STK3100
Introduction to Generalized Linear Models

OBLIG 1

Egil Furnes
Studentnummer: 693784

Problem 1

a)

We have data $(x_i, Y_i), \quad i = 1, \dots, n \quad \text{for} \quad n = 16$, where model M_1 is the standard linear regression model $\mu_i = \beta_0 + \beta_1 x_i$ and M_0 is the same model without the intercept.

The model matrix for M_0:

$$\mu = X_0 \beta^*, \quad X_0 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \beta^* = [\beta_1^*]$$

The model matrix for M_1:

$$\mu = X_1 \beta, \quad X_1 = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Here X_1 has two columns where x is not constant such that the columns are not linearly dependeant, and $\text{rank}(X_1) = 2$. For X_0 we have one column and if we set $x_i \neq 0 \quad \forall i \in n$ the column is nonzero and as such $\text{rank}(X_0) = 1$.

Here the model M_0 could be nested within M_1 if we can get M_0 by a special case of M_1 by setting restrictions top its parameters. If we set $\beta = 0$ then M_0 can be obtained from M_1 and as such we could say $C(X_0) \subseteq C(X_1)$.

b)

Here we seek to find the projection matrices $\mathbf{P}_0$ and $\mathbf{P}_1$.

If we let $\mathbf{1}_m$ be the $n \times 1$ vector of ones $x = (x_1, \dots, x_n)^\top$ and set the following:

$$X_0 = x \quad (n \times 1) \quad X_1 = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} 1_n & x \end{bmatrix} \quad (n \times 2)$$

Then we find the projection matrices by:

$$\mathbf{P}_0 = X_0 (X_0^\top X_0)^{-1} X_o^\top = \frac{x x^\top}{x^\top x}$$

$$\mathbf{P}_1 = X_1 (X_1^\top X_1)^{-1}$$

c)

First we set $x = (x_1, \dots, x_n)^\top$ to $\mathbf{1}_n$ and $Y = (Y_1, \dots, Y_n)^\top$.

For M_0 we have:

$$X_0 = x \quad \mathbf{P}_0 = X_0(X_0^\top X_0)^{-1}X_o^\top$$
$$\hat{\mu_0} = P_0 Y = x(x^\top x)^{-1}x^\top Y = \hat{\beta}_1^* x$$
$$\hat{\beta}_1^* = \frac{x^\top Y}{x^\top x} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$$

For M_1 we have:

$$X_1 = \begin{bmatrix} \mathbf{1}_n & x \end{bmatrix} \quad \mathbf{P}_1 = X_1(X_1^\top X_1)^{-1}X_1^\top$$
$$\hat{\mu_1} = P_1 Y = X_1\hat{\beta}_1 = \hat{\beta}_0\mathbf{1}_n + \hat{\beta}_1 x$$

Then from normal equations we can set

$$X_1^\top(Y - X_1\hat{\beta}) = 0$$

And solve for

$$\mathbf{1}_n^\top(Y - \hat{\beta}_0\mathbf{1}_n - \hat{\beta}_1 x) = 0 \Rightarrow n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum Y_i$$
$$x^\top(Y - \hat{\beta}_0\mathbf{1}_n - \hat{\beta}_1 x) = 0 \Rightarrow \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i Y_i$$

Solving for $\hat{\beta}_0$:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x}$$
$$\overline{Y} = \frac{1}{n}\sum Y_i$$
$$\overline{x} = \frac{1}{n}\sum x_i$$

From this we can find the formula for the centered slope:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})Y_i}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

As such:

$$\hat{\mu}_1 = \hat{\beta}_0\mathbf{1}_n + \hat{\beta}_1 x = (\overline{Y} - \hat{\beta}_1\overline{x})\mathbf{1}_n + \hat{\beta}_1 x = \overline{Y}\mathbf{1}_n + \hat{\beta}_1(x - \overline{x}\mathbf{1}_n)$$

As such we find:

$$\hat{\mu}_0 = \hat{\beta}_1^* x$$
$$\hat{\beta}_1^* = \frac{\sum x_i Y_i}{\sum x_i^2}$$
$$\hat{\mu}_1 = \overline{Y}\mathbf{1}_n + \hat{\beta}_1(x - \overline{x}\mathbf{1}_n)$$
$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})Y_i}{\sum (x_i - \overline{x})^2}$$

d)

We are to show that these are independent and determine their distribution:

$$\mathbf{Y}^T(\mathbf{P}_1 - \mathbf{P}_0)\mathbf{Y}/\sigma^2$$
$$\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_1)\mathbf{Y}/\sigma^2$$

If we let $Y \sim N(\mu, \sigma^2 I_n)$ we have from before:

$$P_0 Y = x(x^\top x)^{-1} x^\top Y$$
$$\mathbf{P}_1 = X_1(X_1^\top X_1)^{-1} X_1^\top$$

To find independence we need to satisfy the conditions that (i) A and B are symmetric idempotent, or orthogonal projections, (ii) orthogonality with $AB = 0$, (iii) that the ranks give the chi-squared degrees of freedom and (iv) non-centrality.

Since $M_0 \subset M_1$ is nested we have $P_0 P_1 = P_1 P_0 = P_0$

First seeing $\mathbf{P}_0$ and $\mathbf{P}_1$ are symmetric idempotent then:

$$(P_1 - P_0)^2 = P_1 - P_0 P_1 - P_1 P_0 + P_0 = P_1 - P_0$$

Here we also have that

$$\text{rank}(P_1 - P_0) = \text{tr}(P_1) - \text{tr}(P_0) = 2 - 1 = 1$$
$$\text{rank}(I - P_1) = n - \text{tr}(P_1) = n - 2$$

For orthogonality we have

$$(P_1 - P_0)(I - P_1) = P_1 - P_1^2 - P_0 + P_0 P_1 = P_1 - P_1 - P_0 + P_0 = 0$$

Using Cochrans theorem we say that if A and B are symmetric idempotent with $AB = 0$ then the following are independent

$$\frac{Y^\top A Y}{\sigma^2} \quad \text{and} \quad \frac{Y^\top B Y}{\sigma^2}$$

As such with M_0 we have the independent

$$\frac{Y^\top (P_1 - P_0) Y}{\sigma^2} \sim \chi_1^2 \quad \frac{Y^\top (I - P_1) Y}{\sigma^2} \sim \chi_{n-2}^2$$

e)

We have

$$\hat{\mu}_j = P_j Y$$

$$Y - \hat{\mu}_j = (I - P_j)Y$$

As such we can write F as:

$$F = \frac{\|\hat{\mu}_1 - \hat{\mu}_0\|^2}{\|Y - \hat{\mu}_1\|^2/(n-2)} = \frac{Y^\top(P_1 - P_0)Y}{Y^\top(I - P_1)Y/(n-2)}.$$

Using the following

$$\hat{\mu}_0 = \hat{\beta}_1^* x$$

$$\hat{\mu}_1 = \bar{Y}\,1_n + \hat{\beta}_1(x - \bar{x}\,1_n)$$

We can therefore rewrite as

$$F = \frac{\sum_{i=1}^n \left(\bar{Y} - \hat{\beta}_1\bar{x} + (\hat{\beta}_1 - \hat{\beta}_1^*)x_i\right)^2}{\sum_{i=1}^n \left(Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})\right)^2/(n-2)}.$$

Now we look at M_0 under Cochrans theorem we know that these are independent

$$Y^\top(P_1 - P_0)Y/\sigma^2 \sim \chi_1^2$$

$$Y^\top(I - P_1)Y/\sigma^2 \sim \chi_{n-2}^2$$

As such we can find

$$F \sim F_{1,n-2}$$

f)

```r
# f)

library(tidyverse)

wingspan <- read.table("data/wingspan.txt",
                       header = TRUE)
wingspan

M0 <- lm(Wingspan~0+Height,
         data = wingspan)

M1 = lm(Wingspan~Height,
         data = wingspan)

anova(M0,M1)
```

```
Analysis of Variance Table

Model 1: Wingspan ~ 0 + Height
Model 2: Wingspan ~ Height
  Res.Df    RSS Df Sum of Sq
1     15 43.116
2     14 36.271  1    6.8451
       F Pr(>F)
1
2 2.6421 0.1264
```

Looking at the summery from anova we can assess that the residual sum of squares (RSS) is higher for M_0 with 43.116 compared to M_1 with 36.271. Furthermore the F-statistic in this summary is 2.6421 with a p-value of 0.1264 (which is not at the $< 5\%$ which we would like).

Problem 2

a)

```
titanic <- read.table("data/titanic.txt",
                      header = TRUE)

glm(survived~pclass,
    data = titanic,
    family = binomial) %>%
  summary()
```

```
Call:
glm(formula = survived ~ pclass, family = binomial, data = titanic)

Coefficients:
            Estimate Std. Error
(Intercept)  0.06899    0.37161
pclass      -1.33750    0.52958
            z value Pr(>|z|)
(Intercept)   0.186   0.8527
pclass       -2.526   0.0116 *
---
Signif. codes:
  0 ‘’*** 0.001 ‘’** 0.01 ‘’*
  0.05 ‘’. 0.1 ‘’  1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 90.008  on 69  degrees of freedom
Residual deviance: 83.324  on 68  degrees of freedom
AIC: 87.324
```

```

Number of Fisher Scoring iterations: 4
```

Looking at the logistic regression we see a value of $\hat{\beta}_1 =$ `-1.33750` with a z-value of `-2.526` and a p-value of `0.0116 *`.

This indicates that the coefficient is statistically significant at a 1% level and as such we conclude that the ticket class `pclass` has a significant effect on survival probability.

b)

We know that

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

for $\pi = P(\text{survived=1}|\texttt{pclass})$

If we instead look at e^{β_1} this is the odds ratio when comparing `pclass` $= 1$ for 3rd class versus `pclass` $= 0$ for 1st and 2nd class.

From our summary $\hat{\beta}_1 = -1.338$ becomes $e^{\hat{\beta}_1} = \exp(-1.338) \approx 0.26$.

As such we find that the estimated odds of survival for a passenger of `pclass` $= 1$ is approximately 0.26 times the odds for a passenger in 1st or 2nd class.

c)

```r
glm(survived~pclass+age,
    data = titanic,
    family = binomial) %>%
  summary()
```

```
Call:
glm(formula = survived ~ pclass + age, family = binomial, data =
   titanic)

Coefficients:
            Estimate Std. Error
(Intercept)  1.55623    0.87390
pclass      -1.73105    0.59825
age         -0.04018    0.02125
            z value Pr(>|z|)
(Intercept)   1.781  0.07495 .
pclass       -2.894  0.00381 **
age          -1.891  0.05859 .
---
Signif. codes:
  0 ‘’*** 0.001 ‘’** 0.01 ‘’*
```

```
  0.05 ‘’. 0.1 ‘’  1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 90.008  on 69  degrees of freedom
Residual deviance: 79.397  on 67  degrees of freedom
AIC: 85.397

Number of Fisher Scoring iterations: 4
```

From the regression output we now find that age is significant at a .0.05 or 5% level and as such we find it to have a statistically significant effect on the response variable `survived`.

We have

$$\hat{\beta}_2 = \texttt{-0.04018}$$
$$e^{\hat{\beta}_2} = \exp(-0.04018) \approx 0.961$$

As such for every one unit of age it multiplies the odds of survival by 0.96 or about 4% lower odds of survival per extra year.

d)

```r
M0 <- glm(survived~pclass, data = titanic, family = binomial)
M1 <- glm(survived~pclass+age, data = titanic, family = binomial)

wald <- summary(M1)$coefficients["age",]
wald

lrt <- anova(M0,M1,test = "Chisq")
lrt
```

```
> wald
   Estimate  Std. Error
-0.04018375  0.02124713
    z value    Pr(>|z|)
-1.89125548  0.05859025

> lrt
Analysis of Deviance Table

Model 1: survived ~ pclass
Model 2: survived ~ pclass + age
  Resid. Df Resid. Dev Df
1        68     83.324
2        67     79.397  1
  Deviance Pr(>Chi)
1
2   3.9265  0.04753 *
```

```
---
Signif. codes:
  0 ‘’*** 0.001 ‘’** 0.01 ‘’*
  0.05 ‘’. 0.1 ‘’  1
```

Looking at the wald test we get $\hat{\beta}_{\text{age}} = -0.0402$ with SE $= 0.0212$ and $p = 0.059$ which is not significant.

The likelihood test gives us a `Deviance` of 3.93 at $p = 0.0475 < 5\%$ which indicates statistical significance.

As such these tests all indicate and agree that `age` has a negative effect on survival. For the Wald test we have 10% significance and for likelihood test 5%.

e)

```
M0 %>% confint()
M1 %>% confint()
M0 %>% confint() %>% exp()
M1 %>% confint() %>% exp()
```

```
> M0 %>% confint()
                 2.5 %     97.5 %
(Intercept) -0.6643751  0.8085452
pclass      -2.4105375 -0.3199407

> M1 %>% confint()
                  2.5 %        97.5 %
(Intercept) -0.09130037  3.3719740892
pclass      -2.97572548 -0.6066245106
age         -0.08457772 -0.0004218667

> M0 %>% confint() %>% exp()
                 2.5 %    97.5 %
(Intercept) 0.51459500 2.2446400
pclass      0.08976703 0.7261921

> M1 %>% confint() %>% exp()
                 2.5 %     97.5 %
(Intercept) 0.91274351 29.1359874
pclass      0.05101041  0.5451880
age         0.91890024  0.9995782
```

As such we have the confidence intervals for e^{β_1} from b) and e^{β_2} from c)

$$
\begin{aligned}
e^{\beta_1} &\in [0.090,\ 0.726] \\
e^{\beta_2} &\in [0.919,\ 0.999]
\end{aligned}
$$