Assignment 1 STK4100
Hanne Rønning Berg

1a)

M_1: $\mu_i = B_0 + B_1 x_i$

M_1: $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{bmatrix}$ Rank 2, as long as not all x_i are the same

M_0: $\mu_i = B_1^* x_i$

M_0: $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}$ Rank 1.

The models are nested. M_0 is a special case of M_1, but without the intercept term.

Since we get M_0 from M_1 when $B_0 = 0$, M_0 is nested in M_1.

b) $P_X = X(X^TX)^{-1}X^T$

$$P_0 = \mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T$$

$$\mathbf{x}^T\mathbf{x} = [x_1 \ x_2 \ \ldots] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \sum x_i^2$$

$$(\mathbf{x}^T\mathbf{x})^{-1} = \frac{1}{\sum x_i^2}$$

$$= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} \left([x_1 \ x_2 \ x_3 \ldots] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} \right)^{-1} [x_1 \ x_2 \ x_3 \ldots]$$

$$= \frac{1}{\sum x_i^2} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} [x_1 \ x_2 \ x_3 \ldots]$$

$$= \frac{1}{\sum x_i^2} \begin{bmatrix} x_1^2 & x_1x_2 & \ldots & x_1x_n \\ x_1x_2 & x_2^2 & \ldots & x_2x_n \\ \vdots & & \ddots & \\ x_1x_n & & & x_n^2 \end{bmatrix}$$

b) $[\mathbf{1} \ \mathbf{x}] \cdot \left(\begin{bmatrix} 1 & 1 & 1 & \cdots \\ x_1 & x_2 & x_3 & \cdots \end{bmatrix} [\mathbf{1}_k \ \mathbf{x}] \right)^{-1} \cdot \begin{bmatrix} 1 & 1 & \cdots \\ x_1 & x_2 & \cdots \end{bmatrix}$

$$\left(\begin{bmatrix} 1 & 1 & 1 & \cdots \\ x_1 & x_2 & x_3 & \cdots \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{bmatrix} \right)^{-1} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1}$$

as long as not all x_i are equal, the matrix is invertible

$$P_1 = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{bmatrix} \cdot \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & \cdots \\ x_1 & x_2 & \cdots \end{bmatrix}$$

$$= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} \sum x_i^2 - x_1 \sum x_i & \sum x_i^2 - x_2 \sum x_i & \cdots \\ -\sum x_i + n x_1 & -\sum x_i + n x_2 & \cdots \end{bmatrix}$$

$$= \frac{1}{n\sum x_i - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 - x_1 \sum x_i - x_i(\sum x_i + n x_i) & \cdots & \sum x_i^2 - x_n \sum x_i - (\sum x_i + n x_n) x_i \\ \sum x_i^2 - x_i \sum x_i - x_2(\sum x_i + n x_i) & \ddots & \\ \vdots & & \\ \sum x_i^2 - x_1 \sum x_i - x_n(\sum x_i + n x_1) & \cdots & \sum x_i^2 - x_n \sum x_i - (\sum x_i + n x_n) \end{bmatrix}$$

c) $\hat{\mu}_0 = \hat{\beta}_1^* \mathbf{x}$, $\hat{\beta}_1^* = \dfrac{\sum x_i y_i}{\sum x_i^2}$

$\hat{\mu}_0 = P_0 y$

$$\frac{1}{\sum x_i^2} \begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1 x_n \\ x_1 x_2 & x_2^2 & & x_2 x_n \\ x_1 x_3 & & \ddots & \vdots \\ \vdots & & & \vdots \\ x_1 x_n & & & x_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \frac{1}{\sum x_i^2} \begin{bmatrix} x_1^2 y_1 + x_1 x_2 y_2 + x_1 x_3 y_3 \cdots + x_1 x_n y_n \\ x_1 x_2 y_1 + x_2^2 y_2 + x_2 x_3 y_3 \cdots + x_2 x_n y_n \\ \vdots \\ x_1 x_n y_1 + x_n x_2 y_2 \cdots x_n^2 y_n \end{bmatrix}$$

Each row is $\left(\sum x_i y_i\right) \cdot x_i$

i: for each row.

So $\hat{\mu}_0 = \frac{1}{\sum x_i^2} \cdot \sum x_i y_i \cdot \mathbf{x}$

$= \hat{\beta}_1^* \mathbf{x}$

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

The hat matrix for the bivariate model with centered data:

$$\begin{bmatrix} \frac{1}{n}\frac{(x_1-\bar{x})^2}{\Sigma(x_i-\bar{x})^2} & \cdots & \frac{1}{n}\frac{(x_1-\bar{x})(x_n-\bar{x})}{\Sigma(x_i-\bar{x})^2} \\ \vdots & \ddots & \\ \frac{1}{n}\frac{(x_n-\bar{x})(x_1-\bar{x})}{\Sigma(x_i-\bar{x})^2} & & \frac{1}{n}+\frac{(x_n-\bar{x})^2}{\Sigma(x_1-\bar{x})^2} \end{bmatrix}$$

$$\hat{\mu}_1 = Hy = \begin{bmatrix} \frac{1}{n}\frac{(x_1-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}y_1 + \dots + \frac{1}{n}\frac{(x_1-\bar{x})(x_n-\bar{x})}{\Sigma(x_1-\bar{x})^2}y_n \\ \vdots \\ \frac{1}{n}\frac{(x_n-\bar{x})(x_1-\bar{x})}{\Sigma(x_i-\bar{x})^2}y_1 \dots + \frac{1}{n}\frac{(x_n-\bar{x})^2}{\Sigma(x_1-\bar{x})^2}y_n \end{bmatrix}$$

For each row in the matrix: $\frac{\Sigma y_i}{n} + \frac{n \cdot (x_i-\bar{x}) \cdot \Sigma(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2} y_i$

So $\hat{\mu}_1 = \bar{Y}\mathbf{1}_n + \frac{\Sigma(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2}(\mathbf{x} - \bar{x}\mathbf{1}_n)\, Y_i$

d) $Y^T(P_1 - P_0)Y/\sigma^2$ and $Y^T(I - P_a)Y/\sigma^2$

For cochrans theorem $\Sigma P_i = I$

From orthogonal decomposition:

$P_0 + (P_1 - P_0) + (I - P_1) = I$

Then $\{Y^T P_i Y\}$ are independent and

$\frac{1}{\sigma^2} Y^T P_i Y \sim \chi^2_{r_i, \lambda_i}$

r_i is $Rank(P_i)$

$Rank(P_1 - P_0) = 1$, because $Rank(P_1) = 2$

and $Rank(P_0) = 1$.

$\lambda_i = \frac{1}{\sigma^2} \mu^T P_i \mu$

e)

$$F = \frac{Y^T(P_1 - P_0)Y/1}{Y^T(1-P_1)Y/(n-2)}$$

$$= \frac{Y^T(P_1-P_0)(P_1-P_0)Y}{Y^T(1-P_1)Y/(n-2)}$$

$P^2 = P$ for projection matrices

$$= \frac{((P_1-P_0)y)^T((P_1-P_0)y)}{Y^T(1-P_1)Y/(n-2)}$$

$(BA)^T = A^T B^T$ and projection matrices are symmetrical

$$= \frac{\|\hat{\mu}_1 - \hat{\mu}_0\|^2}{Y^T(1-P_1)Y/(n-2)}$$

same procedure for denomenator

$$= \frac{\|\hat{\mu}_1 - \hat{\mu}_0\|^2}{\|Y - \hat{\mu}_1\|^2/(n-2)}$$

Problem 2

Hanne Rønning Berg

2024-09-27

Problem 1

f)

```
data = read.table("https://www.uio.no/studier/emner/matnat/math/STK3100/data/wingspan.txt", h
eader = T)

fit.m0 = lm(Height ~ Wingspan -1, data = data)
fit.m1 = lm(Height ~ Wingspan, data = data)

anova(fit.m0, fit.m1)
```

```
## Analysis of Variance Table
##
## Model 1: Height ~ Wingspan - 1
## Model 2: Height ~ Wingspan
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     15 42.243
## 2     14 22.580  1    19.663 12.191 0.003595 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Very low p-value, and the output from R has stars (**) that indicates its significant level. This indicates that M1 is a better model.

Problem 2

a)

```
titanic = read.table("https://www.uio.no/studier/emner/matnat/math/STK3100/data/titanic.txt",
header = T)
fit.1 = glm(survived ~ pclass, family=binomial(link='logit'),data=titanic)
summary(fit.1)
```

```
##
## Call:
## glm(formula = survived ~ pclass, family = binomial(link = "logit"),
##     data = titanic)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.06899    0.37161   0.186   0.8527
## pclass      -1.33750    0.52958  -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
```

Summary() does help us with stars indicating if a covariate is significant. Here it is a low p-score (P>|z|) = 0.0116, means that it is significant.

b)

```
B_1 = coefficients(fit.1)[2]
B_1
```

```
##    pclass
## -1.337504
```

```
exp(B_1)
```

```
## pclass
## 0.2625
```

In normal linear regression the beta_j-coefficients is the increase when we have one more unit of the j_covariate. Here, in the logistic model, e^B_1 is the relative effect on the odds of one unit's increase in the j-th covariate when the other covariates remain the same (from slides section 4).

e^B_1 = 0.2625, means that people with 3rd class tickets have a higher chance of surviving.

c)

```
fit.2 = glm(survived ~ pclass + age, family=binomial(link='logit'),data=titanic)
exp(coefficients(fit.2)[3])
```

```
##       age
## 0.9606129
```

e^B_i > 1 means that there is a decrease in probability of the person surviving.

e^B_i < 1 means an increase.

An odds ratio of 1 indicates no effect

Since e^B_2 is 0.9606129, which is close to 1 there is not much of an effect, but the odds ratio still changes a bit.

d)

Wald test, likelihood ratio test and score test

Wald test

```
summary(fit.2)
```

```
##
## Call:
## glm(formula = survived ~ pclass + age, family = binomial(link = "logit"),
##     data = titanic)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.55623    0.87390   1.781  0.07495 .
## pclass      -1.73105    0.59825  -2.894  0.00381 **
## age         -0.04018    0.02125  -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
```

Summary(fit.2) gives us the Wald test results with the p-values to the right of the table.

The p-value 0.05859 is the result of the Wald test.

Likelihood ratio test

```
anova(fit.1,fit.2,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: survived ~ pclass
## Model 2: survived ~ pclass + age
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        68     83.324
## 2        67     79.397  1   3.9265  0.04753 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is 0.04753 (so below 0.05)

Score test

```
anova(fit.1,fit.2,test="Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: survived ~ pclass
## Model 2: survived ~ pclass + age
##   Resid. Df Resid. Dev Df Deviance    Rao Pr(>Chi)
## 1        68     83.324
## 2        67     79.397  1   3.9265 3.7332  0.05334 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value from the score-test is 0.05334. >0.05.

If the p-values are small (typically under 0.05 for a 95% significant level), it indicates that the coefficient is different than 0.

In the likelihood-ratio test, we would keep the B_2 coefficient, but in the other tests we would set B_2 = 0.

e)

95% confident interval for e^B_1 from b)

```
ci.wald=confint.default(fit.1)
ci.wald
```

```
##                  2.5 %     97.5 %
## (Intercept) -0.6593526  0.7973384
## pclass      -2.3754642 -0.2995442
```

```
exp(-2.3754642)
```

```
## [1] 0.09297132
```

```
exp(-0.2995442)
```

```
## [1] 0.741156
```

95% confident interval for e^B_2 from c)

```
ci.wald=confint.default(fit.2)
ci.wald
```

```
##                   2.5 %       97.5 %
## (Intercept) -0.15657729  3.269035764
## pclass      -2.90359941 -0.558496592
## age         -0.08182736  0.001459859
```

```
exp(-0.08182736)
```

```
## [1] 0.921431
```

```
exp( 0.001459859)
```

```
## [1] 1.001461
```