

Panel 1

Last time

Contingency tables

- row % : $30 \rightarrow \frac{30}{70} \cdot 100\%$
- col % : $30 \rightarrow \frac{30}{100} \cdot 100\%$
- expected values : $30 \rightarrow \frac{(\text{row total}) \cdot (\text{col total})}{\text{total}} = \frac{70 \cdot 40}{100}$

10	20	
30	40	70
40		100

Expected values are computed assuming vars. are independent

Chi Square Test. if $p < 0.05 \Rightarrow$ dependent vars
 χ^2 $p > 0.05 \Rightarrow$ independent

Note: Rule of Expected 5: All expected values should be 5 or more for test to be valid!

Panel 2

Example

Rows: MODERN SCI DOES MORE HARM THAN GOOD
 Columns: NUMBER OF COLLEGE SCI COURSES

Cell format																			
Count																			
Expected count																			
	1	2	3	4	5	6	7	8	9	10	12	15	20	21	23	25	28	30	32
1 - strongly agree	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 - agree	2	2	5	2	0	1	1	0	0	0	3	0	0	0	0	0	0	0	0
3 - neutral	7	11	8	7	5	1	2	1	1	1	2	0	1	0	0	0	0	0	0
4 - disagree	23	38	22	19	9	9	3	3	0	10	1	3	4	1	0	1	0	0	2
5 - strongly disagree	9	18	14	11	8	5	1	2	0	1	6	1	2	0	1	2	1	2	0
Total	42	69	49	39	22	16	7	6	1	12	12	4	7	1	1	3	1	3	2

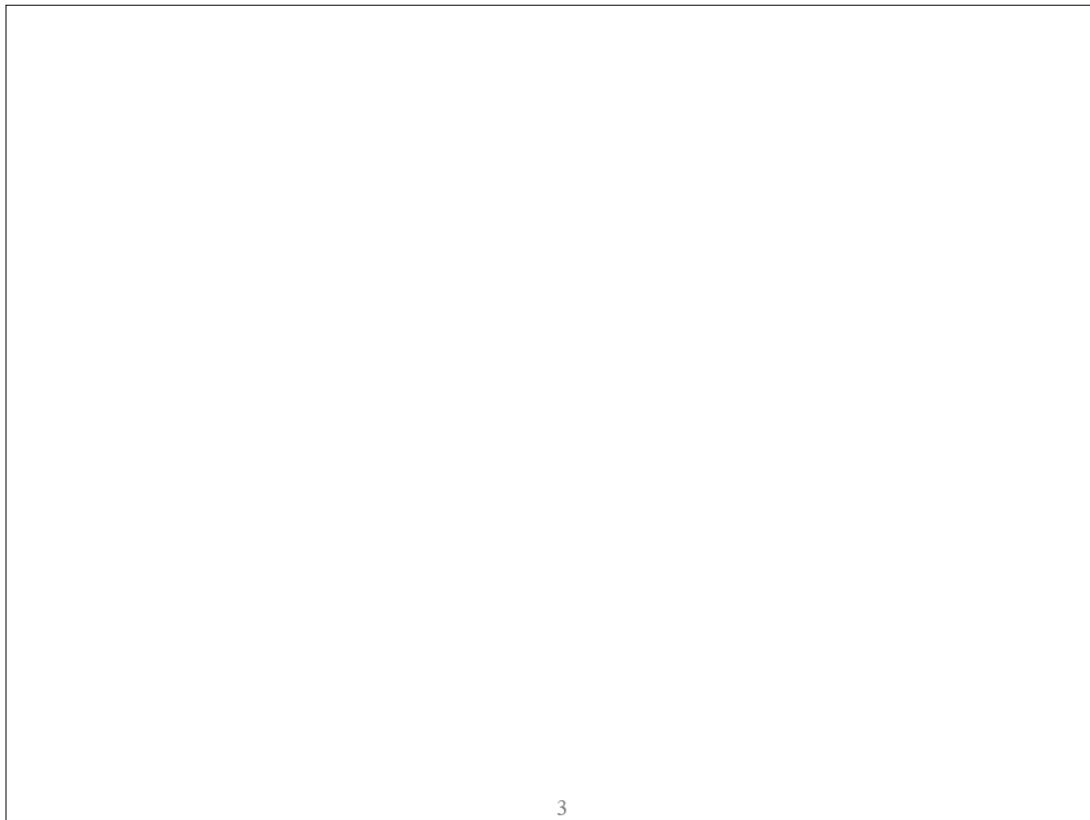
Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	88	67.09697	0.9525

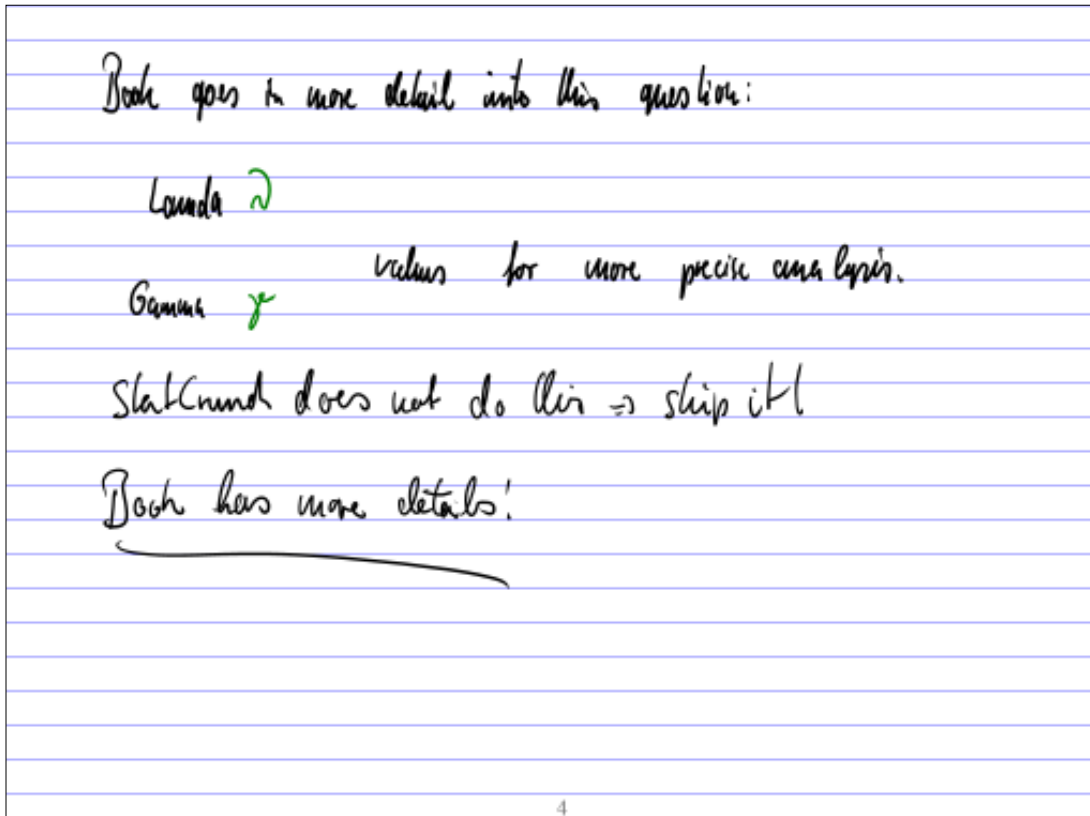
$p = 0.925 > 0.05 \Rightarrow$ not related

Most expected values less than 5 \Rightarrow invalid test

Panel 3



Panel 4



Panel 5

Numeric Variables

Want to know is there a relationship, and if so, make predictions

Ex:

HS GPA	vs	College GPA
3.2		2.8
3.1		2.2
4.0		3.5
2.5	✓	1.9
3.3		2.5

a) Are they related
 b) Predict College GPA based on HS GPA

Compute:
 Correlation coefficient
 Linear Regression line

5

Panel 6

Def. Correlation Coefficient r.

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

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

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

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

6

Panel 7

x is independent var (HS)

y is the dependent var (College)

x	y	x ²	y ²	xy
38	28	1444	784	1064
31	22	961	484	682
40	35	1600	1225	1400
27	19	729	361	513
33	25	1089	625	825
<u>167</u>	<u>129</u>	<u>5719</u>	<u>3479</u>	<u>4446</u>
$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$

Correlation coeff →

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 5719 - \frac{(167)^2}{5} = 1412$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 3479 - \frac{(129)^2}{5} = 1509$$

$$S_{xy} = \sum xy - \frac{\sum x \cdot \sum y}{n} = 4446 - \frac{167 \cdot 129}{5} = 1374$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{1374}{\sqrt{1412 \cdot 1509}} = 0.9426$$

Panel 8

The Correlation Coefficient r:

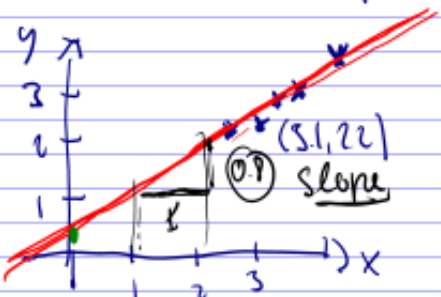
- always between -1 and 1
- if close to -1 or 1, strong relation between x, y vars.
- if close to zero: no relation
- if positive: positive correlation (x goes up → y goes up)
- if negative: negative correlation (x goes up → y goes down)

In our example: $r = 0.9426 \Rightarrow$ strong positive relation between HS GPA + College GPA

Panel 9

Scatter Plot

- To visualize x,y relation, plot them as a scatter plot



x	y
2.8	2.1
3.1	2.2
4.0	3.5
2.7	1.9
3.3	2.5

- Draw line close to most points (may not hit any of them) → Least Square Regression Line
- Find equation of that line:

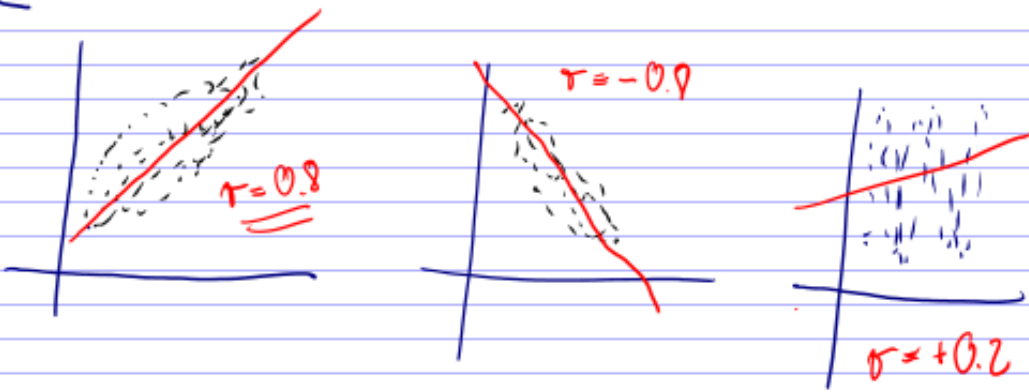
$$y = mX + b$$

$$= 0.8X + 0.5$$

slope
y-intercept

Panel 10

Ex: Scatter Plot



10