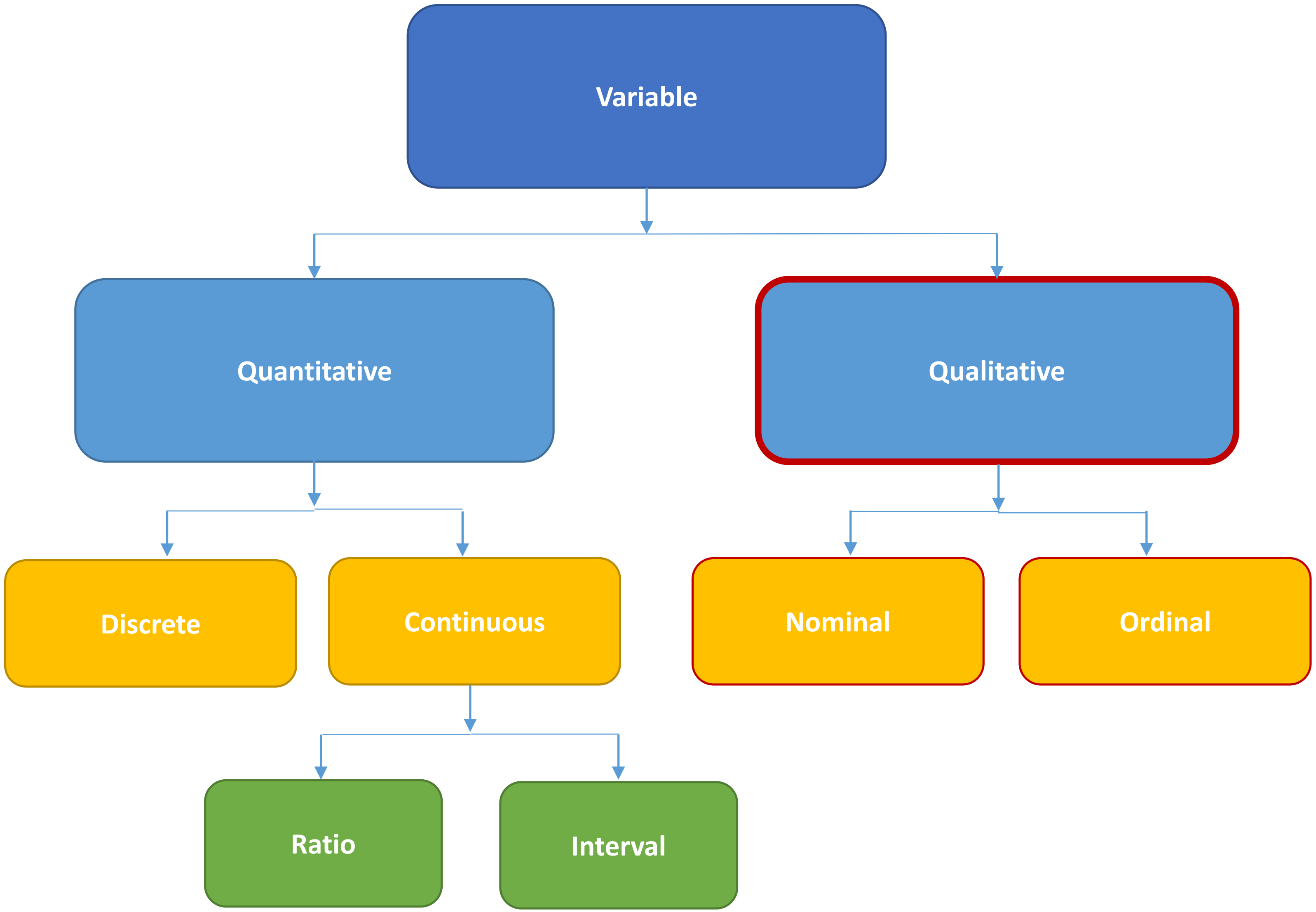
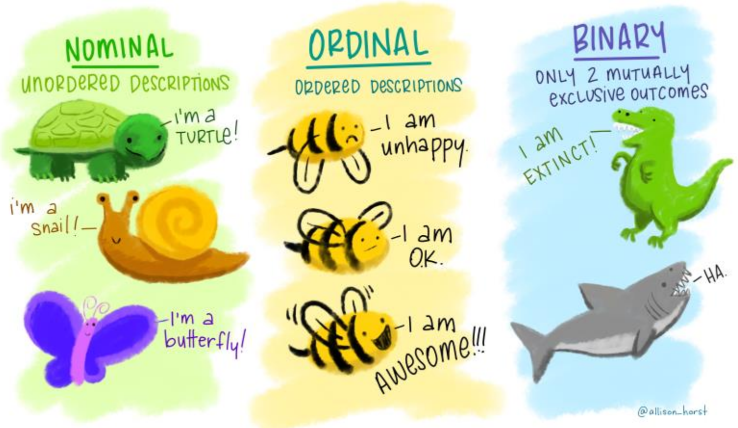
Analysis of Qualitative data

Anne Segonds-Pichon
v2020-08


Variable
Quantitative
Qualitative
Discrete
Continuous
Nominal
Ordinal
Ratio
Interval

Qualitative data

- = not numerical
- = values taken = usually names (also *nominal*)
  - e.g. genotypes
- Values can be numbers but not numerical
  - e.g. group number = numerical label but not unit of measurement
- Qualitative variable with intrinsic order in their categories = *ordinal*
- Particular case: qualitative variable with 2 categories: ***binary*** or *dichotomous*
  - e.g. alive/dead or presence/absence



https://github.com/allisonhorst/stats-illustrations#other-stats-artwork

Fisher's exact and Chi^2

Example: **cats and dogs.xlsx**

- Cats and dogs trained to line dance
- 2 different rewards: food or affection
- **Question**: Is there a difference between the rewards?

- **Is there a significant relationship between the 2 variables?**
  - does the reward significantly affect the likelihood of dancing?

- To answer this type of question:
  - **Contingency table**
  - **Fisher's exact or Chi^2 tests**

	Food	Affection
Dance	?	?
No dance	?	?

But first: **how many animals** do we need?

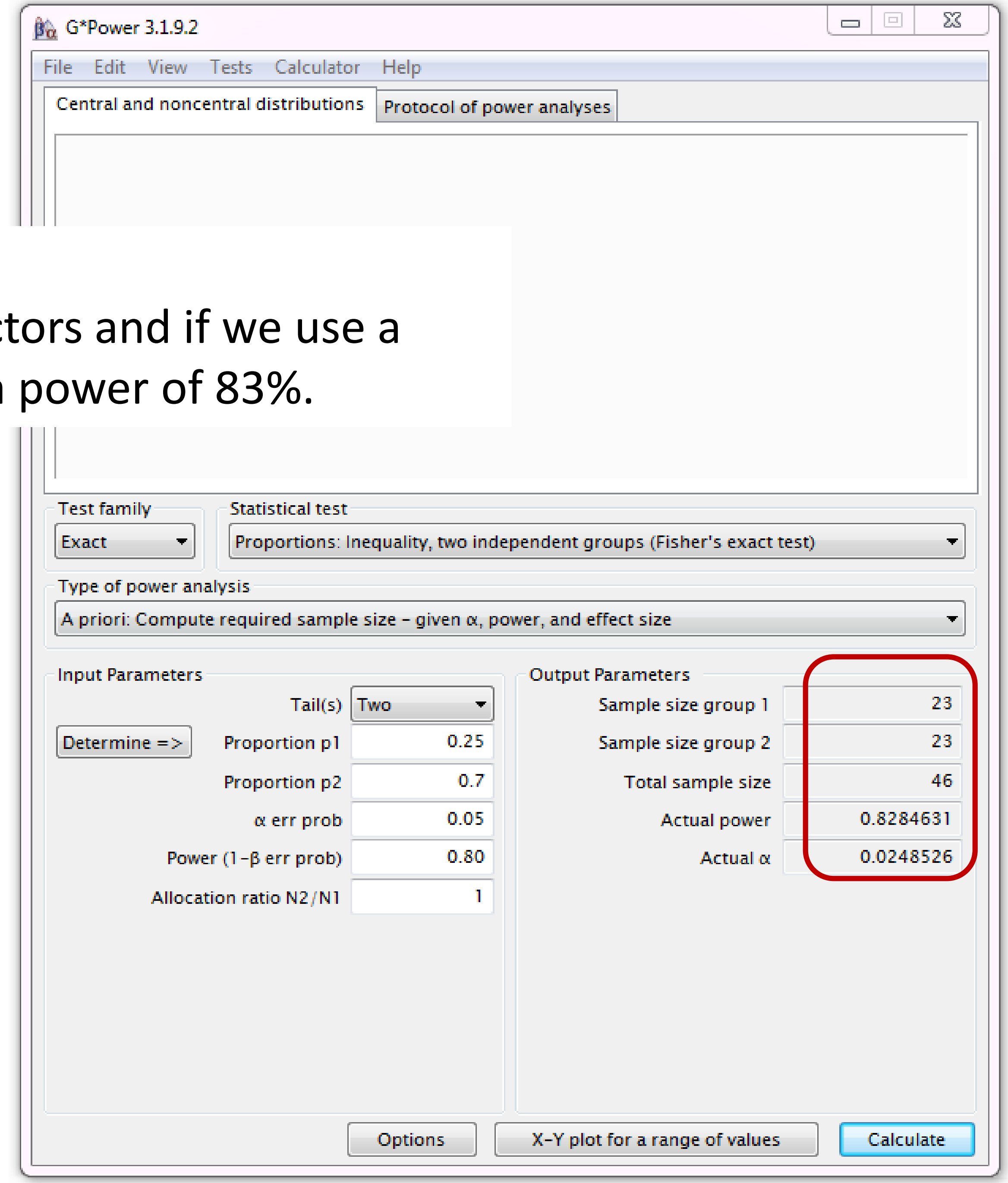
Exercise: Power calculation

- Preliminary results from a pilot study: **25%** line-danced after having received affection as a reward vs. **70%** after having received food.
  - **How many cats** do we need?

Exercise: Power calculation

Output:

If the values from the pilot study are good predictors and if we use a sample of **n=23 for each group**, we will achieve a power of 83%.



Chi-square and Fisher's tests

- Chi^2 test very easy to calculate by hand but Fisher's very hard

- Many software will not perform a Fisher's test on tables > 2x2

- **Fisher's test more accurate** than Chi^2 test on **small samples**
- **Chi^2 test more accurate** than Fisher's test on **large samples**

- Chi^2 test assumptions:
  - 2x2 table: no expected count <5
  - Bigger tables: all expected > 1 and no more than 20% < 5

Chi-square test

- In a chi-square test, the observed frequencies for two or more groups are compared with expected frequencies by chance.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

  - O = Observed frequencies
  - E = Expected frequencies

- **Example with ‘cats and dogs’**

How are the expected frequencies calculated?

Example: expected frequency of cats line dancing after having received food as a reward.

Direct counts approach:

Expected frequency = (row total)*(column total)/grand total

= **32*32/68** = **15.1**

Probability approach: The Multiplicative Rule

Probability of line dancing: **32/68**
Probability of receiving food: **32/68**

Expected frequency:(**32/68**)*(**32/68**)=0.22: **22% of 68 = 15.1**

Observed frequencies

	Food	Affection	Total
Dance	**26**	6	**32**
No dance	6	30	36
Total	**32**	36	**68**

Expected frequencies

	Food	Affection
Dance	**15.1**	16.9
No dance	16.9	19.1

Chi^2 test

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Observed frequencies

	Food	Affection
Dance	26	6
No dance	6	30

Expected frequencies

	Food	Affection
Dance	15.1	16.9
No dance	16.9	19.1

$Chi^2 = (26-15.1)^2/15.1 + (6-16.9)^2/16.9 + (6-16.9)^2\ /16.9 + (30-19.1)^2/19.1 = 28.4$

Is 28.4 big enough for the test to be significant?

Is 28.4 big enough for the test to be significant?

The old fashion way

Degree of freedom: df

df = (row-1)(col-1)=1

	Food	Affection
Dance	26	6
No dance	6	30

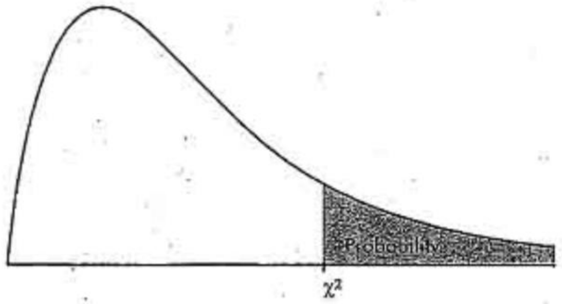


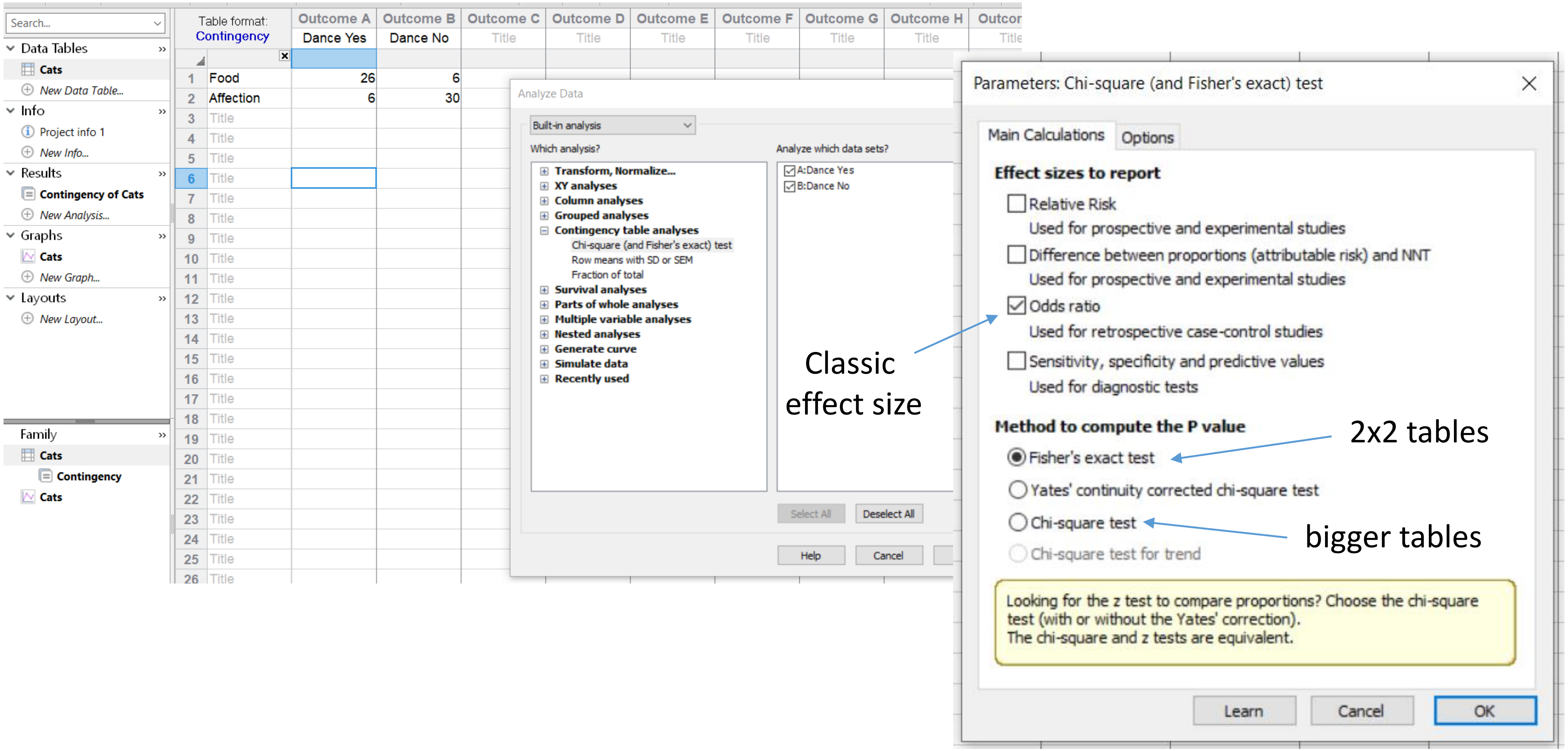
TABLE C: χ^2 CRITICAL VALUES

	Tail probability p								
df	.25	.20	.15	.10	.05	.025	.02	.01	.005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19

Critical value

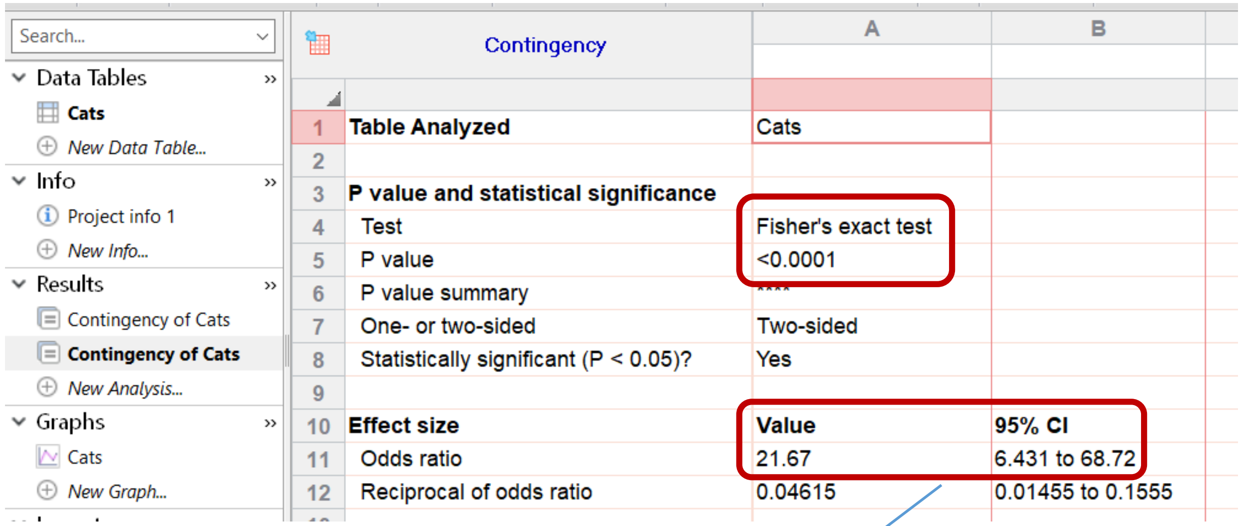
χ^2 = 28.4 > 3.84 so Yes!

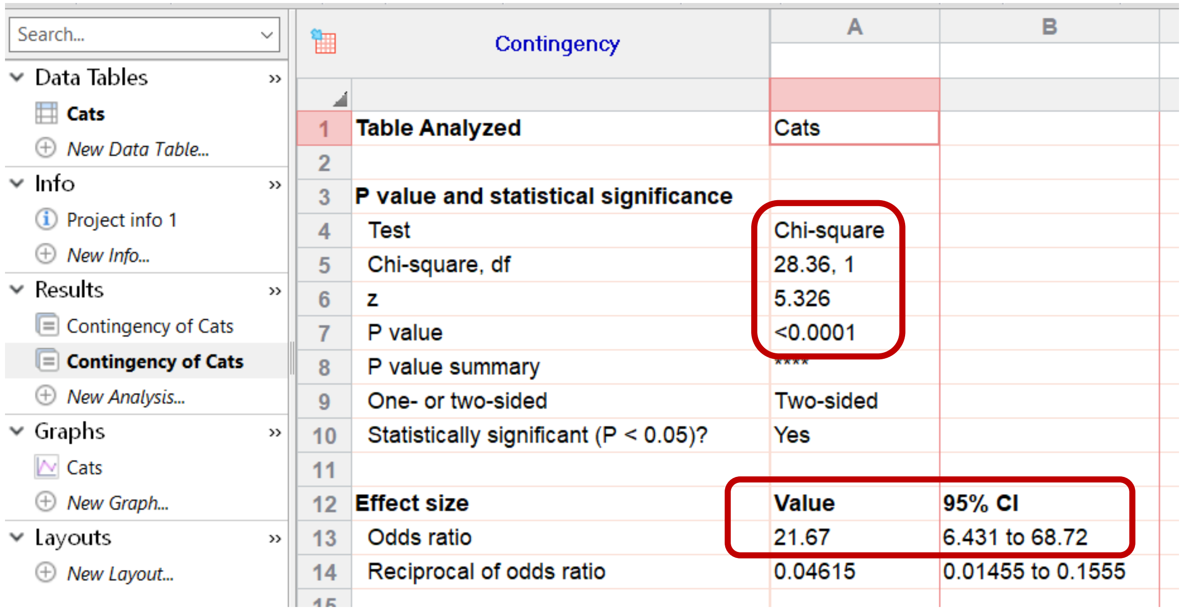
Fisher's exact and Chi^2 tests with Prism 8



Fisher's exact and Chi^2 tests

Results





	Food	Affection
Dance	26	6
No dance	6	30

Odds of dancing on Food group = **26/6**

Odds of dancing on Affection group = **6/30**

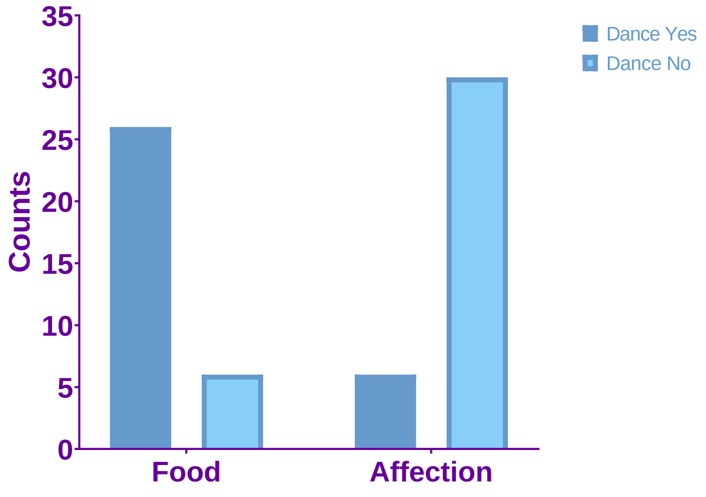
O.R. = ratio of the odds = 26/6**/**6/30 = **21.7**

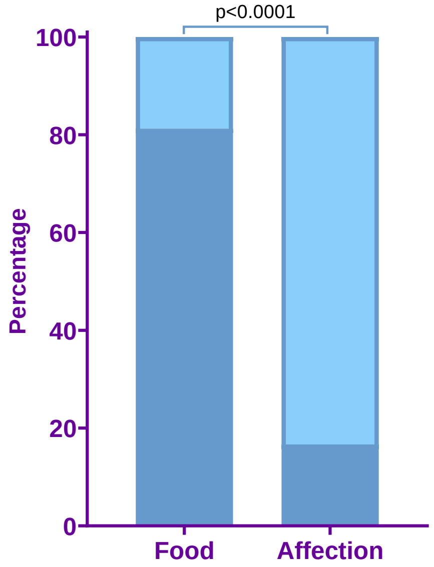
Odds Ratio = 21.7

If you are a dancing cat, you are almost 22 times more likely to have received food than affection as a reward (p<0.0001).

Fisher's exact and Chi^2 tests with Prism 8
Beyond significance

- Two super important things to keep in mind:
    - ❖ Qualitative data can be presented as percentages but the **tests should always be run on actual counts.**
        - ❖ Power!
    - ❖ A **p-value should always be interpreted in the context of the experiment.**
        - ❖ Power!





Let's do it with the dogs

Results for cats and dogs

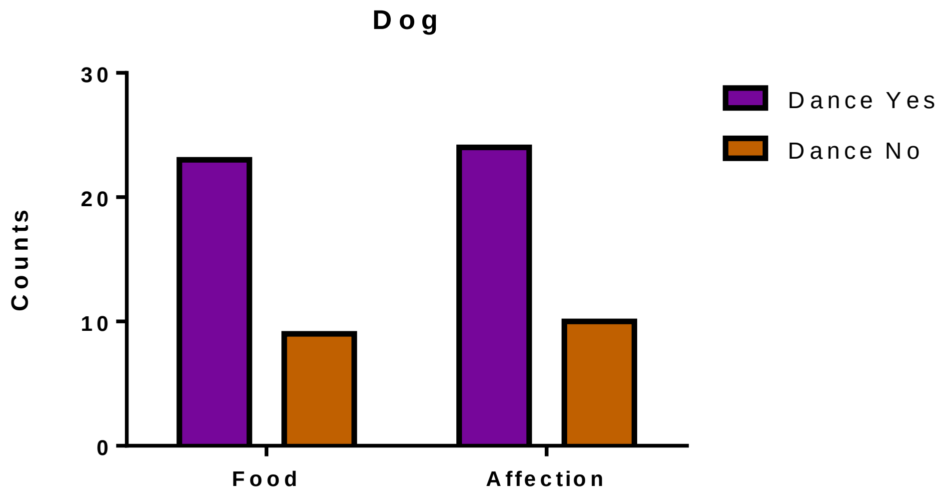
Table Analyzed	Cat
P value and statistical significance	
Test	Chi-square
Chi-square, df	28.36, 1
z	5.326
P value	<0.0001
P value summary	****
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	Yes

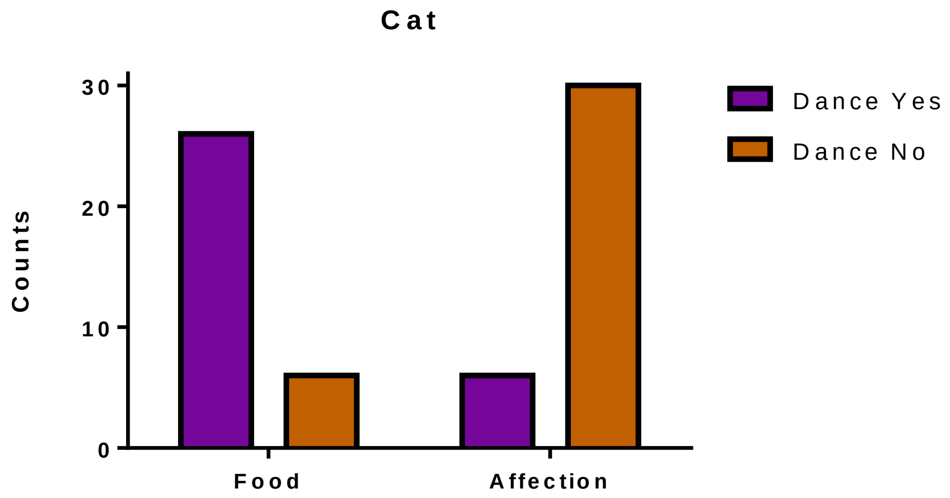
1	Table Analyzed	Cat
2		
3	Fisher's exact test	
4		
5	P value	< 0.0001
6	P value summary	***
7	One- or two-sided	Two-sided
8	Statistically significant? (alpha<0.05)	Yes

Table Analyzed	Dog
P value and statistical significance	
Test	Chi-square
Chi-square, df	0.01331, 1
z	0.1154
P value	0.9081
P value summary	ns
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	No

Table Analyzed	Dog
P value and statistical significance	
Test	Fisher's exact test
P value	>0.9999
P value summary	ns
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	No

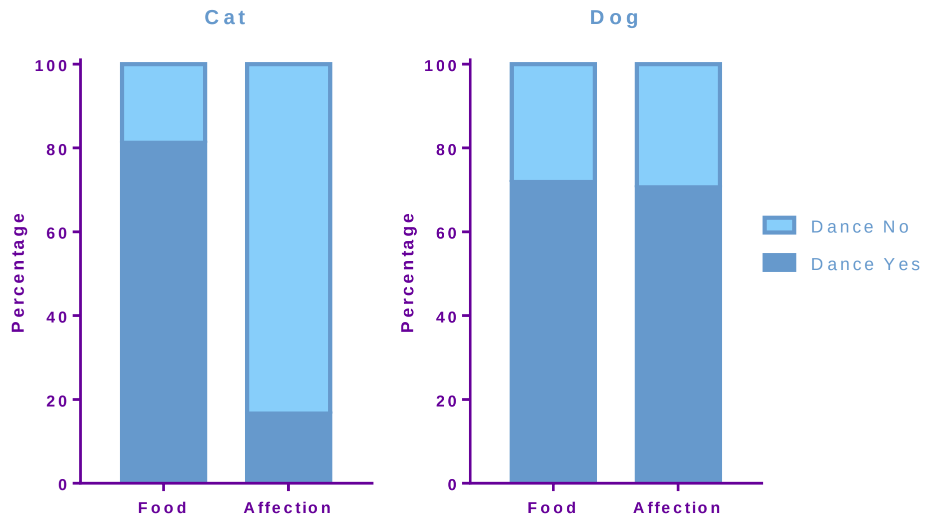
Fisher's exact test: results





- **In our example:**

cats are more likely to line dance if they are given food as reward than affection (p<0.0001) whereas dogs don't mind (p>0.99).



Exercise: Cane toads

	Infected	Uninfected
Rockhampton	12	8
Bowen	4	16
Mackay	15	5

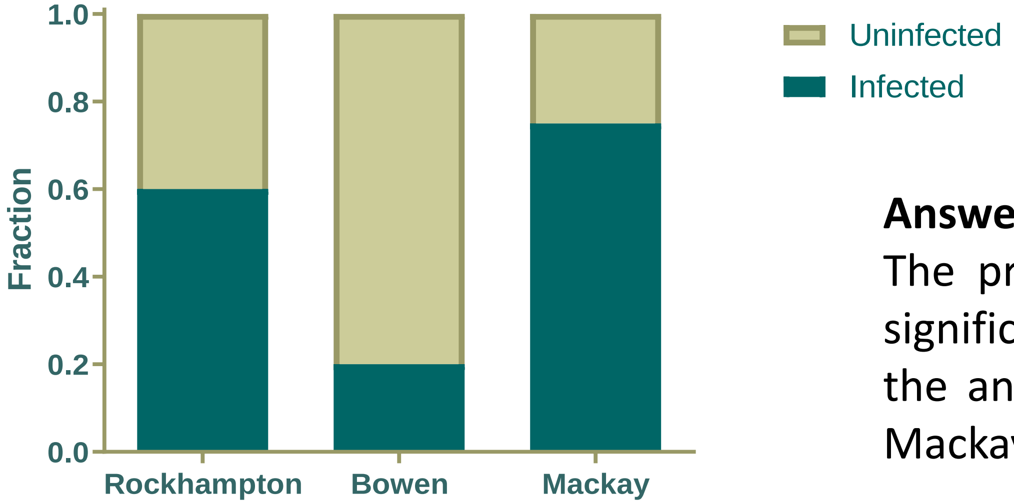
- A researcher decided to check the hypothesis that the proportion of cane toads with intestinal parasites was the same in 3 different areas of Queensland.

From Statistics Explained by Steve McKillup

- **Question**: Is the proportion of cane toads infected by intestinal parasites the same in 3 different areas of Queensland?

Exercise: Cane toads

Table Analyzed	Cane toad
Chi-square	
Chi-square, df	12.95, 2
P value	0.0015
P value summary	**
One- or two-tailed	NA
Statistically significant? (alpha<0.05)	Yes
Data analyzed	
Number of rows	3
Number of columns	2

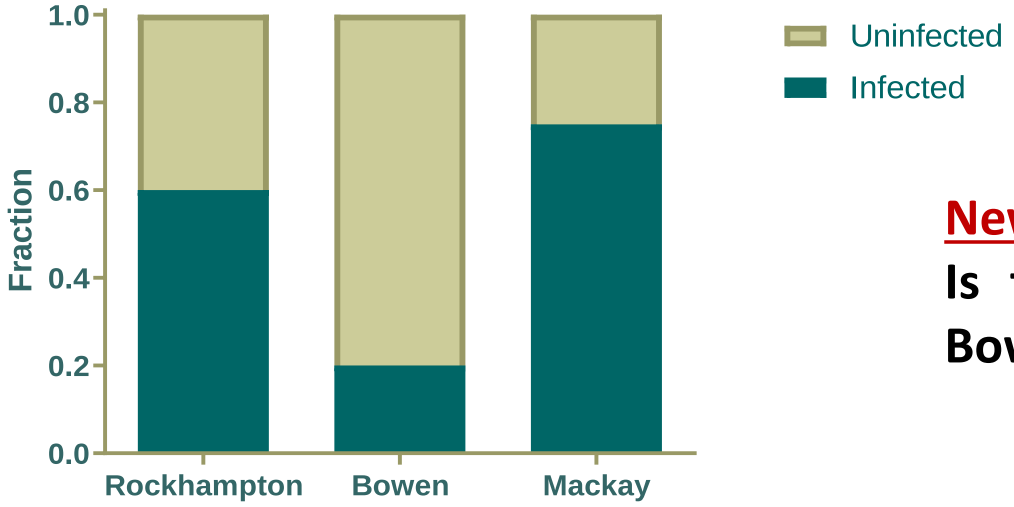


Answer:

The proportion of cane toads infected by intestinal parasites varies significantly between the 3 different areas of Queensland (p=0.0015), the animals being more likely to be parasitized in Rockhampton and Mackay than in Bowen.

Exercise: Cane toads

Table Analyzed	Cane toad
Chi-square	
Chi-square, df	12.95, 2
P value	0.0015
P value summary	**
One- or two-tailed	NA
Statistically significant? (alpha<0.05)	Yes
Data analyzed	
Number of rows	3
Number of columns	2



New question:

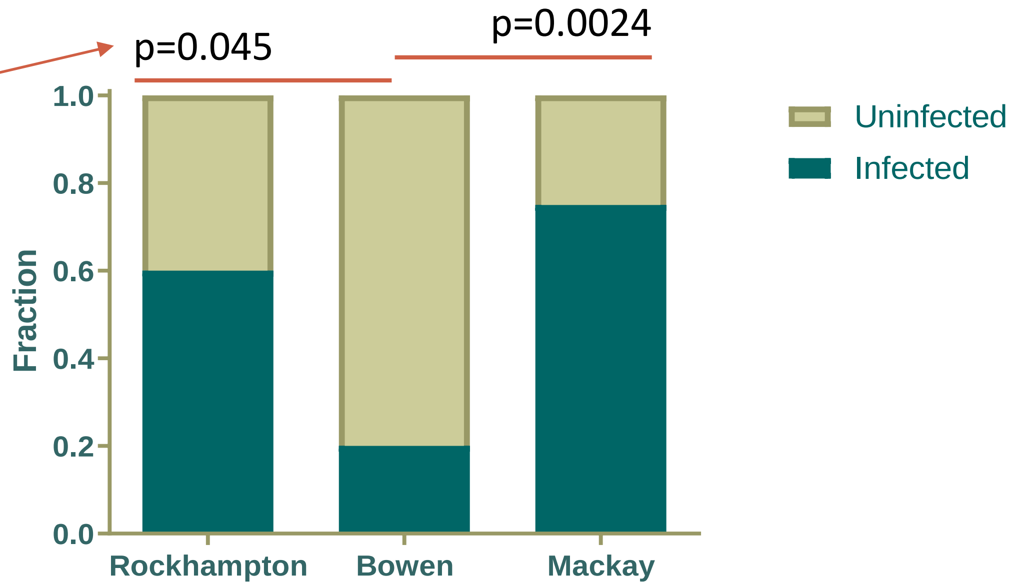
Is the proportion of infected cane toads different in Bowen than in the other 2 areas?

Exercise: Cane toads

P value and statistical significance	
Test	Fisher's exact test
P value	0.0225

P value and statistical significance	
Test	Fisher's exact test
P value	0.0012

Bonferroni correction



Is the proportion of infected cane toads different in Bowen than in the other 2 areas? Yes, it is.