I have added PA and alcohol to the diet score. Variable representing PA is g6* and variable for alcohol is alkosum1*.

Since we have been observing a significant negative association of alcohol with bmi, I had checked the relationship of alcohol and socioeconomic status, represented with the variable utbild for education. Reasons for this are, education level can be a good representation of the socioeconomic status, which has been shown to be negatively associated with bmi, while being positively associated with alcohol consumption, therefor the observed negative association of alcohol and bmi could be due to confounding of the socioeconomic status.

I had checked the association with bmi results for alcohol, when adding utbild and when not.
I had checked the association with bmi results for alcohol, stratified in levels of utbild.
I had checked spearmans rank correlation results for alkosum1 and utbild and utbild and bmi.
I had obtained the residuals from modeling alkosum1 with utbild and check the association with bmi results with the residuals.

Tests did not confirm confounding with utbild.

I had added PA, alcohol and socioeconomic status to the diet score and checked the predictive ability of several models, including or excluding any of these three variables.

Best results were when adding all three variables to the diet score, although difference in AUC was non-significant for some variable exclusions.

*How often have you been training or exercising in exercise outfit during the last three months with the purpose to enhance your condition and/or to feel good?;1 = Never; 2 = Every now and then - not regularly;3 = 1-2 times/week; 4 = 2-3 times/week;5 = More than 3 times/week
*Alcohol (g/day)

**- Associating alcohol with bmi, adjusting for basic covariates without utbild:**

```
associations<-lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+alkosum1)
```

alkosum1:
beta : -1.331e-02, p-value :  < 2e-16, vif : 1.109547

**- Associating alcohol with bmi, adjusting for basic covariates with utbild:**

```
associations<-
lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+alkosum1+utbild)
```

alkosum1:
beta : -9.848e-03, p-value :  < 2e-16, vif : 1.141107

utbild:
beta : -1.042e-01, p-value :  < 2e-16, vif : 1.331501

The effect size of alcohol did not change much when adding utbild and vif did not increase by much and is still very small.

**- Associating alcohol with bmi, stratified in levels of utbild:**

```
associations<-
lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+alkosum1,data=VIP_data_inde
pendant[utbild==1,])
```

alkosum1:
beta : -8.759e-03, p-value :  0.000486

```
associations<-
lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+alkosum1,data=VIP_data_inde
pendant[utbild==2,])
```

alkosum1:
beta : -9.854e-03, p-value :  9.93e-05

```
associations<-
lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+alkosum1,data=VIP_data_inde
pendant[utbild==3,])
```

alkosum1:
beta : -1.033e-02, p-value :  2.15e-09

```
associations<-
lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+alkosum1,data=VIP_data_inde
pendant[utbild==4,])
```

alkosum1:
beta : -0.0108333, p-value :  5.98e-13

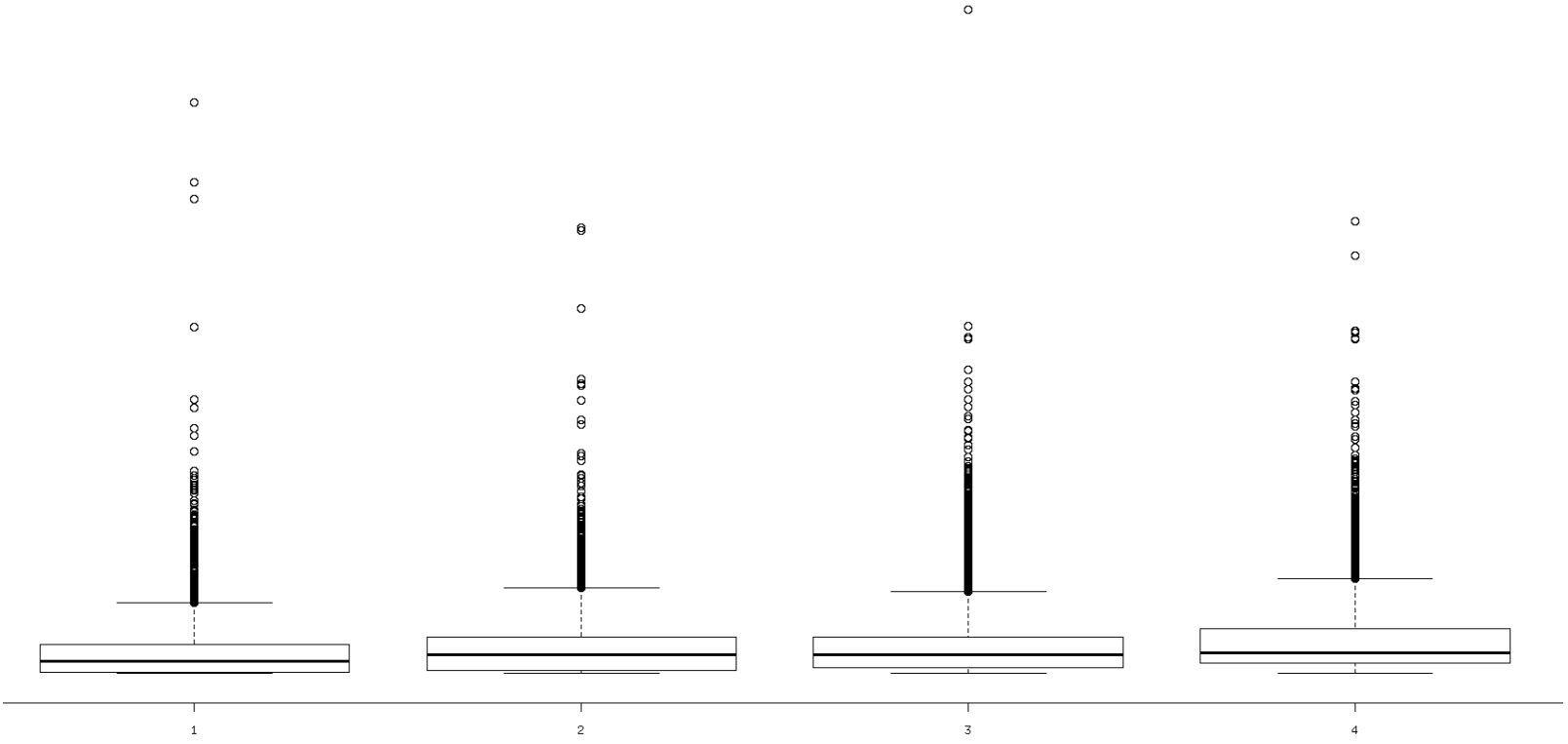
In each level alcohol is still significantly negatively associated with bmi. Although a pattern of increasing effect size and decreasing p-value can be observed, this could be due to sample size of the utbild level.

**- Correlating alcohol with socioeconomic status:**

```
rcorr(alkosum1,utbild,type="spearman")
```

R : 0.13, p-value : < 2e-16

boxplot, alcohol per utbild level:

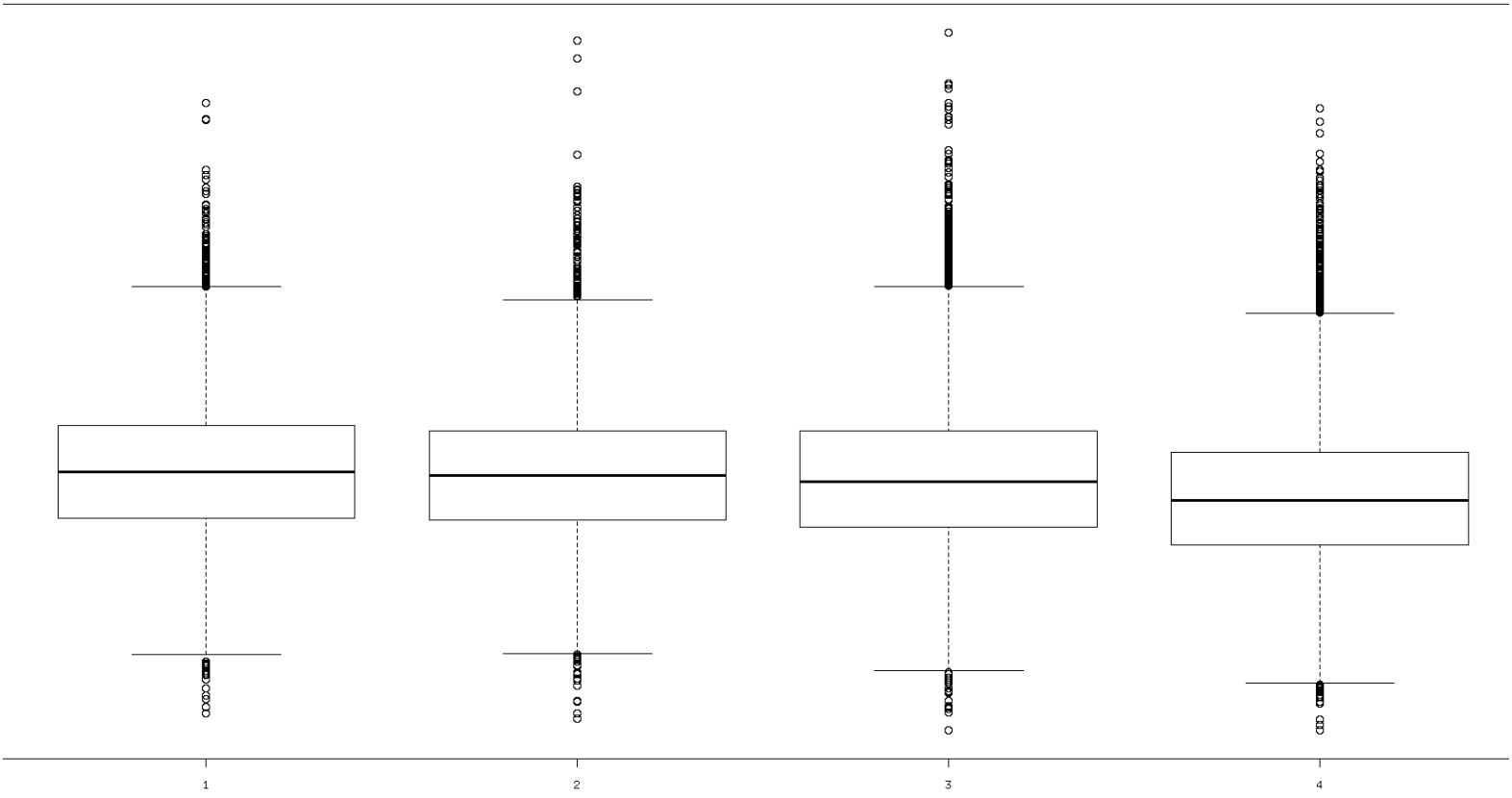


**- Correlating bmi with socioeconomic status:**

```
rcorr(bmi_norm_sd,utbild,type="spearman")
```

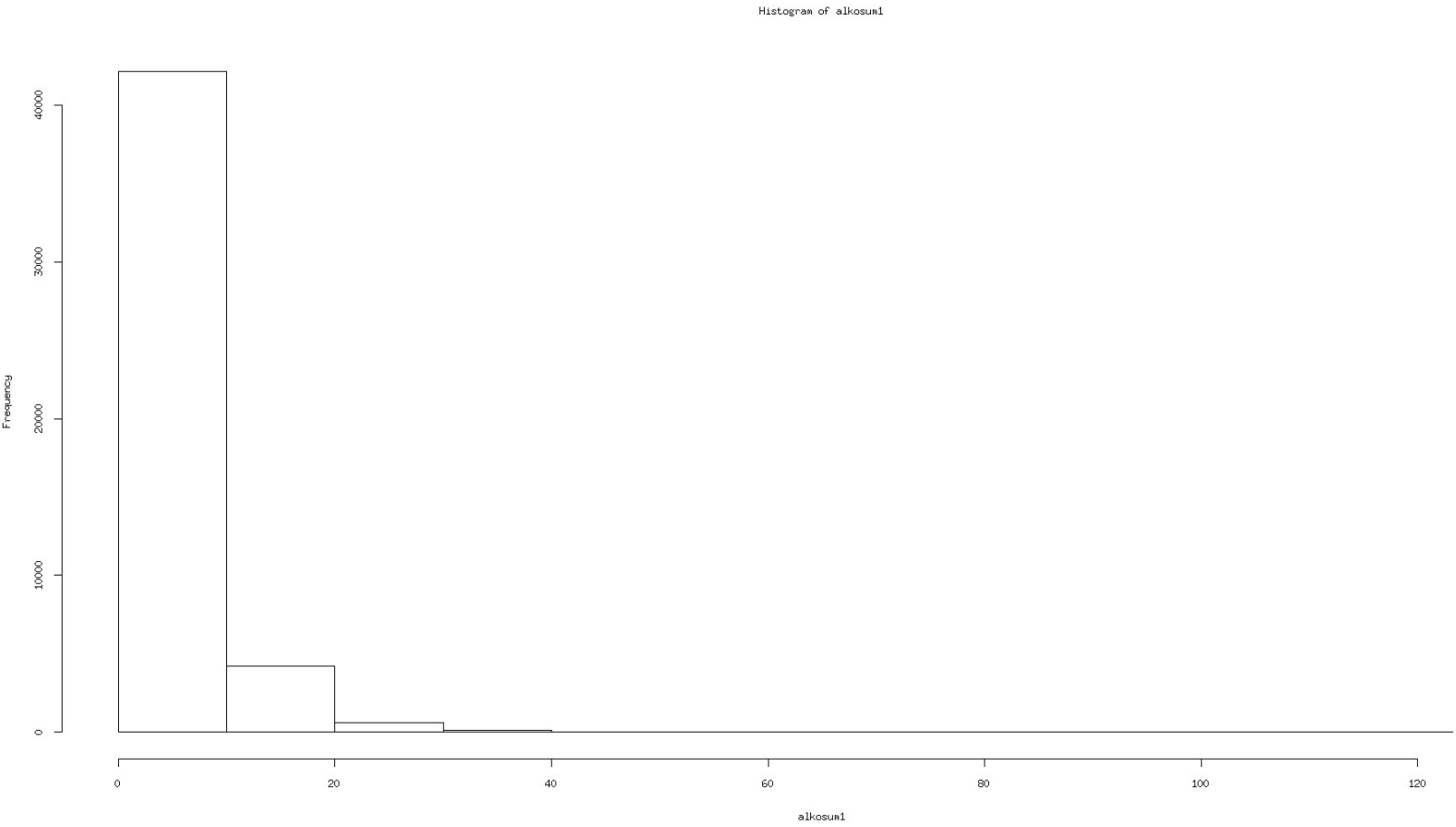
R : -0.14, p-value : < 2e-16

boxplot, bmi per utbild level:



**- Obtaining the residuals of model of utbild for alcohol:**

Alcohol has a very messy, zero inflated distribution:




Hard to transform it to normal, so I multiplied it with 1000 to get mg per day, round it up and modeled it as counts with the zero inflated negative binomial distribution:

```
alkosum1_mg<-round(1000*alkosum1)
model_znb<-zeroinfl(alkosum1_mg~utbild, dist="negbin")

alkosum1_mg_residuals[!is.na(VIP_data_independant$alkosum1) & !
is.na(VIP_data_independant$utbild)]<-model_znb$residuals
```

Associating residuals with bmi, standardizing in order to be comparable with standardized alkosum1 and standardized alkosum1 in model with utbild:

```
associations<-
lm(bmi_norm_sd~age+agesq+gender_factor+year+ffq_factor+scale(alkosum1_mg_residuals
))
```

scale(alkosum1_mg_residuals):
beta : -5.561e-02, p-value : < 2e-16

scale(alkosum1):
beta : -6.686e-02, p-value : < 2e-16

scale(alkosum1) in model with utbild:
beta : -4.948e-02, p-value : < 2e-16

Not much difference between the effect sizes, direction stays the same.

**- Adding PA, alcohol, socioeconomic status to diet score, modeling bmi categories:**

Results here are focused on the inclusion/exclusion of any of these three variables, for comparison with the basic model and model with diet only, check previous update results, would have to do a significance test against the null model and the null model with diet to confirm improvement, although I am sure the difference in AUC will be significant.

Visit 1:

| | AUC | $R^2$ |
|---|---|---|
| PA + alcohol | 0.6553299 s. | 0.0498074 |
| alcohol + socioeconomic status | 0.6530721 s. | 0.04863714 |
| socioeconomic status + PA | 0.6576916 n.s. | 0.05162163 |
| PA + alcohol + socioeconomic status | 0.6580812 | 0.05211517 |

Visit 2:

| | AUC | $R^2$ |
|---|---|---|
| PA + alcohol | 0.6499973 s. | 0.04580637 |
| alcohol + socioeconomic status | 0.6456903 n.s. | 0.04210992 |
| socioeconomic status + PA | 0.6519583 s. | 0.04672227 |
| PA + alcohol + socioeconomic status | 0.6524016 | 0.04756398 |

p-value indication for difference significance when compared to the model including all variables. p-value was checked for each of the three classes, n.s. is reported if in at least one class out of three p-value turns out to be n.s. and s. is reported if in all three classes p-value is s.

Tables when modeling bmi with diet, PA, alcohol and socioeconomic status:

Green=”right”, Red=”wrong” = “our interest”

visit 1:

| true / predicted | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 14090 | 7122 | 1771 |
| 1 | 2688 | 3632 | 809 |
| 2 | 5 | 0 | 1 |

visit 2:

| true / predicted | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 7998 | 5144 | 1899 |
| 1 | 4818 | 7524 | 2523 |
| 2 | 46 | 77 | 89 |