

EXERCISE 3**A)**

The first model ($\text{points} = \beta_0 + \beta_1 \cdot \text{price} + \beta_2 \cdot \text{alcohol} + \epsilon$) has $R^2 = 0.4706$ and $R^2_{\text{adjust}} = 0.4686$,

While the second one has $R^2 = 0.5256$ and $R^2_{\text{adjust}} = 0.5239$. So, it seems that the second model is better.

For the rest of the diagnostics:

- they seem both quite normal looking at the qqplot.
- They are both not affected by collinearity looking at the vif results.
- The first model has a worse plot residual vs fitted and there are some points that have a leverage effect, while the second model doesn't have these problems.
- They are both homoscedastic.

I would choose the second one.

B)

The coefficients are:

(Intercept) log(price) alcohol

4.3758959052 0.0282611175 0.0008299285

While the value of the estimator for the variance is:

0.0003071755

C)

Even if we don't know which one of the two (doing only this test), they have an effect on the model, so cannot discard both.

D)

By looking at the one-at-the-time tests of the summary, we can see that the p-value of alcohol is 0.205, so we remove it.

Model: $\log(\text{points}) \sim \log(\text{price})$

The coefficients are:

(Intercept) log(price)

4.38571235 0.02825545

While the value of the estimator for the variance is:

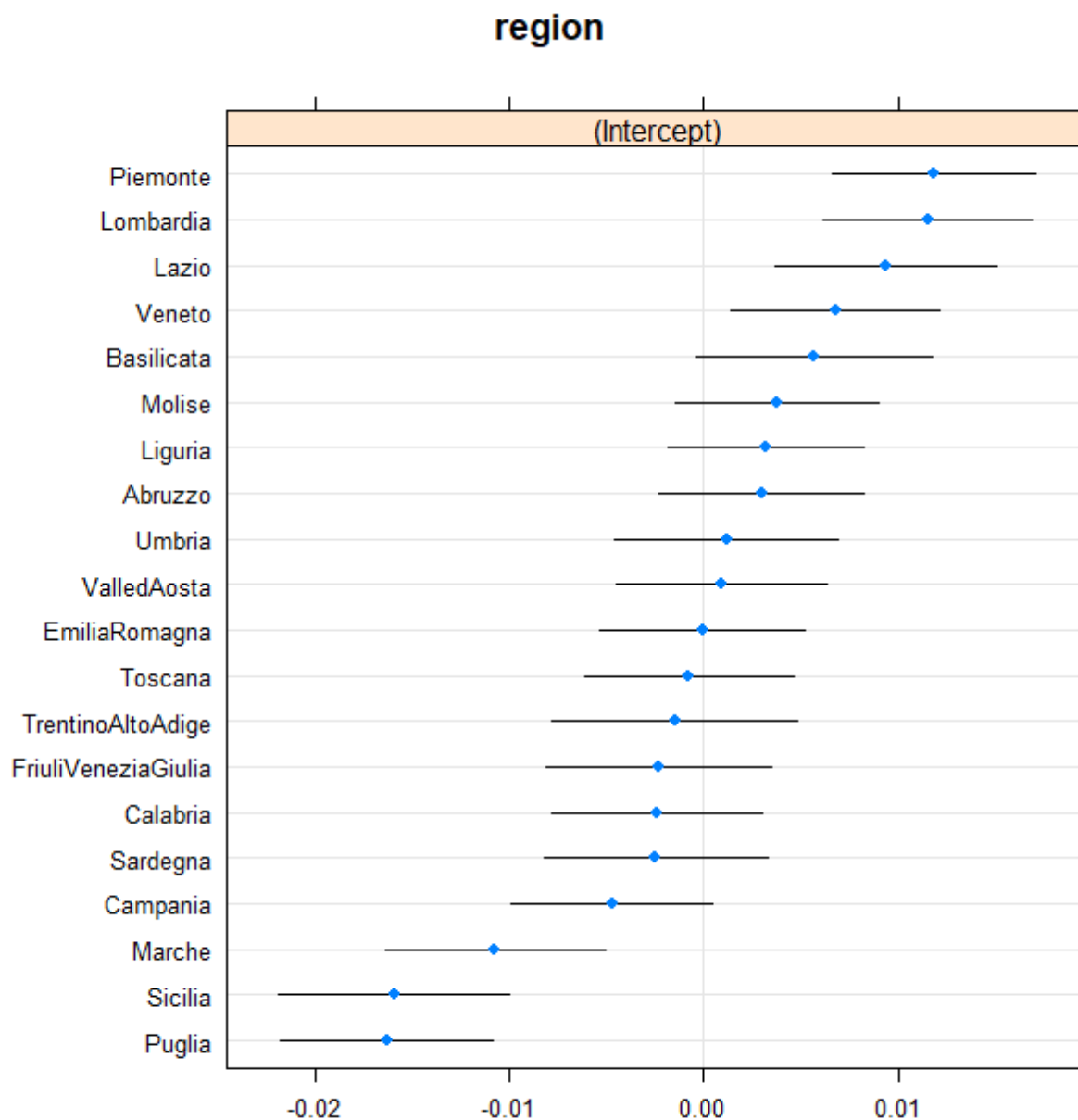
0.0003075201

E)

The new model is: $\log(\text{points}) \sim \log(\text{price}) + (1 | \text{region})$, which takes into account the regions on the intercept.

The PRVE index is 0.2198255.

F)



From the dot plot, it seems that the province with the lowest number of points is Puglia.