Today: 2-D Caterogical Data
2x2 Contingency Tables
Odds Ratios
Four Fold Plots

Sam Ventura
36-315

Department of Statistics
Carnegie Mellon University

February 3, 2016

→ Pearson Residuals: $r_{ij} = \dfrac{c_{ij} - \hat{c}_{ij}}{\sqrt{\hat{c}_{ij}}}$

$\hat{c}_{ij} \rightarrow$ see notes for formula

mosaic plots: see notes

Last Time: Contingency Tables

→ row sums

→ Marginal distn of Var1

whole thing: joint distn

single row (row i)
cond distn of
Var2 | Var1 = i

Single column (column j):
cond distn of
Var1 | Var2 = j

marginal distn
of Var2
(col. sums)

	cat			
	1	2	— k_2	
category 1	c_{11}	c_{12}	— c_{1k_2}	$n_{1\bullet}$
Var1 2	c_{21}	c_{22}	— c_{2k_2}	$n_{2\bullet}$
k_1	$c_{k_1 1}$	$c_{k_1 2}$	— $c_{k_1 k_2}$	$n_{k_1 \bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	— $n_{\bullet k_2}$	n

Var2

odds
=
success : failure

Var 1

	Var 2	
	Yes	No
Yes	A	B
no	C	D

$A = C_{11}, \; B = C_{12} \ldots$

$prob = \dfrac{\# \text{ of ways to succeed}}{\text{total} \# \text{ of ways}}$

$odds = \dfrac{\# \text{ of ways to Suc}}{\# \text{ of ways to Fail}}$

Special Case: 2x2 Contingency Tables & Odds Ratios

Two Categorical Variables: Two categories each $Var 1 \begin{cases} C1 \\ C2 \end{cases} Var 2 \begin{cases} C1 \\ C2 \end{cases}$

Odds Ratio: Measures the association in a 2x2 Contingency Table

give the odds of "success" in Var2 if we had "success" in Var1
vs. if we had "failure" in Var 1

two types of association:

1) High #s in A,D ⎬ positive assoc.
   Low #s in B,C

2) Low #s in A,D ⎬ negative assoc.
   High #s in B,C

$OR = \dfrac{\text{odds of success in Var2 if Var1=Y}}{\text{odds of Success in Var2 if Var1=N}}$

↳ odds ratio

No Association:

1) $P(Var2 = Y \mid Var1 = Y) = P(Var2 = Y \mid Var1 = N)$

2) $P(Var2 = N \mid Var1 = Y) = P(Var2 = N \mid Var1 = N)$

$\dfrac{\dfrac{P(Var2 = Y \mid Var1 = Y)}{P(Var2 = N \mid Var1 = Y)}}{\dfrac{P(Var2 = Y \mid Var1 = N)}{P(Var2 = N \mid Var1 = N)}} = \dfrac{\dfrac{A}{A+B} \Big/ \dfrac{B}{A+B}}{\dfrac{C}{C+D} \Big/ \dfrac{D}{C+D}}$

$= \dfrac{A/B}{C/D} = \dfrac{AD}{BC}$

Four Fold Plot: Displays Assoc. / Odds Ratio in 2x2 Table

each cell in 2x2 table gets a "fold" (quarter circle) of the plot

— radius $\propto \sqrt{\text{freq.} / \text{cell count } (C_{ij})}$

— Area $\propto$ freq. / cell count / C_{ij}
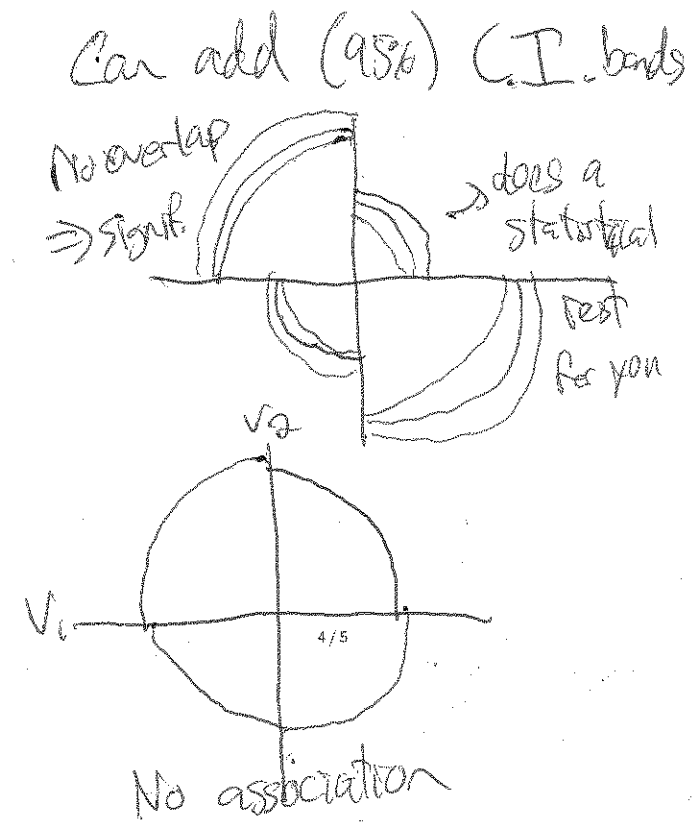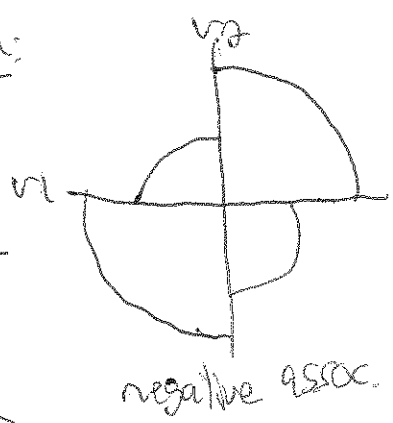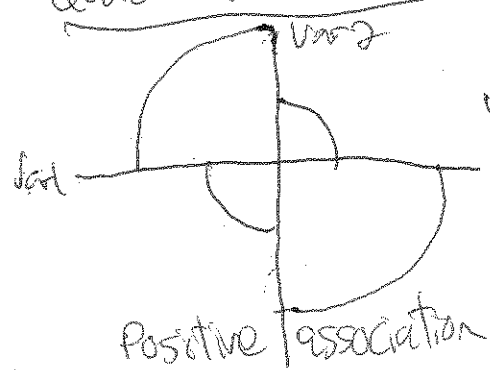
Can add (95%) C.I. bands

No overlap ⇒ signif.

⇒ does a statistical test for you

Quick interpretation:

Positive association

negative assoc.

No association