Handout: Three way tables

Data from IMDb (January 11, 2011)
Y= rating for movies X = horror(Alien) or animated movie (Wall-e) controlling for Z=sex (age group).
Male

	Rating					
Movie	1-2	3-4	5-6	7-8	9-10	Total
Alien	1502	1404	7090	49158	77097	136251
Wall-e	5654	2261	8199	43116	94079	153309
Total	7156	3665	15289	92274	171176	289560

Female

	Rating					
Movie	1-2	3-4	5-6	7-8	9-10	Total
Alien	430	356	1313	5075	6777	13951
Wall-e	1104	441	1274	5998	19106	27923
Total	1534	797	2587	11073	25883	41874

The partial tables describe the conditional joint distribution of X and Y (conditional on the different levels of Z).
Given that the rater was female the probability for rating Alien 9 or 10 equals 6777/41874.

movie * rating * sex Crosstabulation

sex				rating					Total
				12	34	56	78	910	
f	movie	Alien	Count	430	356	1313	5075	6777	13951
			% within movie	3.1%	2.6%	9.4%	36.4%	48.6%	100.0%
			% within rating	28.0%	44.7%	50.8%	45.8%	26.2%	33.3%
		Wall-e	Count	1104	441	1274	5998	19106	27923
			% within movie	4.0%	1.6%	4.6%	21.5%	68.4%	100.0%
			% within rating	72.0%	55.3%	49.2%	54.2%	73.8%	66.7%
	Total		Count	1534	797	2587	11073	25883	41874
			% within movie	3.7%	1.9%	6.2%	26.4%	61.8%	100.0%
			% within rating	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
m	movie	Alien	Count	1502	1404	7090	49158	77097	136251
			% within movie	1.1%	1.0%	5.2%	36.1%	56.6%	100.0%
			% within rating	21.0%	38.3%	46.4%	53.3%	45.0%	47.1%
		Wall-e	Count	5654	2261	8199	43116	94079	153309
			% within movie	3.7%	1.5%	5.3%	28.1%	61.4%	100.0%
			% within rating	79.0%	61.7%	53.6%	46.7%	55.0%	52.9%
	Total		Count	7156	3665	15289	92274	171176	289560
			% within movie	2.5%	1.3%	5.3%	31.9%	59.1%	100.0%
			% within rating	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Ignoring Z entirely results in the marginal table for X an Y. To obtain the marginal table just add the equivalent table entries of the partial tables.

Rows: movie Columns: rating

	12	34	56	78	910	All
Alien	1932	1760	8403	54233	83874	150202
Wall-e	6758	2702	9473	49114	113185	181232
All	8690	4462	17876	103347	197059	331434

Cell Contents -- Count

We can now use these tables to test for conditional independence and marginal independence.
From SPSS:

Test	χ^2	df	P
conditional independence (female)	1793.102	4	< 0.001
conditional independence (male)	3778.475	4	< 0.001
marginal independence	4692.375	4	< 0.001

At significance level of 5% the data provide sufficient evidence that the rating is not independent from the movie (not marginally, and not conditionally).

We can also do conditional and marginal trend tests (I used scores 1,2,3,4,5) (This is a Wilcoxon Test): Use SPSS (1 = Alien, 2 = Wall-e) to find the correlations, and $M_0^2 = (r^2(n-1)$:

Test	r	M_0^2	df	P
trend (conditional female)	0.188	1480.0	1	< 0.001
trend (conditional male)	-0.024	166.79	1	< 0.001
marginal trend	-0.006	11.93	1	< 0.001

At significance level of 5% the data provide sufficient evidence that the median ratings for Alien and Wall-e conditional for women and men and combined are not the same.
The sign of r indicates that the median rating by women for Wall-e is higher than for Alien, and Alien is higher rated by men.
The marginal result indicates that the median rating for Alien is higher overall than for Wall-e. (We are missing an important fact here, and this result is only due to the fact that men are much more likely to give a rating on IMDb).

Simpson's Paradox

In a paper by Charig et al. (British Medical Journal, Clinical Research Ed., March 1986, 292(6524): 879-882) two different treatments for kidney stones are compared. Call the treatments A and B.
It is reported that treatment A is successful for 78% of patients and treatment B is successful for 83% of patients. So one would assume that treatment B is "better" than treatment A.
BUT they also report that for small kidney stones treatment A is successful for 93% of all cases, but B only in 87% of all cases
and that for large kidney stones treatment A is successful for 73% of all cases, but B only in 69% of all cases.
So treatment A is "better" than treatment B?

Treatment A

	Success	No Success	Total
small	81	6	87
large	192	71	263
Total	273	77	350

Treatment B

	Success	No Success	Total
small	234	46	270
large	55	25	80
Total	289	71	350

Treatment A is better than treatment B, because it has a higher success rate for small and large kidney stones, i.e. 93% instead of 87%, and 73% instead of 69%.
How come, that overall treatment B seems to be the better treatment?
Both treatments are much more successful for small kidney stones than for large ones, and the sample for treatment B mostly consists of small kidney stones, but the sample for treatment A mostly consists of patients treated for large kidney stones.
As a result the overall success rate for treatment B is closer to the success rate for small kidney stone (looks good), but the overall success rate for treatment A is closer to the success rate for large kidney stones (looks not as good).