In [96]:
```
library(stats)
library(MASS)
```

```
Attaching package: 'MASS'


The following object is masked from 'package:dfidx':

    select


The following object is masked from 'package:dplyr':

    select
```

Question 1

We are interested in how governments' management of public resources impacts economic prosperity. Our data come from Alvarez, Cheibub, Limongi, and Przeworski (1996) and is labelled gdpChange.csv on GitHub. The dataset covers 135 countries observed between 1950 or the year of independence or the first year forwhich data on economic growth are available ("entry year"), and 1990 or the last year for which data on economic growth are available ("exit year"). The unit of analysis is a particular country during a particular year, for a total > 3,500 observations.

• Response variable:

– GDPWdiff: Difference in GDP between year t and t−1. Possible categories include: "positive", "negative", or "no change"

• Explanatory variables:

– REG: 1=Democracy; 0=Non-Democracy – OIL: 1=if the average ratio of fuel exports to total exports in 1984-86 exceeded 50%; 0= otherwise

1. Construct and interpret an unordered multinomial logit with GDPWdiff as the output and "no change" as the reference category, including the estimated cutoff points and coefficients.

In [103…
```
data <- read.csv("gdpChange.csv")


# we create the classes
data$GDPWdiff <- ifelse(data$GDPWdiff > 0, "positive",
                        ifelse(data$GDPWdiff < 0, "negative", "no change"))
data$GDPWdiff <- ifelse(data$GDPWdiff == "positive", 1,
                 ifelse(data$GDPWdiff == "negative", 2, 3))

# we create the factor
data$GDPWdiff <- factor(data$GDPWdiff)
```

```r
# Ajustar el modelo de regresión logística multinomial no ordenado
gdp_model <- multinom(GDPWdiff ~ REG + OIL, data = data)

# Imprimir los resultados del modelo
print(summary(gdp_model)$coefficients)
```

```
# weights:  12 (6 variable)
initial  value 4087.936326
iter  10 value 2363.684951
iter  20 value 2339.374916
final  value 2339.363006
converged
```

```
Warning message in sqrt(diag(vc)):
"Se han producido NaNs"
  (Intercept)        REG         OIL
2  -0.7284133 -0.3899074   0.2076476
3  -4.5297594 -1.7427883 -35.5665728
```

In [104…

```r
#The REG coefficient (-0.389) indicates the change in the probabilities of category
#in democratic countries compared to non-democratic countries, holding constant the

#The OIL coefficient (0.208) indicates the change in the probabilities of category
#in countries with significant oil exports compared to countries without significar

#In category 2 ("negative"), the coefficient of the intercept is -4.530,
#which indicates the natural logarithm of the probabilities that the difference in
```

1. Construct and interpret an ordered multinomial logit with GDPWdiff as the outcome variable, including the estimated cutoff points and coefficients.

In [107…

```r
gdp2_model <- polr(GDPWdiff ~ REG + OIL, data = data)
# Imprimir los resultados del modelo
print(summary(gdp2_model)$coefficients)
```

```
Re-fitting to get Hessian
```

```
         Value Std. Error   t value
REG -0.4101680  0.0751802 -5.455798
OIL  0.1788450  0.1154582  1.549002
1|2  0.7035892  0.0475622 14.793032
2|3  5.3198478  0.2522797 21.087105
```

In [108…

```r
#The REG coefficient in the first column (-0.410) is interpreted as the change
#in the probabilities that the difference in GDP is from category 1 to category 2,
#compared to non-democratic countries, holding constant the effect of the OIL varia

#The OIL coefficient in the second column (0.179) indicates the change in the proba
#in GDP will be from category 1 to category 2, for countries with significant oil e
#without significant oil exports. , keeping constant the effect of the variable REG

#The coefficient in the third column (0.704) is interpreted as the natural logarith
#the difference in GDP will be category 1 ("positive"), compared to category 3 ("ne
#holding constant the effect of the variables REG and OIL.

#The coefficient in the fourth column (5.320) is interpreted as the natural logarit
#the difference in GDP will be category 2 ("negative"), compared to category 3 ("ne
#holding constant the effect of the variables REG and OIL.
```

Question 2

Consider the data set MexicoMuniData.csv, which includes municipal-level information from Mexico. The outcome of interest is the number of times the winning PAN presidential candidate in 2006 (PAN.visits.06) visited a district leading up to the 2009 federal elections, which is a count. Our main predictor of interest is whether the district was highly contested, or whether it was not (the PAN or their opponents have electoral security) in the previous federal elections during 2000 (competitive.district), which is binary (1=close/swing district, 0="safe seat"). We also include marginality.06 (a measure of poverty) and PAN.governor.06 (a dummy for whether the state has a PAN-affiliated governor) as additional control variables.

In [109…
```
datamex = read.csv("MexicoMuniData.csv")
```

(a) Run a Poisson regression because the outcome is a count variable. Is there evidence that PAN presidential candidates visit swing districts more? Provide a test statistic and p-value.

In [118…
```
# we fit the model
model <- glm(PAN.visits.06 ~ competitive.district + marginality.06 + PAN.governor.(
             data = datamex, family = "poisson")

# we adjust the we obtain the test statistic and the value of p
test_stat <- summary(model)$coef["competitive.district", "z value"]
p_value <- summary(model)$coef["competitive.district", "Pr(>|z|)"]
cat("test statistic:", test_stat, "\n")
cat("p value:", p_value, "\n")
alpha <- 0.05
if (p_value < alpha) {
  cat("There is evidence to reject the null hypothesis. PAN presidential candidates
} else {
  cat("There is not enough evidence to reject the null hypothesis. It cannot be cor
}
```

```
test statistic: -0.4766106
p value: 0.6336394
There is not enough evidence to reject the null hypothesis. It cannot be concluded
that PAN presidential candidates visit contested districts more frequently than sa
fe districts.
```

(b) Interpret the marginality.06 and PAN.governor.06 coefficients.

In [120…
```
print("the coefficients are:")
print(summary(model)$coefficients)
```

```
[1] "the coefficients are:"
                        Estimate Std. Error     z value     Pr(>|z|)
(Intercept)          -3.81023498  0.2220947 -17.1559048 5.677880e-66
competitive.district -0.08135181  0.1706882  -0.4766106 6.336394e-01
marginality.06       -2.08014361  0.1173386 -17.7277058 2.562806e-70
PAN.governor.06      -0.31157887  0.1667306  -1.8687569 6.165665e-02
```

In [121…
```
#The marginality coefficient (marginality.06) is -2.080,
#which means that an increase in marginality by one unit is associated with an 87.8
#decrease in the expected number of visits by a PAN presidential candidate to a mun

#The coefficient for PAN.governor.06 is -0.312,
#which means that in municipalities where the PAN-affiliated governor was in power,
#the number of visits by the PAN presidential candidate would be expected to decrea
```

(c) Provide the estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district that was competitive (competitive.district=1), had an average poverty level (marginality.06 = 0), and a PAN governor (PAN.governor.06=1).

In [125…
```
pre <- data.frame(competitive.district = 1, marginality.06 = 0, PAN.governor.06 = 1
```

In [128…
```
predicted <- predict(model, pre, type="response")
print(paste("The estimated number of visits is:",predicted))
```

```
[1] "The estimated number of visits is: 0.0149481810354763"
```

In [ ]: