Chapter 6
Contingency tables

Contingency tables

- Summarize counts in categories defined by two (or more) variables (predictors)
- May be analyzed by a χ^2-test
- H0: observed frequencies of combinations correspond to proportions derived from marginal sums
- DF = (*n* rows – 1) * (n colums -1)

		Hair color			
		black	brown	blonde	marginal sums
Eye color	blue	12	45	14	71
	brown	51	256	84	391
	marginal sums	63	301	98	grand total: 462

2 x 2 Tables

- Special simple case of a contingency table
- 2 rows, 2 columns
- Strength of the association between the two predictors can be computed
- Phi – coefficient $\varphi = \frac{f11f22 - f12f21}{\sqrt{R1R2C1C2}} = \pm\sqrt{\frac{\chi^2}{n}}$
  - 0 = independence
  - -1 = full negative correlation
  - 1 = full positive correlation
  - Corresponds to Pearson *r* for quantitiative data

		Var2		
		level 1	level 2	Sums
Var 1	level 1	f11	f12	R1
	level 2	f21	f22	R2
	Sums	C1	C2	n

Odds and Odds-ratios

- Allow a detailed analysis of contingency tables beyond the rejection of independence of the two variables
- Useful for *n* x 2 tables
- Based on a subdivision of the table into several 2 x 2 tables
- Odds = probability of one outcome of a variable within one category of the other variable
- Odds ratio = ratio of two odds
  - Can be tested for significance

		Hair color			
		black	brown	blonde	marginal sums
Eye color	blue	12	45	14	71
	brown	51	256	84	391
	marginal sums	63	301	98	grand total: 462

Odds blue eyes in black-haired people = 12/63 = 0.19

Odds blue eyes in brown-haired people = 45/301 = 0.15

Odds ratio (blue-eyed in black vs. brown-haired people)= 0.19/0.15 = 1.27