Contingency tables

Contingency tables

- Contingency tables show frequencies produced by cross-classifying observations
- Example, aspects of career development and group age– is career development related to group age?

	Learning new skills	Pay increases	Career path	Total
18-39	7	20	16	43
40-60	18	22	8	48
Total	25	42	24	91

Contingency tables

- Baseline comparison: counts that would occur by random chance if the variables are independent
- Each cell E_{ij} = (i-th col total * j-th row total)/table total

	Learning new skills	Pay increases	Career path	Total
18-39	11.8	19.8	11.3	43
40-60	13.2	22.2	12.7	48
Total	25	42	24	91

Contingency tables

- Significance is the probability of obtaining, by chance, a table as or more deviant than the observed table, if the variables are independent
- Note: no causality is necessarily implied by the outcome

	Learning new skills	Pay increases	Career path	Total
18-39	(7) 11.8	(20) 19.8	(16) 11.3	43
40-60	(18) 13.2	(22) 22.2	(8) 12.7	48
Total	25	42	24	91

Contingency tables

- Chi-square is a test statistic that aggregates the information in table – the greater the deviation from expected values, the larger is the statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

	Learning new skills	Pay increases	Career path	Total
18-39	(7) 11.8	(20) 19.8	(16) 11.3	43
40-60	(18) 13.2	(22) 22.2	(8) 12.7	48
Total	25	42	24	91

Contingency tables

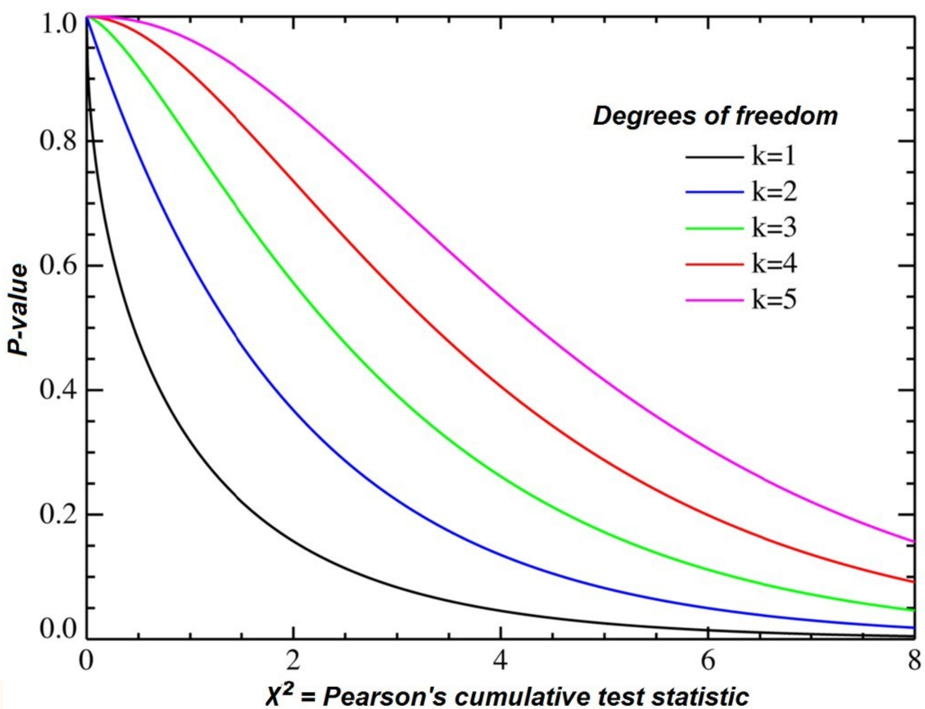
- Chi-square is a test statistic that aggregates the information in table – the greater the deviation from expected values, the larger is the statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 7.3$$

	Learning new skills	Pay increases	Career path	Total
18-39	2.00	0.0	1.9	
40-60	1.8	0.0	1.7	
Total				

Contingency tables

- Chi-square is distributed *approximately* according to the Chi-square probability distribution
- Chi-square probabilities are found in tables. Needs to know the degrees of freedom: df = (r-1)*(c-1)



eases	Career path	Total
	1.9	
	1.7	

X^2 = 7.3   P-value = .025

Contingency tables

- Chi-squared test in R

```
> f <- matrix(c(7,20,16,18,22,8),nr=2,byrow=T)
> f
     [,1] [,2] [,3]
[1,]    7   20   16
[2,]   18   22    8
> chisq.test(f)

    Pearson's Chi-squared test

data:  f
X-squared = 7.3494, df = 2, p-value = 0.02536
```

Contingency tables

- Chi-square must be based only on counts and each subject must contribute only to a single cell
- Rules of thumb: No expected counts less than 5
- Can combine columns/rows to increase expected counts that are too low – but it may reduce interpretability
- It says nothing about which parts of the table are responsible for a of association
- It a measure of *significance* but it is not a good measure of *strength* large chi-square statistic

Contigency Tables

Evaluation of a new user interface

	Male	Female
Like	165	300
Don’t like	176	81

H_0: The opinion with respect to the new interface is not related to gender

H_1: The opinion with respect to the new interface is related to gender

Contingency Tables

Evaluation of a new user interface

	Male	Female	Total
Like	165 (219.6)	300 (245.4)	465
Don't like	176 (121.4)	81 (135.6)	257
Total	341	381	722

Chi-square value: 72.3

Critical value in the Chi-square distribution for 5% significance level and df=1: 3.84

H_0 is rejected

in R: qchisq(.95,df=1)

Contigency Tables

In the case of 2x2, Fisher test gives an exact value

	Male	Female
Like	165	300
Don't like	176	81

```
> f <- matrix(c(165,300,176,81),nr=2,byrow=T)
> fisher.test(f)
        Fisher's Exact Test for Count Data
data:  f
p-value < 2.2e-16
```