HS 267 LECTURE NOTES 1/31/08

2-BY-2 TABLE

- INDEPENDENT SAMPLES ≐ BINARY EXPLANATORY VARIABLE
- BINARY RESPONSE

NOTATION

EXPL. VAR.	RESP +	RESP −	
+	a_1	b_1	n_1
−	a_2	b_2	n_2
	m_1	m_2	N

"COUNT DATA"
OBSERVED FREQUENCY
"CROSS-TABULATION"

EXAMPLE: RCT OF ESTROGEN, "COMBINED OUTCOME", WHI

	+	−	
+	751	7755	8506
−	623	7479	8102
	1374	15234	16608

$$\hat{P}_1 = \frac{a_1}{n_1} = \frac{751}{8506} = .08829$$

$$\hat{P}_2 = \frac{a_2}{n_2} = \frac{623}{8102} = .07689$$

TEST $H_0: p_1 = p_2$ VIA z TEST, χ^2 TEST, EXACT TEST $\longrightarrow$ P-VALUE; WILL DO χ^2 TESTS NEXT WEEK

z TEST = χ^2 TEST AS WILL BE EXPLAINED IN CH 18

BOTH z & χ^2 COME IN "UNCORRECTED" & CONTINUITY CORRECTED FORMS

ESTIMATE OF "EFFECT" OF THE EXPOSURE

ABSOLUTE EFFECT $\longrightarrow$ RISK DIFFERENCE $= \hat{P}_1 - \hat{P}_2$ (§17.2 LAST SEMESTER)

RELATIVE EFFECT $\longrightarrow$ RELATIVE RISK $= \frac{\hat{P}_1}{\hat{P}_2}$ (§17.4)

ILLUSTRATIVE EXAMPLE: WHI ESTROGEN TRIAL

$$\widehat{RR} = \frac{.08829}{.07689} = 1.1483 \doteq 1.15$$

INTERPRETATION $\longrightarrow$ DISCUSS

CONFIDENCE INTERVAL FOR RR

$$e^{\ln \widehat{RR} \pm z \cdot SE_{\ln \widehat{RR}}}$$

WHERE $SE_{\ln \widehat{RR}} = \sqrt{\frac{1}{a_1} - \frac{1}{n_1} + \frac{1}{a_2} - \frac{1}{n_2}}$

$$= \sqrt{\frac{1}{751} - \frac{1}{8506} + \frac{1}{623} - \frac{1}{8102}}$$

$$= 0.051920$$

FOR 95% CONFIDENCE USE $z = 1.96$

$$e^{\ln(1.1483) \pm (1.96)(.051920)} = e^{0.1382 \pm 0.0857}$$

$$= e^{.0528, .2236} = (1.05, 1.25)$$

SYSTEMATIC ERRORS IN INFERENCE

- EMPHASIZE IMPORTANCE, ESP IN NON-EXPERIMENTAL STUDIES
- CONFOUNDING → EXTRANEOUS FACTORS EXPLAIN THE ASSOCIATION
- WHY DID THE NON-EXPERIMENTAL STUDIES ON POSTMENOPAUSAL ESTROGEN GET IT WRONG? → Biometrics, 2005, 61(4)

INTRODUCE NOTIONS OF INFORMATION BIAS & SELECTION BIAS; HAVE STUDENTS READ TEXT