

SOME ADDITIONAL DATA IF YOU'D LIKE TO EXPLORE MORE

Here's some other data about the Titanic you could explore if you wanted to geek out more and deepen your own familiarity with the content. The statistics vary slightly from source to source so these numbers may deviate slightly from the data we explored earlier in the presentation.

Adults	Survivors		Non-Survivors	
	Male	Female	Male	Female
1st Class	57	140	118	4
2nd Class	14	80	154	13
3rd Class	75	76	387	89
Crew	192	20	670	3

Children	Survivors		Non-Survivors	
	Male	Female	Male	Female
1st Class	5	1	0	0
2nd Class	11	13	0	0
3rd Class	13	14	35	17
Crew	0	0	0	0

Using the data above, we can create some examples of frequency tables.

	Survivors	Non-Survivors	Total
Passenger	499	817	1316
Crew	212	673	885
Total	711	1490	2201

	Survivors	Non-Survivors	Total
1 st class	197	122	319
2 nd class	94	167	261
3 rd class	151	476	627
Total	442	765	1207

What patterns of association do you see?

What do the relative frequency tables suggest about any possible patterns of association?

What might be good titles for these tables?

What other patterns of association might exist in the data at the top of the page?

What kind of raw and relative frequency tables would you need to construct in order to see if those conclusions might be true?

SIMPSON'S PARADOX: A CONCEPTUAL EXTENSION FOR YOU TO PONDER

In 1999 at the University of Calizona at Los Phoenix (UCLP), 393 men applied to the graduate school and 294 were admitted for an admission rate of 74.8%. The same year 444 women applied and 135 were admitted for an admission rate of 30.4%. To UCLP administrators this strongly suggested sex bias favoring men in graduate admissions. To track down the source of the bias, administrators ordered the individual graduate programs at UCLP to report their admission rates for men and women in 1999. This was a simple task in that UCLP has graduate programs in only four fields: English, Physics, Psychology, and Materials Science.

	Men			Women		
	Accepted	Applied	Rate	Accepted	Applied	Rate
English	12	43	27.9%	48	130	36.9%
Physics	119	124	96.0%	8	8	100.0%
Psychology	7	60	11.7%	59	285	20.7%
Materials Scien	156	166	94.0%	20	21	95.2%
	294	393	74.8%	135	444	30.4%

The baffled administration at UCLP is at a loss to explain the situation. Every department admits women at a higher rate than men, but overall the university admits men at a much higher rate than women. This counterintuitive situation is a concrete example of Simpson's Paradox. This paradox asserts that event A may be positively relevant to event B in every block of some partition of the population and yet be negatively relevant in the population as a whole (e.g., being female is positively relevant to being admitted in every program yet negatively relevant to being admitted overall).

In our example the source of the problem is clear. The two departments that have low admission rates overall also have high numbers of applications from women and low numbers from men. In the two departments with high admission rates overall the situation is reversed. Thus a large percentage of men are admitted and a large percentage of women are not, even though every department admits women at a higher rate than men.

Here are some other scenarios in which Simpson's Paradox might arise:

–Baseball player Willie may have a higher batting average than player Hank every year of their careers, and yet Hank may have the higher lifetime batting average.

–Cancer treatment A may produce a higher recovery rate than treatment B at every hospital in which both are tested, and yet treatment B may have the higher overall recovery rate when the data are combined.

(A quick note on the source: I pulled this example of Simpson's Paradox from the internet. It's from a college course at the University of Kentucky. I could not find the author. The document can be easily found by copying a sentence into a search engine.)