Exercise 6

3^{rd} - 5^{th} December 2024

Luca Danese - l.danese1@campus.unimib.it

Exercise 1

To meet competition or cope with economic slowdowns, corporations sometimes undertake a "reduction in force" (RIF), in which substantial numbers of employees are terminated. Federal and various state laws require that employees be treated equally regardless of their age. In particular, employees over the age of 40 years are in a "protected" class, and many allegations of discrimination focus on comparing employees over 40 with their younger coworkers. Here are the data for a recent RIF:

Terminated	**Over40: No**	**Over40: Yes**
Yes	17	71
No	564	835

Exercise 1.1

(a) Choose an appropriate response variable and then a regression model. Justify your answer.

(b) Further, find the estimated regression model.

Exercise 1.2

Software gives the estimated slope $\hat{\beta}_2 = 1.0371$ and its standard error $SE(\hat{\beta}_2) = 0.2755$. Transform the results to the odds scale. Summarize the results and write a short conclusion.

1.1)

a)

We make use of Terminated as response variable.

$$\eta_i = g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_1 + \beta_2 D_{1i} \text{ , where } D_{1i} = \begin{cases} 1, & \text{if over40} \\ 0, & \text{otherwise} \end{cases}$$

where $g: [0,1] \longrightarrow \mathbb{R}$

π_i is the mean of $Y_i \sim \text{Bernoulli}(\pi_i) \Rightarrow \mathbb{E}[\pi_i] = \pi_i = P(Y_i = 1)$

$$\frac{\pi_i}{1-\pi_i} \rightarrow \text{odds} = \frac{\text{prob. of success.}}{\text{prob. of failure}}$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) \rightarrow \text{logarithm of the odds is the logit function}$$

b)

let's invert the relationship between η_i and π_i:

$$\pi_i = g^{-1}(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}} \in (0,1)$$

$$\Rightarrow \pi_i = \frac{e^{\beta_1+\beta_2 D_{1i}}}{1+e^{\beta_1+\beta_2 D_{1i}}} = P(Y_i = 1)$$

Since D_{1i} is a dummy variable we have the following possibilities

- $(\pi_i \mid D_{1i} = 1) = P(Y_i = 1 \mid D_{1i} = 1) = \dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}}$

- $(1-\pi_i \mid D_{1i} = 1) = P(Y_i = 0 \mid D_{1i} = 1) = 1 - \dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} = \dfrac{1}{1+e^{\beta_1+\beta_2}}$

$$\Rightarrow \left(\frac{\pi_i}{1-\pi_i} \mid D_{1i} = 1\right) = e^{\beta_1+\beta_2} \quad \text{ODDS for } D_{1i} = 1 \text{ (Over 40)}$$

- $(\pi_i \mid D_{1i} = 0) = P(Y_i = 1 \mid D_{1i} = 0) = \dfrac{e^{\beta_1}}{1+e^{\beta_1}}$

- $(1-\pi_i \mid D_{1i} = 0) = P(Y_i = 0 \mid D_{1i} = 0) = \dfrac{1}{1+e^{\beta_1}}$

$$\Rightarrow \left(\frac{\pi_i}{1-\pi_i} \mid D_{1i} = 0\right) = e^{\beta_1} \quad \text{ODDS for } D_{1i} = 0 \text{ (Under 40)}$$

We can easily find an estimate for β_1 and β_2

$$e^{\beta_1} = \frac{P(Y_i = 1 \mid D_{1i} = 0)}{P(Y_i = 0 \mid D_{1i} = 0)} = \frac{17}{564} \Rightarrow \hat{\beta}_1 = -3.501841$$

$$e^{\beta_1 + \beta_2} = \frac{P(Y_i = 1 \mid D_{1i} = 1)}{P(Y_i = 0 \mid D_{1i} = \cancel{0}\ 1)} = \frac{71}{835} \Rightarrow \hat{\beta}_2 = 1.037089$$

Then we have the following model:

$$\log(\text{odds}) = -3.501841 + 1.037089\, D_{1i}$$

1.2)

$\hat{\beta}_2 = 1.0371 \qquad S.E.(\hat{\beta}_2) = 0.2755$

The ODDS RATIO according 1.1) is equal to:

$$\frac{\left(\frac{\pi_i}{1-\pi_i} \mid D_{1i} = 1\right)}{\left(\frac{\pi_i}{1-\pi_i} \mid D_{1i} = 0\right)} = e^{\beta_1+\beta_2} / e^{\beta_1} = e^{\beta_2} = 2.82$$

-> The odds of Under 40 are multiplied by 2.82 to get the odds for Over 40

-> Employees that are over 40 are 2.82 times more likely to be terminated than those that are under 40

EXERCISE 2

The acquisition literature suggests that takeovers occur either due to conflicts between managers and shareholders or to create a new entity that exceeds the sum of its previously separate components. Other research has offered managerial hubris as a third option, but it has not been studied empirically. Recently, some researchers revisited acquisitions over a 10-year period in the Australian financial system. A measure of CEO overconfidence was based on the CEO's level of media exposure, and a measure of dominance was based on the CEO's remuneration relative to the firm's total assets. They then used logistic regression to see whether CEO overconfidence and dominance were positively related to the probability of at least one acquisition in a year. To help isolate the effects of CEO hubris, the model included explanatory variables of firm characteristics and other potentially important factors in the decision to acquire. The following table summarizes the results for the two key explanatory variables:

Covariates	Estimates	SE
Overconfidence	0.0878	0.0402
Dominance	1.5067	0.0057

Exercise 2.1

Write the estimated regression model and interpret the coefficients estimates.

Exercise 2.2

Perform the significance tests and determine whether the variables are significant at the 0.05 level.

Exercise 2.3

Estimate the odds ratio for each variable.

2.1)

We consider the following response variable

$$Y_i = \begin{cases} 1, \text{ if firm } i \text{ made at least one acquisition} \\ 0, \text{ otherwise} \end{cases}$$

The estimated regression, using a logit model, is:

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = \hat{\beta}_1 + 0.0878\,\text{Overconfidence}_i + 1.5067\,\text{Dominance}_i$$

- the log odds increases by 0.0878 if Overconfidence increases by 1 unit, while keeping constant the other covariates

- the log odds increases by 1.5067 if Dominance increases by 1 unit, while keeping constant the other covariates

2.2)

- Significance of $\hat{\beta}_2$

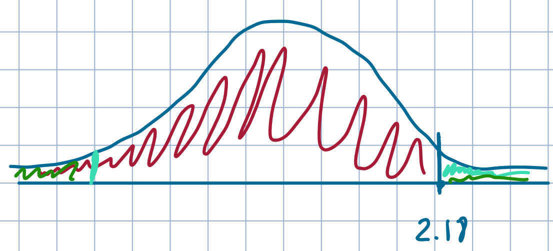
$$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases} \qquad Z_2 = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\left[j(\hat{\underline{\beta}})^{-1}\right]_{jj}}} \overset{H_0}{\sim} N(0,1)$$

↳ j^{th} element of the diagonal of the matrix $X^T V(\hat{\underline{\beta}}) X$

$$Z_2^{obs} = \frac{0.0878}{0.0402} = 2.18408$$

$$\alpha^{obs} = P_{H_0}(|Z_2| \geq |Z_2^{obs}|) = 2(1 - \Phi(|2.18408|)) \overset{or}{=} 2\Phi(-2.18408)$$

$$P(|Z| \geq |z|) = P(Z \leq -z) + P(Z \geq z) =$$



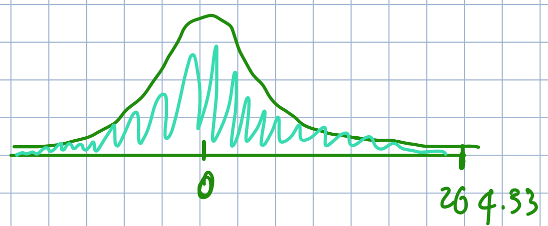
$$Z_{0.9855} = 2.18408$$

$\Rightarrow 2 \cdot (1 - 0.9855) = 0.02896 = \alpha^{obs}$ $\Rightarrow$ we reject H_0 at a 5% confidence level

- Significance of $\hat{\beta}_3$

$$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases} \qquad Z_3^{obs} = \frac{1.5067}{0.0057} = 264.333$$

$$\alpha^{obs} = 2\left(1 - \Phi\left(264.333\right)\right) \approx 0$$ => we reject H_0 at a 5% confidence level



2.3)

- ODDS RATIO OVERCONFIDENCE

Let's consider

$$\log\left(\frac{\pi_0}{1-\pi_0}\right) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$$

$$\log\left(\frac{\pi_1}{1-\pi_1}\right) = \beta_1 + \beta_2\left(X_2 + 1\right) + \beta_3 X_3 = \beta_1 + \beta_2 + \beta_2 X_2 + \beta_3 X_3$$

The odds ratio is the following:

$$\frac{\left(\frac{\pi_1}{1-\pi_1}\right)}{\left(\frac{\pi_0}{1-\pi_0}\right)} = e^{\beta_1 + \beta_2 + \beta_2 X_2 + \beta_3 X_3 - \beta_1 - \beta_2 X_2 - \beta_3 X_3} = e^{\beta_2} = 1.097177$$

=> An increase of one unit in the overconfidence correspond to an increase of 1.1 in the odds

- ODDS RATIO DOMINANCE

Here we have:

$$\log\left(\frac{\pi_0}{1-\pi_0}\right) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$$

$$\log\left(\frac{\pi_1}{1-\pi_1}\right) = \beta_1 + \beta_2 X_2 + \beta_3 (X_3+1) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_3$$

Then the odds ratio is:

$$\frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}} = e^{\beta_3} = 4.511817$$

=> If we increase the dominence by 1 unit then the odds increases by 4.511817

EXERCISE 3

Let consider a dataset on the number of research articles published by 915 graduate students in biochemistry PhD programs. The variables for this dataframe are

- **art**: count of articles produced during last 3 years of PhD
- **fem**: factor indicating gender of student, with levels Men and Women
- **mar**: factor indicating marital status of student, with levels Single and Married
- **kid5**: number of children aged 5 or younger
- **phd**: prestige of PhD department
- **ment**: count of articles produced by PhD mentor during last 3 years

Exercise 3.1
Choose an appropriate response variable and then a regression model. Justify your answer.

Exercise 3.2
Complete the following table.

Coefficients	Estimates	SE	Z-obs	P-value
β_1	0.304617	0.102981	2.958	0.0031
β_2	-0.224594	0.054613		
β_3	0.155243			0.0114
β_4		0.040127	-4.607	
β_5		0.026397		0.6271
β_6	0.025543	0.002006	12.733	< 2e-16

Then, interpret the coefficient β_6.

Exercise 3.3
Knowing that the null deviance is equal to 1817.4 while the residual deviance is equal to 1634.4. Perform a statistical test about the overall significance. Specify the hypothesis, the test statistic and the p-value.

Exercise 3.4
Knowing the value of the following quantity

$$\sum_{i=1}^{n} y_i \log(\hat{\mu}_i) - \hat{\mu}_i = -642.0261$$

Find the log-likelihood of the saturated model. Further, interpret the results in terms of goodness of fit.

3.1)

Considering the nature of variable art we assume:

- $Y_i \sim \text{Poisson}(\mu_i)$ where $Y_i = \text{art}_i$
- $\eta_i = x_i^T \underline{\beta}$
- $\log(\mu_i) = \eta_i$ logarithmic link function

$$\Rightarrow \mu_i = \exp\{ \underline{x}_i^T \underline{\beta} \}$$

$$\log(\mu_i) = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 kids_i + \beta_5 phd_i + \beta_6 ment_i$$

where
$$D_{1i} = \begin{cases} 1, & \text{if fem = "Woman"} \\ 0, & \text{otherwise} \end{cases}$$

$$D_{2i} = \begin{cases} 1, & \text{if mar = "Married"} \\ 0, & \text{otherwise} \end{cases}$$

3.2)

- β_2

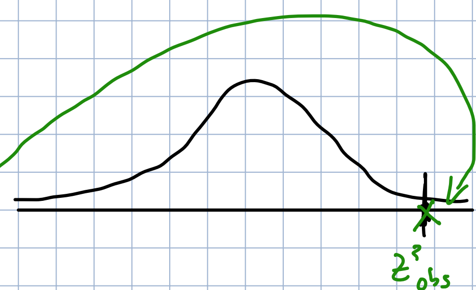
$$z_2^{obs} = \frac{\hat{\beta}_2 - \beta_2}{S.E.(\hat{\beta}_2)} = \frac{-0.224594}{0.054613} = -4.11264$$

$$\alpha_2^{obs} = P_{H_0}\left(|Z_2| \geq |z_2^{obs}|\right) = 2\left(1 - \Phi(4.11264)\right) \approx 0$$

- β_3

$$\alpha_3^{obs} = 0.0144 = 2\left(1 - \Phi\left(|z_3^{obs}|\right)\right)$$

$$1 - \Phi\left(|z_3^{obs}|\right) = \frac{0.0144}{2} = 0.0057$$



$$\Rightarrow z_3^{obs} = 2.530192$$

$$\Rightarrow \Phi(2.530192) = 1 - 0.0057$$

$$S.E.(\hat{\beta}_3) = \frac{\hat{\beta}_3}{z_3^{obs}} = \frac{0.155243}{2.530192} = 0.06136$$

- β_4

$$\hat{\beta}_4 = z_4^{obs} \cdot S.E(\hat{\beta}_4) = -4.607 \cdot 0.040127 \approx -0.18487$$

$$\alpha_4^{obs} = P(|Z_4| \geq |Z_4^{obs}|) = 2(1-\Phi(|Z_4^{obs}|))$$

$$= 2(\underbrace{1-\Phi(4.607)}_{\approx 1}) \approx 0$$

- β_5

$$\alpha_5^{obs} = 0.6271 = 2(1-\Phi(|Z_5^{obs}|)) \Rightarrow Z_5^{obs} = 0.4858127$$

because $\Phi(0.4858127) = 1 - 0.6271$

$$\hat{\beta}_5 = S.E.(\hat{\beta}_5) \cdot Z_5^{obs} = 0.4858127 \cdot 0.026397 = 0.012824$$

- β_6

If we increase the number of articles produced by the mentore by 1 unit, the log of the expected number of articles produced by the PhD student increases by one.

3.3)

We need to do a likelihood ratio test

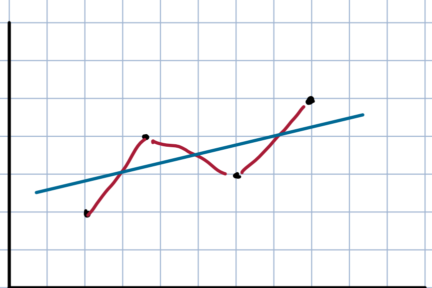
$$W = 2 \log \frac{\hat{L}(\text{model})}{\hat{L}(\text{restricted})} = 2\{\hat{\ell}(\text{model}) - \hat{\ell}(\text{restricted})\}$$

We know that

$$2\{\hat{\ell}(\text{model}) - \hat{\ell}(\text{restricted})\} = 2\{\hat{\ell}(\text{model}) + \hat{\ell}(\text{saturated}) - \hat{\ell}(\text{saturated}) - \hat{\ell}(\text{restricted})\}$$

$$= 2\{[\hat{\ell}(\text{saturated}) - \hat{\ell}(\text{restricted})] - [\hat{\ell}(\text{saturated}) - \hat{\ell}(\text{model})]\} =$$

$$= D(\text{restricted}) - D(\text{model}) = 1817.4 - 1634.4 = 183$$



The SATURATED MODEL is the most elaborated model one can estimate, it is a model with n parameters. It is a model with a perfect fit, it interpolates the data. But

- it does not model the variability
- doesn't provide any semplification

$$\alpha^{obs} = P(W \geq 183) = 1 - P(W < 183) = 1 - \Phi(183) \approx 0$$

$$W \sim N(0,1)$$

3.4)

We know that

$$\Delta(\text{model}) = 2\{\hat{\ell}(\text{saturated}) - \hat{\ell}(\text{model})\} = 1634.4$$

$$\hat{\ell}(\text{model}) = \hat{\ell}(\hat{\mu}) = \sum y_i \log \hat{\mu}_i - \hat{\mu}_i = -642.0261$$

Hence,

$$\hat{\ell}(\text{saturated}) = \frac{1634.4}{2} - 642.0261 = 175.1739$$

To see if the model is good we should check the following

- $\hat{\ell}(\text{model})$ should not be so far from $\hat{\ell}(\text{saturated})$
- $\text{Dev}(\text{model}) \leq n - p = 915 - 6 = 909$

$\Rightarrow$ Overall we don't have a good model

EXERCISE 4

A researcher is interested in how variables, such as **GRE** (Graduate Record Exam scores), **GPA** (grade point average) and prestige of the undergraduate institution (**Rank**), effect admission into graduate school. The response variable (**Admit**), admit/don't admit, is a binary variable. **Rank** takes on the values 1 through 4. The total number of observations is 400.

Exercise 4.1
Write the equation of the regression model using probit. Which other kind of model can we use?

Exercise 4.2
Knowing the estimates of regression coefficients using a probit model, such that

Coefficients	Estimates
β_1	-2.38684
β_2	0.00138
β_3	0.47773
β_4	-0.41540
β_5	-0.81214
β_6	-0.93590

Interpret them.

Exercise 4.3
Now, let consider another link function (should have already been specified in ex. 4.1). Write the regression model.

Exercise 4.4
We are interested in performing a test for comparing models. Let consider the previous model as full model, and the restricted model as a model which just includes **Rank**. Knowing the value of the observed test statistic, $w^{obs} = 16.449$, specify the hypothesis, the theoretical test statistic and p-value. Discuss about the result.

Exercise 4.5
Knowing the deviance of the previous restricted model $D(restricted) = 458.52$, and the deviance of the null model $D(null) = 499.98$, perform a test about overall significance by specifying the hypothesis, the test statistic and the p-value.

Exercise 4.6
Referring to the full model, we know

- $\hat{\beta}_3 = 0.804038$
- $z^{obs} = 2.423$

Interpret the value of β_3. Perform the test of significance and discuss about the result.

Since we have a binary response (admitted / not admitted) which depends on an underlying continuous variable we should consider a PROBIT MODEL.

4.1)

- $y_i \sim \text{Bernoulli}(\pi_i)$
- $\eta_i = \beta_1 + \beta_2 GRE_i + \beta_3 GPA_i + \beta_4 RANK2_i + \beta_5 RANK3_i + \beta_6 RANK4_i$
- $g(\pi_i) = \Phi^{-1}(\pi_i) = \eta_i \quad \text{where } \pi_i = \Phi(\eta_i)$

We are assuming that y_i comes from y_i^* according to the following system:

$$y_i = \begin{cases} 1 \text{ if } y_i^* > K \\ 0 \text{ if } y_i^* \leq K \end{cases} \qquad \text{where } y_i^* \overset{\perp\!\!\!\perp}{\sim} N(\eta_i, 1)$$

Hence in this case we have:

$$y_i^* = \beta_1 + \beta_2 GRE_i + \beta_3 GPA_i + \beta_4 D_{1i} + \beta_5 D_{2i} + \beta_6 D_{3i} + \varepsilon_i, \quad \varepsilon_i \overset{\perp\!\!\!\perp}{\sim} N(0,1)$$

where:

$$D_{1i} = \begin{cases} 1, \text{ if } RANK_i = 2 \\ 0, \text{ otherwise} \end{cases} \qquad D_{2i} = \begin{cases} 1, \text{ if } RANK_i = 3 \\ 0, \text{ otherwise} \end{cases}$$

$$D_{3i} = \begin{cases} 1, \text{ if } RANK_i = 4 \\ 0, \text{ otherwise} \end{cases}$$

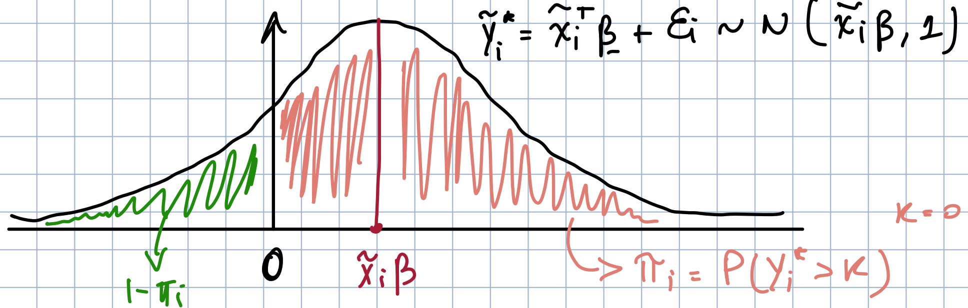
Instead of the PROBIT model we can use the LOGIT model, where the link function is the following:

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$$

4.2)

$$\tilde{y}_i^* = \tilde{x}_i^T \beta + \varepsilon_i \sim N(\tilde{x}_i \beta, 1)$$

-> see note "Threshold models"



If $\beta > 0$ then we have a bigger area for π_i and then the predicted probability of being admitted increases.

- $\beta_2 = 0.00138$

  An increase of GRE score increases the predicted probability of admission.

- $\beta_3 = 0.47773$

  An increase of GPA score increases the predicted probability of admission.

- $\beta_4 = -0.41540$

  If the prestige of the undergraduate institution corresponds to 2, the predicted probability of admission decreases.

- $\beta_5 = -0.81214$

  If the prestige of the undergraduate institution corresponds to 3, the predicted probability of admission decreases.

- $\beta_6 = -0.93590$

  If the prestige of the undergraduate institution corresponds to 4, the predicted probability of admission decreases.

4.3)

- $Y_i \sim \text{Bernoulli}(\pi_i)$
- $\eta_i = \beta_1 + \beta_2 GRE_i + \beta_3 GPA_i + \beta_4 D_{1i} + \beta_5 D_{2i} + \beta_6 D_{3i}$
- $g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i$

We have the following regression model:

$$\pi_i = \frac{\exp(\beta_1 + \beta_2 GRE_i + \beta_3 GPA_i + \beta_4 D_{1i} + \beta_5 D_{2i} + \beta_6 D_{3i})}{1 + \exp(\beta_1 + \beta_2 GRE_i + \beta_3 GPA_i + \beta_4 D_{1i} + \beta_5 D_{2i} + \beta_6 D_{3i})}$$

4.4)

FULL MODEL (M_6):

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_1 + \beta_2 GRE_i + \beta_3 GPA_i + \beta_4 D_{1i} + \beta_5 D_{2i} + \beta_6 D_{3i}$$

RESTRICTED MODEL (M_4):

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \gamma_1 + \gamma_2 D_{1i} + \gamma_3 D_{2i} + \gamma_4 D_{3i}$$

system of hypothesis:

$$\begin{cases} H_0: \beta_2 = \beta_3 = 0 \\ H_1: \overline{H_0} \end{cases}$$

$$W = 2\log\frac{\hat{L}(M_6)}{\hat{L}(M_4)} \overset{H_0}{\sim} \chi^2_{p-p_0}$$

p: number of coefficients in M_6

p_0: number of coefficients in M_4

we know that $w^{obs} = 16.449$

$\alpha^{obs} = P_{H_0}(W \geq 16.449) \approx 0.00026 \Rightarrow$ we reject H_0

4.5)

System of the hypothesis:

$$\begin{cases} H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \\ H_1: \overline{H_0} \end{cases}$$

$$W = 2\log\frac{\hat{L}(model)}{\hat{L}(null)} \overset{H_0}{\sim} \chi^2_{p-1}$$

We know that

$$W = 2\left[\hat{\ell}(model) - \hat{\ell}(null)\right] = D(null) - D(model).$$

We need to find $D(model)$. From 4.4) we have that

$$w^{obs}_{4.4)} = D(restricted) - D(model) = 458.52 - D(model) = 16.449$$

$$\Rightarrow D(model) = 458.52 - 16.449 = 442.071$$

$$\Rightarrow W^{obs} = 499.98 - 442.071 = 57.909$$

$\Rightarrow \alpha^{obs} = P_{H_0}(W \geq 57.909) \approx 0 \Rightarrow$ we reject H_0

4.6) $\hat{\beta}_3 = 0.804038$ $z^{obs} = 2.423$

INTERPRETATION OF β_3

The log odds increases by 0.804038 if the grade point average increases of 1 unit, while keeping fixed the other variables.

TEST OF SIGNIFICANCE

system of hypothesis:

$$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases} \qquad z^{obs} = 2.423$$

$$\alpha^{obs} = 2\left(1 - \Phi(2.423)\right) \approx 0.0075 \Rightarrow \text{we reject } H_0$$

EXERCISE 5

Let $(y_1, \dots, y_5)$ and $(y_6, \dots, y_{10})$ be two independent samples from a Poisson distribution of mean $\exp\{\beta_1\}$ and from a Poisson distribution of mean $\exp\{\beta_1 + \beta_2\}$, respectively.

Exercise 5.1

Formulate an appropriate Poisson regression model for the expected value of Y_i, $i = 1, \dots, 10$.

Exercise 5.2

Write the log-likelihood function of $\underline{\beta} = (\beta_1, \beta_2)$ and the score function. Find the maximum likelihood estimate of (β_1, β_2). Finally, obtain the observed information matrix.

Exercise 5.3

Write the approximate distribution of the maximum likelihood estimator $\underline{\hat{B}}$ of $\underline{\beta} = (\beta_1, \beta_2)$, and an approximate distribution of the maximum likelihood estimator $\hat{B}_1$ of β_1.

Exercise 5.4

Provide the interpretation of the coefficient β_2.

Exercise 5.5

Define the concept of "saturated model" and obtain the expression of maximum of the log-likelihood for this model.

$(y_1, \dots, y_5)$ sample from $Y_i \sim \text{Poisson}(e^{\beta_1})$ indep. $i = 1, \dots, 5$

$(Y_6, \dots, Y_{10})$ sample from $Y_i \sim \text{Poisson}(e^{\beta_1 + \beta_2})$ indep. $i = 6, \dots, 10$

5.1)

- $Y_i \sim \text{Poisson}(\mu_i)$ $i = 1, \dots, 10$ indep.
- $\eta_i = \beta_1 + \beta_2 x_i$ with $x_i = \begin{cases} 0, & i = 1, \dots, 5 \\ 1, & i = 6, \dots, 10 \end{cases}$
- $\log(\mu_i) = \eta_i$

5.2)

$$f(y_i \mid \mu_i) = e^{-\mu_i} \mu_i^{y_i} \cdot \frac{1}{y_i!}$$

$$f(y_1, \dots, y_{10} \mid \underline{\mu}) = \prod_{i=1}^{10} \left(e^{-\mu_i} \mu_i^{y_i} \frac{1}{y_i!} \right)$$

$$L(\underline{\mu}) \propto \prod_{i=1}^{10} \left(e^{-\mu_i} \mu_i^{y_i}\right) \quad \text{likelihood}$$

$$\ell(\underline{\mu}) \propto \sum_{i=1}^{10} \{-\mu_i + y_i \log \mu_i\} \quad \text{loglikelihood}$$

Since we know that $\mu_i = e^{\beta_1 + \beta_2 x_i}$

$$\ell(\underline{\beta}_1, \underline{\beta}_2) = \sum_{i=1}^{10} \left\{- e^{\beta_1 + \beta_2 x_i} + y_i (\beta_1 + \beta_2 x_i)\right\} \leftarrow \text{loglikelihood of } (\beta_1, \beta_2)$$

$$= - \sum_{i=1}^{10} e^{\beta_1 + \beta_2 x_i} + \beta_1 \sum_{i=1}^{10} y_i + \beta_2 \sum_{i=1}^{10} y_i x_i$$

$$= - \sum_{i=1}^{5} e^{\beta_1} - \sum_{i=6}^{10} e^{\beta_1 + \beta_2} + \beta_1 \sum_{i=1}^{10} y_i + \beta_2 \sum_{i=6}^{10} y_i$$

$$= - 5 e^{\beta_1} - 5 e^{\beta_1 + \beta_2} + \beta_1 \sum_{i=1}^{10} y_i + \beta_2 \sum_{i=6}^{10} y_i$$

$$\ell_*(\beta_1, \beta_2) = \begin{cases} \dfrac{\partial \ell(\beta_1, \beta_2)}{\partial \beta_1} = -5e^{\beta_1} - 5e^{\beta_1 + \beta_2} + \sum_{i=1}^{10} y_i \\ \dfrac{\partial \ell(\beta_1, \beta_2)}{\partial \beta_2} = -5e^{\beta_1 + \beta_2} + \sum_{i=6}^{10} y_i \end{cases} \quad \text{score function}$$

$$\ell_{**}(\beta_1, \beta_2) = \begin{bmatrix} -5e^{\beta_1} - 5e^{\beta_1 + \beta_2} & -5e^{\beta_1 + \beta_2} \\ -5e^{\beta_1 + \beta_2} & -5e^{\beta_1 + \beta_2} \end{bmatrix} = -5 \begin{bmatrix} e^{\beta_1} + e^{\beta_1 + \beta_2} & e^{\beta_1 + \beta_2} \\ e^{\beta_1 + \beta_2} & e^{\beta_1 + \beta_2} \end{bmatrix}$$

observed information $j(\underline{\beta}) = - \ell_{**}(\beta_1, \beta_2)$

- FIRST WAY

for sample 1 $(y_1, \dots, y_5) = \underline{y}_1$ the expected value is $\mathbb{E}[\underline{Y}_1] = \mu_1 = e^{\beta_1}$

for sample 2 $(y_6, \dots, y_{10}) = \underline{y}_2$ the expected value is $\mathbb{E}[\underline{Y}_2] = \mu_2 = e^{\beta_1 + \beta_2}$

We have that the function from (μ_1, μ_2) to (β_1, β_2) is bijective.

$$\Rightarrow (\hat{\beta}_1, \hat{\beta}_2) = f(\hat{\mu}_1, \hat{\mu}_2)$$

We know that the MLE for (μ_1, μ_2) are:

$$\hat{\mu}_1 = \bar{y}_1 = \frac{1}{5}\sum_{i=1}^{5} y_i \quad \text{sample mean of } \underline{Y}_1$$

$$\hat{\mu}_2 = \bar{y}_2 = \frac{1}{5}\sum_{i=6}^{10} y_i \quad \text{sample mean of } \underline{Y}_2$$

$$\Rightarrow \begin{cases} \hat{\mu}_1 = e^{\hat{\beta}_1} \\ \hat{\mu}_2 = e^{\hat{\beta}_1 + \hat{\beta}_2} \end{cases} \Rightarrow \begin{array}{l} \hat{\beta}_1 = \log \hat{\mu}_1 = \log \bar{y}_1 \\ \hat{\beta}_1 + \hat{\beta}_2 = \log \hat{\mu}_2 \Rightarrow \hat{\beta}_2 = \log \hat{\mu}_2 - \log \hat{\mu}_1 = \log \frac{\bar{y}_2}{\bar{y}_1} \end{array}$$

- ALTERNATIVE WAY

$$\begin{cases} -5e^{\beta_1} - 5e^{\beta_1+\beta_2} + \sum_{i=1}^{10} y_i = 0 \rightarrow 5e^{\beta_1} = \sum_{i=1}^{10} y_i - \sum_{i=6}^{10} y_i \rightarrow e^{\beta_1} = \frac{1}{5}\left(\sum_{i=1}^{10} y_i - \sum_{i=6}^{10} y_i\right) = \bar{Y}_1 \\ -5e^{\beta_1+\beta_2} + \sum_{i=6}^{10} y_i = 0 \rightarrow e^{\beta_1+\beta_2} = \frac{\sum_{i=6}^{10} y_i}{5} = \bar{Y}_2 \end{cases}$$

$$\Rightarrow \hat{\beta}_1 = \log(\bar{Y}_1)$$

$$\Rightarrow \hat{\beta}_1 + \hat{\beta}_2 = \log(\bar{Y}_2) \Rightarrow \hat{\beta}_2 = \log \frac{\bar{Y}_2}{\bar{Y}_1}$$

5.3)

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, j(\hat{\underline{\beta}})^{-1} \right)$$

$\rightsquigarrow$ the marginals are:

$$\hat{\beta}_i \sim N(\beta_i, [j(\hat{\underline{\beta}})^{-1}]_{ii})$$

5.4)

We already noticed that

for sample 1 $\mathbb{E}[Y_i] = \mu_1 = e^{\beta_1}$

for sample 2 $\mathbb{E}[Y_i] = \mu_2 = e^{\beta_1+\beta_2}$

$$\frac{\mu_2}{\mu_1} = \frac{e^{\beta_1+\beta_2}}{e^{\beta_1}} = e^{\beta_2} \Rightarrow \beta_2 = \log \frac{\mu_2}{\mu_1} = \log \frac{\mathbb{E}[Y_i \mid \text{sample } 1]}{\mathbb{E}[Y_i \mid \text{sample } 2]}$$

=> β_2 represents the log of the ratio between the means of the two samples.

=> Noting that $\mu_2 = e^{\beta_2} \cdot \mu_1$, the mean of sample 2 is obtained by multiplying the mean of sample 1 of a coefficient e^{β_2}

5.5)

The saturated model is a model with n parameters $\mu_1, \dots, \mu_n$. In this case we have one parameter for each observation and the estimates are:

$$\hat{\mu}_i = y_i \quad \forall i$$

Hence the log-likelihood of this model is:

$$\tilde{\ell}\,(\text{saturated}) = \sum_{i=1}^{10} \{-\hat{\mu}_i + y_i \log \hat{\mu}_i\} = \sum_{i=1}^{10} \{-y_i + y_i \log y_i\}$$