UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: STK3100/STK4100 — Introduction to generalized linear models. SOLUTIONS TO PROBLEMS

Day of examination: Friday 14 December 2018.

Examination hours: 09.00 – 13.00.

This problem set consists of 9 pages.

Appendices: Formulas in STK3100/4100.

Permitted aids: Approved calculator and collection of formulas for STK1100/STK1110 and STK2120.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

We have $V \sim \text{bin}(n, \pi)$ and $Y = V/n$. The pmf of Y is given by

$$P(Y = y) = \binom{n}{ny} \pi^{ny} (1-\pi)^{n-ny} \tag{1}$$

for $y = 0, \frac{1}{n}, \frac{2}{n}, \ldots \frac{n-1}{n}, 1$.

a) We may rewrite (1) as

$$\begin{aligned} P(Y = y) &= \exp\left\{ ny\log(\pi) + (n - ny)\log(1-\pi) + \log\binom{n}{ny} \right\} \\ &= \exp\left\{ \frac{y\log\left(\frac{\pi}{1-\pi}\right) - [-\log(1-\pi)]}{1/n} + \log\binom{n}{ny} \right\}. \end{aligned}$$

This is of the form (2) in the exam problems with $\theta = \log\left(\frac{\pi}{1-\pi}\right)$. Hence $\pi = e^\theta/(1+e^\theta)$, and we have that $b(\theta) = -\log(1-\pi) = \log(1+e^\theta)$, $a(\phi) = 1/n$ and $c(y,\phi) = \log\binom{n}{ny}$.

b) We have that

$$\mu = E(Y) = b'(\theta) = \frac{d}{d\theta}\log(1+e^\theta) = \frac{e^\theta}{1+e^\theta} = \pi,$$

and

$$\begin{aligned} \text{Var}(Y) &= b''(\theta) \cdot a(\phi) = \frac{d}{d\theta}\left(\frac{e^\theta}{1+e^\theta}\right) \cdot \frac{1}{n} = \frac{e^\theta}{(1+e^\theta)^2} \cdot \frac{1}{n} \\ &= \frac{1}{n} \cdot \frac{e^\theta}{1+e^\theta} \cdot \frac{1}{1+e^\theta} = \frac{1}{n}\pi(1-\pi). \end{aligned}$$

(Continued on page 2.)

We then assume that $V_1, V_2, \ldots, V_N$ are independent with $V_i \sim \text{bin}(n_i, \pi_i)$, and let $Y_i = V_i/n_i$ for $i = 1, 2, \ldots, N$. We consider a generalized linear model (GLM) for $Y_1, Y_2, \ldots, Y_N$ with canonical link function $\text{logit}(\pi_i) = \log\{\pi_i/(1-\pi_i)\} = \eta_i$ and linear predictor $\eta_i = \beta_0 + \beta_1 x_i$. Here $x_1, \ldots, x_N$ are known covariate values.

c) Let $y_1, y_2, \ldots, y_N$ be the observed values of $Y_1, Y_2, \ldots, Y_N$. Then the likelihood is given by

$$\ell(\beta_0, \beta_1) = \prod_{i=1}^{N} \binom{n_i}{n_i y_i} \pi_i^{n_i y_i} (1-\pi_i)^{n_i - n_i y_i} = C \prod_{i=1}^{N} \left(\frac{\pi_i}{1-\pi_i}\right)^{n_i y_i} (1-\pi_i)^{n_i},$$

where $C = \prod_{i=1}^{N} \binom{n_i}{n_i y_i}$. Now $\pi_i = e^{\eta_i}/(1+e^{\eta_i})$, so the log-likelihood becomes

$$\begin{aligned} L(\beta_0, \beta_1) &= \log C + \sum_{i=1}^{N} \left\{ n_i y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i) \right\} \\ &= \log C + \sum_{i=1}^{N} \left\{ n_i y_i \eta_i + n_i \log\left(\frac{1}{1+e^{\eta_i}}\right) \right\} \\ &= \log C + \sum_{i=1}^{N} \left\{ n_i y_i (\beta_0 + \beta_1 x_i) - n_i \log(1 + e^{\beta_0 + \beta_1 x_i}) \right\} \end{aligned}$$

We differentiate the log-likelihood function, and find

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{N} \left(n_i y_i - n_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = \sum_{i=1}^{N} n_i (y_i - \pi_i),$$

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{N} \left(n_i y_i x_i - n_i \frac{x_i e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = \sum_{i=1}^{N} n_i x_i (y_i - \pi_i).$$

We obtain the maximum likelihood estimates by solving the equation we obtain by setting the partial derivatives equal to zero. Expressed in terms of the random Y_i's we therefore have that the maximum likelihood estimators are the solutions of the equations

$$\sum_{i=1}^{N} n_i (Y_i - \pi_i) = 0 \qquad \text{and} \qquad \sum_{i=1}^{N} n_i x_i (Y_i - \pi_i) = 0.$$

d) We first note that

$$\begin{aligned} \frac{\partial \pi_i}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \\ &= \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \cdot \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} = \pi_i (1 - \pi_i), \end{aligned}$$

(Continued on page 3.)

and

$$\begin{aligned}\frac{\partial \pi_i}{\partial \beta_1} &= \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right) = \frac{x_i e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2} \\ &= x_i \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \cdot \frac{1}{1+e^{\beta_0+\beta_1 x_i}} = x_i \pi_i (1-\pi_i).\end{aligned}$$

By differentiating the log-likelihood function one more time, we then find that

$$\frac{\partial^2 L(\beta_0,\beta_1)}{\partial \beta_0^2} = \frac{\partial}{\partial \beta_0}\sum_{i=1}^N n_i(y_i-\pi_i) = -\sum_{i=1}^N n_i \frac{\partial \pi_i}{\partial \beta_0} = -\sum_{i=1}^N n_i \pi_i(1-\pi_i),$$

$$\frac{\partial^2 L(\beta_0,\beta_1)}{\partial \beta_1^2} = \frac{\partial}{\partial \beta_1}\sum_{i=1}^N n_i x_i(y_i-\pi_i) = -\sum_{i=1}^N n_i x_i \frac{\partial \pi_i}{\partial \beta_1} = -\sum_{i=1}^N n_i x_i^2 \pi_i(1-\pi_i),$$

and

$$\frac{\partial^2 L(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} = \frac{\partial^2 L(\beta_0,\beta_1)}{\partial \beta_1 \partial \beta_0} = \frac{\partial}{\partial \beta_1}\sum_{i=1}^N n_i(y_i-\pi_i) = -\sum_{i=1}^N n_i \frac{\partial \pi_i}{\partial \beta_1} = -\sum_{i=1}^N n_i x_i \pi_i(1-\pi_i).$$

The second order patrial derivatives do not depend on the y_i's. Therefore the observed and the expected information matrices coincided, and we have that

$$\mathcal{J} = \left\{-E\left(\frac{\partial^2 L(\beta_0,\beta_1)}{\partial \beta_h \partial \beta_j}\right)\right\}_{h,j=0,1} = \begin{pmatrix} \sum_{i=1}^N n_i \pi_i(1-\pi_i) & \sum_{i=1}^N n_i x_i \pi_i(1-\pi_i) \\ \sum_{i=1}^N n_i x_i \pi_i(1-\pi_i) & \sum_{i=1}^N n_i x_i^2 \pi_i(1-\pi_i) \end{pmatrix}.$$

Problem 2

a) The analysis reported in question a is based on a logistic regression model. To describe the model, we let $Y_i = 1$ if passenger i survived the disaster, $Y_i = 0$ if passenger i died, and we let $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, x_{i3}, x_{i4})$ denote its covariates (including $x_{i0} = 1$ for the intercept) defined as follows:

$x_{i1} = 1$ if passenger i is female; $x_{i1} = 0$ if passenger i is male,

$x_{i2} = a_i - 30$, where a_i is the age (in years) of passenger i,

$x_{i3} = 1$ if passenger i travelled on second class; $x_{i3} = 0$ otherwise,

$x_{i4} = 1$ if passenger i travelled on third class; $x_{i4} = 0$ otherwise.

The model assumes that the Y_i's are independent and that $\pi_i = P(Y_i = 1)$ is given as

$$\pi_i = \frac{\exp\left(\sum_{j=0}^4 \beta_j x_{ij}\right)}{1+\exp\left(\sum_{j=0}^4 \beta_j x_{ij}\right)}.$$

(Continued on page 4.)

To give an interpretation of the estimated intercept, we consider a 30 years old male passenger who travelled on first class. Such a passenger has $x_{ij} = 0$ for $j = 1, 2, 3, 4$, so (according to the given model) his probability of surviving the disaster is $\pi_i = e^{\beta_0}/(1 + e^{\beta_0})$, and this is estimated by

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0}}{1 + e^{\widehat{\beta}_0}} = \frac{e^{-0.007567}}{1 + e^{-0.007567}} = 0.498.$$

This gives an interpretation of the estimated intercept.

To interpret the estimate for (centered) age, we consider two passengers, k and i, of the same sex who travelled on the same class, but where passenger k is one year older than passenger i. Then the odds ratio for the two passengers is

$$\mathrm{OR}(k,i) = \frac{\pi_k/(1-\pi_k)}{\pi_i/(1-\pi_i)} = \frac{\exp\left(\sum_{j=0}^{4} \beta_j x_{kj}\right)}{\exp\left(\sum_{j=0}^{4} \beta_j x_{ij}\right)} = e^{\beta_2},$$

and this is estimated by

$$\widehat{\mathrm{OR}}(k,i) = e^{\widehat{\beta}_2} = e^{-0.034393} = 0.966.$$

Thus the odds of surviving is reduced by 3.4% when the age is increased by one year.

b) The model in question a has (residual) deviance 982.45 with 1041 degrees of freedom, while the (residual) deviance of the model in question b is 931.99 with 1039 degrees of freedom. The difference in (residual) deviance between the two models is $982.45 - 931.99 = 50.46$, while the difference in the degrees of freedom is $1041 - 1039 = 2$. The difference between the deviances equals minus two times the likelihood ratio of the model in a to the model in b. So it is approximately chi-squared distributed with 2 degrees of freedom if the model in a holds. Comparing the difference 50.46 between the deviances with a chi-squared distribution with 2 degrees of freedom, we see that the model in b gives an improved fit compared to the model in a.

For the model in question b there is interaction between sex and passenger class. So in addition to the covariates in question a, we for this model also have the covariates:

$x_{i5} = 1$ if passenger i is female who travelled on second class; $x_{i5} = 0$ otherwise,

$x_{i6} = 1$ if passenger i is female who travelled on third class; $x_{i6} = 0$ otherwise.

The expression for $\pi_i = P(Y_i = 1)$ now becomes

$$\pi_i = \frac{\exp\left(\sum_{j=0}^{6} \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=0}^{6} \beta_j x_{ij}\right)}.$$

In order to describe the effects of sex and passenger class, we consider a passenger that is 30 years old (which makes $x_{i2} = 0$).

(Continued on page 5.)

For a 30 years old male the probability of surviving the disaster is estimated to be:

– For a male at first class:

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0}}{1+e^{\widehat{\beta}_0}} = \frac{e^{-0.234083}}{1+e^{-0.234083}} = 0.442.$$

– For a male at second class:

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0+\widehat{\beta}_3}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_3}} = \frac{e^{-0.234083-1.600280}}{1+e^{-0.234083-1.600280}} = 0.138.$$

– For a male at third class:

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0+\widehat{\beta}_4}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_4}} = \frac{e^{-0.234083-1.576159}}{1+e^{-0.234083-1.576159}} = 0.141$$

For a 30 years old female the probability of surviving the disaster is estimated to be:

– For a female at first class:

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0+\widehat{\beta}_1}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1}} = \frac{e^{-0.234083+3.886388}}{1+e^{-0.234083+3.886388}} = 0.975.$$

– For a female at second class:

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0+\widehat{\beta}_1+\widehat{\beta}_3+\widehat{\beta}_5}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1+\widehat{\beta}_3+\widehat{\beta}_5}} = \frac{e^{-0.234083+3.886388-1.600280+0.070407}}{1+e^{-0.234083+3.886388-1.600280+0.070407}} = 0.893.$$

– For a female at third class:

$$\widehat{\pi}_i = \frac{e^{\widehat{\beta}_0+\widehat{\beta}_1+\widehat{\beta}_4+\widehat{\beta}_6}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1+\widehat{\beta}_4+\widehat{\beta}_6}} = \frac{e^{-0.234083+3.886388-1.576159-2.488805}}{1+e^{-0.234083+3.886388-1.576159-2.488805}} = 0.398.$$

We see that for all three classes, the probability that a female will survive is higher than for a male. For males the probability for surviving is highest for a first class passenger (44.2%) and much lower for a second or third class passenger (13.8% and 14.1 %, respectively). For females the probability of surviving is high both for first and second class (97.5% and 89.3%, respectively) and much lower for a third class passenger (39.8%).

c) The analysis of deviance table with numbers filled in for the question marks is given below. The numbers that have been filled in are underlined.

```
Model 1: Survived~Sex+Cage+Class+Sex:Class
Model 2: Survived~Sex+Cage+Class+Sex:Class+Cage:Class
Model 3: Survived~Sex+Cage+Class+Sex:Class+Cage:Class+Sex:Cage
Model 4: Survived~Sex+Cage+Class+Sex:Class+Cage:Class+Sex:Cage+Sex:Cage:Class
     Resid. Df    Resid. Dev    Df    Deviance    Pr(>Chi)
1         1039        931.99
2         1037        922.17     2        9.82     0.00739
3         1036        917.84     1      4.3308     0.03743
4         1034        915.97     2      1.8661     0.39335
```

(Continued on page 6.)

In order to describe how we have arrived at the underlined numbers, we first describe the content of columns two to five in the analysis of deviance table:

- `Resid.Df` is the residual degrees of freedom and equals $N - p$, where N is the number of observations (here $N = 1046$) and p is the number of parameters in the model.
- `Resid.Dev` is the deviance for the fitted model. (Note that what is denoted 'deviance' in the text book, is denoted 'residual deviance' by `R`.)
- `Df` is the difference in residual degrees of freedom for the model on the line above and the actual model. Note that this is the same as the difference in the number of parameters of the actual model and the model on the line above.
- `Deviance` is the difference in deviance for the model on the line above and the actual model.

Using this, we find the underlined numbers as follows:

- <u>`Resid.Df` for model 3:</u> Model 3 has 1 parameter for the intercept, 1 parameter for the main effect of `Sex`, 1 parameter for main effect of `Cage`, 2 parameters for the main effect of `Class`, 2 parameters for the interaction `Sex:Class`, 2 parameters for the interaction `Cage:Class`, and 1 parameter for the interaction `Sex:Cage`. In total this gives 10 parameters, so `Resid.Df` equals $1046 - 10 = 1036$. [Alternatively (and easier) we may use that `Resid.Df` for model 2 is 1037 and the difference in residual degrees of freedom between models 2 and 3 is 1, so `Resid.Df` for model 3 is $1037 - 1 = 1036$.]
- <u>`Resid.Dev` for model 4:</u> From the output, the (residual) deviance of model 3 is 917.84 and the difference in (residual) deviance between models 3 and 4 is 1.8661. Hence the (residual) deviance for model 4 is $917.84 - 1.87 = 915.97$.
- <u>`Df` for model 2:</u> The difference in residual degrees of freedom for models 1 and 2 is $1039 - 1037 = 2$.
- <u>`Deviance` for model 2:</u> The difference in (residual) deviance between models 1 and 2 is $931.99 - 922.17 = 9.82$.

To compare the models, we may look at the P-values in the last column of the analysis of deviance table. (The P-value for a line in the table, is the P-value for a likelihood ratio test that tests the null hypothesis that the model on the previous line of the table holds, assuming that the model on the given line holds.) From the P-values, we see that model 2 gives a significantly better fit than model 1, and that model 3 gives a significantly better fit than model 2. But model 4 does not significantly improve the fit compared to model 3. Therefore we prefer model 3.

(Continued on page 7.)

Problem 3

a) The projection matrix is given by $\boldsymbol{P}_1 = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$. To evaluate this, we first note that

$$\begin{aligned}
\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} &= \begin{bmatrix} \mathbf{1}_n^{\mathrm{T}} \\ (\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}} \end{bmatrix} [\mathbf{1}_n \,,\, \boldsymbol{x} - \bar{x}\mathbf{1}_n] \\
&= \begin{bmatrix} \mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n & \mathbf{1}_n^{\mathrm{T}}(\boldsymbol{x} - \bar{x}\mathbf{1}_n) \\ (\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}\mathbf{1}_n & (\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}(\boldsymbol{x} - \bar{x}\mathbf{1}_n) \end{bmatrix} \\
&= \begin{bmatrix} n & \sum_{i=1}^n (x_i - \bar{x}) \\ \sum_{i=1}^n (x_i - \bar{x}) & \sum_{i=1}^n (x_i - \bar{x})^2 \end{bmatrix} \\
&= \begin{bmatrix} n & 0 \\ 0 & M \end{bmatrix},
\end{aligned}$$

and therefore

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} = \begin{bmatrix} n^{-1} & 0 \\ 0 & M^{-1} \end{bmatrix}.$$

We then find that the projection matrix may be given as

$$\begin{aligned}
\boldsymbol{P}_1 &= \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}} \\
&= [\mathbf{1}_n \,,\, \boldsymbol{x} - \bar{x}\mathbf{1}_n] \begin{bmatrix} n^{-1} & 0 \\ 0 & M^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}_n^{\mathrm{T}} \\ (\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}} \end{bmatrix} \\
&= \left[n^{-1}\mathbf{1}_n \,,\, M^{-1}(\boldsymbol{x} - \bar{x}\mathbf{1}_n)\right] \begin{bmatrix} \mathbf{1}_n^{\mathrm{T}} \\ (\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}} \end{bmatrix} \\
&= n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} + M^{-1}(\boldsymbol{x} - \bar{x}\mathbf{1}_n)(\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}
\end{aligned}$$

b) From the result in a, we find the vector of fitted values may be given as

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \boldsymbol{P}_1\boldsymbol{Y} \\
&= \left[n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} + M^{-1}(\boldsymbol{x} - \bar{x}\mathbf{1}_n)(\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}\right]\boldsymbol{Y} \\
&= n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\boldsymbol{Y} + M^{-1}(\boldsymbol{x} - \bar{x}\mathbf{1}_n)(\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}\boldsymbol{Y} \\
&= \mathbf{1}_n n^{-1}\sum_{i=1}^n Y_i + (\boldsymbol{x} - \bar{x}\mathbf{1}_n)M^{-1}\sum_{i=1}^n Y_i(x_i - \bar{x}) \\
&= \overline{Y}\mathbf{1}_n + \widehat{\beta}_1(\boldsymbol{x} - \bar{x}\mathbf{1}_n).
\end{aligned}$$

(Continued on page 8.)

The projection matrix for the null model $\mu_i = E(Y_i) = \beta_0$ is $\boldsymbol{P}_0 = n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$. We consider the orthogonal decomposition

$$\boldsymbol{Y} = \boldsymbol{P}_0\boldsymbol{Y} + (\boldsymbol{P}_1 - \boldsymbol{P}_0)\boldsymbol{Y} + (\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}$$

with corresponding sum of squares decomposition

$$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_0\boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_1 - \boldsymbol{P}_0)\boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}.$$

c) Using the result in question a, we have that

$$\begin{aligned}
\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_1 - \boldsymbol{P}_0)\boldsymbol{Y} &= \boldsymbol{Y}^{\mathrm{T}}[M^{-1}(\boldsymbol{x} - \bar{x}\mathbf{1}_n)(\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}]\boldsymbol{Y} \\
&= M^{-1}[\mathbf{Y}^{\mathrm{T}}(\boldsymbol{x} - \bar{x}\mathbf{1}_n)]\,[(\boldsymbol{x} - \bar{x}\mathbf{1}_n)^{\mathrm{T}}\mathbf{Y}] \\
&= M^{-1}\left[\sum_{i=1}^{n} Y_i(x_i - \bar{x})\right]^2 \\
&= M\widehat{\beta}_1^2,
\end{aligned}$$

which shows the first result. For the second result, we note that since $\boldsymbol{I} - \boldsymbol{P}_1$ is a projection matrix, it is idempotent and symmetric. Hence we have that

$$\boldsymbol{I} - \boldsymbol{P}_1 = (\boldsymbol{I} - \boldsymbol{P}_1)(\boldsymbol{I} - \boldsymbol{P}_1) = (\boldsymbol{I} - \boldsymbol{P}_1)^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1).$$

Then by the result in question b, we have that

$$\begin{aligned}
\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y} &= \boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y} \\
&= [(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}]^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y} \\
&= (\boldsymbol{Y} - \boldsymbol{P}_1\boldsymbol{Y})^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{P}_1\boldsymbol{Y}) \\
&= [\boldsymbol{Y} - \overline{Y}\mathbf{1}_n - \widehat{\beta}_1(\boldsymbol{x} - \bar{x}\mathbf{1}_n)]^{\mathrm{T}}[\boldsymbol{Y} - \overline{Y}\mathbf{1}_n - \widehat{\beta}_1(\boldsymbol{x} - \bar{x}\mathbf{1}_n)] \\
&= \sum_{i=1}^{n}\Big[Y_i - \overline{Y} - \widehat{\beta}_1(x_i - \overline{x})\Big]^2,
\end{aligned}$$

which shows the second result.

d) By Cochran's theorem we have that the terms on the right-hand side of (4) in the exam problems are independent, and when divided by σ_2 they are (non-central) chi-squared distributed. In particular, it follows that

$$M\widehat{\beta}_1^2/\sigma^2 = \boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_1 - \boldsymbol{P}_0)\boldsymbol{Y}/\sigma^2$$

and

$$\sum_{i=1}^{n}\Big[Y_i - \overline{Y} - \widehat{\beta}_1(x_i - \overline{x})\Big]^2/\sigma^2 = \boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}/\sigma^2$$

(Continued on page 9.)

are independent and (non-cental) chi-squared distributed.

Further, for (5) in the exam problems the degrees of freedom is given by (using that the rank of a projection matrix equals is trace)

$$\begin{aligned}\mathrm{df}_1 &= \mathrm{rank}(\boldsymbol{P}_1 - \boldsymbol{P}_0) = \mathrm{trace}(\boldsymbol{P}_1 - \boldsymbol{P}_0)\\ &= \mathrm{trace}(\boldsymbol{P}_1) - \mathrm{trace}(\boldsymbol{P}_0) = \mathrm{rank}(\boldsymbol{P}_1) - \mathrm{rank}(\boldsymbol{P}_0)\\ &= 2 - 1 = 1,\end{aligned}$$

while the degrees of freedom for (6) equals

$$\begin{aligned}\mathrm{df}_2 &= \mathrm{rank}(\boldsymbol{I} - \boldsymbol{P}_1) = \mathrm{trace}(\boldsymbol{I} - \boldsymbol{P}_1)\\ &= \mathrm{trace}(\boldsymbol{I}) - \mathrm{trace}(\boldsymbol{P}_1) = \mathrm{trace}(\boldsymbol{I}) - \mathrm{rank}(\boldsymbol{P}_1)\\ &= n - 2.\end{aligned}$$

The mean vector $\boldsymbol{\mu} = \boldsymbol{X\beta}$ is in the model space $C(\boldsymbol{X})$, so $\boldsymbol{P}_1\boldsymbol{\mu} = \boldsymbol{\mu}$. Therefore, by Cochran's theorem, the non-centrality parameter of (6) is

$$\lambda_2 = \frac{\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{\mu}}{\sigma^2} = \frac{\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\mu} - \boldsymbol{P}_1\boldsymbol{\mu})}{\sigma^2} = \frac{\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\mu} - \boldsymbol{\mu})}{\sigma^2} = 0$$

e) For testing the the null hypothesis $H_0 : \beta_1 = 0$ versus the alternative hypothesis $H_A : \beta_1 \neq 0$, we may use the test statistic

$$F = \frac{\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_1 - \boldsymbol{P}_0)\boldsymbol{Y}/1}{\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}/(n-2)} = \frac{(n-2)M\widehat{\beta}_1^2}{\sum_{i=1}^n \left[Y_i - \overline{Y} - \widehat{\beta}_1(x_i - \overline{x})\right]^2}.$$

By the results of question d (including the one given in parenthesis), this has a non-central F-distribution with 1 and $n-2$ degrees of freedom and non-centrality parameter $\lambda_1 = M\beta_1^2/\sigma^2$. In particular, under H_0, we have $\lambda_1 = 0$ and then F has a central F-distribution with 1 and $n-2$ degrees of freedom.