Problem set 9

Dr. McNamara

Exercises to hand in: 9.4, 9.6, 9.22, 9.27, 9.28

9.4 Probability to odds

(no dataset)

a. If the probability of an event occurring is 0.8, what are the odds?

probability $= \pi = 0.8$

odds $= \pi/1 - \pi = 4$

b. If the probability of an event occurring is 0.25, what are the odds?

probability $= \pi = 0.25$

odds $= \pi/1 - \pi = 0.33$

c. If the probability of an event occurring is 0.6, what are the odds?

probability $= \pi = 0.6$

odds $= \pi/1 - \pi = 1.5$

9.6 Odds to probabilities

(no dataset)

a. If the odds of an event occurring are 1:3, what is the probability?

odds $= \pi/1 - \pi = 1{:}3$

probability $= \pi = 0.25 = 1/4$

b. If the odds of an event occurring are 5:2, what is the probability?

odds $= \pi/1 - \pi = 5{:}2$

probability $= \pi = 0.71 = 5/7$

c. If the odds of an event occurring are 1:9, what is the probability?

odds $= \pi/1 - \pi = 1{:}9$

probability $= \pi = 0.1 = 1/10$

9.22 Dementia: Odds and probability

(no dataset)

a. Estimate odds

$log(\hat{\pi}/1 - \hat{\pi})$ = -0.742 - 0.294 * MMSE

MMSE = -4

So, $log(\hat{\pi}/1 - \hat{\pi}) = 0.434$

Taking exponential on both sides

```
exp(0.434)
```

```
[1] 1.543419
```

Taking exponential on both sides. So, the odds of Alzheimer's disease if a patient's MMSE is -4 is 1.543419.

b. Estimate probability

$\hat{\pi}/1 - \hat{\pi} = 1.543419$

$\hat{\pi} = 1.543419(1 - \hat{\pi})$

$2.543419\hat{\pi} = 1.543419$

So, $\hat{\pi} = 1.543419/2.543419$

```
1.543419/2.543419
```

```
[1] 0.6068284
```

The probability of Alzheimer's disease if a patient's MMSE is -4 is 0.6068284.

c. How much do the estimated odds change if the MMSE changes from -4 to -3?

$log(\hat{\pi}/1-\hat{\pi})$ = -0.742 - 0.294 * MMSE

MMSE = -3

So, $log(\hat{\pi}/1-\hat{\pi}) = 0.14$

Taking exponential on both sides

$(\hat{\pi}/1-\hat{\pi}) = 1.150274$

```
-0.742 - (0.294 * -3)
```

```
[1] 0.14
```

```
exp(0.14)
```

```
[1] 1.150274
```

```
exp(0.14) - exp(0.434)
```

```
[1] -0.3931451
```

The estimated odds changes by 0.3931451 if the MMSE changes from -4 to -3.

d. How much does the estimate of probability change if the MMSE changes from -4 to -3?

$\hat{\pi}/1-\hat{\pi} = 1.150274$

$\hat{\pi} = 1.150274(1-\hat{\pi})$

$2.150274\hat{\pi} = 1.150274$

So, $\hat{\pi} = 1.150274/2.150274$

```
1.150274/2.150274
```

```
[1] 0.534943
```

```r
0.534943 - 0.6068284
```

```
[1] -0.0718854
```

The estimate of probability will change by 0.0718854 if the MMSE changes form -4 to -3.

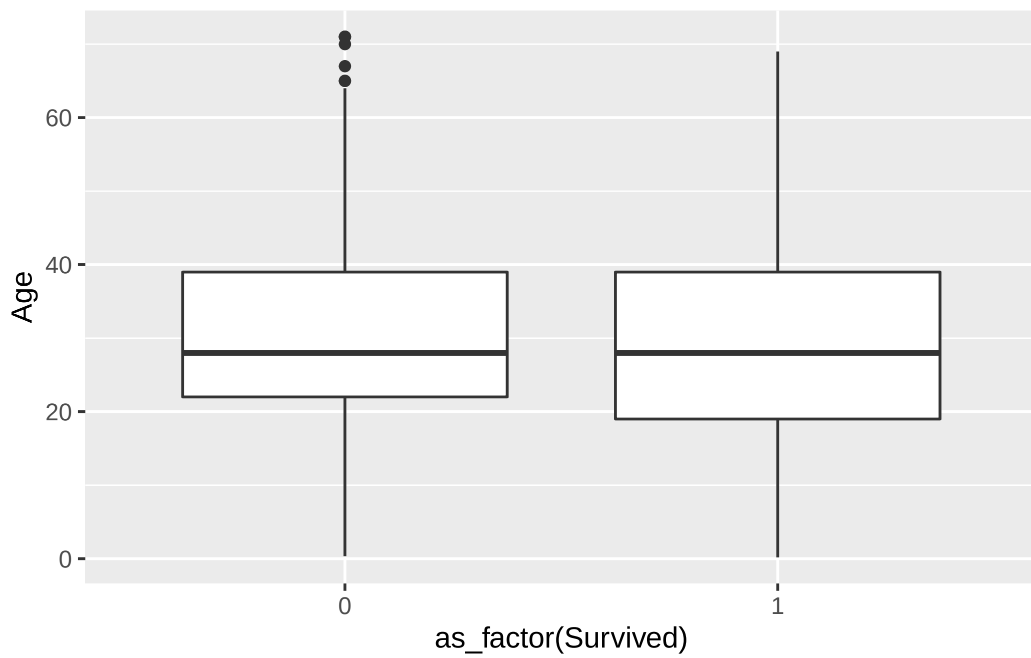
9.27 Titanic: Survival and age, CHOOSE and FIT

```
data("Titanic")
```

a. Plot

```
ggplot(Titanic) + geom_boxplot(aes(x=as_factor(Survived), y=Age))
```

```
Warning: Removed 557 rows containing non-finite values (stat_boxplot).
```



From the above box plot, both appears to be similar. This means that age is similarly effective on both box plots.

b. Logistic model

```r
model1<-glm(Survived~Age, data=Titanic,family=binomial)
summary(model1)
```

```
Call:
glm(formula = Survived ~ Age, family = binomial, data = Titanic)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1418  -1.0489  -0.9792   1.3039   1.4801

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.081428   0.173862  -0.468   0.6395
Age         -0.008795   0.005232  -1.681   0.0928 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1025.6  on 755  degrees of freedom
Residual deviance: 1022.7  on 754  degrees of freedom
  (557 observations deleted due to missingness)
AIC: 1026.7

Number of Fisher Scoring iterations: 4
```

Null Hypothesis: Beta 1 =0

Alternative Hypothesis: Beta 1 not equals to 0.

Reject the null hypothesis when the p value is less than the level of significance. This means that Age and survived variables are not independent to each other.

9.28 Titanic: Survival and sex, CHOOSE and FIT.

a. Two way table

```r
mosaic::tally(Survived~Sex, data = Titanic)
```

```
Registered S3 method overwritten by 'mosaic':
  method                           from
  fortify.SpatialPolygonsDataFrame ggplot2
```

```
        Sex
Survived female male
       0    154  709
       1    308  142
```

```r
mosaic::tally(Survived~Sex, data = Titanic, format = "proportion")
```

```
        Sex
Survived    female      male
       0 0.3333333 0.8331375
       1 0.6666667 0.1668625
```

According to the two way table above, 33.33% out of the total females on the ship died whereas 83.31% out of the total males on the ship died. This means that a greater percentage of male passengers died as compared to female showing a possible relation between sex of the passenger and survival.

b. Logistic model

```r
logm1 <- glm(Survived ~ SexCode, data = Titanic, family = binomial)
summary(logm1)
```

```
Call:
glm(formula = Survived ~ SexCode, family = binomial, 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
```

```r
exp(coef(logm1))
```

```
(Intercept)     SexCode
  0.2002821   9.9859155
```

Since the p-value is less than $\alpha = 0.05$, we reject the null hypothesis. Therefore, we can say that there is a relationship between the sex of the passenger and their survival.

$$\log\left(\frac{\pi}{1-\pi}\right) = logit(\pi) = \beta_0 + \beta_1 X$$

"'

So β_0 is -1.60803 and β_1 is 2.30118.

$$\log\left(\frac{\pi}{1-\pi}\right) = logit(\pi_{survived}) = -1.60803 + 2.30118 * SexCode$$

"'

Log Odds: A one unit increase in SexCode is associated with a 2.3 increase in the log odds of surviving

Odds space: A one unit increase in SexCode is associated with multiplying the odds of surviving by 9.98

Probability space: No nice sentence.

Therefore, we can say that the interpretation provided by the logistic model confirms the descriptive analysis.