Panel 1

Contingency Tables

Do Males / Females vote differently.

	Male	Female	
Dems	4 (4/7, 4/12)	3 (3/7)	7
Ind.	4 (40%, 4/12)	6 (60%)	10
Repub.	4 (4/12)	2 (6/13, 2/25)	6
Others	2 (2/12 = .096, 2/25)	2	2
	12	13	25

How many of the Dems are male? 4/7 (Row %)

How many women are indep. 6/13 (Col %)

1

Panel 2

What is the point of the expected values?

They are those counts that are to be expected if two variables have nothing to do with eachother, i.e. if there is no relation between them.

2 Tables: actual counts:  what is actually the case

expected counts:  what are expected

Ex:

10	20
25	10

actual

11	19
24	11

expected

=) No Relation

2

Panel 3

10	20
25	10

actual

5	33
6	19

expected

→ Related

In other words, you compute

χ = sum of (difference between actual and expected counts)^2

"chi"

If sum is large ⇒ is relation

If sum is small ⇒ no relation

Work is done by Stat Crunch:

3

Panel 4

Ex: Gender vs. Pres. Candidate in 2004

Contingency table results:

Rows: SEX

Columns: WOULD HAVE VOTED FOR IN 2004

Cell format
Count Expected count

	Bush	Kerry	Nader	Total
Female	106 105.1	170 166.5	63 67.37	339
Male	89 89.9	139 142.5	62 57.63	290
Total	195	309	125	629

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	2	0.78769374	0.6745

If p-value of Chi-Square Test is

$p \geq 0.05$ ⇒ no relation

$p < 0.05$ ⇒ is relation

$p = 0.6745$ ⇒ No relation

Caveat: This procedure (chi-square test) may be invalid if any expected value is less than 5

4

Panel 5

Contingency table results:

Rows: DEATH PENALTY FOR MURDER

Columns: LIBERAL OR CONSERVATIVE

Cell format
Count Expected count

	1 - extremely liberal	2 - liberal	3 - slightly liberal	4 - moderate	5 - slightly conservative	6 - conservative	7 - extremely conservative	Total
favor	28 45.06	105 151.8	126 143.8	468 453.9	181 169.6	253 205.4	53 44.4	1214
oppose	40 22.94	124 77.25	91 73.2	217 231.1	75 86.36	57 104.6	14 22.6	618
Total	68	229	217	685	256	310	67	1832

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	6	109.52265	<0.0001

all expected values ≥ 5 ✓

$p < 0.0001$ ⇒ There is a relation between variables

5

Panel 6

Book goes in more detail into this question:

Lambda λ

Gamma γ values for more precise analysis.

Stat Crunch does not compute those λ, γ values, so

we'll skip them for now.

6

Panel 7

Correlation for Numeric Variables

Q1: Are they related?
Q2: How strong is association
Q3: Make predictions

HS GPA	College GPA
3.8	2.8
3.1	2.2
4.0	3.5
2.5	1.9
3.3	2.5

Compute Correlation Coefficient r:

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

7

Panel 8

X	Y	X^2	Y^2	XY
3.8	2.8	14.44	7.84	3.8 · 2.8 = 10.64
3.1	2.2	9.61	4.84	6.82
4.0	3.5	16.00	12.25	14.00
2.5	1.9	6.25	3.61	4.75
3.3	2.5	10.89	6.25	8.25
16.7	12.9	57.19	34.79	44.46

$$S_{xx} = \sum x^2 - \frac{(\sum x)^3}{n} =$$

$$= 57.19 - \frac{16.7^2}{5} = 1.412$$

$$S_{yy} = 34.79 - \frac{(12.9)^2}{5} = 1.509$$

$$S_{xy} = 44.46 - \frac{167 \cdot 129}{5} = 1.734$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{1.334}{\sqrt{1.412 \cdot 1.509}} = 0.9416$$

X is independent var (HS)
y is the dependent var (college)

8

Panel 9

The Correlation Coefficient r:

- is always between −1 and 1
- if close to 1 or −1, there is a strong relation between variables
  if close to 0, no relation
- if positive ⇒ positive relation (if x gets bigger, y gets bigger)
  if negative ⇒ negative relation (if x gets bigger, y gets smaller)

In our example: $r = 0.9$ ⇒ strong positive relation between X, Y,

i.e. HS GPA (x) is a good predictor of College GPA (y)

9