Hypothesis Testing

Hypothesis testing

- An **objective** method of making decisions or **inferences** from sample data (evidence)
- Sample data used to choose between two choices i.e. **hypotheses** or statements about a population
- We typically do this by comparing what we have observed to what we expected if one of the statements (**Null Hypothesis**) was true

Hypothesis testing Framework
What the text books might say!

- Always two hypotheses:

  H_A: Research (Alternative) Hypothesis

  - What we aim to gather evidence of
  - Typically that there **is** a difference/effect/relationship etc.

  H_0: Null Hypothesis

  - What we assume is true to begin with
  - Typically that there is **no** difference/effect/relationship etc.

Discussion

- How could you help a student understand what hypothesis testing is and why they need to use it?

Could try explaining things in the context of “The Court Case”?

- Members of a jury have to decide whether a person is guilty or innocent based on evidence

  Null: The person is innocent

  Alternative: The person is not innocent (i.e. guilty)

- The null can only be rejected if there is enough evidence to doubt it
- i.e. the jury can only convict if there is beyond reasonable doubt for the null of innocence
- They do not know whether the person is really guilty or innocent so they may make a mistake

Types of Errors

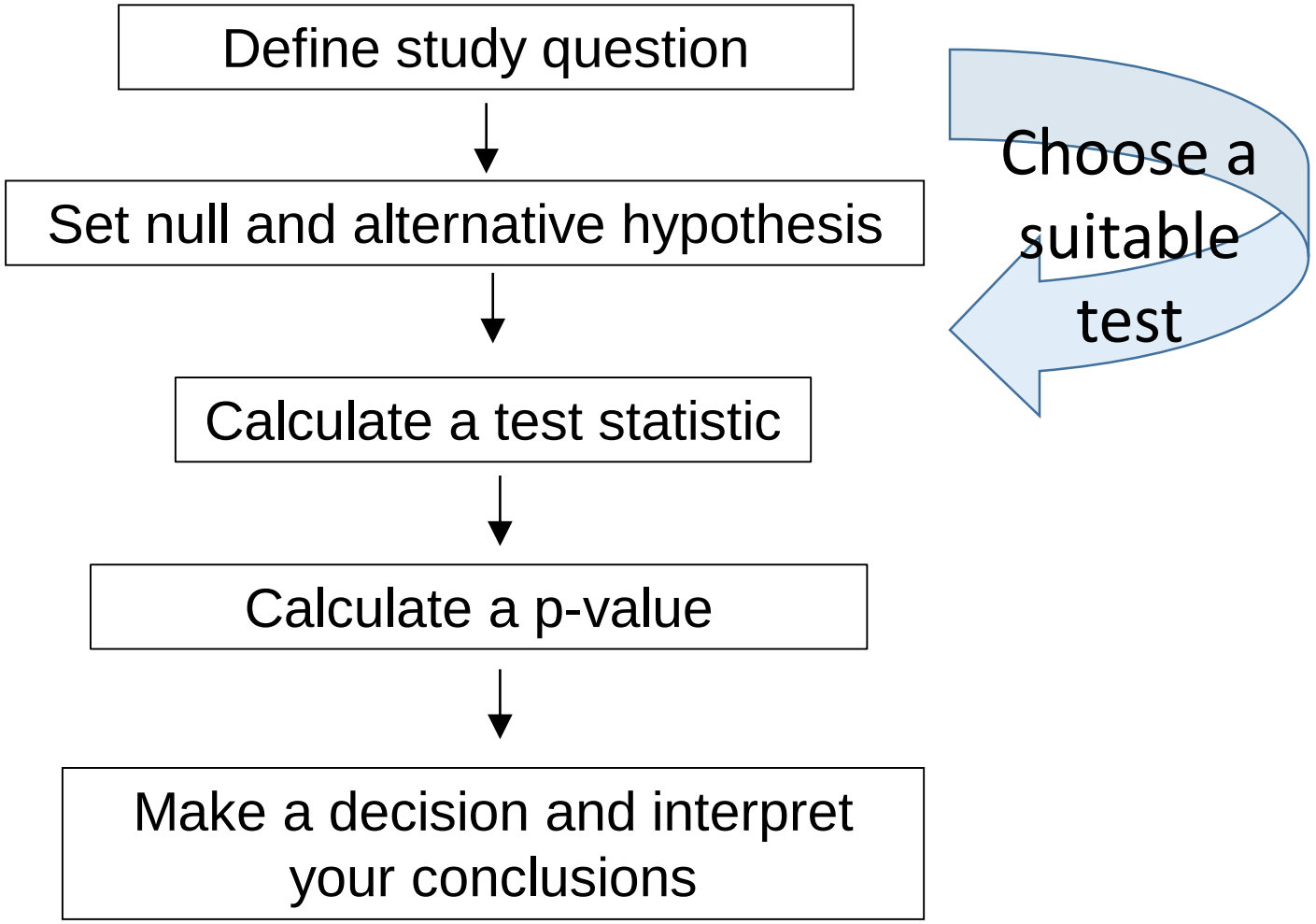
Controlled via sample size (=1-Power of test)

Typically restrict to a 5% Risk = level of significance

	Study reports **NO** difference (Do not reject H_0)	Study reports **IS** a difference (Reject H_0)
H_0 is true Difference Does **NOT** exist in population	✓	X Type I Error
H_A is true Difference **DOES** exist in population	X Type II Error	✓

Prob of this = Power of test

Steps to undertaking a Hypothesis test



Example: Titanic

- The ship Titanic sank in 1912 with the loss of most of its passengers
- 809 of the 1,309 passengers and crew died

  = 61.8%
- **Research question:** Did class (of travel) affect survival?

Chi squared Test?

- **Null:** There is **NO** association between class and survival
- **Alternative:** There **IS** an association between class and survival

3 x 2 contingency table

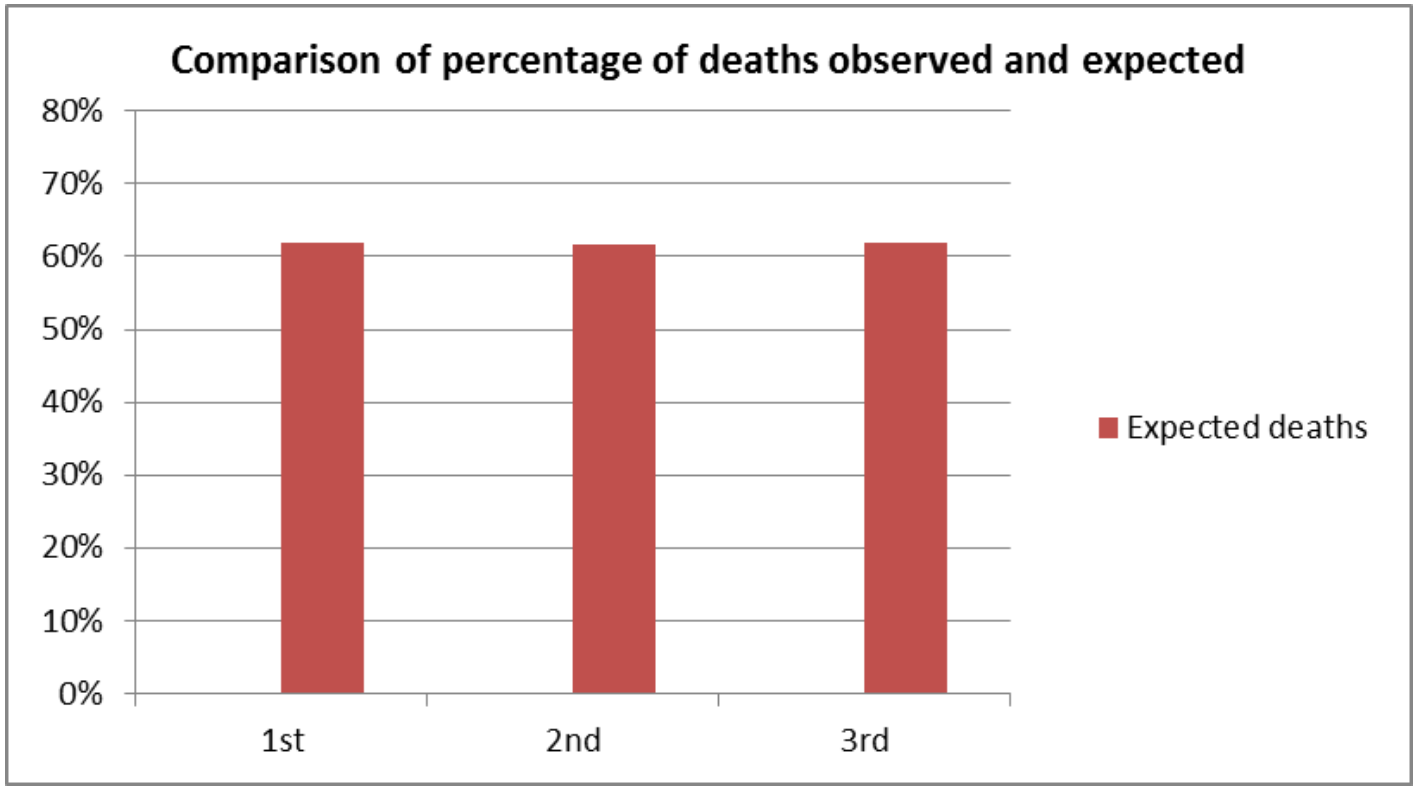
Class * Survived? Crosstabulation

Count

		Survived?		
		Died	Survived	Total
Class	1st	123	200	323
	2nd	158	119	277
	3rd	528	181	709
Total		809	500	1309

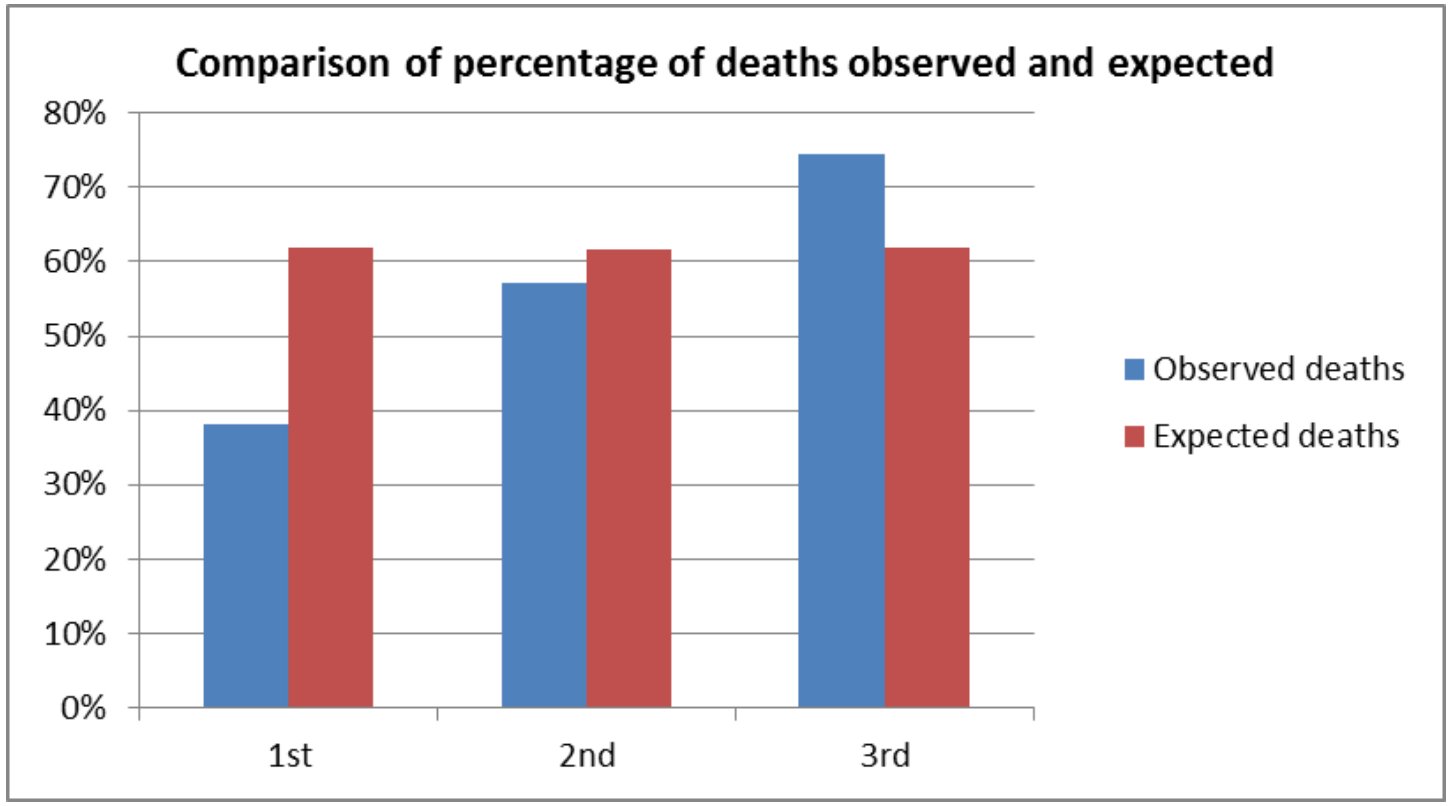
What would be expected if the null is true?

- Same proportion of people would have died in each class!
- Overall, 809 people died out of 1309 = 61.8%

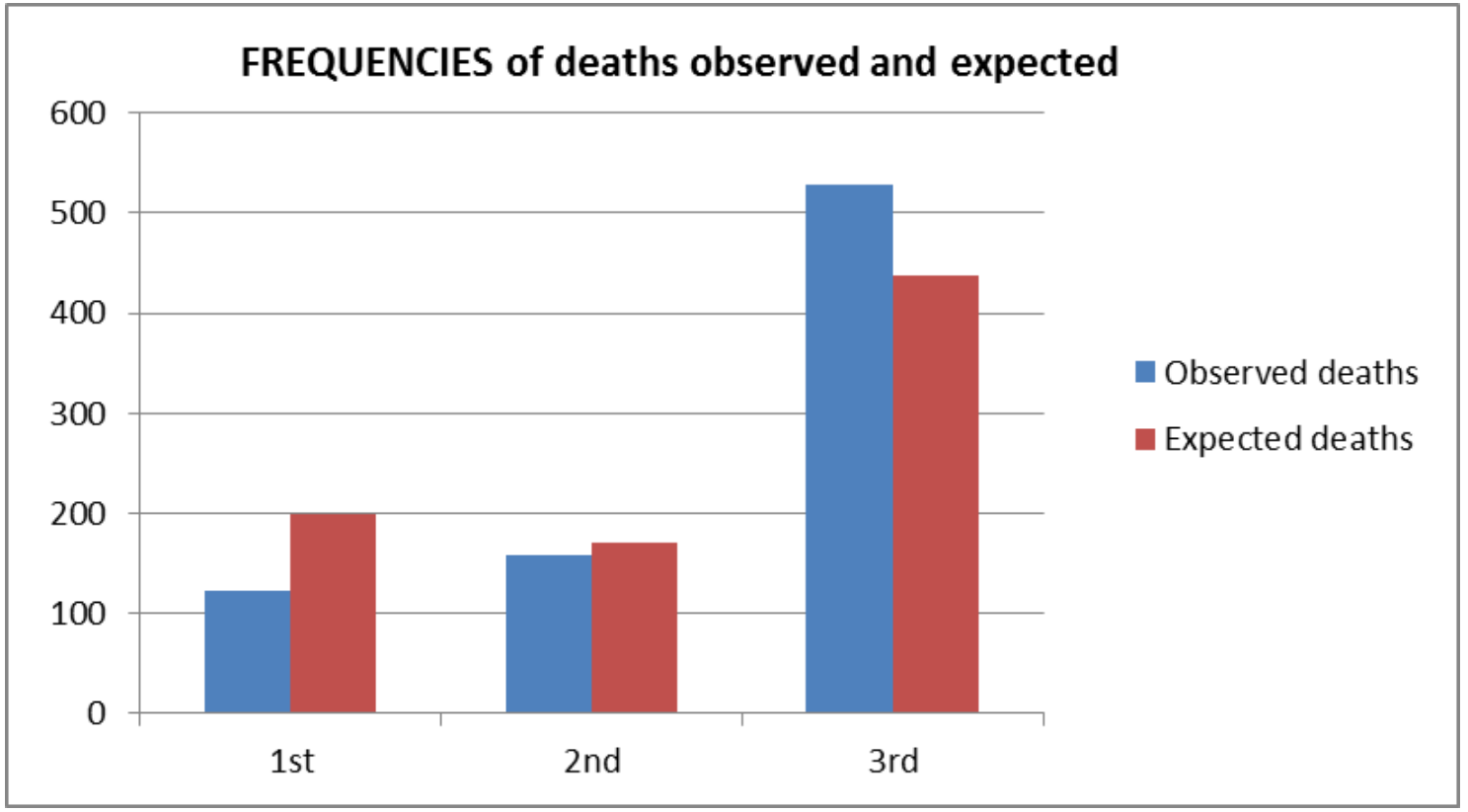


What would be expected if the null is true?

- Same proportion of people would have died in each class!
- Overall, 809 people died out of 1309 = 61.8%



Chi-Squared Test Actually Compares Observed and Expected Frequencies



Expected number dying in each class = 0.618 * no. in class

Chi-squared test statistic

- The chi-squared test is used when we want to see if two categorical variables are related
- The test statistic for the Chi-squared test uses the sum of the squared differences between each pair of observed (O) and expected values (E)

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(O_i - E_i\right)^2}{E_i}$$

Using SPSS

Analyse → Descriptive Statistics → Crosstabs

Click on ‘Statistics’ button & select Chi-squared

Test Statistic = 127.859

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	127.859[a]	2	.000
Likelihood Ratio	127.765	2	.000
Linear-by-Linear Association	127.709	1	.000
N of Valid Cases	1309		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 105.81.

p- value
p < 0.001

Note: Double clicking on the output will display the p-value to more decimal places

Hypothesis Testing: Decision Rule

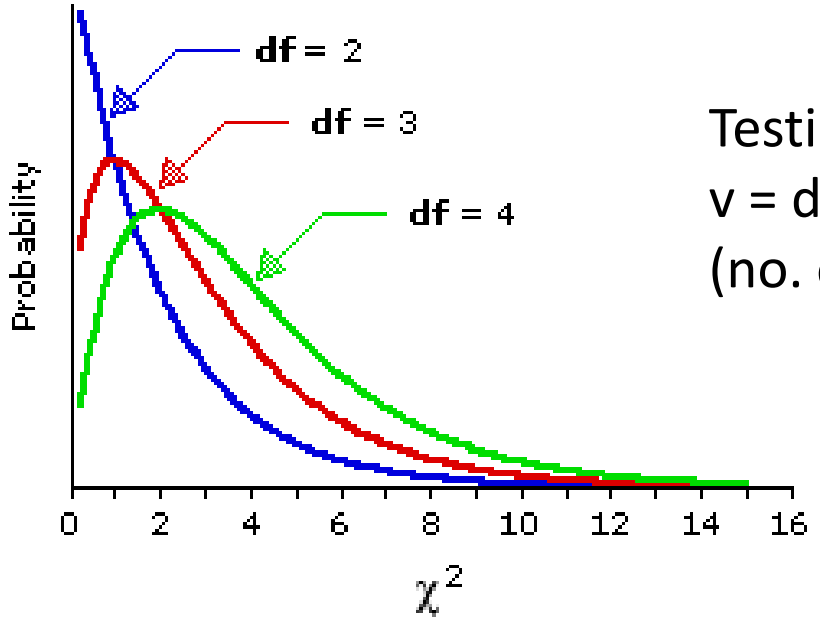
- We can use statistical software to undertake a hypothesis test e.g. SPSS
- One part of the output is the p-value (P)

- If P **< 0.05 reject** H_0 => **Evidence** of H_A being true (i.e. **IS** association)

- If P **> 0.05 do not** reject H_0 (i.e. **NO** association)

Chi squared distribution

- The p-value is calculated using the Chi-squared distribution for this test
- Chi-squared is a skewed distribution which varies depending on the degrees of freedom



Testing relationships between 2:
v = degrees of freedom
(no. of rows – 1) x (no. of columns – 1)

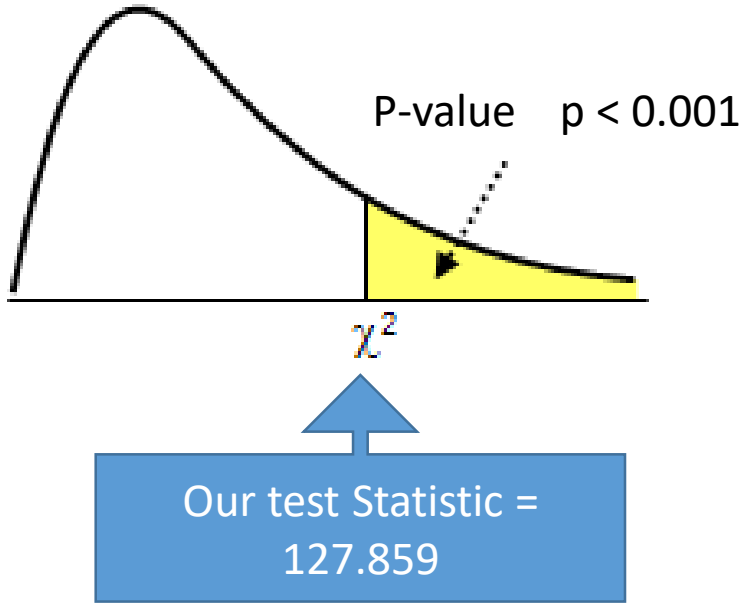
Note: One sample test:
v = df = outcomes – 1

What's a p-value?
The technical answer!

Probability of getting a test statistic at least as extreme as the one calculated **if the null is true**

In Titanic example, the probability of getting a test statistic of 127.859 or above (**if the null is true**) is < 0.001

Distribution of test statistics



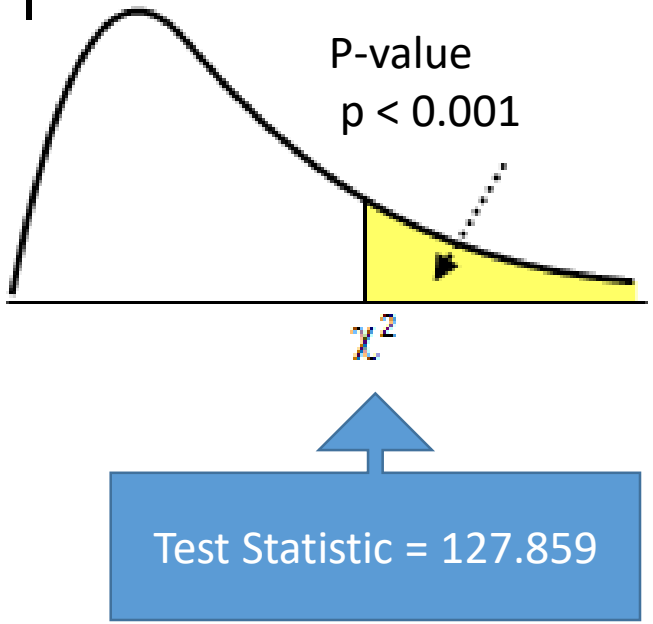
Interpretation

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	127.859[a]	2	.000

Since $p < 0.05$ we reject the null

There is evidence (χ^2_2=127.86, $p < 0.001$) to suggest that there is an association between class and survival

But… what is the nature of this association/relationship?



Titanic exercise

Were ‘wealthy’ people more likely to survive on board the Titanic?

Option 1:

- Choose the right percentages from the next slide to investigate
- Fill in the stacked bar chart with the chosen %’s
- Write a summary to go with the chart

Contingency tables exercise

Which percentages are better for investigating whether class had an effect on survival?

Column

Class * Survived? Crosstabulation

			Survived?		
			Died	Survived	Total
Class	1st	Count	123	200	323
		% within Survived?	15.2%	40.0%	24.7%
	2nd	Count	158	119	277
		% within Survived?	19.5%	23.8%	21.2%
	3rd	Count	528	181	709
		% within Survived?	65.3%	36.2%	54.2%
Total		Count	809	500	1309
		% within Survived?	100.0%	100.0%	100.0%

Row

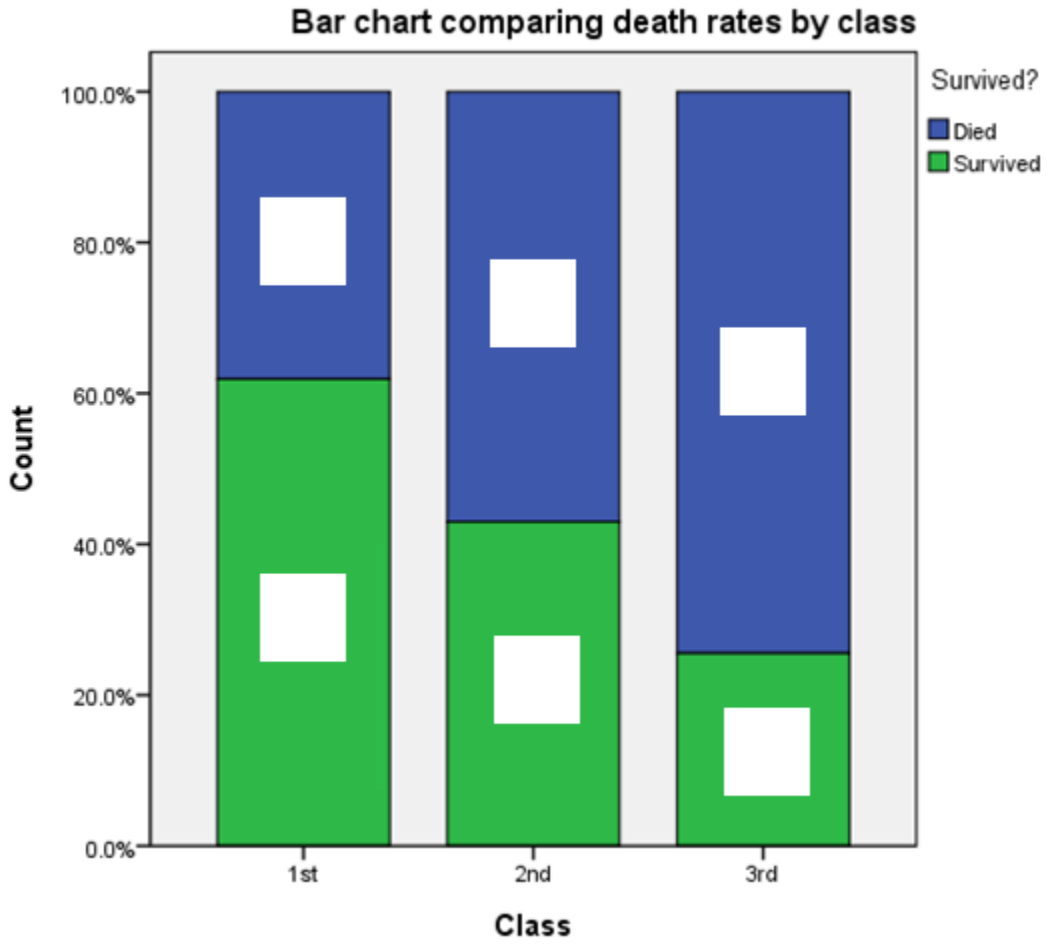
Class * Survived? Crosstabulation

			Survived?		
			Died	Survived	Total
Class	1st	Count	123	200	323
		% within Class	38.1%	61.9%	100.0%
	2nd	Count	158	119	277
		% within Class	57.0%	43.0%	100.0%
	3rd	Count	528	181	709
		% within Class	74.5%	25.5%	100.0%
Total		Count	809	500	1309
		% within Class	61.8%	38.2%	100.0%

65.3% of those who died were in 3rd class 74.5% of those in 3rd class died

Did class affect survival? Question

Fill in the %'s on the stacked bar chart and interpret



Did class affect survival? Solution

%'s within each class are preferable due to different class frequencies

pclass * survived Crosstabulation

			survived		Total
			Died	Survived	
pclass	1st	Count	123	200	323
		% within pclass	38.1%	61.9%	100.0%
	2nd	Count	158	119	277
		% within pclass	57.0%	43.0%	100.0%
	3rd	Count	528	181	709
		% within pclass	74.5%	25.5%	100.0%
Total		Count	809	500	1309
		% within pclass	61.8%	38.2%	100.0%

Did class affect survival? Solution

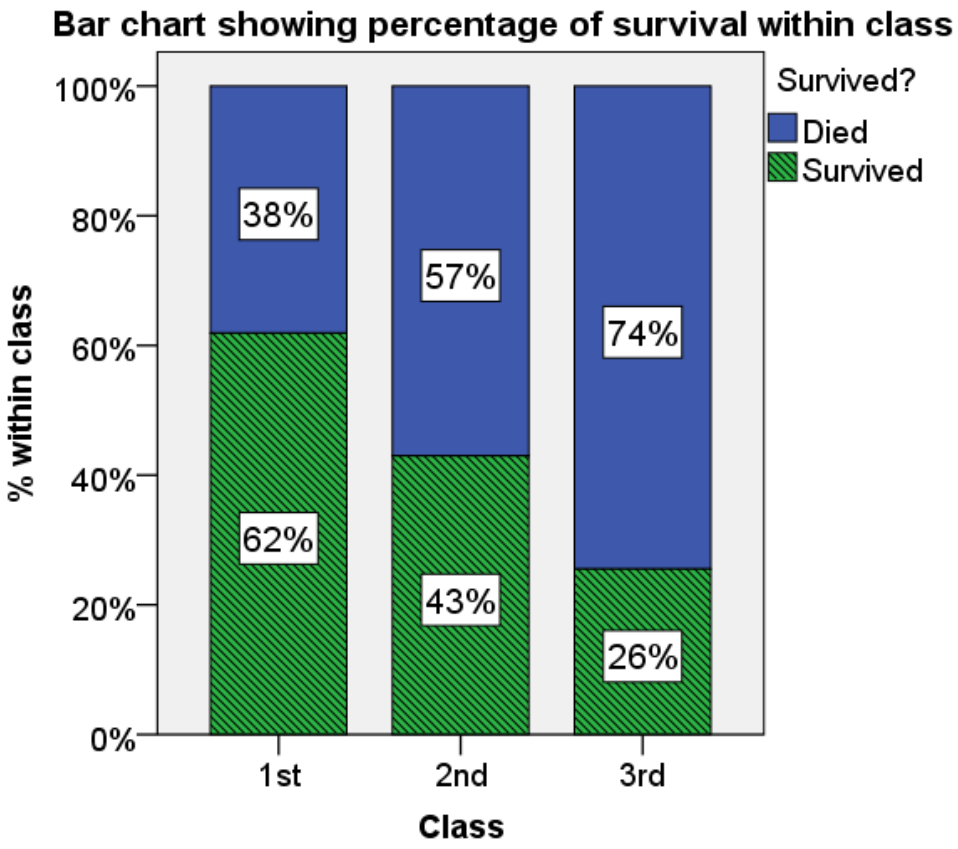


Figure 1: Bar chart showing % of passengers surviving within each class

Data collected on 1309 passengers aboard the Titanic was used to investigate whether class had an effect on chances of survival. There was evidence (χ^2_2=127.86, $p < 0.001$) to suggest that there is an association between class and survival.

Figure 1 shows that class and chances of survival were related. As class decreases, the percentage of those surviving also decreases from 62% in 1st Class to 26% in 3rd Class.

Low EXPECTED Cell Counts with the Chi-squared test

We have no cells with expected counts below 5

	Died	Survived	Total
1st Class	**200**	123	323
2nd Class	**171**	106	277
3rd Class	**438**	271	709
Total	809	500	1,309

SPSS Output

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	127.859[a]	2	.000
Likelihood Ratio	127.765	2	.000
Linear-by-Linear Association	127.709	1	.000
N of Valid Cases	1309		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 105.81.

Low Cell Counts with the Chi-squared test

- Check no. of cells with EXPECTED counts less than 5
- SPSS reports the % of cells with an expected count <5
- If more than 20% then the test statistic does not approximate a chi-squared distribution very well
- If any expected cell counts are <1 then cannot use the chi-squared distribution
- In either case if have a 2x2 table use **Fishers' Exact test** (SPSS reports this for 2x2 tables)
- In larger tables (3x2 etc.) combine categories to make cell counts larger (providing it's meaningful)