1.4 Three-way Tables

Example 1

In this example we will illustrate how the results in the analysis of the movie rating example from class were obtained.

We consider a three way table defined by categorical variables, X, Y, and Z.

Here:

Y response variable (Rating(score as 1,2,3,4,5)

X explanatory variable (Movie: Alien (1) or Wall-e(2))

Z explanatory (confounding) variable (Gender: Female (1) and Male (2))

The data for this example is in the file "alienwalle.sav".

The tools presented here do not permit to analyse the effect of Movie and Gender on the Rating simultaneously, but they allow to test for a first indication if gender has an effect on the relationship between Movie and Rating.

A χ^2 test for conditional independence and test for conditional trend, based on the correlation, r, using the chosen scores 1, 2, 3, 4, 5 considers the data for each gender (female, male) separately.

A χ^2 test for marginal independence and Correlation, r , to test marginal trend ((for chosen scores: 1,2,3,4,5) considers the data for both genders (ignoring that such a distinction exists) together.

SPSS Commands and test for independence output:

1. Data>Weight Cases
   Weight by Frequency Variable: Freq
   OK

2. Analysis>Descriptive Statistics>Crosstabs
   Row(s): `movie_n`
   Column(s): `score`
   Layer 1 of 1: `sex_n`

3. Statistics popup:
   Check Chi-square
   Check Correlation

4. Cells popup: Counts: check Observed
   check Expected

Chi-Square Tests

sex_n		Value	df	Asymptotic Significance (2-sided)
1.00	Pearson Chi-Square	1793.102[b]	4	.000
	Likelihood Ratio	1758.419	4	.000
	Linear-by-Linear Association	584.354	1	.000
	N of Valid Cases	41874		
2.00	Pearson Chi-Square	3778.475[c]	4	.000
	Likelihood Ratio	3927.323	4	.000
	Linear-by-Linear Association	160.043	1	.000
	N of Valid Cases	289560		
Total	Pearson Chi-Square	4692.375[a]	4	.000
	Likelihood Ratio	4823.106	4	.000
	Linear-by-Linear Association	10.671	1	.001
	N of Valid Cases	331434		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 2022.13.

b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 265.53.

c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 1724.55.

The contingency table:

movie_n * score * sex_n Crosstabulation

sex_n				score 1.00	2.00	3.00	4.00	5.00	Total
1.00	movie_n	1.00	Count	430	356	1313	5075	6777	13951
			Expected Count	511.1	265.5	861.9	3689.1	8623.3	13951.0
		2.00	Count	1104	441	1274	5998	19106	27923
			Expected Count	1022.9	531.5	1725.1	7383.9	17259.7	27923.0
	Total		Count	1534	797	2587	11073	25883	41874
			Expected Count	1534.0	797.0	2587.0	11073.0	25883.0	41874.0
2.00	movie_n	1.00	Count	1502	1404	7090	49158	77097	136251
			Expected Count	3367.2	1724.5	7194.2	43419.1	80546.0	136251.0
		2.00	Count	5654	2261	8199	43116	94079	153309
			Expected Count	3788.8	1940.5	8094.8	48854.9	90630.0	153309.0
	Total		Count	7156	3665	15289	92274	171176	289560
			Expected Count	7156.0	3665.0	15289.0	92274.0	171176.0	289560.0
Total	movie_n	1.00	Count	1932	1760	8403	54233	83874	150202
			Expected Count	3938.2	2022.1	8101.2	46835.6	89304.8	150202.0
		2.00	Count	6758	2702	9473	49114	113185	181232
			Expected Count	4751.8	2439.9	9774.8	56511.4	107754.2	181232.0
	Total		Count	8690	4462	17876	103347	197059	331434
			Expected Count	8690.0	4462.0	17876.0	103347.0	197059.0	331434.0

Symmetric Measures

sex_n			Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
1.00	Interval by Interval	Pearson's R	.118	.005	24.344	.000[c]
	Ordinal by Ordinal	Spearman Correlation	.181	.005	37.564	.000[c]
	N of Valid Cases		41874			
2.00	Interval by Interval	Pearson's R	-.024	.002	-12.654	.000[c]
	Ordinal by Ordinal	Spearman Correlation	.029	.002	15.343	.000[c]
	N of Valid Cases		289560			
Total	Interval by Interval	Pearson's R	-.006	.002	-3.267	.001[c]
	Ordinal by Ordinal	Spearman Correlation	.048	.002	27.571	.000[c]
	N of Valid Cases		331434			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

In all three output tables the results on the conditional analysis for each gender according to the label is given first, and underneath one can find the result on the marginal analysis (labeled: Total).

- In the **contingency table** the first section shows the rating by women, the second section shows the rating by men, and the third section shows the rating for all. The expected cell counts illustrate where the biggest difference in rating can be found.

- **Chi-Square Tests:**
  For women and men the results show that separately at significance level of 5% the rating is not independent from the movie, which means that the rating depends on the movie (women: $X^2(4) = 1793.102, P - value < 0.001$, men: $X^2(4) = 3778.475, P - value < 0.001$).

  The output also indicates that at significance level of 5% movie and rating are also not marginally independent (collapsed across gender) ($X^2(4) = 4692.375, P - value = 0.001$).

- **Symmetric Measures** (Test for trends):
  Use Pearson's correlation. At significance level of 5% the marginal table and both conditional distributions demonstrate a trend for movie and rating (women: $M_0^2(1) = 24.344^2, P-value < 0.001$, men: $M_0^2(1) = (-12.654)^2, P-value < 0.001$, marginal: $M_0^2(1) = -3.264^2, P-value = 0.001$), with the correlation being positive for women's rating ($r = .118$) but negative for men ($r = -0.24$) and the marginal table ($r = -0.006$). This is indicating that the median rating by women was higher for Wall-e(2) than for Alien(1), but men and men and women combined rate Alien higher.

  Seeing that the trends for women and men are different should inform us not combine the sample for a marginal analysis because there were many more men in the sample than women.