☼ RNR 613 — Proportions and Odds

So far: tools for continuous response variables and continuous or categorical explanatory variables

Now: tools for categorical response variables

Common when the response can only take on a limited number of possibilities.

Example. Investigate suitability of 3 host plants for an insect.
Of 50 eggs placed on each host, survival / mortality was 40/10, 35/15, and 25/25.

Host is the explanatory variable (3 levels), and the response variable has 2 levels, survived or died.

Host Plant	Alive	Dead
A	40	10
B	35	15
C	25	25

Types of categorical responses:

• Binomial or Binary (2 categories): present or absent; sick or healthy; male or female.

• Multinomial (>2 categories): race (black, white, etc.); results of a taste test (4 alternatives); hair color.

For categorical responses, we can calculate a probability for each response represented and they will sum to 1.

Binary response variables

For continuous responses, we can describe populations by their means.

For binary responses, we can describe populations with a proportion that is analogous role to the mean.

Define the response as 1 for a yes or a success and 0 for a no or a failure.

Average of a binary response variable in a population is the proportion of members in the population classified as a yes.

Example: A population with 90 yes and 10 no.

Average = π = (90 x 1) + (10 x 0) / 100 = 0.90

So π is the proportion of "yes" responses in a population, which is the mean response.

Tools for analysis of continuous response variables assume that the mean response and its variance are independent:

• The distribution of the residuals is homogenous across levels of the explanatory variable(s)

• Average size of these residuals (Error MS) provides a basis for assessing the effect of explanatory variables.

In contrast, variance of binary variables is determined completely by its mean: Variance{π} = $\pi(1 - \pi)$.

• Variance of π is a function of the mean; it is highest at 0.5 and lowest near 0 or 1.

• There is no paramater analagous to σ^2.

Example: A population of 100 has 50 black observations (Y = 1) and 50 blue observation (Y = 0).

The mean (π) for color is 0.50.

The variance of π is: $[\ 50\ (1 - 0.5)^2 + 50\ (0 - 0.5)^2\ ] / 100 = 0.25$ $[\ 0.5(1 - 0.5) = 0.25)\ ]$

Example: Another population of 100 has 90 black and 10 blue.

Here $\pi = 0.9$ and variance(π) = 0.09.

Odds

Can describe a binary variable using odds instead of the proportion of yes outcomes. The corresponding odds of yes outcomes are:

$$\omega = \frac{\pi}{(1-\pi)} = \frac{p(yes)}{p(no)} = \frac{n(yes)}{n(no)}$$

Example: Proportion of colds in the placebo group for the vitamin C experiment (π) was 335/411 = 0.82. The odds of getting a cold in the placebo group (ω) = 335/76 = 4.4.

So the odds of getting a cold in the placebo were about 4 to 1; about 4 people got sick to every 1 that did not.

Example: An event with chances of 0.95 has odds of 19 to 1 (0.95 / 0.05 = 19) in favor of its occurrence.

An event with chances of 0.05 has odds of 19 to 1 (0.05 / 0.95 = 1/19) against its occurrence.

Numerical factoids about odds:

1. Odds vary between 0 and ∞; proportions vary between 0 and 1.
2. A proportion of ½ corresponds to odds of 1; these are called “equal odds."
3. Odds are undefined for proportions equal to 0 or 1.
4. If the odds of a yes outcome are ω, the odds of a no are $1/\omega$.

   ω_{yes} = 3 to 1; ω_{no} = 1 to 3

5. If the odds of a yes outcome are ω, then the probability (or population proportion) of yes is $\pi = \omega / (1+ \omega)$.

   ω_{yes} = 3 to 1; π = probability of yes = 3 / (3+1) = 0.75.
   ω_{no} = 1 to 3; 1 - π = probability of no = 0.33 / (0.33 + 1) = 0.25.

Ratio of Two Odds

A useful alternative to the difference in proportions between two populations π_1 and π_2 is the *odds ratio*, $\phi = \omega_2/\omega_1$

Odds are advantageous for several reasons (p. 540), one of which is that they lends themselves to regression-type analyses.

Example: Proportion of colds in the vitamin C group was 302/407 = 0.74, so the odds of getting a cold were 302/105 = 2.87 or about 3 to 1; that is, about 3 people got sick to every 1 that did not.

The odds ratio between placebo and vitamin C groups was estimated to be 4.408/2.876 = 1.53.

The odds of getting a cold in the placebo group were 1.53 times as large as the odds of getting a cold in the vitamin C group.

Larger odds are traditionally placed in the numerator, which yields odds *in favor* of an event, rather than odds *against* the event.

Differences Between Proportions or Odds Ratios

For large samples, two porportions or two odds can be compared with a Z-test.

A test for equal proportions is:

$$Z = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - 0}{SE_0(\hat{\pi}_2 - \hat{\pi}_1)}$$

where SE_0, which is appropriate when testing that the null of equal proportions:

$$SE_0(\hat{\pi}_2 - \hat{\pi}_1) = \sqrt{\frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_1} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_2}}$$

where π_c is the sample proportion from the combined sample.

The appropriate SE for a confidence interval for the difference in proportions is based on a model for estimating separate proportions:

$$SE_0(\hat{\pi}_2 - \hat{\pi}_1) = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

A test for equality of odds is:

$$Z = \frac{\log\left(\frac{\hat{\omega}_1}{\hat{\omega}_2}\right)}{SE_0 \log\left(\frac{\hat{\omega}_1}{\hat{\omega}_2}\right)}$$

where SE_0, which is appropriate when testing the null of equal proportions:

$$SE_0 \log\left(\frac{\hat{\omega}_1}{\hat{\omega}_1}\right) = \sqrt{\frac{1}{n_1\hat{\pi}_c(1-\hat{\pi}_c)} + \frac{1}{n_2\hat{\pi}_c(1-\hat{\pi}_c)}}$$

For odds, the log of the ratios has better distributional properties.

Example: Vitamin C and colds.

Placebo group: $\pi_1 = 0.82$; $\omega_1 = 4.408$

Vitamin C group: $\pi_2 = 0.74$; $\omega_2 = 2.876$

$\phi = 4.408/2.876 = 1.533$; $\log(\phi) = 0.427$

SE $\log(\phi) = 0.170$

95% CI for $\log(\phi) = 0.427 \pm 1.96\ (0.170) = 0.094$ to 0.760

95% CI for $\phi = e^{0.094}$ to $e^{0.760} = 1.10$ to 2.14

$Z = 0.427/0.170 = 2.51$, $P = 0.0059$

The odds of a cold for the placebo group were 1.53 times the odds of a cold for the vitamin C (95% CI = 1.10 to 2.14), $Z = 2.51$, $P = 0.0059$.

☼ RNR 613 — Tables of Counts

A single, binary response from each subject (0 or 1) can be viewed as a tables of counts.

Number of subjects falling in each cross-classification of a row factor and a column factor.

Example Donner Party. Trapped by early snow in October 1846 in the Sierra Nevadas. Total of 87 people, 40 of whom died by April 1847. We consider the 45 people aged ≥15, 25 (55%) of which died.

	Survived		
Female	0 (no)	1 (yes)	
0 (male)	10	20	30
1 (female)	10	5	15
	20	25	45

With binary counts, often interested in *proportions* of events and state conclusions as *odds* or *odds ratios*.

Recall that the mean of a *binary response* is π, the proportion of positive responses; its *odds* are: $\omega = \pi/(1-\pi)$.

When two populations have proportions π_1 and π_2, with corresponding odds ω_2 and ω_1, a useful alternative to the difference in proportions $\pi_2 - \pi_1$ is the odds ratio ω_2/ω_1.

2 x 2 Tables of Counts

Relevant null hypotheses are either: H_o: $\pi_2 - \pi_1 = 0$ or H_o: $\omega_2/\omega_1 = 1$

where π_2 and π_1 are population proportions and ω_2 and ω_1 are the corresponding *odds*.

Two types of hypotheses from tables of counts:

Homogeneity and *Independence*.

1. *Hypotheses of homogeneity* are appropriate when you can think of <u>two</u> populations, each with two categories of response; one factor is a clear response.

   The question is whether the distribution of the response is the same across populations.

   E.g. Whether gender affected whether or not an individual was promoted.

2. *Hypotheses of independence* are appropriate when you can think of <u>one</u> population with four categories of response; neither factor is a response.

   The question is whether there is an association between row and column factors where there is no response.

   E.g. Whether or not beer preference (ale or lager) was associated with snack preference (chips or pretzels).

* Choice of hypothesis affects only the interpretation of results but not the calculations, statistics, or *P*-values.

Anatomy of a 2 x 2 Table

Row factor *(explanatory)*	Column factor *(response)*		
	Level 1	Level 2	*Row totals*
Level 1	n_{11}	n_{12}	R_1
Level 2	n_{21}	n_{22}	R_2
Column totals	C_1	C_2	*Grand total* = T

Marginal totals are the *Column totals* and *Row totals*, both of which sum to the *Grand total*.

Example: Coffee Drinking and Sexual Activity

Responses of 225 married women (>60 yrs) to the questions:

1. *Are you sexually active?*
2. *Do you drink coffee?*

Drink Coffee?	Sexually Active?		
	Yes	No	
Yes	15	25	40
No	115	70	185
	130	95	225

• Sample proportion ($\hat{\pi}$) of coffee-drinking women that are sexually active = 15/40 = 0.375

• Sample variance = $\hat{\pi}(1 - \hat{\pi}) / n$ = 0.375(1 – 0.375)/40 = 0.0059
• Standard error = $\sqrt{}$(Var [$\hat{\pi}$]) = 0.0766
• 95% Confidence Interval for $\hat{\pi}$ = $\hat{\pi}$ ± 1.96 SE($\hat{\pi}$) =

0.375 ± 1.96(0.076) = (0.225, 0.525)

Sample proportions of women who <u>do</u> drink coffee and are sexually active and those that <u>do not</u> drink coffee and are sexually active:

$$\hat{\pi}_1 = 15/40 = 0.375 \qquad \hat{\pi}_2 = 115/185 = 0.622$$

Odds: $\hat{\omega}_1 = 0.375/(1 - 0.375) = 15/25 = 0.60$; $\hat{\omega}_2 = 0.622/(1–0.622) = 115/70 = 1.64$

R*atio*: $\hat{\omega}_2/\hat{\omega}_1 = 1.64/0.60 = 2.74$

The estimated odds of a woman being sexually active is 2.74 times higher if they do *not* drink coffee.

Sampling Schemes Leading to 2 x 2 Tables

Whether a model for 1 or 2 populations is appropriate depends on the question and on how the data were collected.

1. **Poisson.** A fixed amount of time devoted to collecting a random sample from one population; each subject falls into 1 of the 4 table cells. Marginal totals not known in advance.

Example: For the "coffee drinking" study of older women, researchers spent a fixed amount of time searching for older women in Washtenaw County, Michigan.

2. **Multinomial.** Each subject in the sample is categorized into one of several cells. Similar to *Poisson* except that the total sample size (*Grand total*) is known in advance.

Example: If researchers had a list of all older women in the county, from which they selected a sample of women.

3. **Product Binomial.** Each subject in each population falls into 1 of 2 categories. *Product binomial* indicates there are >1 populations from which samples have been collected.

a. **Prospective.** Populations defined by the levels of the *explanatory* factor; random samples collected from each. If separate lists of all coffee-drinking and non-coffee drinking women in the county existed, a random sample would then be collected from each population.

b. **Retrospective.** Random samples are selected from subpopulations as defined by each level of the *response* factor. The researcher specifies the sample totals to be obtained at each level of their response factor.

4. **Randomized Binomial Experiment.** The subjects, regardless of how they were obtained, are allocated randomly to the two levels of the *explanatory* factor. Besides the randomization, the “setup” is essentially the same as in prospective product binomial sampling. The totals for the explanatory factor are fixed by the researcher.

Knowing how data were collected, study objectives, and which marginal totals were fixed determines which hypothesis (*Independence* or *Homogeneity*) is appropriate.

	Marginal Totals	Usual Hypotheses	
Sampling Scheme	Fixed in Advance	*Independence*	*Homogeneity*
Poisson	*None*	✘	✘
Multinomial	*Grand total*	✘	✘
Product binomial			
Prospective	*Row (explanatory) totals*		✘
Retrospective	*Column (response) totals*		✘
Randomized experim.	*Row (explanatory) totals*		✘

Both types of null hypotheses can be stated as:

Row categorization is independent of *Column categorization*.

Again, if one category is a response then the hypothesis to test is *homogeneity*; if not, the hypothesis to test is *independence*.

Chi-Squared Test for Goodness-of-Fit

A method of comparing observation with theory; approach assumes that the sampled units fall randomly into cells and that the chance that a unit falls into a particular cell can be estimated from the theory (model) being tested.

The number of units that fall into a cell is the cell's *observed* count; the number predicted by theory (model) is the cells *expected* count.

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

The test statistic is calculated for each cell then summed across all cells in the table.

Has a χ^2 distribution with *df* = (No. cells) – (No. parameters) – 1.

Chi-Squared Test of Independence in a 2 x 2 Table

When neither category is a response the hypothesis is one of *independence*.

A 2 x 2 table has 4 cells; the proportion of counts that fall in column 1 is C_1/T; if the column proportion is independent of row, this same proportion should fall in both row 1 and row 2.

Because there are R_1 subjects in row 1, R_1 x (C_1/T) of them are expected to fall into column 1.

By extension, the expected cell count in cell (*i, j*), if rows (*i*) and columns (*j*) are independent, is estimated by R_iC_j/T as follows:

Cell (*i, j*)	Observed	Expected
(1, 1)	n_{11}	R_1C_1/T
(1, 2)	n_{12}	R_1C_2/T
(2, 1)	n_{21}	R_2C_1/T
(2, 2)	n_{22}	R_2C_2/T

The pattern for determining an estimate for an expected cell count is to multiply the *marginal totals* in the *same row* and the *same column* together, then divide by the *grand total*.

So testing the hypothesis of independence involves:

1. Estimate expected cell counts.
2. Compute the χ^2 using the previous formula.
3. Find the *p*-value from a χ^2 distribution with 1 *df*.

Example: Death penalty and Race of victim (ignoring aggravation level of crime)

Observed

Race of victim	Death Penalty		*Totals*
	Yes	No	
White	45	85	130
Black	14	218	232
Totals	59	303	362

Expected [from marginals, e.g., 21.9 = (130)(59)/(362)]

Race of victim	Death Penalty	
	Yes	No
White	21.19	108.81
Black	37.81	194.19

$\chi^2 = (45 - 21.19)^2/21.19 + (85 - 108.81)^2/108.81 + (14 - 37.81)^2/37.81 + (218 - 194.19)^2/194.19 = 49.89$

$\chi^2_1 = 49.89$, $P < 0.0001$, so conclude convincingly that the death penalty sentence and race of the victim are not independent.

The expected cell counts obtained under the *hypothesis of independence* are the same as expected cell counts under the *hypothesis of homogeneity*.

Therefore, the χ^2 statistic and *P*-values from the χ^2 *test of independence* and the χ^2 *test of homogeneity* are identical.

Limitations

χ^2 tests of independence and homogeneity extend logically from 2 x 2 to *r x c* tables of counts; however, they are uninformative.

1. They produce only a *P*-value and have no associated parameter to describe the degree of dependence (effect).

2. The alternative hypothesis—that the rows and columns are not independent—is very general.

When >2 rows or columns are involved, there may be a more specific form of dependence to explore using logistic regression for binomial counts or log-linear regression for Poisson counts.

Fisher's Exact Test for 2 x 2 Tables

Another limitation of χ^2 tests occurs if sample sizes in any cell are small (<5). In these cases, Fisher's Exact Test is appropriate as it is for tests of any sample sizes.

This is a permutation test, which generates a *P*-value based on the number of possible regroupings of *N* observations into two groups of size $n_1 + n_2$.

The total number of permutations possible becomes the *P*-value's denominator; the numerator becomes the number of regroupings that yield a statistic (say proportion or mean) $\geq$ what was observed.

So, the one-sided *P*-value from a permutation test is the actual proportion of groupings that leads to values of $\hat{\pi}_1 - \hat{\pi}_2 \geq$ the one observed.

Fisher's Exact Test provides the exact *P*-value for all hypotheses whereas Pearson's provides only an approximate *P*-value.

So why use other tests?

If a computer is available and all that is required is a test, then there is no reason to use any test other than Fisher's.

Other Related Tests

Mantel-Haenszel Test

> a test for equal odds when there are several 2 x 2 tables involved, one for each level of a confounding or blocking factor. Compares the odds in two groups after accounting for the effect of a blocking or confounding variable.

For example, if we examined *Death penalty and Race* data for several *Aggravation Levels* (i.e., there would be as many 2 x 2 tables as aggravation levels), then this would be an appropriate test.

Also, if you instead chose to consider the response to be *Death Penalty*, you could use Logistic Regression to model variation in the binary response by *Race* and *Aggravation*.