Test For Linear Trend

- $H_0 : r = 0$ v.s. $H_1 : r \neq 0$
- use test statistic

  $M = \sqrt{(n-1)}\, r$

  $M^2 = (n-1) r^2 : \sim \chi^2$
- If n $\to \infty$

  $M = \sqrt{(n-1)}\, r \xrightarrow{D} N(0, 1)$ one-sided alternative

  $M^2 \xrightarrow{D} \chi_1^2$ two-sided alternative

Takle2.7: Infant Malformation and Mothers Alcohol Consumption

Alcohol Consumption	Malformation		Total	Percentage Present	Standardized Residual
	Absent	Present			
0	17066	48	17114	0.28	-0.18
< 1	14464	38	14502	0.26	-0.71
1-2	788	5	793	0.63	1.84
3-5	126	1	127	0.79	1.06
6	37	1	38	2.63	2.71

$$\text{Standardized Residual} = \frac{n_{ij} \quad \hat{m}_{ij}}{}$$

Extra Power with Ordinal Tests

- H_0: indep. v.s. H_1: any form of dep.
  - $G^2 = 6.2$, p-value$=0.19$

    $X^2 = 12.1$, p-value$=0.02$

Extra Power with Ordinal Tests

- H_0 :

- $H_0 : r = 0$ v.s. $H_1 : r \neq 0$
  - $M^2 = (32573)(0.0142)^2 = 6.6$, p-value$=0.01$
- one-sided alternative
  - M$=2.56$, p-value$=0.005$

, ! strong evidence to reject H_0

When there truly is a linear trend between variables

, ! M^2 has a power advantage over the tests besed on X^2 or G^2

Choices of Scores

- Different choices of ordered scores ⇒ little effect on the results

  row scores=(0, 0.5, 1.5, 4.0, 7.0), M^2=6.6, p-value=0.01

  row scores=(1, 2, 3, 4, 5), M^2=1.83, p-value=0.18

- equally spaced row scores ⇒ r and M^2 do not change

  ex:
  row scores=(1, 2, 3, 4, 5) or (2, 4, 6, 8, 10) or (10, 20, 30, 40, 50)

Choices of Scores

Choices of Scores

- Example: Alcohol Use and Infant Malformation

	Malformation				
Alcohol Consumption	Absent	Present	Total	Total cumulative	Midrank
0	17066	48	17114	17114	8557.5
< 1	14464	38	14502	31616	24365.5
1-2	788	5	793	32509	32013
3-5	126	1	127	32536	32473
6	37	1	38	32574	32555.5

$\frac{(1+17114)}{2} = 8557.5$, $\frac{(17115+31616)}{2} = 24365.5$,

, ! $M^2 = 0.35$, p-value $= 0.55$ (weaker conclusion)

Choices of Scores

- Example: Alcohol Use and Infant Malformation

	Malformation				
Alcohol Consumption	Absent	Present	Total	Total cumulative	Midrank
0	17066	48	17114	17114	8557.5
< 1	14464	38	14502	31616	24365.5
1-2	788	5	793	32509	32013
3-5	126	1	127	32536	32473
6	37	1	38	32574	32555.5

$\frac{(1+17114)}{2} = 8557.5, \quad \frac{(17115+31616)}{2} = 24365.5,$

, ! $M^2 = 0.35$, p-value $= 0.55$ (weaker conclusion)

I 2 Tables

- Y: binary ! I 2 table

Of interest:

- Proportion of 'successes' of Y varies across the levels of X

Small p-value

- $H_1 : r \neq 0$

Cochran - Armitage trend test

Example: AZT Use and AIDS

- Table4.4 Development of AIDS symptoms by AZT Use and Race

Race	AZT Use	Yes	No
Black	No	12	43

Y: whether AIDS symptoms developed (1=Yes, 0=No)

- $=)\quad logit(p) = a + b_1x + b_2z$

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.0736	0.2629	-4.08	0.0000
AZT0newYes	-0.7195	0.2790	-2.58	0.0099
Race0newWhite	0.0555	0.2886	0.19	0.8475

- Null deviance: 8.33499 on 3 degrees of freedom
- Residual deviance: 1.3835 on 1 degrees of freedom
- $\exp(\hat{b}_1) = \exp(\ \ 0.720) = 0.49$

- Controlling for race, is $H_0 : b_1 = 0$ v.s. $H_1 : b_1 \neq 0$
- The likelihood ratio statistic:
  $4G^2 = G^2(M_0|M_1) = G^2(M_0) \quad G^2(M_1) = \quad 2(L_0 \quad L_1) = 6.87$
  (p-value=0.01)
- Wald statistic: $(\frac{b_1}{ASE})^2 = (\frac{0.720}{}$

ANOVA-Type Model Representation of Factors

A factor having two categories requires only a single indicator variable, a factor having J categories requires $J - 1$ indicator variables.

$$\log[P(Y = 1)] = \alpha + \beta^{X}_{i} + \beta^{Z}_{k}$$

β^{X}_{i}: the effects of X through parameters

β^{Z}_{k}: the effects of Z through parameters

- The term b_i^X in this model then is a simple way of representing

$$b_1^X x_1 + b_2^X x_2 + \qquad + b_I^X x_I$$

where $fx_1, \qquad , x_I g$: are indicator variables for I categories of X

Cochran-Mantel-Haenszel Test for 2 2 K

- X, Y : 2 levels, Z (control variable): K levels
- Model
  - $logit(P(Y = 1)) = a + bx + b_k^Z$
- e^b = common odds ratio between X and Y for each level of Z

  $q_{XY(1)} = \ldots = q_{XY(K)}$: homogeneous association
- $H_0 : X ? Y \; j \; Z$

  $( ) \; b = 0 \; ( ) \; q_{XY(1)} = \ldots = q_{XY(K)} = 1$
- Wald test or LRT

  CMH Test

Conclusion

- CMH test is the score test of XY conditional independence

  for $logit(P(Y = 1)) = a + bx + b_k^Z$
- When the XY association is similar in each partial table
  - The CMH test works best
- Similarity of results for LRT, Wald test, and CMH test as n ! ¥