Lesson 25

Thursday 5/2/24

Chapter 9: Categorical Data

2 x 2 Contingency Table

	Recidivism=No	Recidivism=Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

Calculating Marginal Probabilities from a 2x2 Contingency Table (pp. 267-268)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

p(recidivism=yes | incarcerated=no) = 203/379 = 0.536

p(recidivism=yes | incarcerated=yes) = 246/403 = 0.610

Difference Between 2 Conditional Probabilities (p. 268)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

p(recidivism=yes | incarcerated=no) = 203/379 = 0.536

p(recidivism=yes | incarcerated=yes) = 246/403 = 0.610

Difference = 0.610 - 0.536 = 0.074

Interpretation: Difference between recidivism probability between the 2 groups is 0.610-0.536 = 0.074 (or 7.4 percentage points difference).

Relative Risk Statistic (p. 268)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

p(recidivism=yes | incarcerated=no) = 203/379 = 0.536

p(recidivism=yes | incarcerated=yes) = 246/403 = 0.610

Relative risk = 0.610/0.536 = 1.138

Interpretation: risk of recidivism is 1.138 times greater in the incarcerated group compared to the non-incarcerated group.

Chi-Square Test of Independence

	Recidivism = No	Recidivism = Yes	Total
Incarcerated = No	176 A	203 B	379
Incarcerated = Yes	157 C	246 D	403
Total	333	449	782

Question: are the 2 variables statistically independent of each other? (Table 9.9 in book); testing the independence hypothesis.

cell	obs	exp	obs-exp	(obs-exp)^2	[(obs-exp)^2]/exp
A	176	333*379/782 = 161.39	14.61	213.452	1.323
B	203	449*379/782 = 217.61	-14.61	213.452	0.981
C	157	333*403/782 = 171.61	-14.61	213.452	1.244
D	246	449*403/782 = 231.39	14.61	213.452	0.922

Obtained Chi-Square Statistic = 1.323+0.981+1.244+0.922 = 4.47

degrees of freedom = (rows-1)*(columns-1) = (2-1)*(2-1) = 1

Conduct test at $p < .05$ significance level

Critical Value of Chi-Square with 1 degree of freedom = 3.841

Note: this comes from Table B.4 on p. 537

Obtained Value of Chi-Square > Critical Value

Decision: reject independence hypothesis

Interpreting Correlations From a 2x2 Contingency Table

Dependent Variable

Independent Variable

	Recidivism = No	Recidivism = Yes	Total
Incarcerated = No	176 A	203 B	379
Incarcerated = Yes	157 C	246 D	403
Total	333	449	782

Yule's Q = (AD-BC)/(AD+BC)

AD = 176 x 246 = 43296

BC = 203 x 157 = 31871

AD-BC = 43296-31871 = 11425

AD+BC = 43296+31871 = 75167

Q = 11425/75167 = 0.152

→ A Weak Positive Relationship

Then, a positive correlation means that "yes" on the independent variable tends to be paired with "yes" on the dependent variable.

And, a negative correlation means that "yes" on the independent variable tends to be paired with "no" on the dependent variable; and vice-versa.

Confidence Interval for Yule's Q: Overview

	Recidivism = No	Recidivism = Yes	Total
Incarcerated = No	176 A	203 B	379
Incarcerated = Yes	157 C	246 D	403
Total	333	449	782

Step 1: Decide on the precision of the confidence interval (i.e., 80%, 90%, 95%, 99%, etc.)

Step 2: Use Table B.3 on p. 536 to identify the appropriate two-tailed quantile of the normal distribution (for example, for a 95% confidence interval, we set **α** = 0.05 and choose z = 1.96)

Step 3: Calculate the Lower/ Upper Confidence Limit:

$$Q \pm z \times \sqrt{\frac{(1-Q^2)^2(\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D})}{4}}$$

Source: N. Davidson (1976). Causal inference for dichotomous variables. Monograph #9 in Concepts & Techniques in Modern Geography.

Step 4: Determine whether the confidence interval includes the number zero.

Confidence Interval for Yule's Q: How to Calculate

	Recidivism = No	Recidivism = Yes	Total
Incarcerated = No	176 A	203 B	379
Incarcerated = Yes	157 C	246 D	403
Total	333	449	782

Step 1: Decide on the precision of the confidence interval: 95%

Step 2: Set z = 1.96 for a 95% ($\alpha = 0.05$) confidence interval

Step 3: Calculate the Lower/ Upper Confidence Limits:

$$Q \pm z \times \sqrt{\frac{(1-Q^2)^2(\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D})}{4}}$$

$$Q \pm 1.96 \times \sqrt{\frac{(1-0.152^2)^2(\frac{1}{176}+\frac{1}{203}+\frac{1}{157}+\frac{1}{246})}{4}}$$

Step 4: Determine whether the confidence interval includes the number zero. → Confidence interval is [0.013,0.291] which does not include zero

Another 2x2 Table: Difference Between 2 Probabilities

Dependent Variable

Independent Variable

	Delinq = No	Delinq = Yes	Total
Strain = Low	104 A	47 B	151
Strain = High	83 C	52 D	135
Total	187	99	286

p(Delinq=Yes|Strain=Low) = 47/151 = 0.311

p(Delinq=Yes|Strain=High = 52/135 = 0.385

Difference Between the Two Conditional Probabilities

0.385 - 0.311 = 0.074

Interpretation: the delinquency rate is 7.4 percentage points higher in the high strain group.

Another 2x2 Table: Relative Risk Statistic

Dependent Variable

Independent Variable

	Delinq = No	Delinq = Yes	Total
Strain = Low	104 A	47 B	151
Strain = High	83 C	52 D	135
Total	187	99	286

p(Delinq=Yes|Strain=Low) = 47/151 = 0.311

p(Delinq=Yes|Strain=High = 52/135 = 0.385

Relative Risk Statistic

0.385/0.311 = 1.238

Interpretation: the probability of delinquency involvement is 1.238 times higher in the high strain group compared to the low strain group.

Chi-Square Test of Independence

	Delinq = No	Delinq = Yes	Total
Strain = Low	A 104	B 47	151
Strain = High	C 83	D 52	135
Total	187	99	286

Question: are the 2 variables statistically independent of each other? (Table 9.9 in book); testing the independence hypothesis.

cell	obs	exp	obs-exp	(obs-exp)^2	[(obs-exp)^2]/exp
A	104	187*151/286 = 98.731	5.269	27.762	0.281
B	47	99*151/286 = 52.269	-5.269	27.762	0.531
C	83	187*135/286 = 88.269	-5.269	27.762	0.315
D	52	99*135/286 = 46.731	5.269	27.762	0.594

Obtained Chi-Square Statistic = 0.281+0.531+0.315+0.594 = 1.721

degrees of freedom = (rows-1)*(columns-1) = (2-1)*(2-1) = 1

Conduct test at p < .01 significance level

Critical Value of Chi-Square with 1 degree of freedom = 6.635
Note: this comes from Table B.4 on page 537

Obtained Value of Chi-Square < Critical Value

Decision: fail to reject independence hypothesis

Yule's Q Statistic

	Delinq = No	Delinq = Yes	Total
Strain = Low	104 A	47 B	151
Strain = High	83 C	52 D	135
Total	187	99	286

Yule's Q = (AD-BC)/(AD+BC)

AD = 104 x 52 = 5408

BC = 47 x 83 = 3901

AD-BC = 5408-3901 = 1507

AD+BC = 5408+3901 = 9309

Q = 1507/9309 = 0.162

→ A Weak Positive Relationship

Then, a positive correlation means that "yes" on the independent variable tends to be paired with "yes" on the dependent variable.

And, a negative correlation means that "yes" on the independent variable tends to be paired with "no" on the dependent variable; and vice-versa.

Yule's Q Statistic Confidence Interval

	Delinq = No	Delinq = Yes	Total
Strain = Low	104 A	47 B	151
Strain = High	83 C	52 D	135
Total	187	99	286

Step 1: Decide on the precision of the confidence interval: 99%

Step 2: Set z = 2.576 for a 99% (α = 0.01) confidence interval

Step 3: Calculate the Lower/ Upper Confidence Limits:

$$Q \pm z \times \sqrt{\frac{(1-Q^2)^2(\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D})}{4}} \qquad Q \pm 2.576 \times \sqrt{\frac{(1-0.162^2)^2(\frac{1}{104}+\frac{1}{47}+\frac{1}{83}+\frac{1}{52})}{4}}$$

Step 4: Determine whether the confidence interval includes the number zero. → Confidence interval is [-0.153,0.476] which includes zero