Binary Outcomes: Comparisions of Proportions or Odds Example

Douglas Kondro: March 8th

i) Parameters and Statistics for a Binary Response Variable

Y	A binary response variable for the poputlation of intrest
π	The parameter population proportion or $\sum Y/_n$
π_1	The parameter population proportion of population 1
π_o	The anticipated population proportion for normal distribution test
$\widehat{\pi}_1$	statistic for the sampled population proportion of population 1
$\widehat{\pi}_c$	statistic for the population proportion of the combined populations
n_1	The total number of samples from population 1
$\widehat{\omega}_1$	The odds for population 1 or = $\widehat{\pi_1}/(1-\widehat{\pi_1})$
ϕ	The odds ratio between two populations or = $\widehat{\omega_2}/\widehat{\omega_1}$

ii) Confidence Intervals and Equality Tests

Statistic Comparision	Hypothesis	Standard Error (Confidence Interval)	Standard Error (z score equality test)
Proportion	$H_o : \widehat{\pi}_2 - \widehat{\pi}_1 = 0$	$\sqrt{\frac{\widehat{\pi_1}(1-\widehat{\pi_1})}{n_1} + \frac{\widehat{\pi_2}(1-\widehat{\pi_2})}{n_2}}$	$\sqrt{\frac{\widehat{\pi_c}(1-\widehat{\pi_c})}{n_1} + \frac{\widehat{\pi_c}(1-\widehat{\pi_c})}{n_2}}$
Log odds ratio	$H_o : log\frac{\widehat{\omega_1}}{\widehat{\omega_1}} = 0$	$\sqrt{\frac{1}{n_1\widehat{\pi_1}(1-\widehat{\pi_1})} + \frac{1}{n_2\widehat{\pi_2}(1-\widehat{\pi_2})}}$	$\sqrt{\frac{1}{n_1\widehat{\pi_c}(1-\widehat{\pi_c})} + \frac{1}{n_2\widehat{\pi_c}(1-\widehat{\pi_c})}}$

Normal distribution check for a binomially distributed binary response variable:

1: $n\pi_o > 5$ Do we have more then five occurances of Y=1?

2: $n(1-\pi_o) > 5$ Do we have more then five occurances of Y=0?

iii) Odds ratio as the Only Appropriate Parameter If the Sampling is Retrospective

Odds ratio is idependent of what group is selected the the response (output) or the explanitory (input) variables.

Example: Does Vitamin C work?

We want to know if vitamin C helps prevent the common cold. We look at a stuby conducted in Canada using 818 volunteers during a winter period. The study population is divided randomly into two groups: One that recieves 1000 mg of bitamin C per day and the other that recieves pacebo pills. At the end of the time period all subjects were interviewed by a doctor who determined if they caught a cold during the period.

	Cold	No Cold	Total
Vitamin C (Group 1)	**303**	**105**	407
Placebo (Group 2)	**335**	**76**	411
Totals	637	181	818

a) What is the confidenece interval of the proportion of subjects that caught a common cold that received a placibo?

Statistic Comparision	Statistic	Standard Error (Confidence Interval)
Proportion	$\widehat{\pi}$	$\sqrt{\frac{\widehat{\pi}_1(1-\widehat{\pi}_1)}{n_1}}$
Example work	$\widehat{\pi}$ = 335/411 $\widehat{\pi}$ = 0.815	SE(C.I)= $\sqrt{0.815(1-0.815)/411}$ SE(C.I)= 0.019

1. We check that our statistic follows the normal distribution:

$$n\pi_o > 5 \text{ and } n(1-\pi_o) > 5 \rightarrow 411(0.815) > 5 \text{ and } 411(0.185) > 5$$

   We have at least 5 occurances of each binary output!

2. We know our confidence interval is:

$$\widehat{\pi} \pm SE(C.I)$$

The proportion of the population that got a common cold while receiving a placibo was **0.815 +/- 0.019.**

b) What is the 95% confidenece interval and equality for the diffrence in the two populations proportion: vitamin C and palacibo? How can we interpret this?

Statistic Comparision	Hypothesis	Standard Error (Confidence Interval)	Standard Error (z score equality test)
Proportion	$H_o : \widehat{\pi}_2 - \widehat{\pi}_1 = 0$	$\sqrt{\frac{\widehat{\pi}_1(1-\widehat{\pi}_1)}{n_1} + \frac{\widehat{\pi}_2(1-\widehat{\pi}_2)}{n_2}}$	$\sqrt{\frac{\widehat{\pi}_c(1-\widehat{\pi}_c)}{n_1} + \frac{\widehat{\pi}_c(1-\widehat{\pi}_c)}{n_2}}$
Example work	$\widehat{\pi}_1$=303/407=**0.742** $\widehat{\pi}_2$=335/411=**0.815** $\widehat{\pi}_2 - \widehat{\pi}_1$ = **0.073** (Passes Normal approx check)	$\sqrt{\frac{0.742(1-0.742)}{407} + \frac{0.815(1-0.815)}{411}}$ SE(C.I)= 0.029	$\widehat{\pi}_C$=(303+335)/(407+411) $\widehat{\pi}_C$=**0.780** $\sqrt{\frac{0.78(1-0.78)}{407} + \frac{0.78(1-0.78)}{411}}$ SE(Z.S)= 0.029

1. First lets check if our normal approximation is appropriate:

   $n\pi_1 > 5 \text{ and } n(1-\pi_1) > 5 \rightarrow 407(0.742) > 5 \text{ and } 407(0.258) > 5$

   $n\pi_2 > 5 \text{ and } n(1-\pi_2) > 5 \rightarrow 411(0.815) > 5 \text{ and } 411(0.185) > 5$

   Both populations have more then 5 occurances of each binary output!

2. We find our confidence inteval by using our z multipier atf 95% which is 1.96:

$$\widehat{\pi}_2 - \widehat{\pi}_1 \pm z(0.975)SE(C.I) = (1.96)(0.029) \rightarrow 0.073 \pm 0.057$$

3. Next we use our z score to find our p value using a two tailed test given our hypothesis:

$$z = \frac{\widehat{\pi}_2 - \widehat{\pi}_1}{SE(Z.S)} = \frac{0.073}{0.029} = 2.52 \rightarrow 0.0117$$

The probability of catching a cold after placebo exceeds the probability of catching a cold after vitiman C by 0.073 (95% C.I: **0.016 to 0.130**) (**p = 0.012**).

c) What is the confidenece interval and equality for the diffrence in the two populations odds ratio: vitamin C and palacibo?

Statistic Comparision	Hypothesis	Standard Error (Confidence Interval)	Standard Error (z score equality test)
Log odds ratio	$H_o : log\frac{\widehat{\omega}_1}{\widehat{\omega}_1} = 0$	$\sqrt{\frac{1}{n_1\widehat{\pi}_1(1-\widehat{\pi}_1)} + \frac{1}{n_2\widehat{\pi}_2(1-\widehat{\pi}_2)}}$	$\sqrt{\frac{1}{n_1\widehat{\pi}_c(1-\widehat{\pi}_c)} + \frac{1}{n_2\widehat{\pi}_c(1-\widehat{\pi}_c)}}$
Example work	$\widehat{\pi}_1$=303/407=**0.742** $\widehat{\pi}_2$=335/411=**0.815** $\widehat{\omega}_1 = \frac{\pi_1}{1-\pi_1}$ $\widehat{\omega}_1$ =302/105=**2.876** $\widehat{\omega}_2 = \frac{\pi_2}{1-\pi_2}$ $\widehat{\omega}_2$ =335/76 =**4.408** ϕ = 4.408/2.876 ϕ =1.53 log ϕ = **0.427**	$\sqrt{\frac{1}{407 \times 0.742\,(1-0.742)} + \frac{1}{411 \times 0.815\,(1-0.815)}}$ SE(C.I) = 0.170	$\widehat{\pi}_c$=(303+335)/(407+411) $\widehat{\pi}_c$=**0.780** $\sqrt{\frac{1}{407 \times 0.78\,(1-0.78)} + \frac{1}{411 \times 0.78\,(1-0.78)}}$ SE(Z.S)= 0.170

1. First lets check if our normal approximation is appropriate:

$$n\pi_1 > 5 \text{ and } n(1-\pi_1) > 5 \rightarrow 407(0.742) > 5 \text{ and } 407(0.258) > 5$$

$$n\pi_2 > 5 \text{ and } n(1-\pi_2) > 5 \rightarrow 411(0.815) > 5 \text{ and } 411(0.185) > 5$$

Both populations have more then 5 occurances of each binary output!

2. We find our confidence inteval by using our z multipier atf 95% which is 1.96:

$$log\phi_1 \pm z(0.975)SE(C.I) = (1.96)(0.170) \rightarrow 0.073 \pm 0.057 \rightarrow 0.094 - 0.760$$

3. We covert our log using a exponential

$$e^{0.094} - e^{0.760} \rightarrow 1.10 - 2.14$$

4. Next we use our z score to find our p value using a two tailed test given our hypothesis

$$z = \frac{log\phi}{SE(Z.S)} = \frac{0.427}{0.170} = 2.51 \rightarrow 0.0121$$

The odds of catching a cold for the placebo group are estimated to be 1.53 times the odss of a cold for the vitamin C grou (95% C.I: **1.10-2.14**) (**p = 0.012**). In other words your odds of getting a cold while taking the placebo was 53% more then if you took vitamin C.