Mental Health Analysis

Università Ca'Foscari Venezia

Federico Stragliotto
Elena Palmieri
Jia Shi

Health Data Science (EM14) - a.a. 2022-23

Awareness is the first step

INCLUSION of mental health in the Sustainable Development Goals

Mental Health & PANDEMIC: Rates of already-common conditions such as depression and anxiety went up by more than **25%** in the first year of the pandemic, adding to the nearly one billion people who were already living with a mental disorder.

Suicide accounts for more than one in every 100 deaths. It is a **major cause** of death among young people.

The **economic consequences** of mental health conditions are enormous. **Productivity losses** and other indirect costs to society often far outstrip health care costs. Economically, **schizophrenia** is the most costly mental disorder per person to society. **Depressive** and **anxiety disorders** are much less costly per person; but they are more prevalent, and so majorly contribute to overall national costs.

World mental health report 2022 (WHO)

Problem:

Assesing the issue of mental health in Italy through the analysis of the PASSI (Progressi delle Aziende Sanitarie per la Salute in Italia) survey held in 2019.

Objectives:

Finding possible explicative relationships between mental health and specific lifestyle habits (smoking, drinking) and conditions.

1 Data Cleaning and EDA

2 Contingency Table Analysis

3 Log Linear model

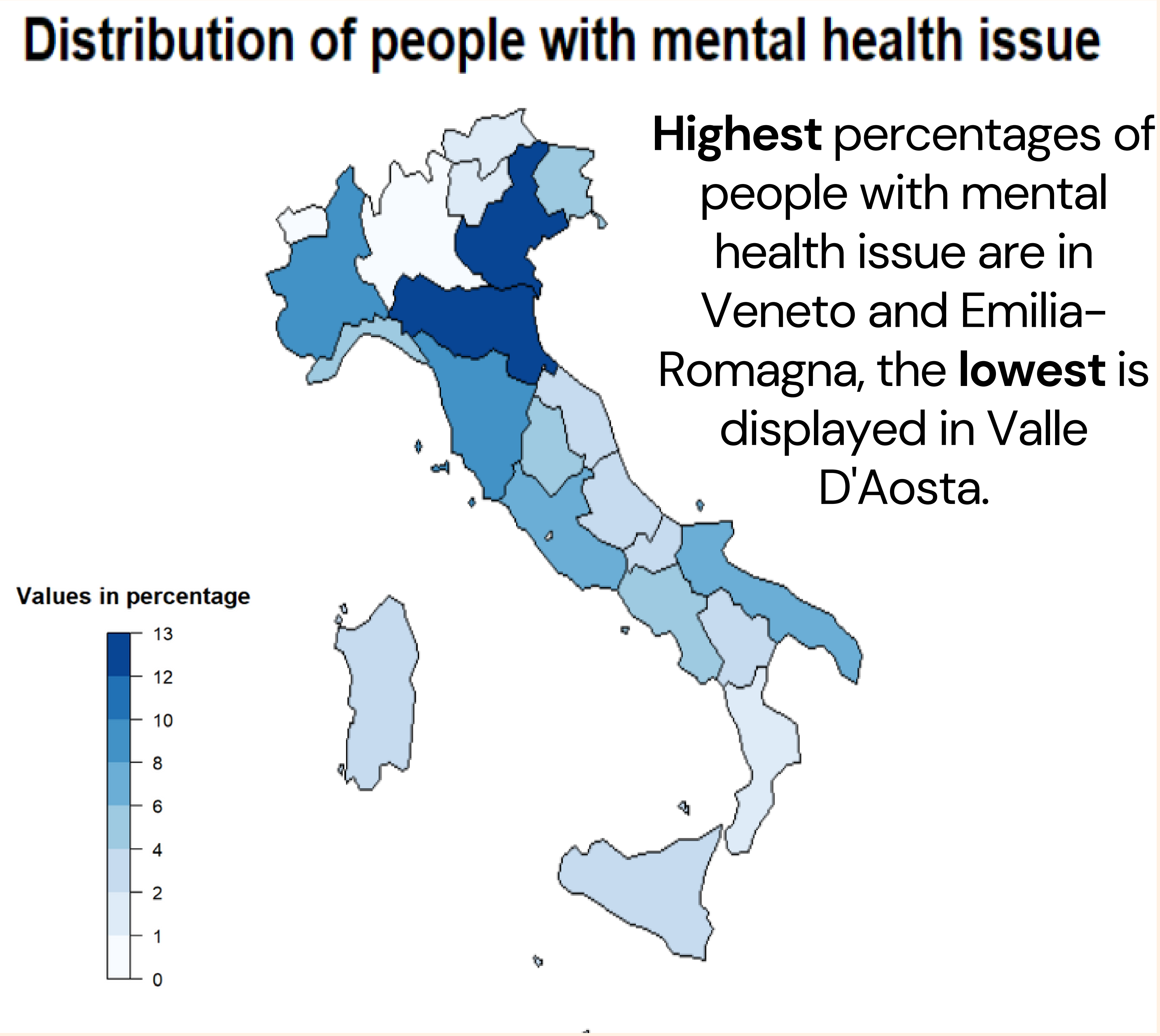
4 Association selection

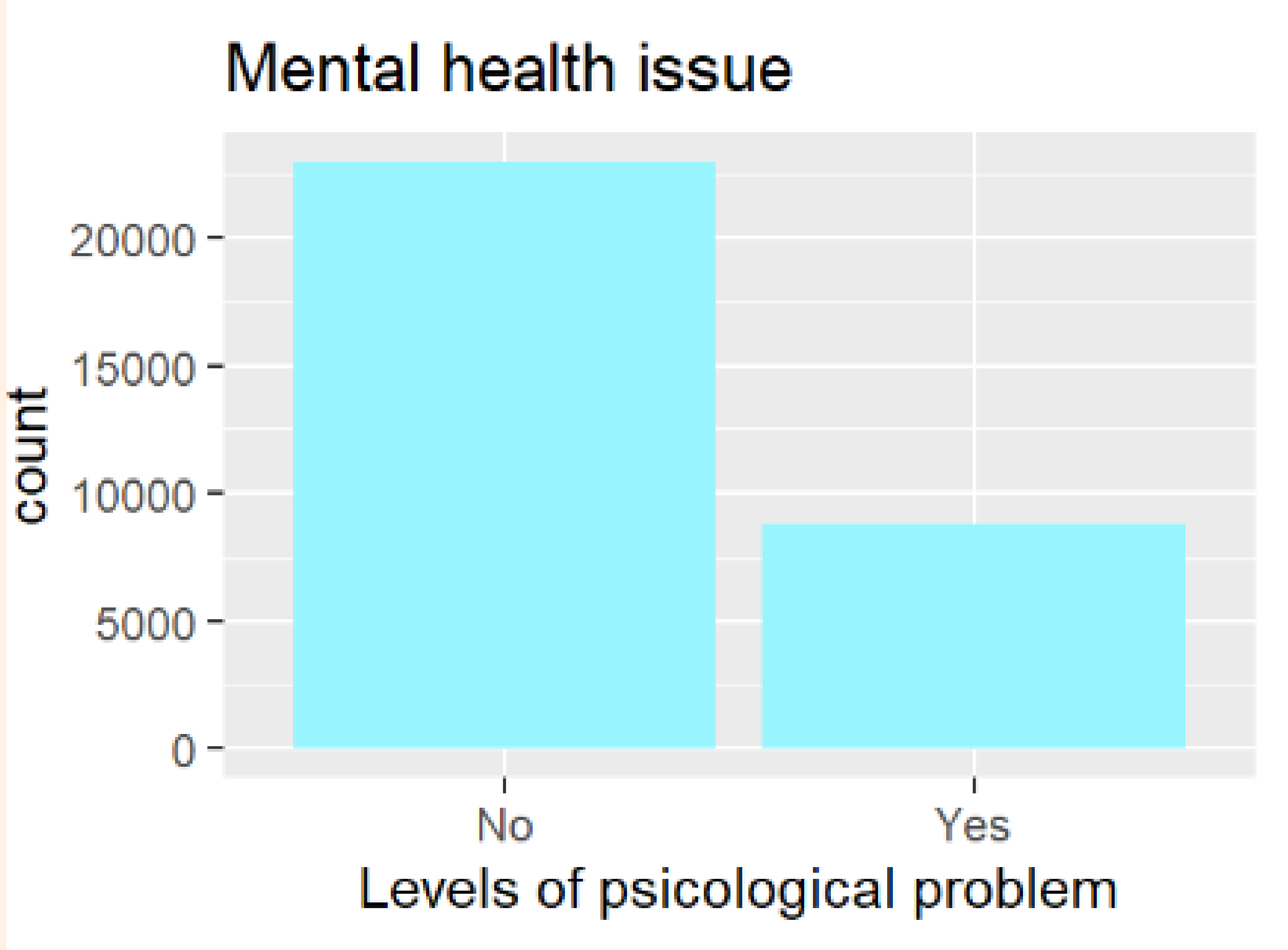
5 Logistic Regression

6 Mixed effects Logistic Regression

1 Data Cleaning and EDA

%= (people with mental health issue per region/ total number of people with mental health) * 100



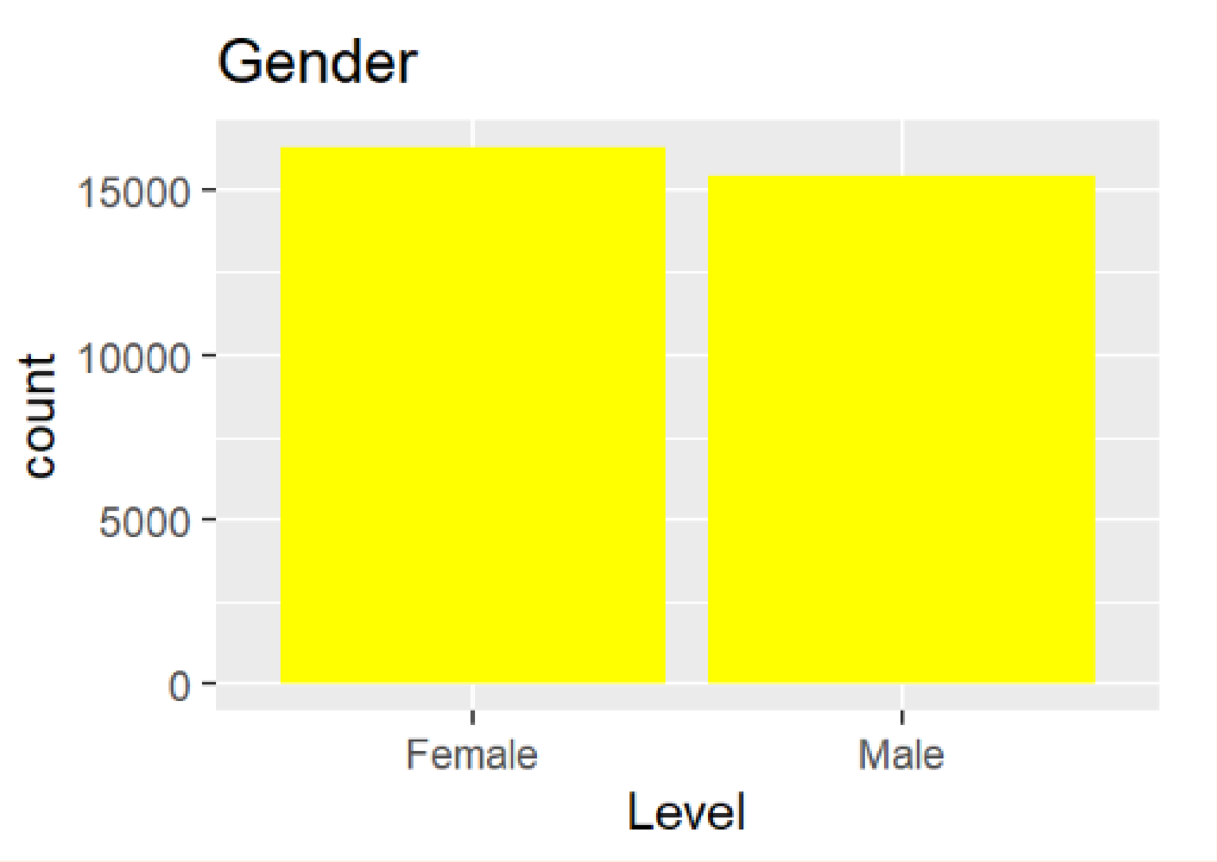


Dependent variable modification:

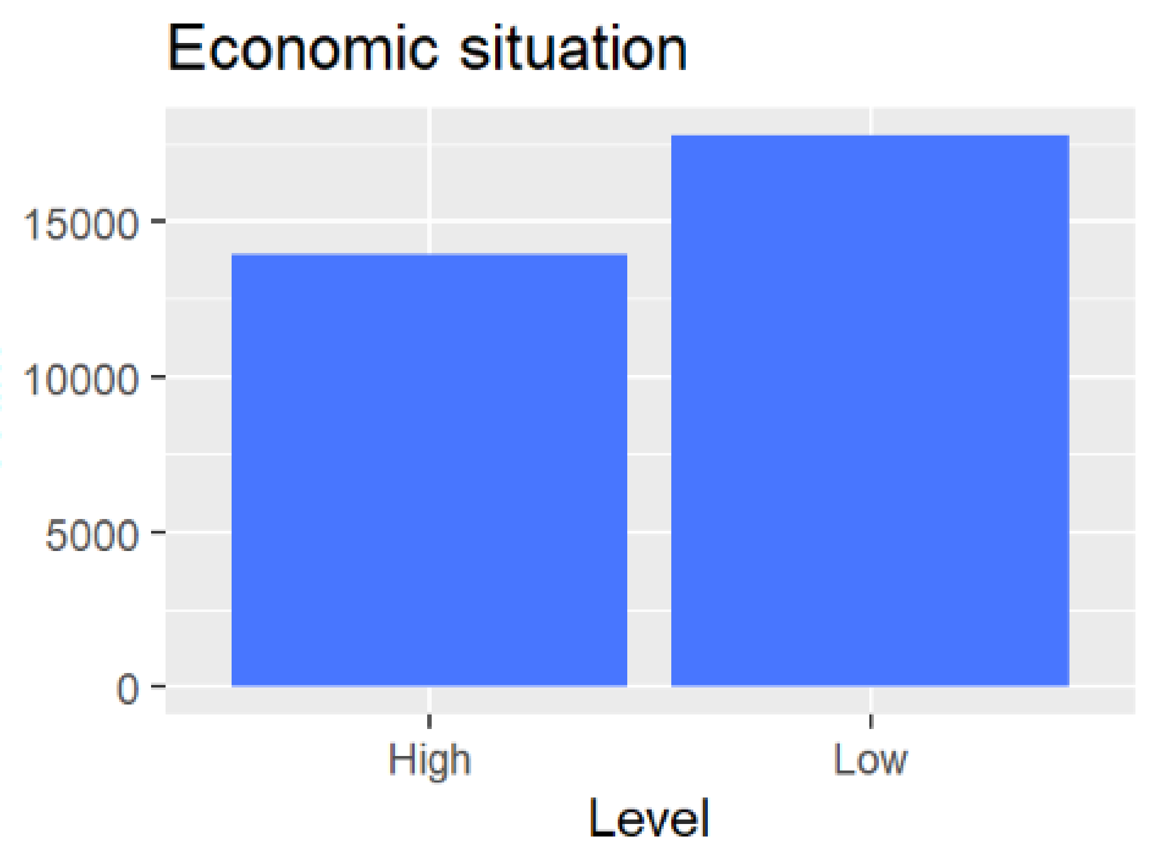
From "In the last 30 days, for how many days did you feel emotional problems, anxiety, depression, stress?" to a BINARY VARIABLE .

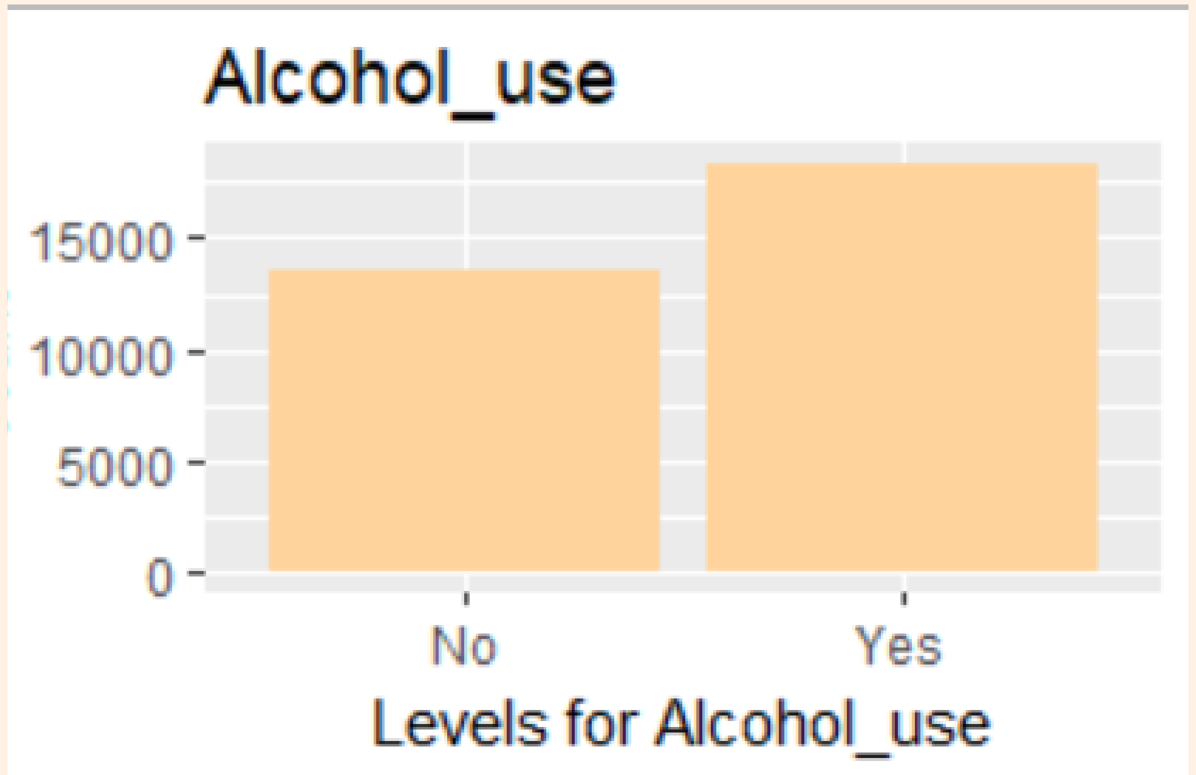
No	Yes
22996	8764

1 Other significant variables:



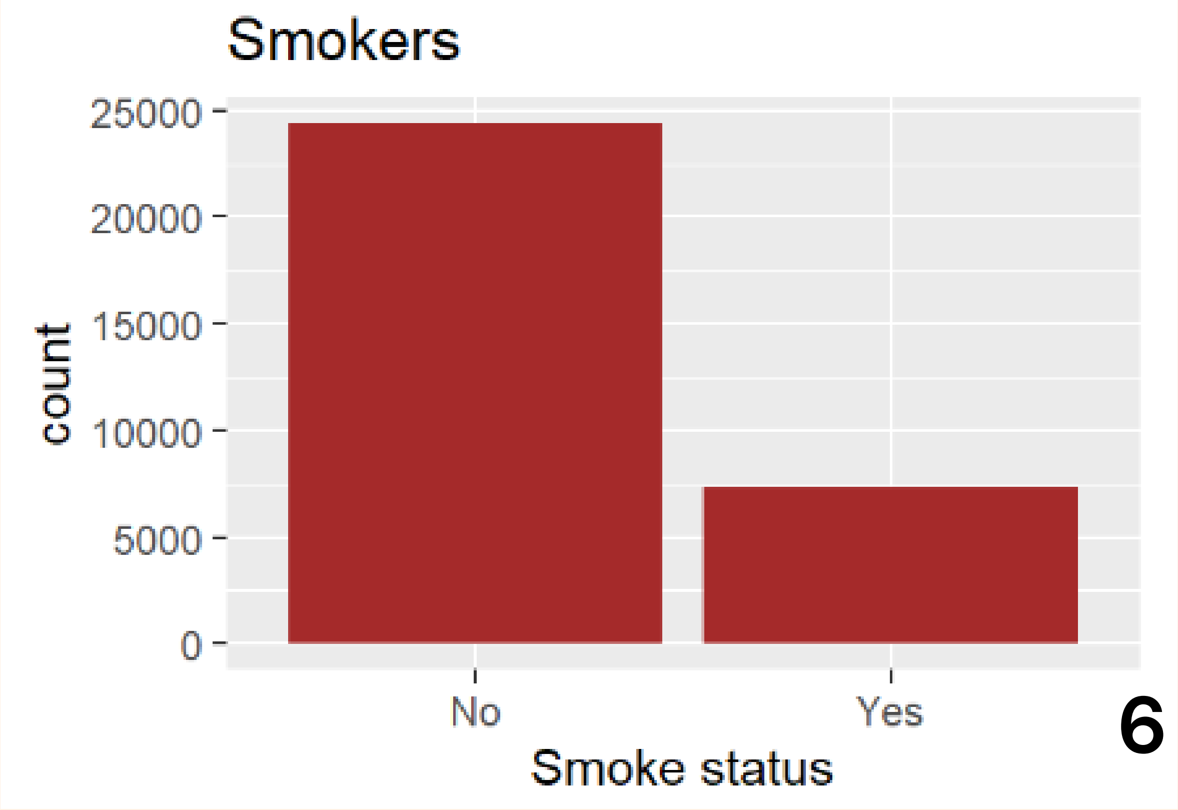
High and Low Economic difficulties

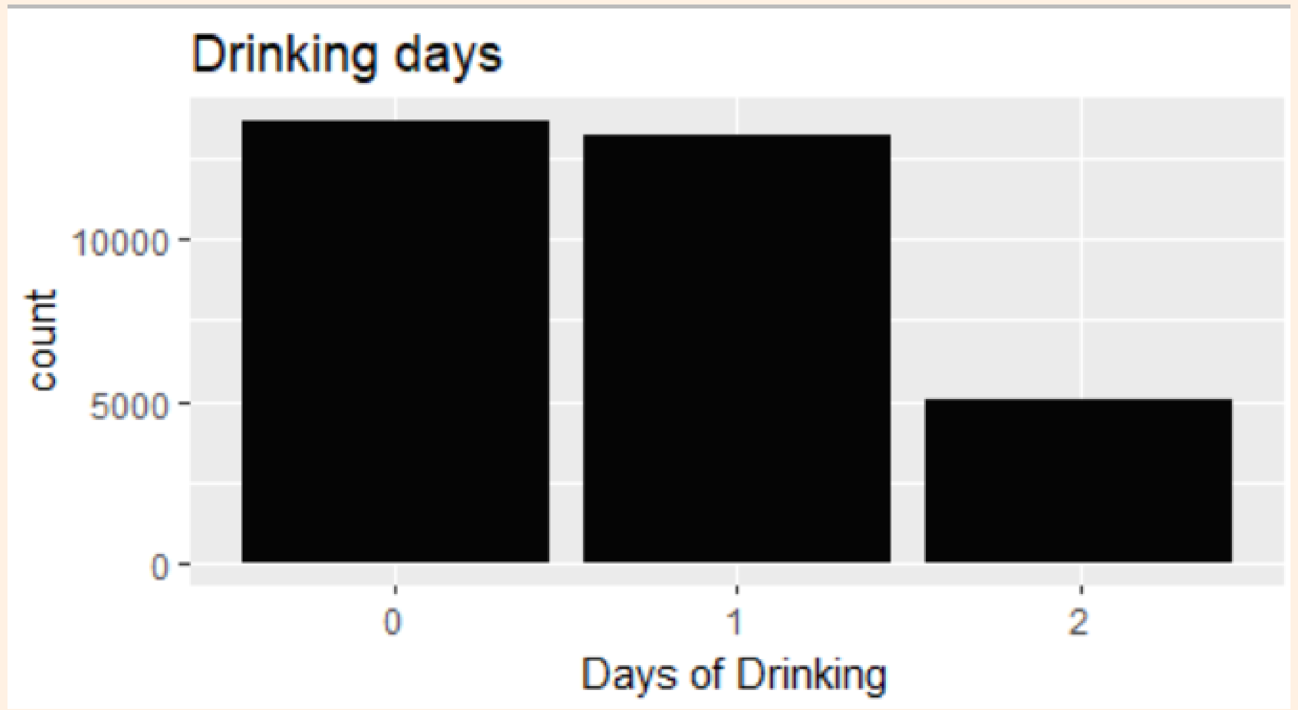




From numerical variable to
"Yes" "No"

Are you currently smoking?





From range -99, 30
to a three levels factor.

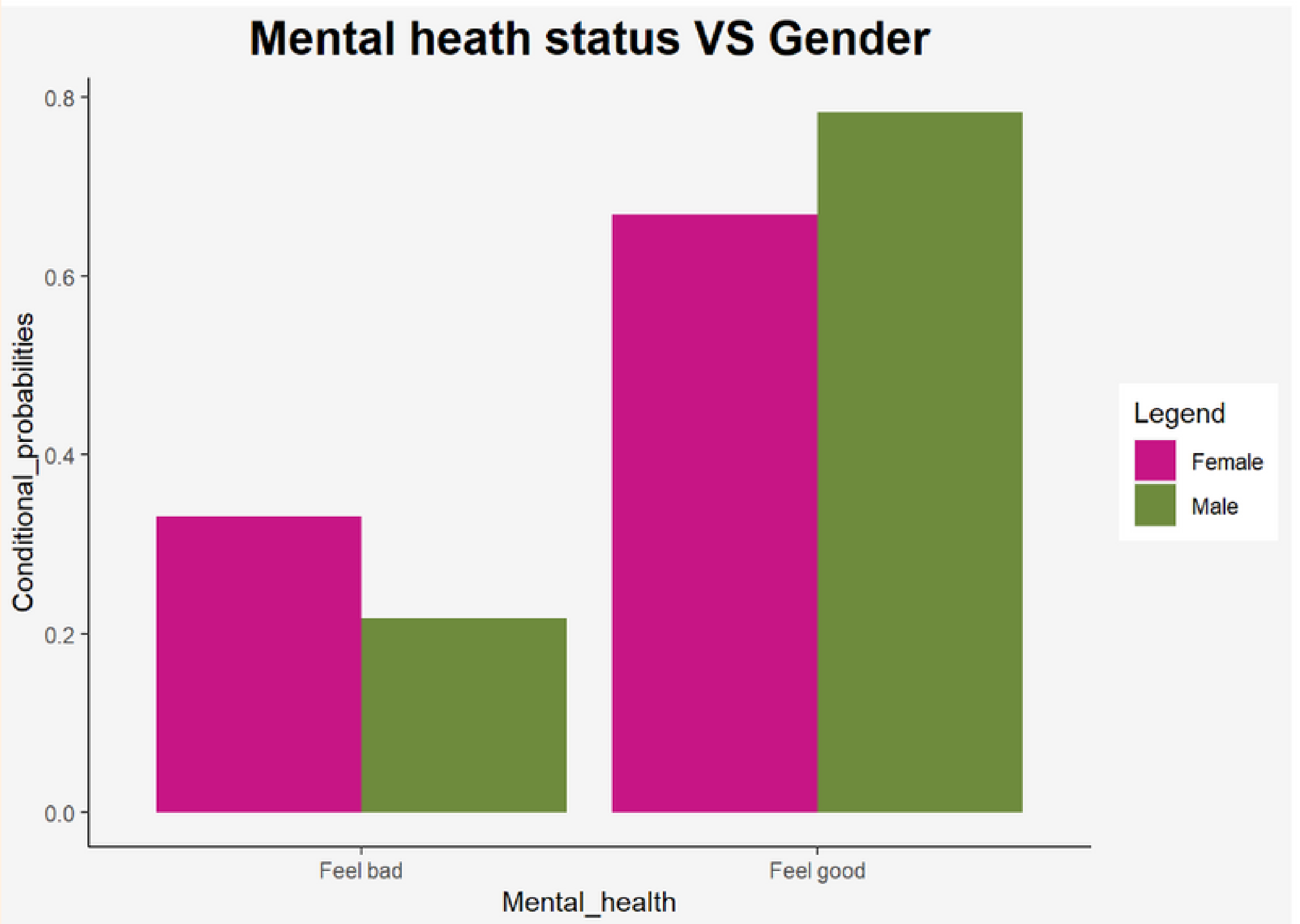
```
case_when(s05_alcool_gg==0~"0",
s05_alcool_gg>0 & s05_alcool_gg<=15 ~"1",
s05_alcool_gg>15 ~"2" )
```

Contingency Table Analysis

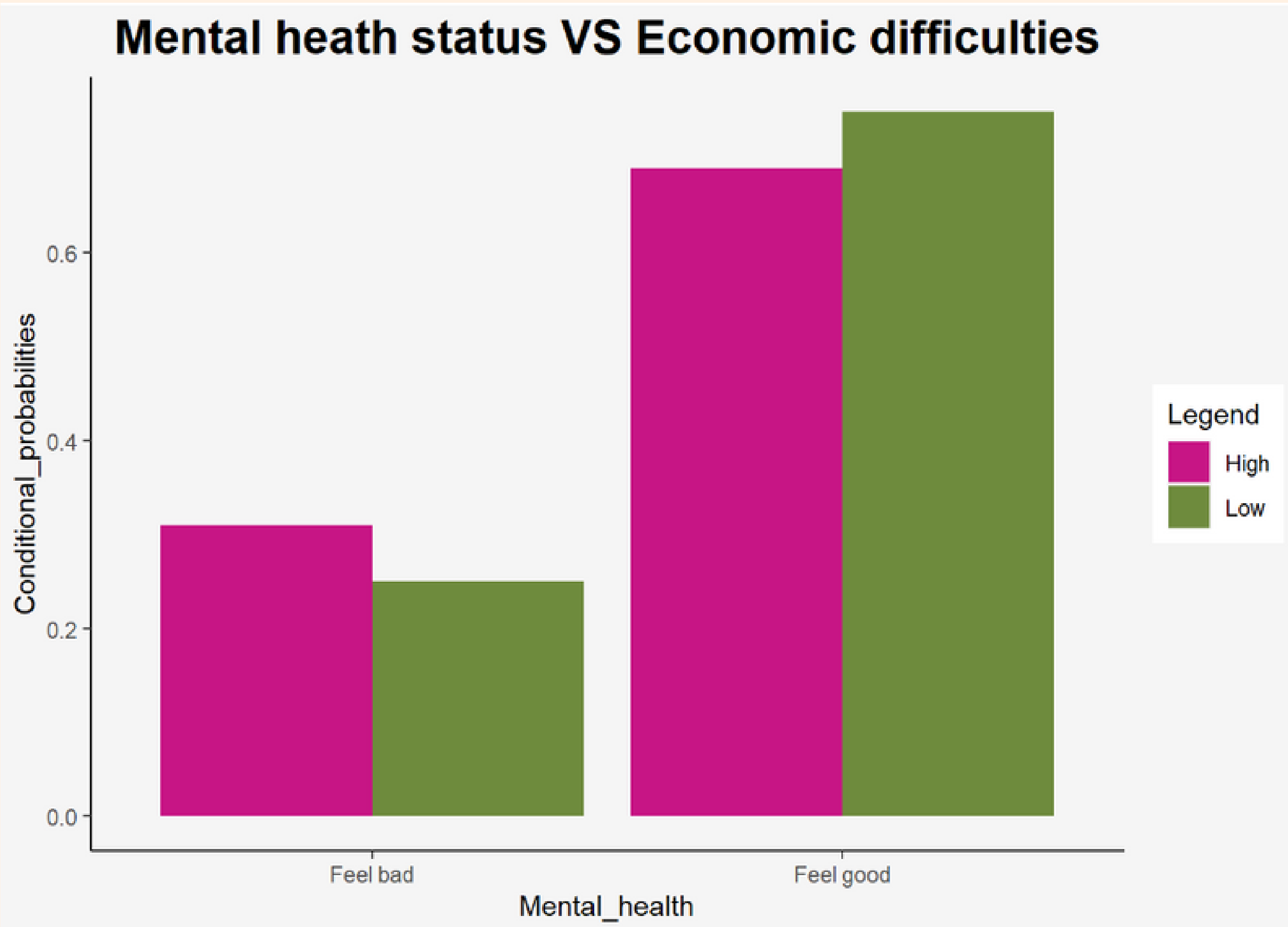
RELATIVE RISK: 1.52
A female is 52% more likely to feel bad mentally than a male.

	Feel Bad	Feel Good
Female	0.331	0.669
Male	0.217	0.783

Probabilities of having mental health issues CONDITIONAL to **Gender**



Pearson's Chi-squared test:
p-value < 0.05



Probabilities of having mental health issues CONDITIONAL to **Economic difficulties**

RELATIVE RISK: 1.25
people with high economic difficulties are 25% more likely to feel bad mentally than the others

	Feel Bad	Feel Good
High	0.310	0.690
Low	0.249	0.751

Pearson's Chi-squared test:
p-value < 0.05

3 Poisson Loglinear model

We created a table with the variables s01_salute_psic, s03_fumo_att and Alcohol use. Our goal was to study how the cell counts depend on the levels of the categorical variables.

alcohol	cigarette	mental issue: yes	no
yes	yes	1479	3450
	no	3760	9528
no	yes	666	1734
	no	2859	8284

, , alcohol = yes

cigarette	mental issue: yes	no
yes	0.3000609	0.6999391
no	0.2829621	0.7170379

, , alcohol = no

cigarette	mental issue: yes	no
yes	0.2775000	0.7225000
no	0.2565736	0.7434264

This table of proportions shows that among people who smoke and drink alcohol, 30% have experienced mental problems in the last 30 days. Similarly, among people who do not smoke and drink alcohol, 74% have no mental problems.

Simplest Model:

$$\log(E(\mathrm{Freq})) = \alpha + \beta_1(\mathrm{cigarette}_{\mathrm{yes}}) + \beta_2(\mathrm{mental.\,issue}_{\mathrm{yes}}) + \beta_3(\mathrm{alcohol}_{\mathrm{yes}})$$

```
Deviance Residuals:
       1         2         3         4         5         6         7         8
  8.9793   -1.7266    7.2078   -6.2026   -6.9688   -0.2937  -11.5986    8.3970

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      8.928383   0.009761  914.67   <2e-16 ***
cigaretteyes    -1.204014   0.013318  -90.40   <2e-16 ***
mental.issueyes -0.964668   0.012553  -76.84   <2e-16 ***
alcoholyes       0.296485   0.011346   26.13   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 17443.78  on 7  degrees of freedom
Residual deviance:   427.72  on 4  degrees of freedom
AIC: 514.2
```

Looking at the summary it appears this is a great model.
We see highly significant coefficients and p-values near 0. But for loglinear models we want to **check the residual deviance** (we'd like it to be close in value to the degrees of freedom). Here we have **427.72** on 4 degrees of freedom. This indicates **a poor fit**!
Our H0 is that the expected frequencies satisfy the given loglinear model but our p-value is 2.844169e-91 so we **reject the null hypothesis**.

0.38 to 1 =

Odds of having mental problem in the last 30 days *regardless* of *alcohol consumption* or *cigarettes smoking*.

A more intuitive way to investigate fit is to compare the fitted values to the observed values.
Our fitted model is not so far from the observed data but it can be improved!

	cigarette	mental.issue	alcohol	Freq	fitted(mod0)
1	yes	yes	yes	1479	1160.0133
2	no	yes	yes	3760	3866.8692
3	yes	no	yes	3450	3043.7774
4	no	no	yes	9528	10146.3401
5	yes	yes	no	666	862.3846
6	no	yes	no	2859	2874.7329
7	yes	no	no	1734	2262.8247
8	no	no	no	8284	7543.0578

Homogeneous model:

$$\log(E(\text{Freq})) = \alpha + \beta_1(\text{cigarette}_{\text{yes}}) + \beta_2(\text{mental. issue}_{\text{yes}}) + \beta_3(\text{alcohol}_{\text{yes}}) + \beta_4(\text{cigarette}_{\text{yes}} \times \text{mental. issue}_{\text{yes}}) + \beta_5(\text{cigarette}_{\text{yes}} \times \text{alcohol}_{\text{yes}}) + \beta_6(\text{mental. issue}_{\text{yes}} \times \text{alcohol}_{\text{yes}})$$

This model fits better. Notice the residual deviance (0.15) compared to the degrees of freedom (1).

Our new p-value is **0.5408396**, so our H0 is true and this means that our expected frequencies satisfy our model!

```
Deviance Residuals:
       1         2         3         4         5         6         7         8
-0.16119   0.10129   0.10575  -0.06358   0.24129  -0.11601  -0.14890   0.06822

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                     9.02133    0.01082 833.814  < 2e-16 ***
cigaretteyes                   -1.55957    0.02390 -65.250  < 2e-16 ***
mental.issueyes                -1.06094    0.02033 -52.196  < 2e-16 ***
alcoholyes                      0.14131    0.01458   9.691  < 2e-16 ***
cigaretteyes:mental.issueyes    0.09110    0.02962   3.076   0.0021 **
cigaretteyes:alcoholyes         0.54126    0.02802  19.315  < 2e-16 ***
mental.issueyes:alcoholyes      0.12882    0.02568   5.016 5.27e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1.7444e+04  on 7  degrees of freedom
Residual deviance: 1.4997e-01  on 1  degrees of freedom
AIC: 92.633
```

Once again we can compare the fitted and observed values and see how well they match.

The model seems to fit very well, but to describe the association between the variables we look at the coefficients of the interactions.

For example, let's look at the coefficient exponentiated for "mental.issueyes:alcoholyes".

	cigarette	mental.issue	alcohol	Freq	fitted(mod1)
1	yes	yes	yes	1479	1485.2077
2	no	yes	yes	3760	3753.7923
3	yes	no	yes	3450	3443.7923
4	no	no	yes	9528	9534.2077
5	yes	yes	no	666	659.7923
6	no	yes	no	2859	2865.2077
7	yes	no	no	1734	1740.2077
8	no	no	no	8284	8277.7923

1.14 =

People who have ***manifested mental problems*** in the last 30 days are 1.14 times **more likely to drink alcohol** than people who have not manifested mental problems (not so high).

It’s a good idea to calculate a confidence interval for the odds ratio estimates:

	2.5 %	97.5 %
cigaretteyes:mental.issueyes	1.033471	1.160721
cigaretteyes:alcoholyes	1.626514	1.815373
mental.issueyes:alcoholyes	1.081685	1.196245

As can be seen, the only notable association is that between alcohol and smoking.
We see that the odds of drinking if you smoke cigarettes is at least 1.63 times higher than the odds of drinking if you don't smoke, and vice versa.

Saturated Model:

$$\log(E(\text{Freq})) = \alpha + \beta_1(\text{cigarette}_{\text{yes}}) + \beta_2(\text{mental. issue}_{\text{yes}}) + \beta_3(\text{alcohol}_{\text{yes}}) + \beta_4(\text{cigarette}_{\text{yes}} \times \text{mental. issue}_{\text{yes}}) + \beta_5(\text{cigarette}_{\text{yes}} \times \text{alcohol}_{\text{yes}}) + \beta_6(\text{mental. issue}_{\text{yes}} \times \text{alcohol}_{\text{yes}}) + \beta_7(\text{cigarette}_{\text{yes}} \times \text{mental. issue}_{\text{yes}} \times \text{alcohol}_{\text{yes}})$$

```
Coefficients:
                                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)                           9.02208    0.01099 821.158  < 2e-16 ***
cigaretteyes                         -1.56390    0.02641 -59.219  < 2e-16 ***
mental.issueyes                      -1.06385    0.02169 -49.047  < 2e-16 ***
alcoholyes                            0.13991    0.01502   9.313  < 2e-16 ***
cigaretteyes:mental.issueyes          0.10696    0.05048   2.119   0.0341 *  
cigaretteyes:alcoholyes               0.54803    0.03305  16.583  < 2e-16 ***
mental.issueyes:alcoholyes            0.13404    0.02901   4.621 3.82e-06 ***
cigaretteyes:mental.issueyes:alcoholyes -0.02415    0.06233  -0.387   0.6984    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance:  1.7444e+04  on 7  degrees of freedom
Residual deviance: -1.3083e-12  on 0  degrees of freedom
AIC: 94.483
```

The deviance of this model is basically 0 on 0 degrees of freedom (as many coef as the number of cells in our table) and the **higher-order interaction** is statistically **not significant**!

The fitted counts match the observed counts.
All things being equal, we prefer a simpler model. We usually don't want to finish with a saturated model that perfectly fits our data. We can verify that <u>the homogeneous association model fits just as well as the saturated model</u> (->H_0) by performing a likelihood ratio test.

```
  cigarette mental.issue alcohol Freq fitted(mod2)
1       yes          yes     yes 1479         1479
2        no          yes     yes 3760         3760
3       yes           no     yes 3450         3450
4        no           no     yes 9528         9528
5       yes          yes      no  666          666
6        no          yes      no 2859         2859
7       yes           no      no 1734         1734
8        no           no      no 8284         8284
```

```
Analysis of Deviance Table

Model 1: Freq ~ (cigarette + mental.issue + alcohol)^2
Model 2: Freq ~ cigarette * mental.issue * alcohol
  Resid. Df Resid. Dev Df Deviance
1         1    0.14997
2         0    0.00000  1  0.14997
```

P-value: 0.698564

This says our null hypothesis is true!

Selected Inter Model:

$$\log(E(\text{Freq})) = \alpha + \beta_1(\text{cigarette}_{\text{yes}}) + \beta_2(\text{alcohol}_{\text{yes}}) + \beta_3(\text{mental. issue}_{\text{yes}}) + \beta_4(\text{cigarette}_{\text{yes}} \times \text{alcohol}_{\text{yes}}) + \beta_5(\text{alcohol}_{\text{yes}} \times \text{mental. issue}_{\text{yes}})$$

Here we are fitting a model with the interactions for cigarette and alcohol, and mental.issue and alcohol, but not cigarette and mental.issue. The implication is that mental.issue and cigarette use are independent of one another, controlling for alcohol use.
Then we performed a LRT between the selected interaction model and the homogeneous one:

```
Analysis of Deviance Table

Model 1: Freq ~ (cigarette * alcohol) + (mental.issue * alcohol)
Model 2: Freq ~ (cigarette + mental.issue + alcohol)^2
  Resid. Df Resid. Dev Df Deviance
1         2     9.5467
2         1     0.1500  1   9.3967
```

P-value: 0.0021

The p-value is tiny. The probability of seeing such a change in deviance (9.3967) if the models really were no different is remote. There appears to be good evidence that the homogeneous association model provides a much better fit than the model that assumes conditional independence between cigarette and mental issue.

Dissimilarity Index

	Simple Mod	Homogeneous	Saturated	Specific Inter
Dissimilarity ind	0.046	0.001	0	0.006

Finally we computed the dissimilarity index for all the model and saw that the data follow the model closely even if the models are not perfect!
As said before the homogeneous model is the best even if the dissimilarity index of the saturated is lower (too complex, so we prefer the simpler)

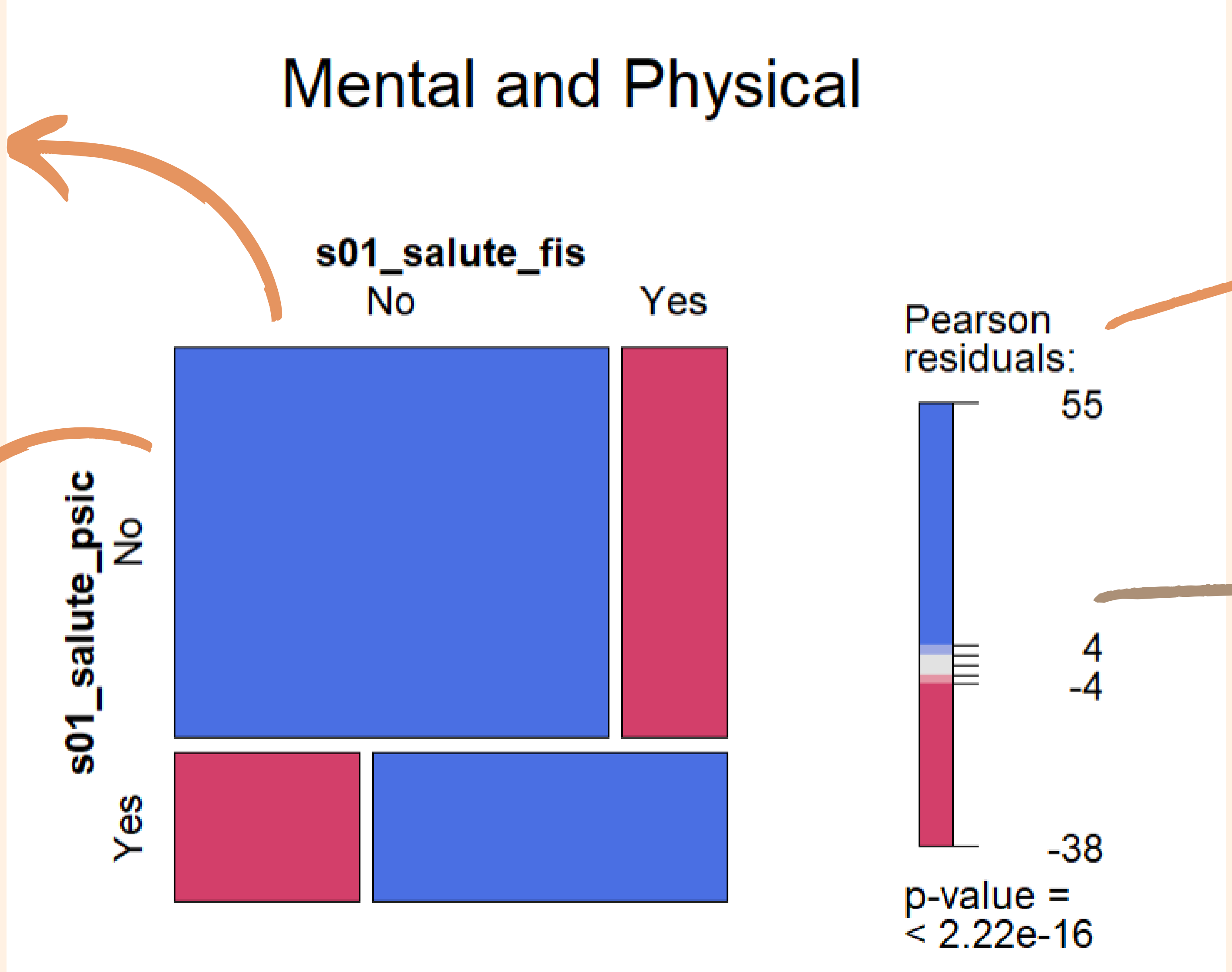
4

Association selection

The mosaic plot is based on conditional probabilities.
It's also called crosstabs or two-way table and it's used to summarize the relationship among several categorical variables.

Proportional to the percentage of being physically healthy and having physical health problem

Proportional to percent of having mental health considering physical health



Pearson residuals show the strength and direction of the association.
A residual greater than 2 or less than -2 represents a significant departure from the independence.

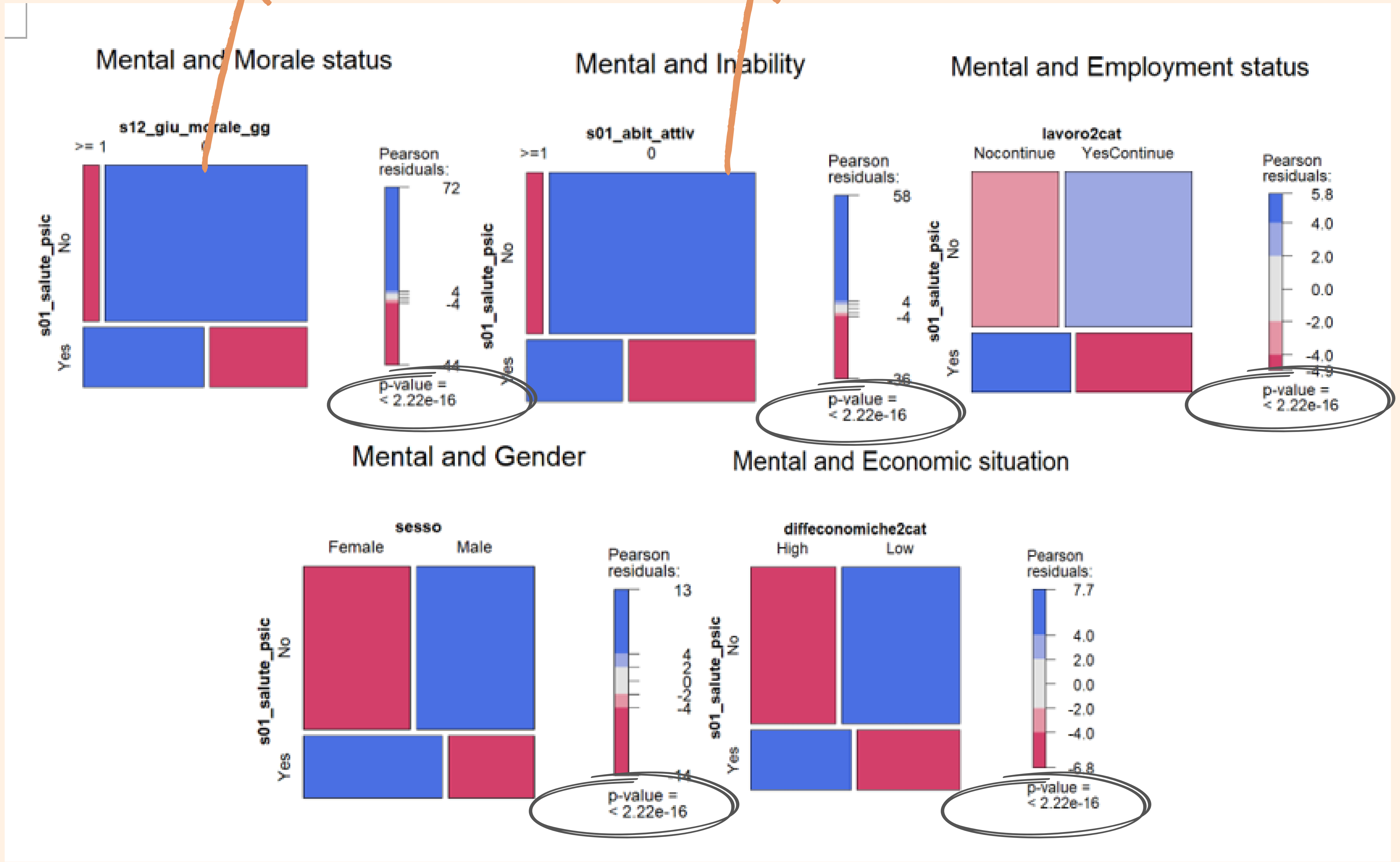
Blue means there are more observations in the cell than would be expected under the null model (independence).
Red means there are fewer observations than would have been expected.

Low p-value: the association between the two variables are statistically significant.

Also for other categorical variables:

93% felt happy

92,4% carried out everyday activities

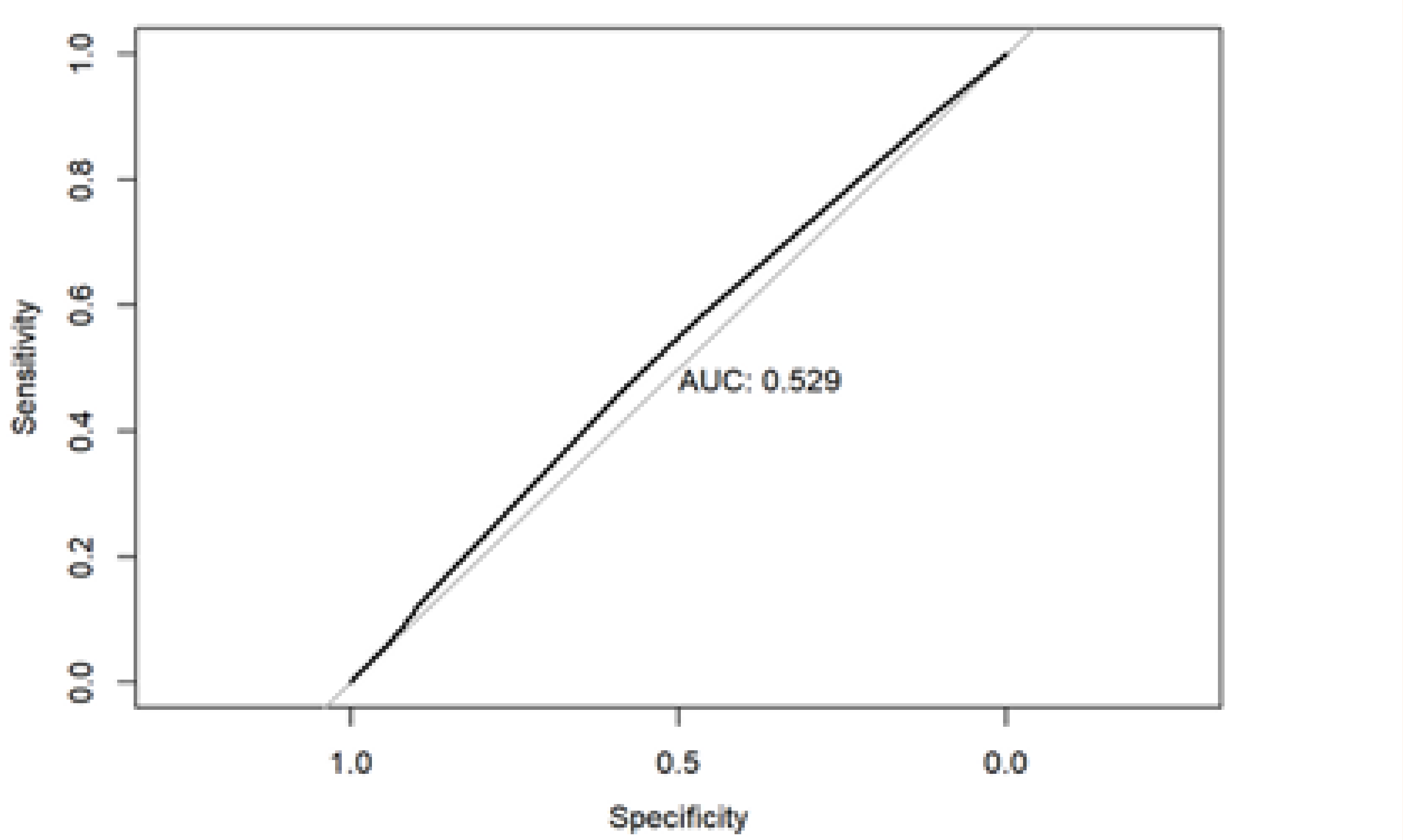


By observing pearson residuals and p-value we can say that moral status, inability, employment status, gender, and economic difficulty are associated to mental health.

5 Logistic Regression 1 → With only variables regarding alcohol and smoking

Unbalanced dataset

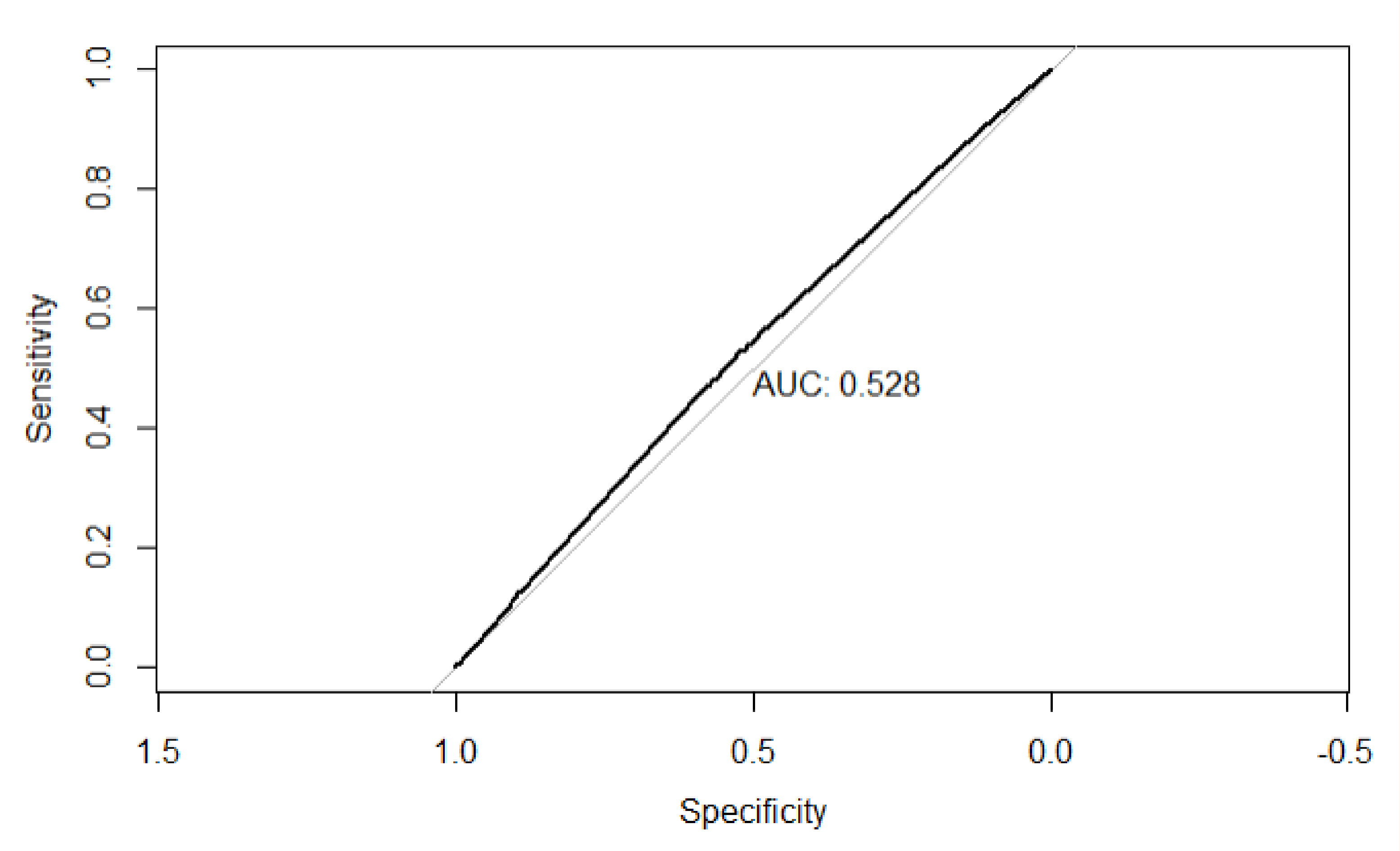
No	Yes
22996	8764



Balanced dataset

No	Yes
15989	15771

We used ovun.sample which random over-samples minority examples



The area under the two ROC curves are really similar, the balanced one does not significantly improve the performance of the model, so we will carry out our analysis with unbalanced dataset

5 Logistic Regression

```
Call:
glm(formula = s01_salute_psic ~ s03_fumo_att + s03_fumo_quanto +
    s05_alcool_gg, family = "binomial", data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8726  -0.8341  -0.7704   1.5204   1.6628

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.0626439  0.0203375 -52.250  < 2e-16 ***
s03_fumo_attYes  0.0985869  0.0483806   2.038   0.0416 *
s03_fumo_quanto  0.0001496  0.0033222   0.045   0.9641
s05_alcool_gg1   0.1857532  0.0274230   6.774 1.26e-11 ***
s05_alcool_gg2  -0.0307845  0.0380437  -0.809   0.4184
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 37419  on 31759  degrees of freedom
Residual deviance: 37346  on 31755  degrees of freedom
AIC: 37356

Number of Fisher Scoring iterations: 4
```

	OR	2.5 %	97.5 %
(Intercept)	0.3455410	0.3319988	0.3595508
s03_fumo_attYes	1.1036103	1.0036199	1.2132250
s03_fumo_quanto	1.0001496	0.9936247	1.0066528
s05_alcool_gg1	1.2041251	1.1411292	1.2706352
s05_alcool_gg2	0.9696845	0.8998212	1.0445423

s03_fumo_attYes and s05_alcool_gg1 are statistically significant.

- All else being equal, people who currently smoke cigarettes have 10% more odds of suffering psychological disturbances than people who do not smoke.
- People who consume alcohol 1 to 15 times in the last 30 days have 20.4% more odds of suffering psychological health disturbances than people didn't drink alcohol lately

Significance of the overall model

```r
#p_value
with(modellog, pchisq(null.deviance - deviance, df.null - df.residual, lower.tail = FALSE))
```

```
## [1] 7.475762e-15
```

Chi-square: 72.28722
Degree of freedom: 4
P-value < 0.001

Our model as a whole fits significantly better than an empty model.

5 Logistic Regression 2

The variables that are associated with s01_salute_psic in the mosaic plots
+
the variables regarding alcohol and smoking

```
Call:
glm(formula = s01_salute_psic ~ s01_salute_fis + s12_giu_morale_gg +
    s05_alcool_gg + s01_abit_attiv + sesso + diffeconomiche2cat +
    lavoro2cat + s03_fumo_att + s03_fumo_quanto + s07_eta, family = "binomial",
    data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3990  -0.4917  -0.4071   0.4199   2.4627

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)            0.673965   0.079847   8.441  < 2e-16 ***
s01_salute_fisYes      1.357824   0.036541  37.159  < 2e-16 ***
s12_giu_morale_gg0    -2.344301   0.036599 -64.054  < 2e-16 ***
s05_alcool_gg1         0.342574   0.036828   9.302  < 2e-16 ***
s05_alcool_gg2         0.209981   0.050468   4.161 3.17e-05 ***
s01_abit_attiv0       -0.925661   0.043335 -21.361  < 2e-16 ***
sessoMale             -0.476728   0.034273 -13.910  < 2e-16 ***
diffeconomiche2catLow -0.009959   0.033383  -0.298  0.76547
lavoro2catYesContinue  0.103769   0.033786   3.071  0.00213 **
s03_fumo_attYes        0.048729   0.061723   0.789  0.42983
s03_fumo_quanto       -0.002025   0.004282  -0.473  0.63621
s07_eta                0.005249   0.001182   4.441 8.93e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 37419  on 31759  degrees of freedom
Residual deviance: 25325  on 31748  degrees of freedom
AIC: 25349

Number of Fisher Scoring iterations: 5
```

+ Positive estimate coefficients

s01_salute_fisYes
(have physical health issue)
s05_alcool_gg1
(1 to 15 times in 30 days)
s05_alcool_gg2
(16 to 30 times in 30 days)
lavoro2catYesContinue
(work continuously)
S07_eta

– Negative estimate coefficients

s13_giu_morale_gg0
(zero day with sad mood)
s01_abit_attiv0
(everyday activities not be influenced by bad mood)
sessoMale

Some differences from the first model:
- positive s05_alcool_gg2
- s03_fumo_attYes not significant

There are variables that are strongly associated both with them and mental health issue

? **Diffeconomiche2cat not significant**

5 Logistic Regression 3

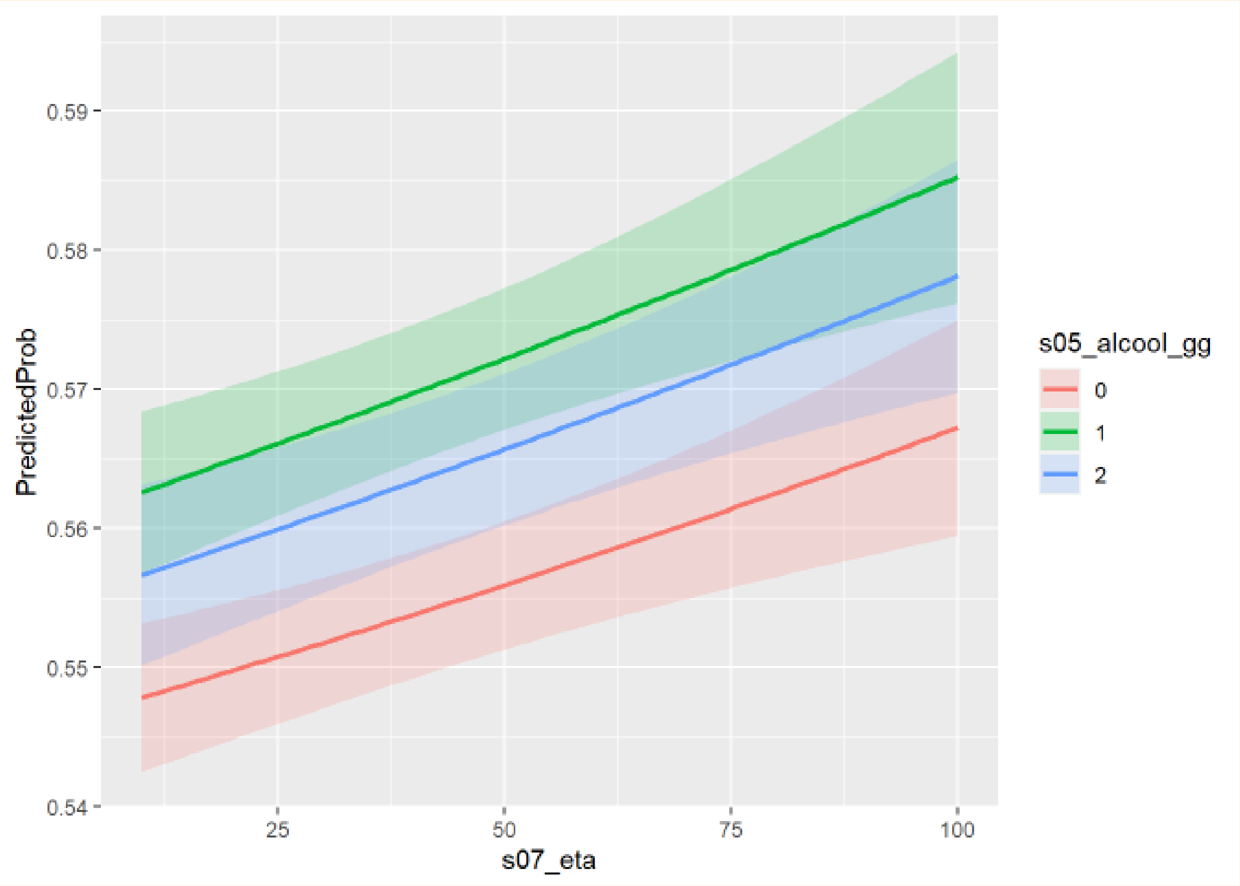
Adding an interaction between s01_salute_fis and diffeconomiche2cat

```
Call:
glm(formula = s01_salute_psic ~ s12_giu_morale_gg + s05_alcool_gg + 
    s01_abit_attiv + lavoro2cat + sesso + s07_eta + s01_salute_fis * 
    diffeconomiche2cat, family = "binomial", data = df)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.4238  -0.4882  -0.4081   0.4067   2.4938  

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)    
(Intercept)                       0.738759   0.082598   8.944  < 2e-16 ***
s12_giu_morale_gg0               -2.344051   0.036642 -63.972  < 2e-16 ***
s05_alcool_gg1                    0.347852   0.036759   9.463  < 2e-16 ***
s05_alcool_gg2                    0.215449   0.050267   4.286 1.82e-05 ***
s01_abit_attiv0                  -0.915033   0.043410 -21.079  < 2e-16 ***
lavoro2catYesContinue             0.103868   0.033803   3.073 0.002121 ** 
sessoMale                        -0.480536   0.034101 -14.092  < 2e-16 ***
s07_eta                           0.004954   0.001176   4.213 2.52e-05 ***
s01_salute_fisYes                 1.207764   0.045699  26.429  < 2e-16 ***
diffeconomiche2catHigh           -0.152091   0.044882  -3.389 0.000702 ***
s01_salute_fisYes:diffeconomiche2catHigh 0.354310 0.065061 5.446 5.16e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 37419  on 31759  degrees of freedom
Residual deviance: 25296  on 31749  degrees of freedom
AIC: 25318

Number of Fisher Scoring iterations: 5
```

All variables are statistically significant

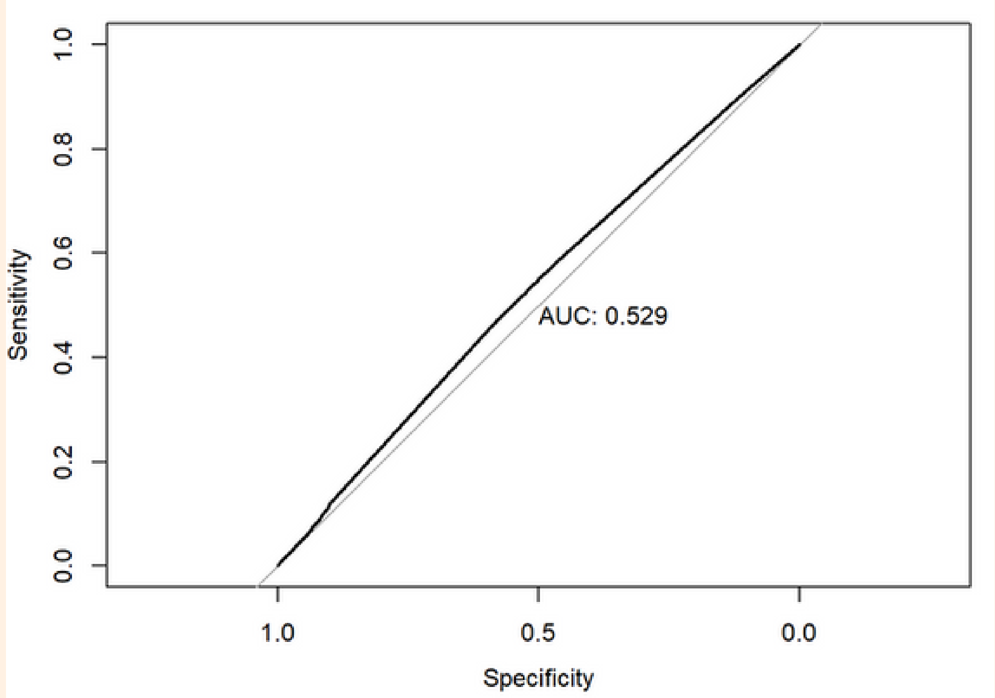
The interaction between the presence of physical health disturbance (s01_salute_fisYes) and high level of economic difficulties is **statistically significant**. From the positive sign we understood that, all else being equal, people having a physical health issue and high level of economic difficulty are more likely to suffer also psychological problem than people without physical health problem and economic difficulty.

How the predicted probabilities of having mental health issues will be with the increase of the age at each level of alcohol consumption in the last 30 days?

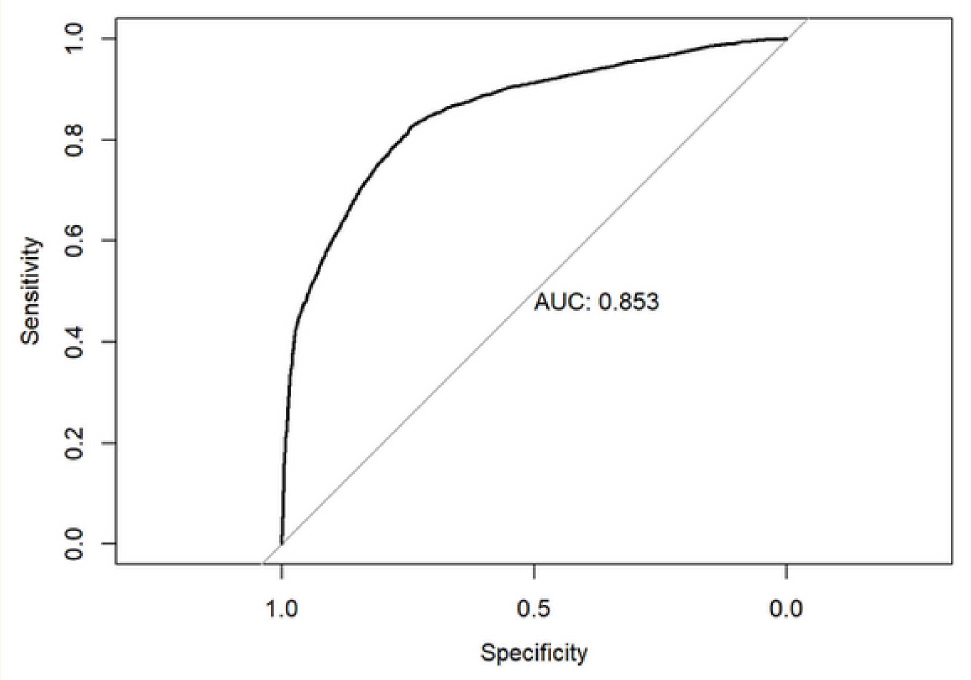


5 Logistic Regression: model selection

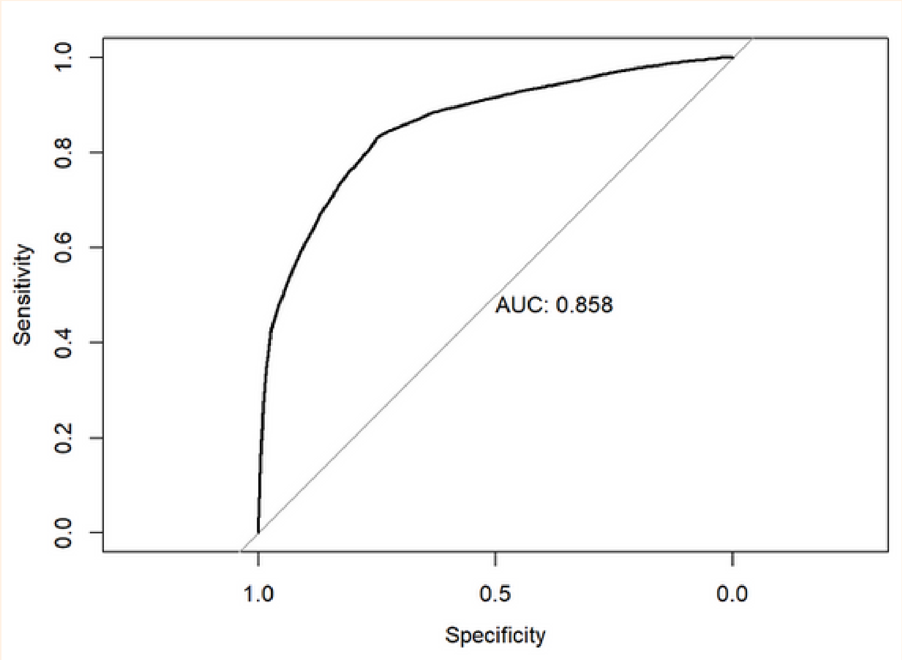
Logistic regression 1



Logistic regression 2



Logistic regression 3




```
Model selection based on AICc:

                      K     AICc Delta_AICc AICcWt Cum.Wt        LL
modellog interaction 11 25318.01       0.00      1      1 -12648.00
modellog selected    12 25349.08      31.07      0      1 -12662.54
modellog pers info    9 36453.05   11135.04      0      1 -18217.52
modellog              5 37356.22   12038.21      0      1 -18673.11
```

K: The number of parameters in the model.

AIC: The information score of the model. The smaller the AIC value, the better the model fit.

LL: Log-likelihood. This is the value describing how likely the model is, given the data. The AIC score is calculated from the LL and K.

From this table we can see that the best model is the **modellog selected with interaction (logistic regression 3)** as it has the lowest AIC.

$$\text{s01\_salute\_psic}_i \sim \text{Binomial}(n = 1, \text{prob}_{\text{s01\_salute\_psic}=1} = \widehat{P})$$

$$\log\left[\frac{\hat{P}}{1-\hat{P}}\right] = \alpha_{j[i]} + \beta_1(\text{s07\_eta}) + \beta_2(\text{s03\_fumo\_att}_{\text{Yes}}) + \beta_3(\text{s05\_alcool\_gg}_1) + \beta_4(\text{s05\_alcool\_gg}_2) + \beta_5(\text{sesso}_{\text{Male}})$$

$$\alpha_j \sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{ for regione j} = 1, \dots, \text{J}$$

6 Mixed Model

	s 01 salute psic		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.75	0.60 – 0.93	**0.010**
s07 eta	1.01	1.01 – 1.01	**<0.001**
s05 alcool gg [1]	1.28	1.21 – 1.35	**<0.001**
s05 alcool gg [2]	1.06	0.98 – 1.14	0.124
sesso [Male]	0.51	0.48 – 0.53	**<0.001**
s03 fumo quanto	1.02	1.01 – 1.02	**<0.001**
Random Effects			
σ^2	3.29		
τ_{00} regione	0.21		
ICC	0.06		
N regione	20		
Observations	31760		

Fixed EFFECTS:

The coefficients' interpretation it's the same of the Logistic Regression model, HOWEVER we're considering the effects of the covariates on the **same cluster**

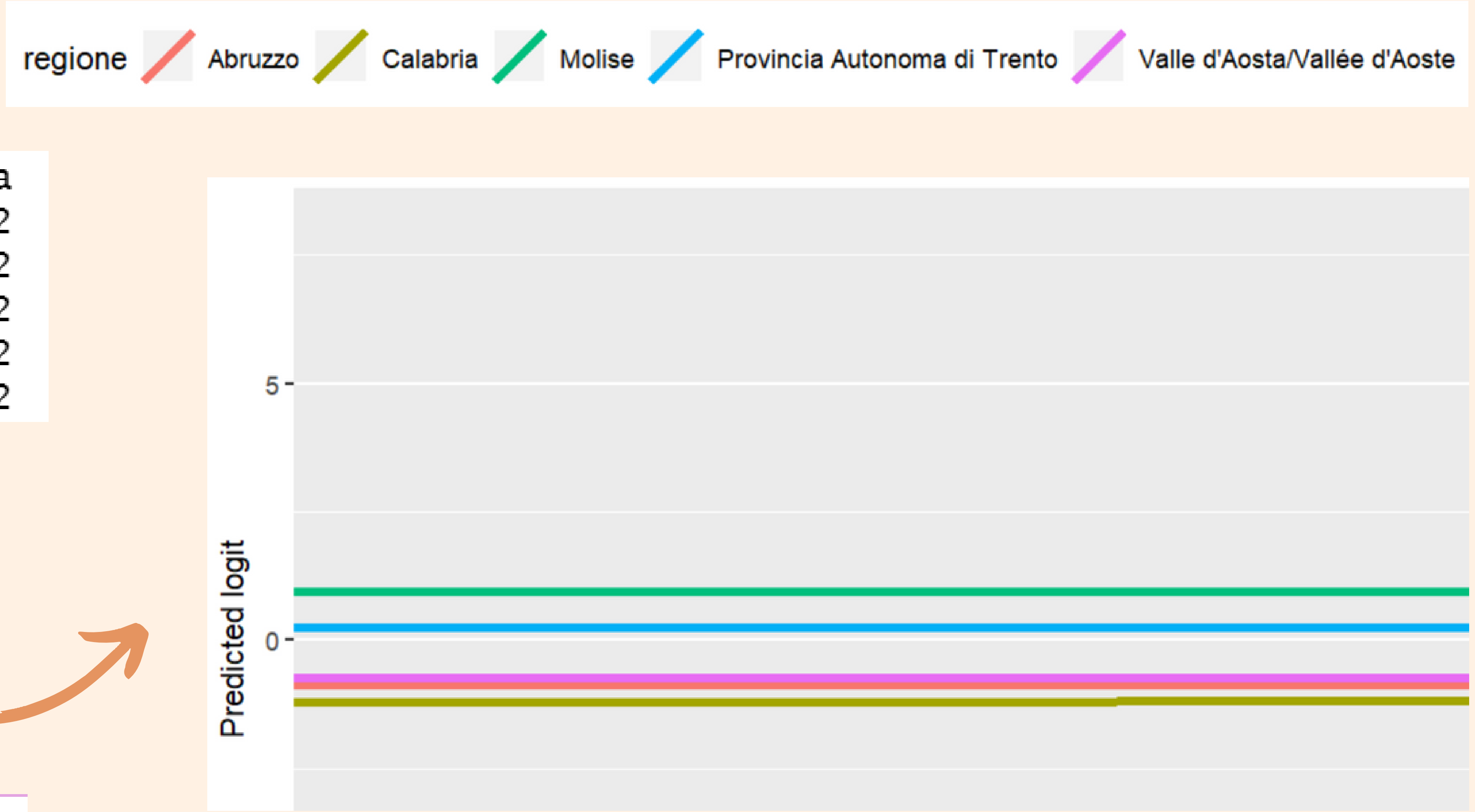
The odds to have mental health problems for a female being classified in the second level of alcool usage is 1.28 times higher than females with no drinking problems, considering the **SAME REGION**

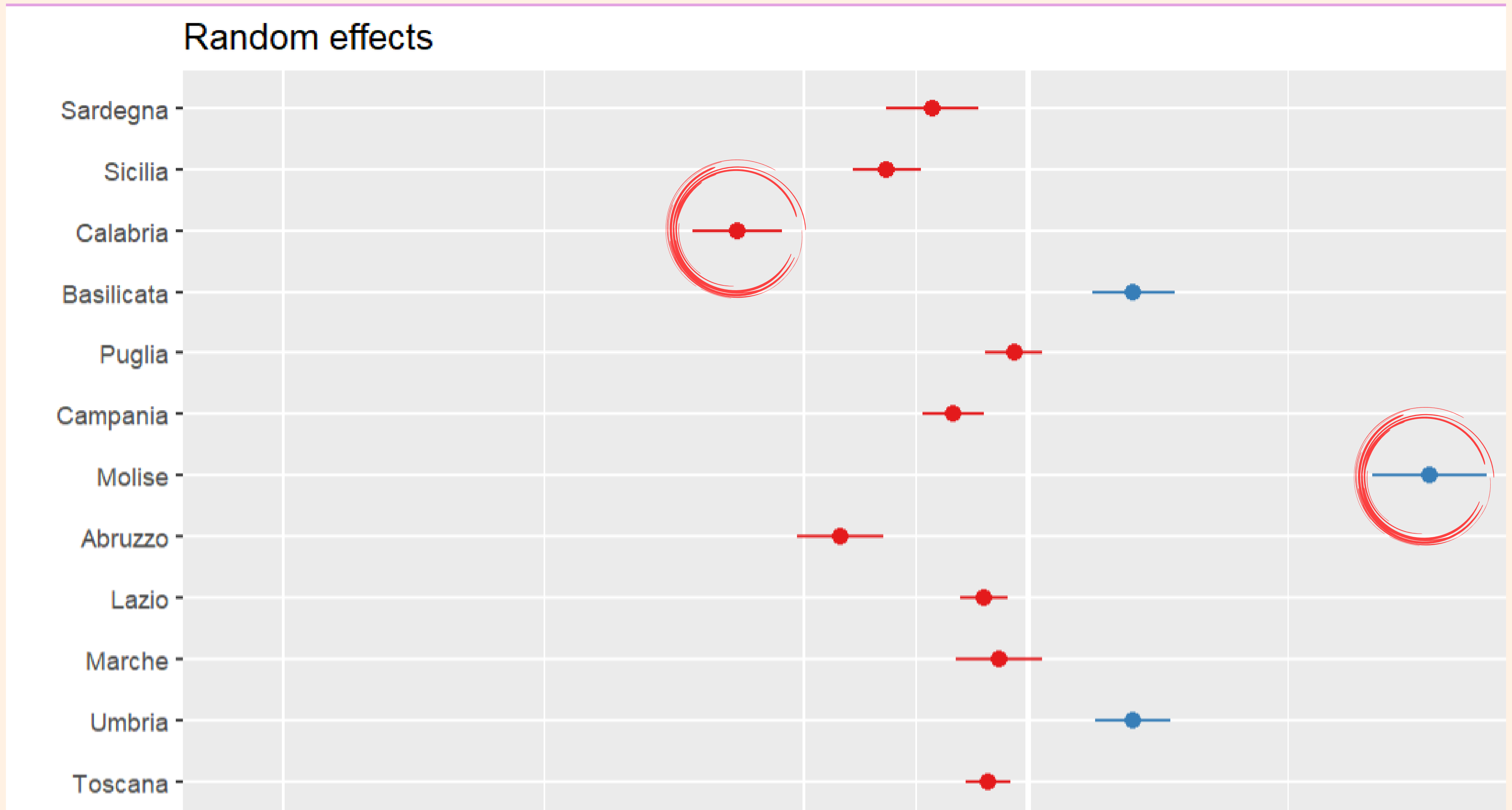
ICC: only 6% of the variance in the model is explained by the Random effect!

Random EFFECTS:

```
                                        (Intercept)     s07_eta
Piemonte                                -0.34060731 0.009269022
Valle d'Aosta/Vallée d'Aoste            -0.74206065 0.009269022
Provincia Autonoma di Trento             0.23482077 0.009269022
Provincia Autonoma di Bolzano/Bozen     -0.48098531 0.009269022
Veneto                                   0.01132592 0.009269022
```

Very little adjustment for the log-odds of having mental health explained by the Age coefficient!






```
Within-Group Variance              3.29 (1.81)
Between-Group Variance
  Random Intercept (regione)       0.21 (0.46)
N (groups per factor)
  regione                            20
Observations                      31760
```

Could it be a matter of different number of observations for each region in the dataset?

Conclusions

- Regarding the variables mental health, alcohol use and smoker status the homogenous model showed a better fit than the one that supposed independence between them . People with medium level of consumption (1 to 15 times in 30 days) are the most likely to suffer mental health issues.

- There is a significant association between economic situation, pysical health and mental health: the presence of physical health issue combined with bad economic situation increases the likelihood of having psychological disturbances.

- We were expecting an higher variability between regions regarding the mental health problem but from the results given by the random intercept model we really can't validate our assumption

"When in a state of good mental health, a person has a general positive outlook, can accomplish daily tasks, maintain relationships and engage in meaningful recreation, this includes a sense of balance and empowerment to set boundaries and address life and work goals, step by step."

Dr. Darleen Dempster, member in the clinical mental health counseling program of the Southern New Hampshire University