SDS 291 HW 7 Answers

Prof Capistrant

4/13/2020

9.18 or 9.28

9.18a or 9.28a

```
gmodels::CrossTable(Titanic$Sex,Titanic$Survived, prop.c = FALSE, prop.t = FALSE, prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1313
##
##
##              | Titanic$Survived
##  Titanic$Sex |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##       female |       154 |       308 |       462 |
##              |     0.333 |     0.667 |     0.352 |
## -------------|-----------|-----------|-----------|
##         male |       709 |       142 |       851 |
##              |     0.833 |     0.167 |     0.648 |
## -------------|-----------|-----------|-----------|
## Column Total |       863 |       450 |      1313 |
## -------------|-----------|-----------|-----------|
##
##
```

The proportion surviving for females is 0.667 (308/(308+154)) and the proportion surviving for males is 0.167 (142/(142+709)). The proportion for females is much higher than for males.

- *check if they did the calculation incorrectly*
- *check-minus if they did the calculating incorrectly*

9.18b / 9.28b

```
m1<-glm(Survived~SexCode, family = binomial(link = "logit"), data=Titanic)
summary(m1)
```

```
##
## Call:
## glm(formula = Survived ~ SexCode, family = binomial(link = "logit"),
##     data = Titanic)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4823  -0.6042  -0.6042   0.9005   1.8924
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.60803    0.09194  -17.49   <2e-16 ***
## SexCode      2.30118    0.13488   17.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1688.1  on 1312  degrees of freedom
## Residual deviance: 1355.5  on 1311  degrees of freedom
## AIC: 1359.5
##
## Number of Fisher Scoring iterations: 4
```

We see that the association is positive - female passengers had a higher log(odds) of survival than males - which is consistent with the higher proportions from a.

- *check: fit the model correctly (used SexCode), interpreted the direction of the slope, may have interpreted it on the log(odds) scale but didn't convert the slope to ORs*
- *check-minus: if they estimate the wrong model, including comparing males to females or using alive as the reference group and predicting whether people would be dead. Especially if they don't connect the answer to the proportions above*

9.19a

```
cbind(OR=exp(coef(m1)), exp(confint(m1)))
```

```
## Waiting for profiling to be done...
```

```
##                    OR     2.5 %     97.5 %
## (Intercept) 0.2002821 0.1666302  0.2389966
## SexCode     9.9859155 7.6867382 13.0454834
```

From the coefficient of SexCode we estimate an odds ratio to be $\text{OR} = e^{2.30118} = 9.986$.

This means that the odds for a female surviving are, on average, about 10 times higher than the odds for a male surviving.

- *check: for an correct calculation of the OR from the model*
- *check-minus: for an incorrect calculation and incorrect interpretation*

##9.19b

Using the output from the previous answer, we find a 95% confidence interval for the odds ratio to be (7.67, 13.01). We are 95% confident that, on average, the odds for females surviving are between 7.7 and 13.0 times the odds for males surviving.

- *check: for an correct calculation* and *interpretation of the OR from the model*

- *check-minus: for an incorrect calculation or incorrect interpretation*

##9.19c

Based on the two-way table, we can find the odds ratio for survival of females compared to males. OR = $\frac{\frac{308}{154}}{\frac{142}{709}}$ = $\frac{308 \times 709}{142 \times 154} = 9.986$

To find the answer we got in 18.b., the log of the odds ratio gives the estimated coefficient for SexCode in the model, $\log(9.986) = 2.301$.

- *check-plus: for an correct calculation – that they calculated the log(odds) based on what they had from the two-by-two table*
- *check: they calculate the OR from the two-by-two table and illustrated how it matched the OR from the model*
- *check-minus: for an incorrect calculation* and *incorrect interpretation*

##9.19d

$\hat{\pi} = \frac{e^{-1.60803+2.301(1)}}{1+e^{-1.60803+2.301(1)}} = \frac{2}{3} = 0.667$

The probability that a woman survived is 0.667, which is the same value we saw in the proportions from the 2-by-2 table.

- *check: they calculate the probability from the model estimate and connect it to their answer from 9.18a.*
- *check-minus: for an incorrect calculation* and *incorrect / incomplete interpretation*

9.19e

Linear: This is easy to say is true, since two points make a line – the relationship sort of *must* be linear

Random: It is highly unlikely that female survival was random. They weren't handing out lottery tickets to get on a life boat.

Independent: It is unlikely that these data are independent – if a mother got on the life boat, likely so did her children.

- *check: they mention each of these conditions and give some reasonable explanation of why each would apply*

- *check-minus: if they seem to really misunderstand more than 1 of the assumptions/conditions*

Recommended problems

Just note that they did them and say good job!

9.20 / 9.31

9.17 / 9.27

9.30 / 9.32

Just note that they did it on the grading sheet