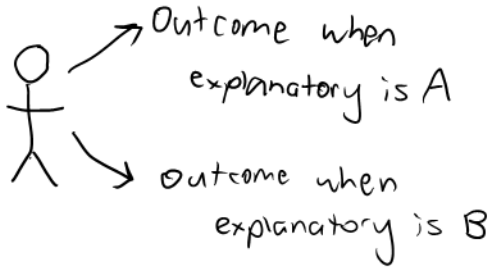
NOTE: Odds ratio is about the PRESENT

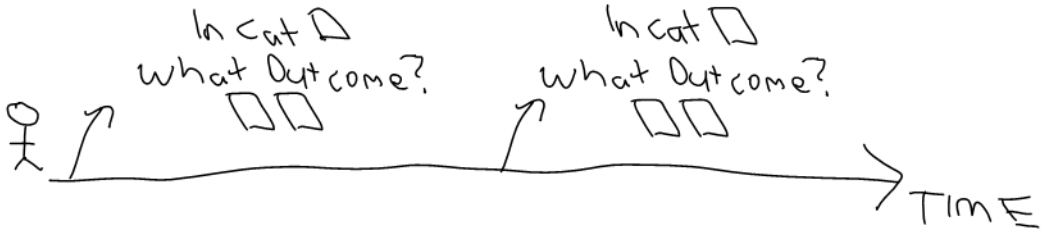
risk ratio is about the FUTURE

Details about statistical methods

McNemar's test

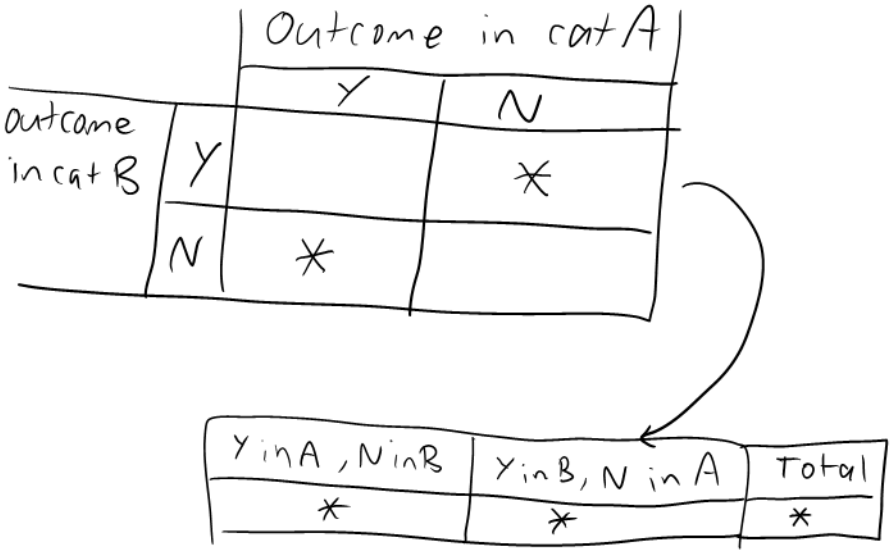






Still use χ^2 but calculated a different way.

"PAIRED CHI-SQUARED TEST"

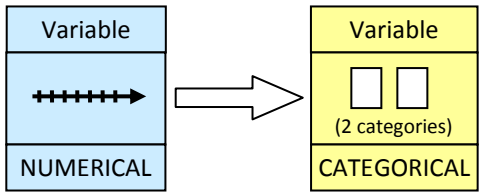


Y in A, N in B	Y in B, N in A	Total
*	*	*

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}$$

Details about statistical methods

Logistic regression



Is there a relationship?

→ unpaired t-test.

Formula for chances of Y outcome based on explanatory variable.

→ logistic regression

$$\underbrace{\log(\text{odds}(\text{Yes}))}_{\text{"logit"}} = \beta_0 + \beta_1 X$$