
DS 705
Statistical Methods
Inference for Categorical Data – Part 3

Odds Ratios

Let's go back to the text book for an example of odds ratios. Example 10.16 on pp. 533-535 of the Ott textbook uses the following data.

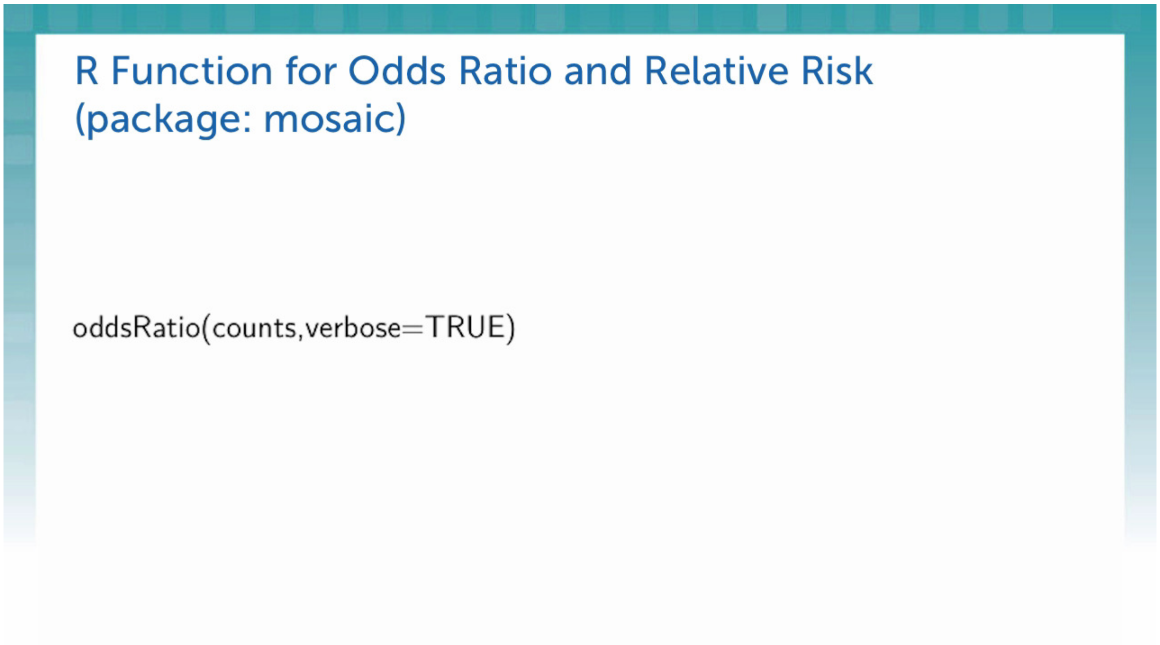
	Employee Response		
Job Stress	**Favorable**	**Unfavorable**	**Total**
Low	250	750	1,000
High	400	1,600	2,000
Total	650	2,350	3,000

Odds Ratios

The R code for entering the data in Example 10.16 on pp. 533-535 of the Ott textbook is

```
counts=matrix(c(250,400,750,1600),nrow=2)
rownames(counts) <- c("Low","High")
colnames(counts) <- c("Favorable","Unfavorable")
counts
```

```
##      Favorable Unfavorable
## Low        250         750
## High       400        1600
```



Using the odds ratio function will require you to install the package called Mosaic first. But it does a nice job of computing the proportions, relative risk, odds, and odds ratio, as well as the confidence intervals for the relative risk and odds ratio.

R Output for `oddsRatio(counts,verbose=TRUE)`

```
Proportions
        Prop. 1:   0.25   <-- "risk" of favorable response for Low Stress Job
        Prop. 2:   0.2    <-- "risk" of favorable response for High Stress Job
     Rel. Risk:   0.8    <-- 0.2 / 0.25

Odds
         Odds 1:   0.3333
         Odds 2:   0.25
    Odds Ratio:   0.75

95 percent confidence interval:
     0.6965 < RR < 0.9189
     0.6263 < OR < 0.8981
```

With the option for both equals true, we get all the output we want here. We get the proportions of a favorable response for both the low and high stress jobs, along with their ratio, the relative risk with row 2 in the numerator, 0.2 divided by 0.25 equals 0.8.

R Output for `oddsRatio(counts,verbose=TRUE)`

```
Proportions
       Prop. 1:   0.25
       Prop. 2:   0.2
     Rel. Risk:   0.8

Odds
        Odds 1:   0.3333
        Odds 2:   0.25
    Odds Ratio:   0.75

95 percent confidence interval:
     0.6965 < RR < 0.8189
     0.6263 < OR < 0.8981
```

250 / 750 = 0.3333

odds of favorable response for Low Stress Jo

R Output for `oddsRatio(counts,verbose=TRUE)`

```
Proportions
       Prop. 1:   0.25
       Prop. 2:   0.2
     Rel. Risk:   0.8

Odds
        Odds 1:   0.3333
        Odds 2:   0.25
    Odds Ratio:   0.75

95 percent confidence interval:
     0.6965 < RR < 0.9189
     0.6263 < OR < 0.8981
```

400 / 1600 = 0.25

odds of favorable response for High Stress Jo

R Output for oddsRatio(counts,verbose=TRUE)

```
Proportions
       Prop. 1:  0.25
       Prop. 2:  0.2
     Rel. Risk:  0.8

Odds
        Odds 1:  0.3333
        Odds 2:  0.25
    Odds Ratio:  0.75  <--  0.25 / 0.3333

95 percent confidence interval:
     0.6965 < RR < 0.9189
     0.6263 < OR < 0.8981
```

We get the odds of a favorable response for the low stress job as 250 divided by 750, which is one third, or as displayed here, 0.3333, and the odds of a favorable response for the high stress job, which is 400 divided by 1600, giving us 0.25.

And of course, we get the ratio of those odds with the odds for row 2 in the numerator as 0.25 divided by 0.3333 to get 0.7595.

R Output for `oddsRatio(counts,verbose=TRUE)`

```
Proportions
       Prop. 1:   0.25
       Prop. 2:   0.2
     Rel. Risk:   0.8

Odds
        Odds 1:   0.3333
        Odds 2:   0.25
    Odds Ratio:   0.75

95 percent confidence interval:
     0.6965 < RR < 0.9189
     0.6263 < OR < 0.8981
```

R Output for `oddsRatio(counts,verbose=TRUE)`

```
Proportions
       Prop. 1:   0.25
       Prop. 2:   0.2
     Rel. Risk:   0.8

Odds
        Odds 1:   0.3333
        Odds 2:   0.25
    Odds Ratio:   0.75  <-- reciprocal of 1.333 (textbook example value)

95 percent confidence interval:
     0.6965 < RR < 0.9189
     0.6263 < OR < 0.8981
```

95% confidence intervals for the relative risk and the odds ratio are also displayed. The level of confidence can be adjusted in the same manner as other functions in R. The odds ratio function processes the two by two table with the expectation that the successes are located in column 1 and the treatment of interest as in row 2.

The treatment of interest is the one that occupies the numerator in the odds ratio. It's the one that leads in the interpretation. This is why the odds ratio you see here, 0.75, does not match the one computed in the textbook, 1.333. Well, actually it does match. It's just that the comparison is from the perspective of low stress job rather than the high stress job. 0.75, or 3/4, is the reciprocal of 1.333, which is 4/3.

Odds Ratio: Interpretation Options When the OR = 0.75

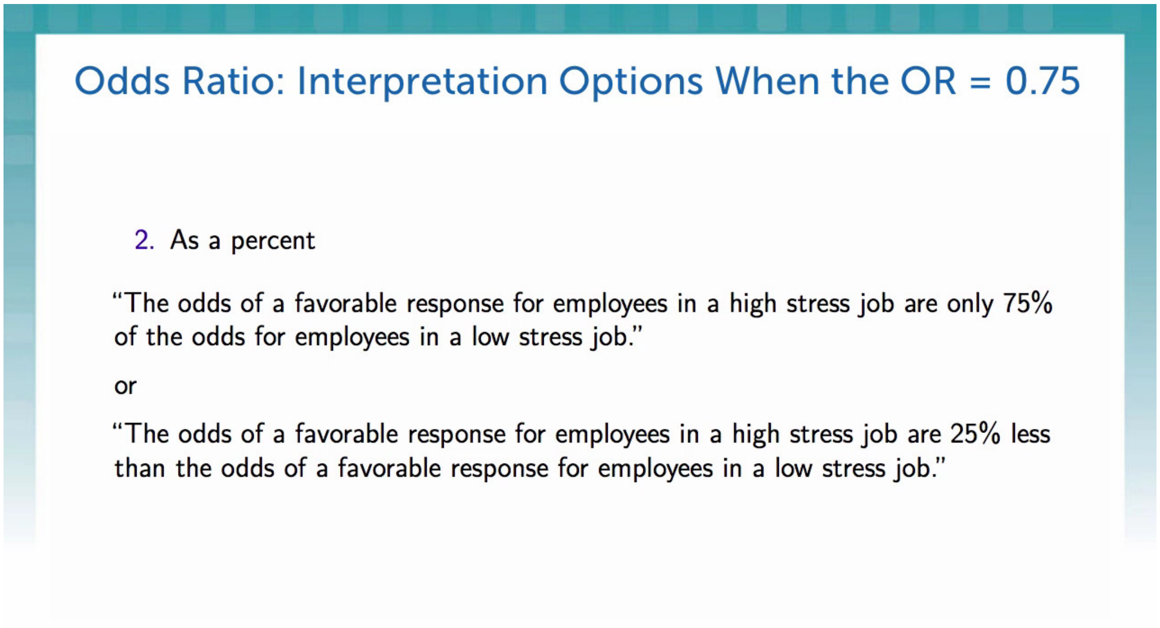
1. As a multiple

"The odds of a favorable response for employees in a high stress job are 0.75 times as large as the odds of a favorable response for employees in a low stress job."

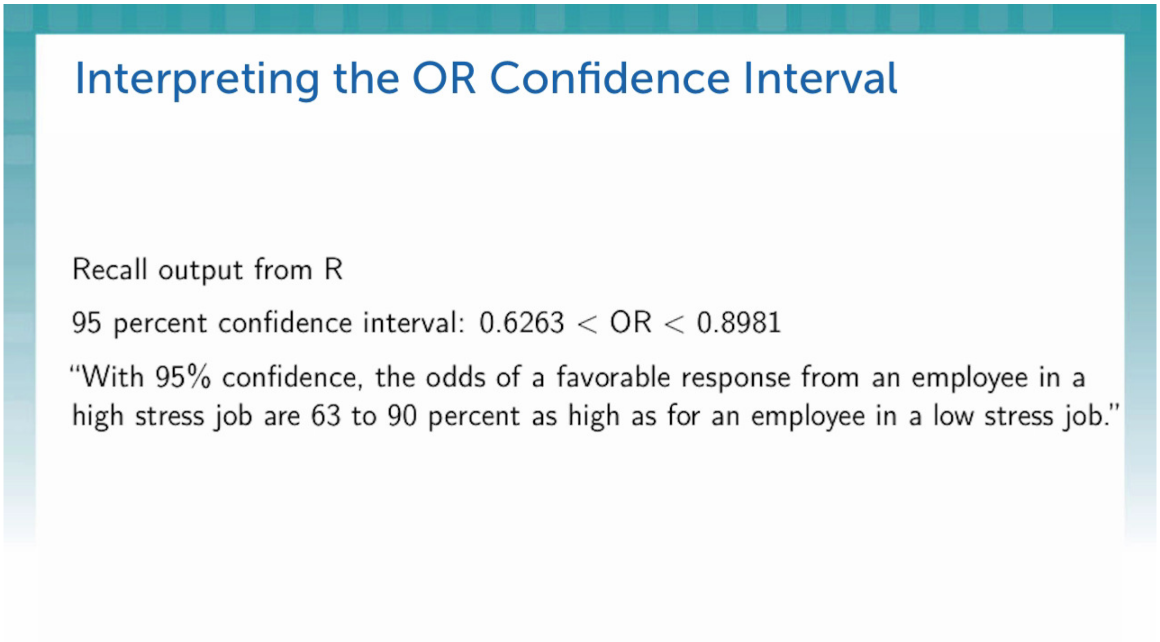
or

"The odds of a favorable response for employees in a high stress job are only three-fourths of the odds for employees in a low stress job."

Odds ratios can be interpreted in a variety of ways. In any case, one must proceed with caution when interpreting odds ratios, because they can so easily be misrepresented or misunderstood. Take your time to read these interpretations carefully.



Notice again here that we lead with the employees in the high stress job. And this is the odds ratio where you have the odds of the employees with a high stress job in the numerator. And we end up with an odds ratio that's less than 1. And here it is interpreted as a percent. I like this option, because I believe it's more common to express an odds ratio as a percent when it's less than 1. So take your time and read through these and let it sink in.



An odds ratio of 1 would tell us that the odds of an event for the first group are identical to the odds for the second group. When we see a confidence interval that does not contain 1, we can conclude that there is a statistically significant relationship between the two categorical factors. We could have equally said, with 95% confident, that the odds of a favorable response from an employee in a high stress job are 10% to 37% less than for an employee in a low stress job, rather than 63% to 90% as high.

Let's Reconstruct the 2x2 Table So Our Output Matches the Textbook Example Output

```
counts=matrix(c(400,250,1600,750),nrow=2)
rownames(counts) <- c("High","Low")
colnames(counts) <- c("Favorable","Unfavorable")
counts
```

```
##      Favorable Unfavorable
## High       400        1600
## Low        250         750
```

row 2

By entering the 2 by 2 table into R, such that the frequencies for the low stress job are now in row 2, so that R puts them in the numerator of the odds ratio, we can replicate the output, for example, 10.16, in the textbook.

R Output for Example 10.16 (Again)

```
Proportions
       Prop. 1:  0.2
       Prop. 2:  0.25
     Rel. Risk:  1.25

Odds
        Odds 1:  0.25
        Odds 2:  0.3333
    Odds Ratio:  1.333

95 percent confidence interval:
     1.088 < RR < 1.436
     1.113 < OR < 1.597
```

Notice now that the odds ratio and confidence interval bounds for the odds ratio now match the values given on page 534 of textbook.

Odds Ratio: Interpretation Options When the OR = 1.333

1. As a multiple

"The odds of a favorable response for employees in a low stress job are 1.33 times the odds of a favorable response for employees in a high stress job."

Odds Ratio: Interpretation Options When the OR = 1.333

2. As a percent

"The odds of a favorable response for employees in a low stress job are 133% of the odds for employees in a high stress job."

or

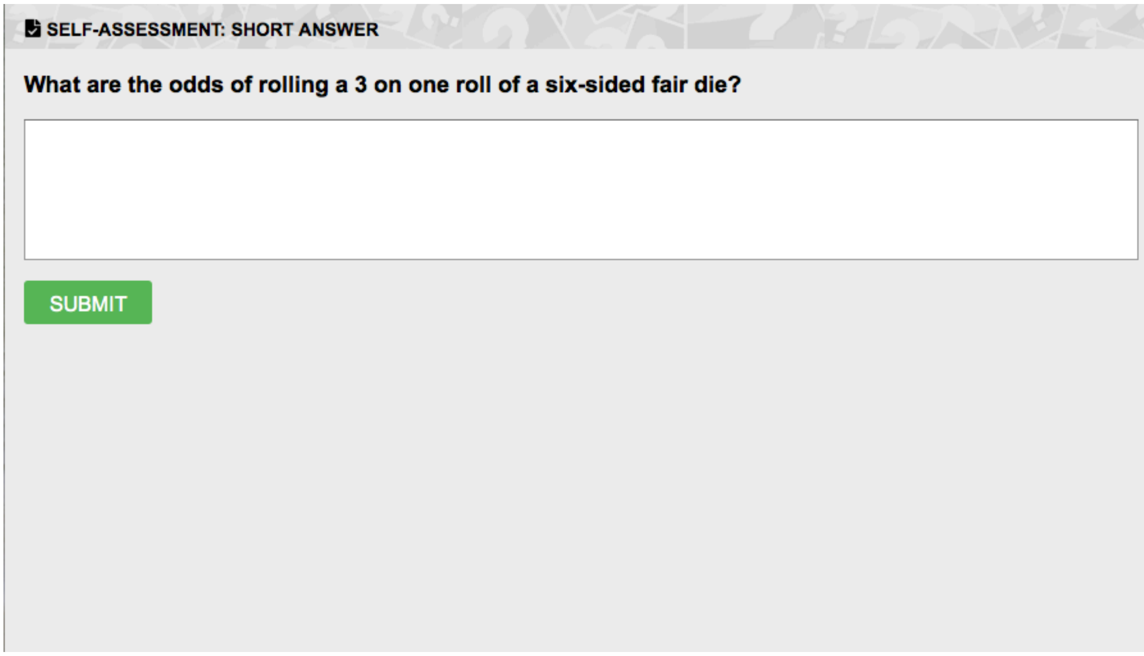
"The odds of a favorable response for employees in a low stress job are 33% more than the odds of a favorable response for employees in a high stress job."

Interpreting the OR Confidence Interval

Recall output from R

95 percent confidence interval: $1.113 < OR < 1.597$

"With 95% confidence, the odds of a favorable response from an employee in a low stress job are 11 to 60 percent higher than for an employee in a low stress job."


SELF-ASSESSMENT: SHORT ANSWER

What are the odds of rolling a 3 on one roll of a six-sided fair die?

SUBMIT

Question 1

See correct answer at the end of the transcripts.

SELF-ASSESSMENT: TRUE OR FALSE

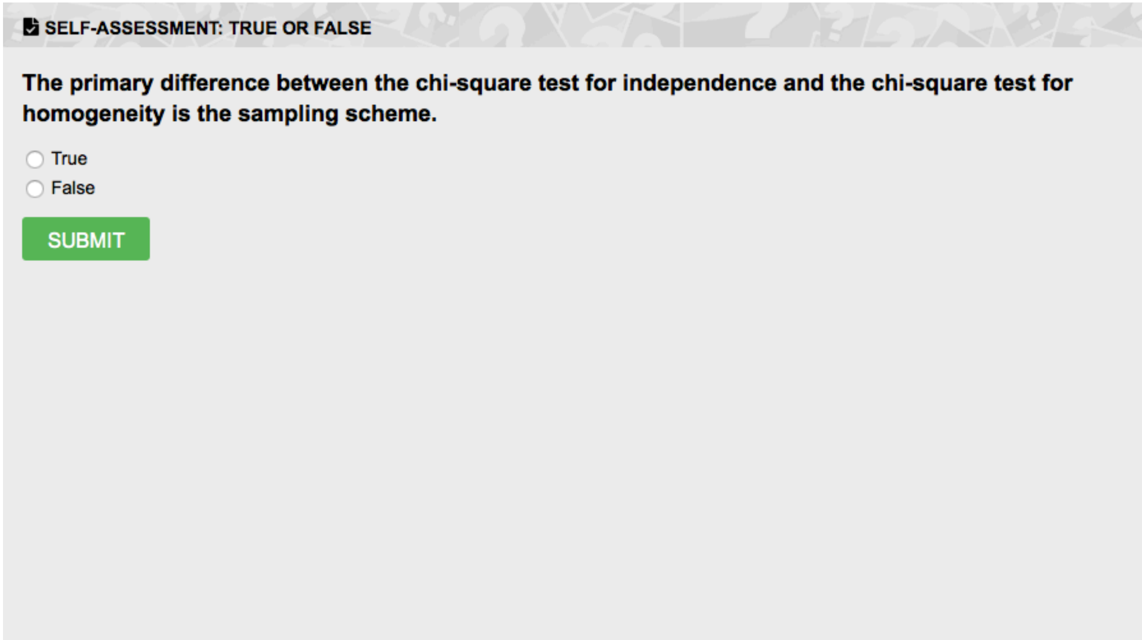
Chi-square tests for independence and homogeneity only can detect of a relationship exists between two categorical variables, but cannot provide a measure of the strength of the relationship.

- True
- False

SUBMIT

Question 2

See correct answer at the end of the transcripts.



Question 3

See correct answer at the end of the transcripts.

Question 1 answer:

SELF-ASSESSMENT FEEDBACK

What are the odds of rolling a 3 on one roll of a six-sided fair die?

Your answer:
–

Correct answer:
0.2

Other acceptable answers:
1/5
1:5

Question 2 answer:

SELF-ASSESSMENT FEEDBACK

CORRECT

Chi-square tests for independence and homogeneity only can detect of a relationship exists between two categorical variables, but cannot provide a measure of the strength of the relationship.

Your answer:
true

Correct answer:
true

Feedback:
You must have read the book!

Question 3 answer:

SELF-ASSESSMENT FEEDBACK

CORRECT

The primary difference between the chi-square test for independence and the chi-square test for homogeneity is the sampling scheme.

Your answer:
true

Correct answer:
true

Feedback:
Right. The test statistic and p-value are exactly the same. However, this difference will carry over into how the hypotheses are stated and how the conclusion is interpreted.