

New results, by excluding the extremes of the variables. Numbers for when not excluding extremes are in previous documents.

For each variable, the extremes are removed:

- by making a transformed variable, by log transforming and standardizing
- then in the original variable, set NA to those values that are below -4 and above 4 in the transformed variable
- log transform and standardize the clipped variables

This kind of exclusion has been criticized, Yan send a paper on this, where there is a different proposed way, I can try with that kind of exclusion as well.

Beta coefficients are calculated as before and for several different models.

Mainly we are planning to work with variables expressed in % of total energy intake (except for fiber and salt) and categorized based on guidelines, but nevertheless, I had put together details of several associations, just to get a clearer picture.

I checked the association with bmi, using:

- raw continuous variables(log transformed and standardized), fitted together or separately,
- continuous variables expressed in % of total energy intake, fitted together or separately,
- variables categorized based on guidelines, also fitted together or separately (most of variables have to be expressed in % of TEI to be able to apply recommendations, except for fiber and salt).

Regarding the variables that have only two categories, we discussed with Alaitz and decided it is best and used in other papers, that we set the highest score, so just 1 and 2.

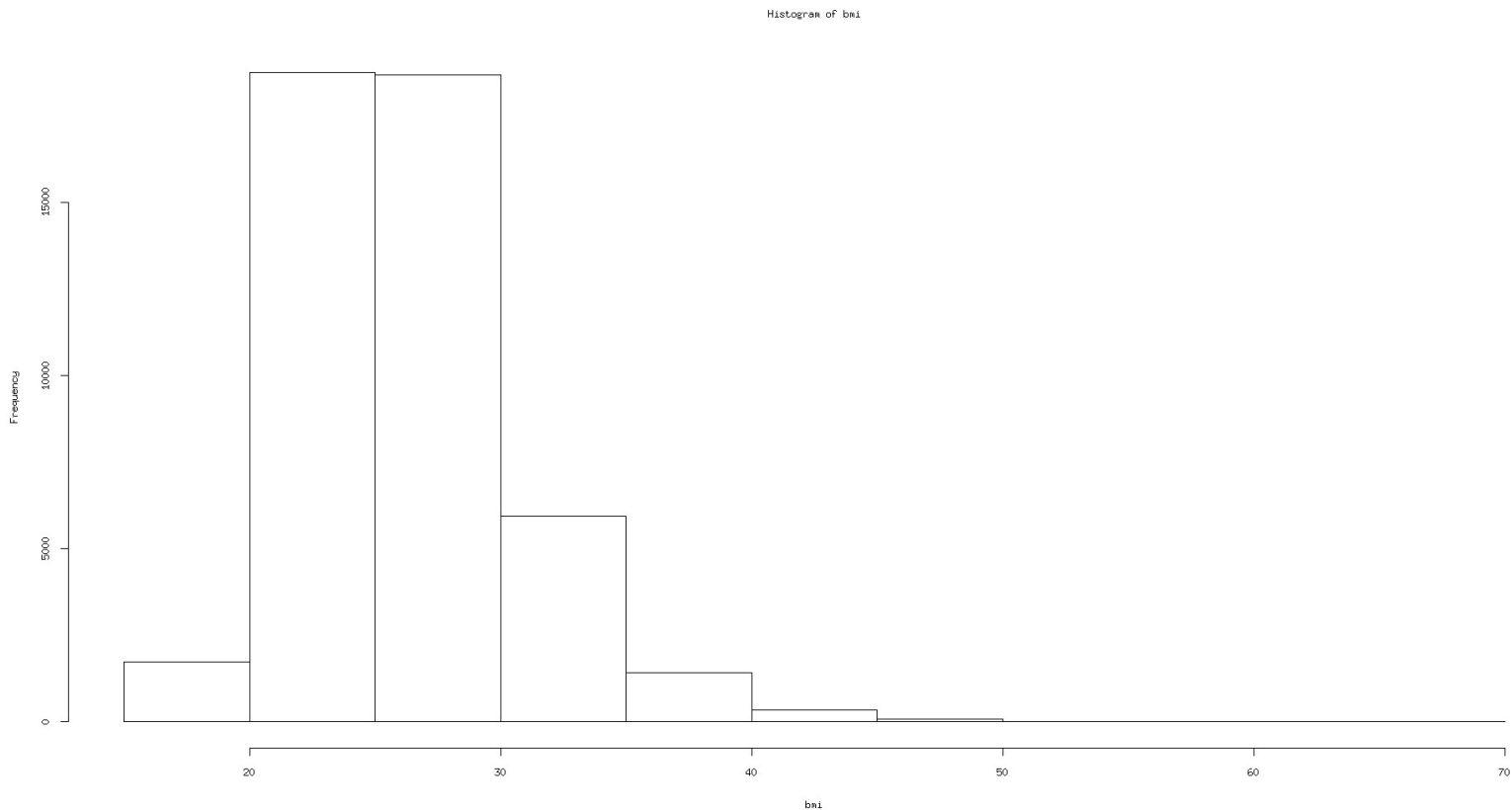
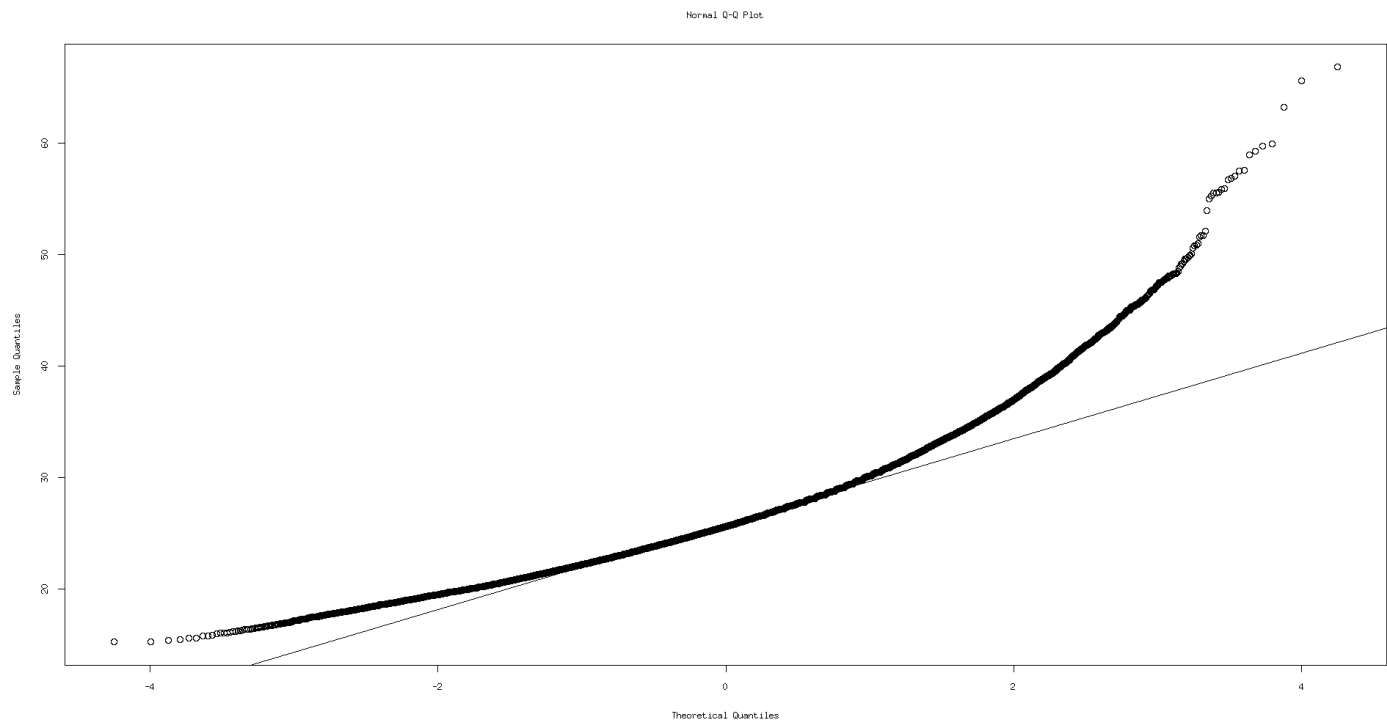
Same as before, I have marked some numbers/variables:

	non-significant
	wrong direction of association
	direction of association different in other data or model

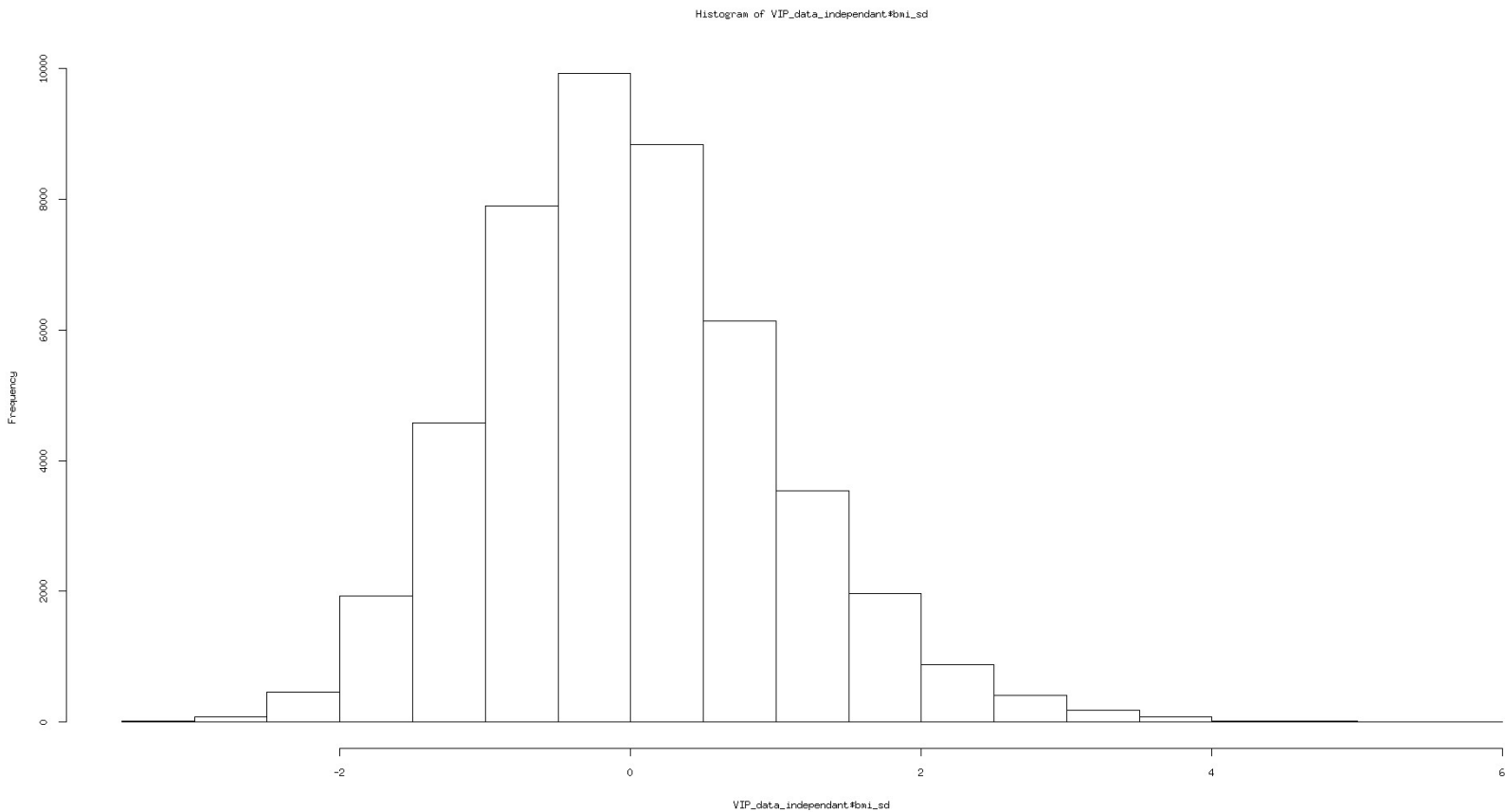
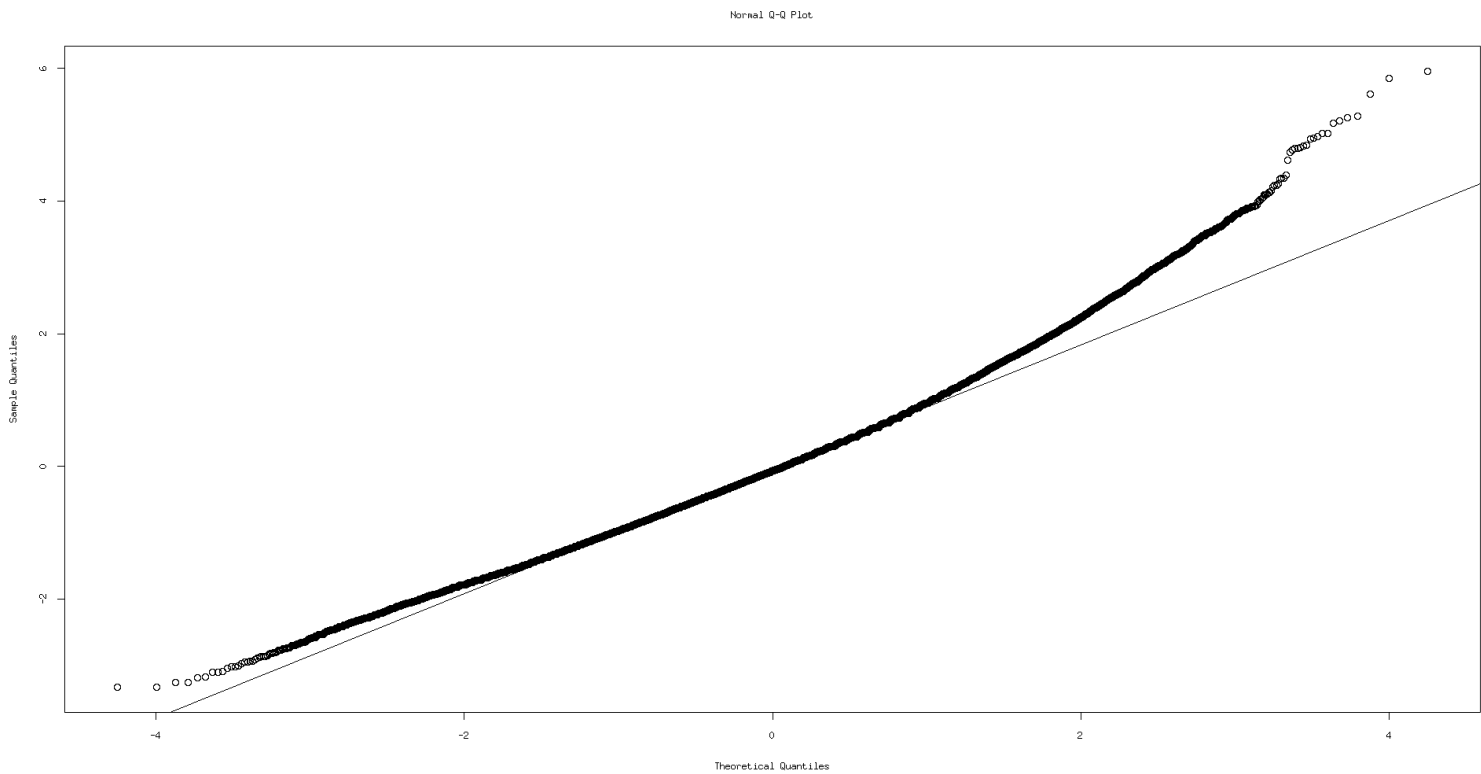
I had added the vif value for each of the nutrient variables. Might be a problem with looking just at vif, since it will not explain the relationship between nutrients and bmi, but only between nutrients, it is only providing us with the information on how much of the variation in one independent variable can be explained by the variation in others, similar as PCA, where only the variability in the independent variables is considered. Although it shows a lot of information, we might need an evaluation of how much does one independent variable influence on the effect of another independent variable on bmi. Sugar always has the lowest vif, always below 4, so the main reason why we went to look at vif is not solved, I made a specific analysis just for sugar, with carbs and fiber, but it seems that the “real” model is far more complicated as it is expected with diet.

Plots for bmi after extremes are removed, taking several different SD cutpoints:

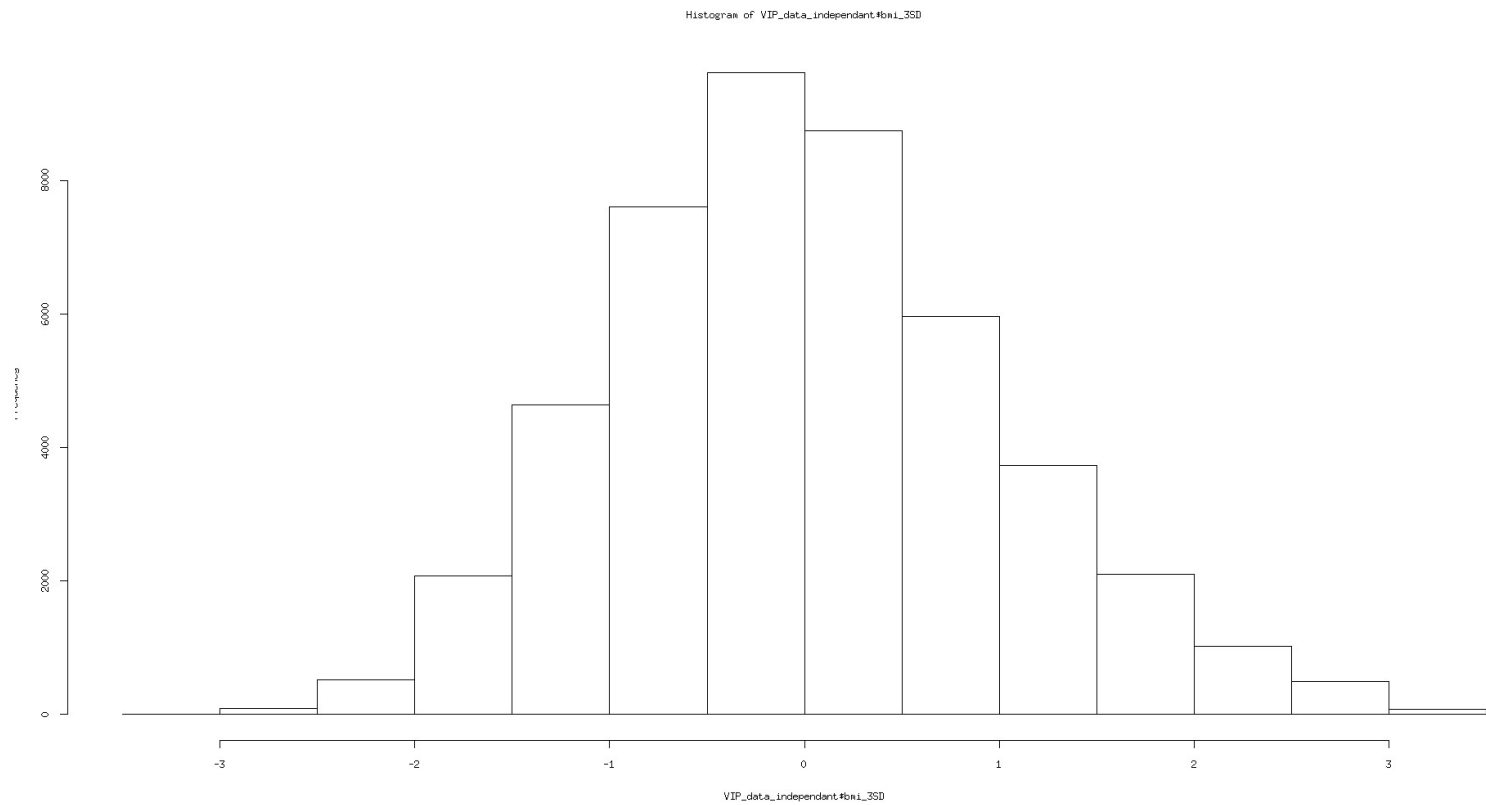
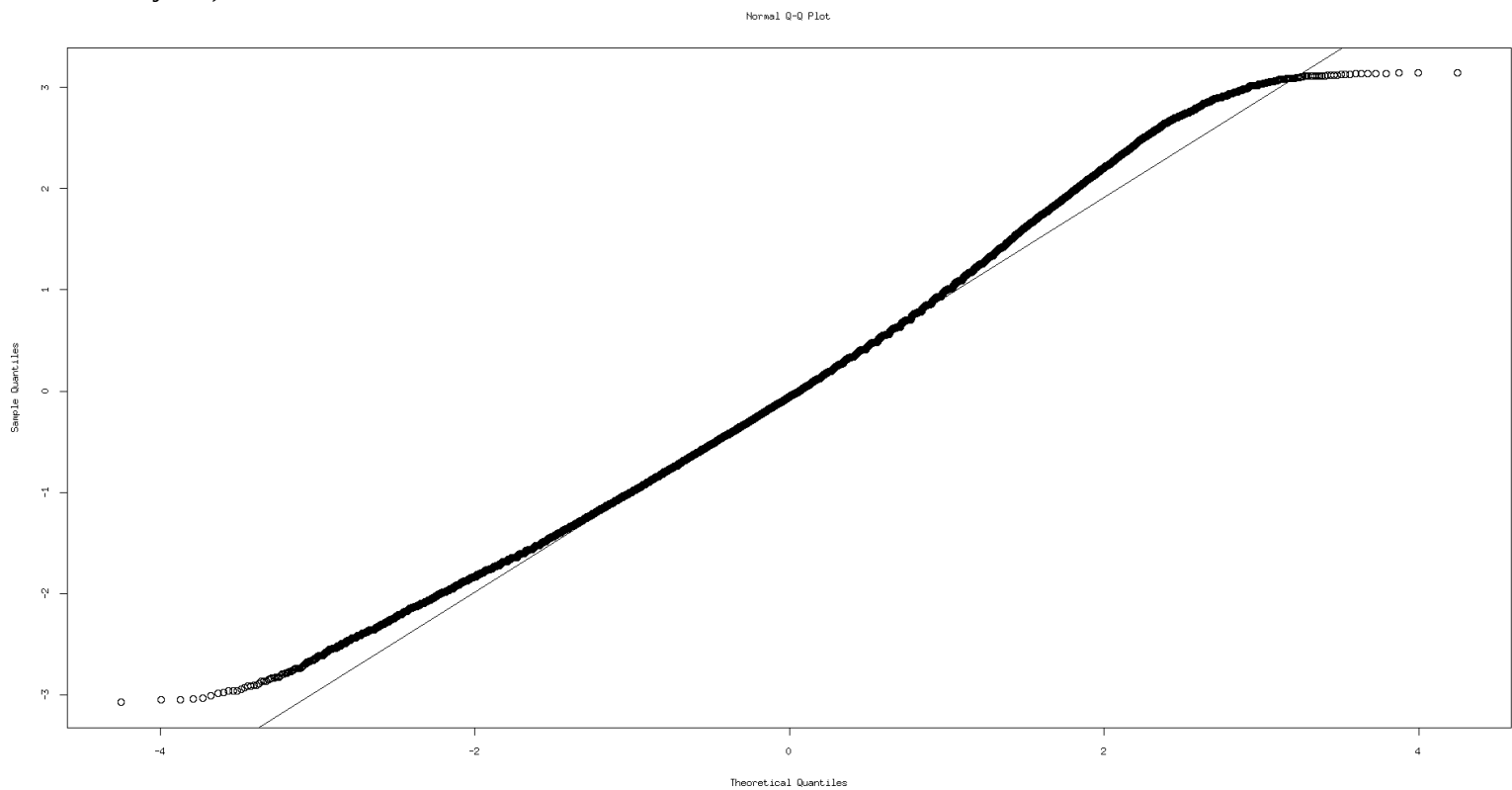
original bmi, not log transformed and standardized, without any extremes excluded



log transformed and standardized bmi, without any exclusion of extremes

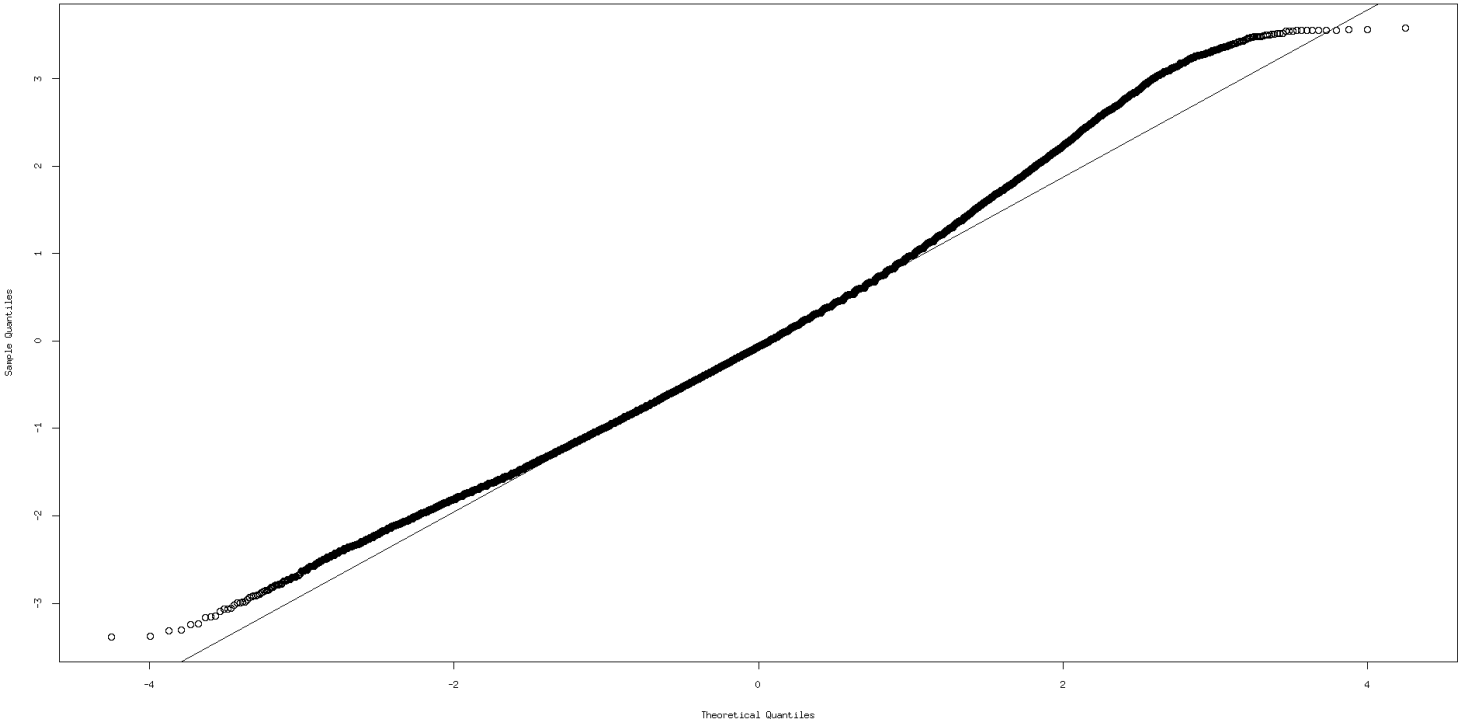


log transformed and standardized bmi, after excluding those that are below and above 3SD (313 subjects):

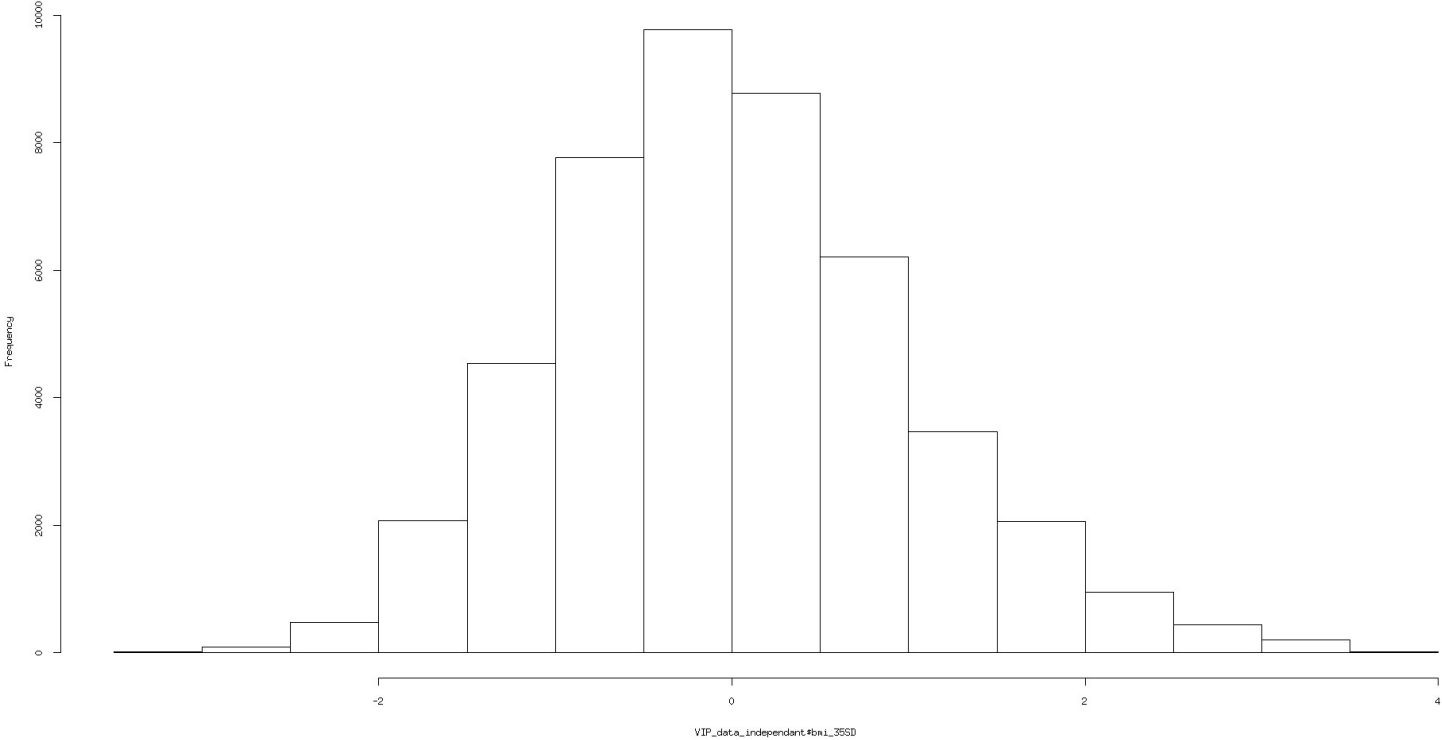


log transformed and standardized bmi, after excluding those that are below and above 3.5SD (116 subjects):

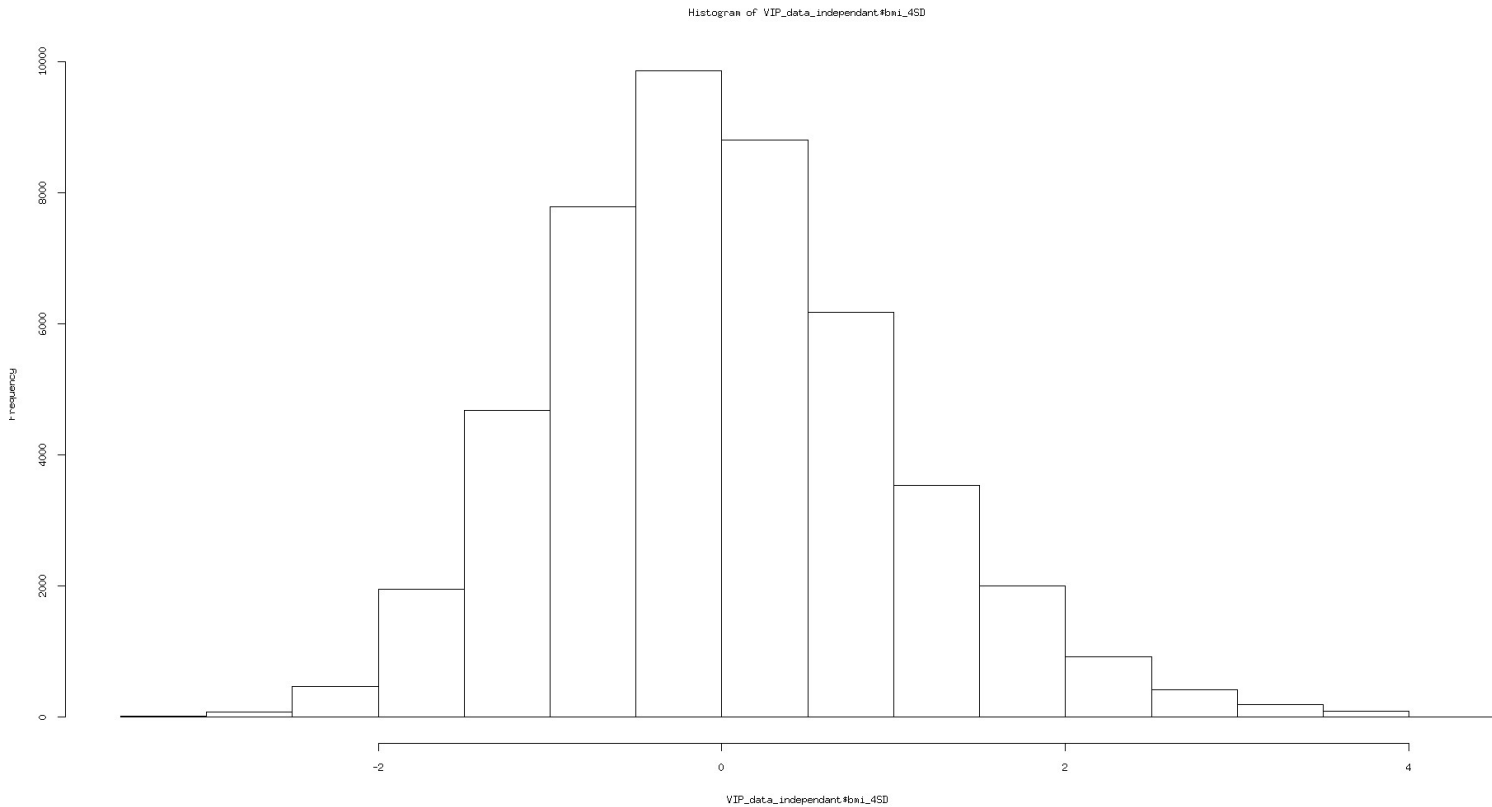
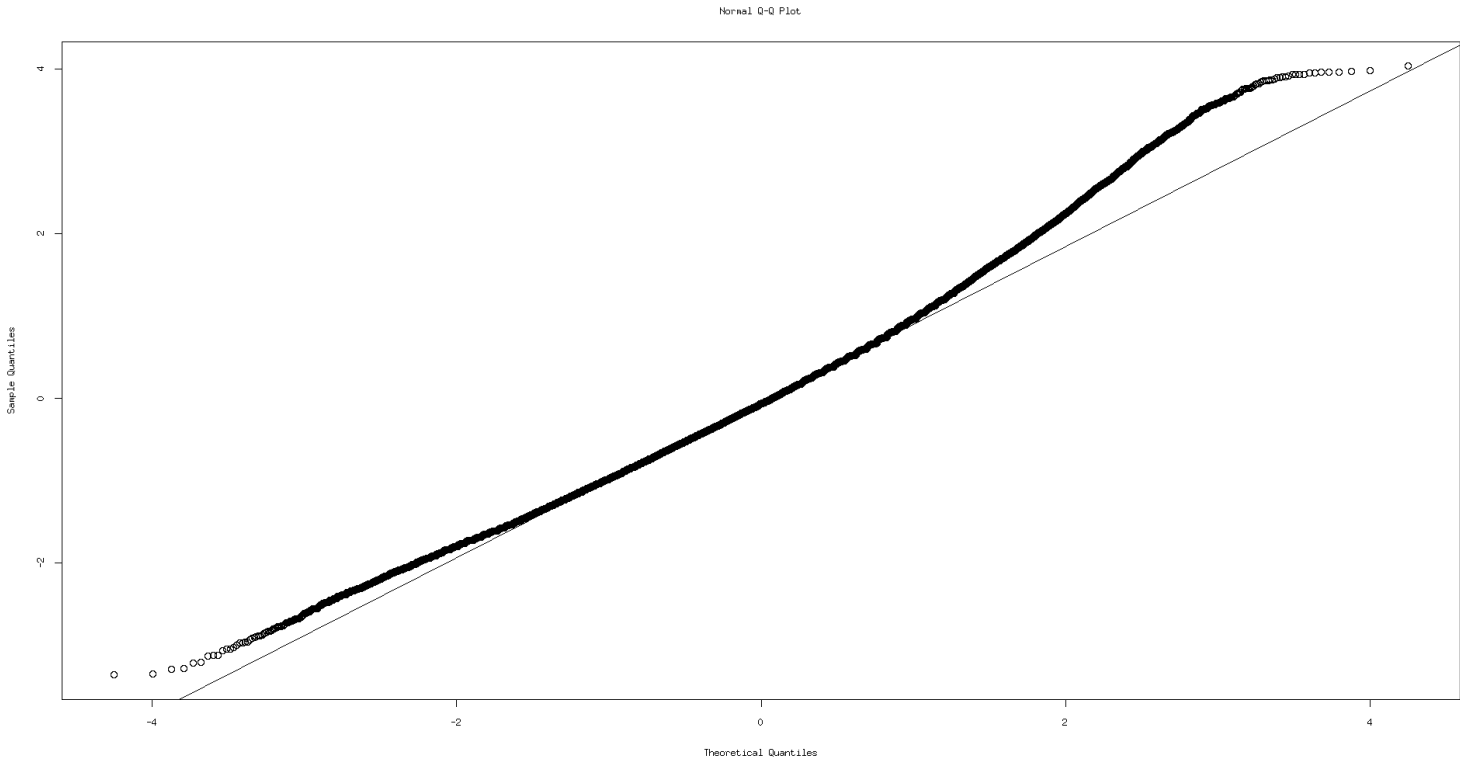
Normal Q-Q Plot



Histogram of VIP_data_independent#bmi_35SD



log transformed and standardized bmi, excluding those that are below and above 4SD (38 subjects):



Independent dataset of Swedish only (47107 subjects):

Pairwise Pearson correlation coefficient for all variables , raw, as in g per day, included in the model:

```
rcorr( cbind(bmi,POLYsum1_transformed,MONOsum1_transformed,mfetsum1_transformed,fettsum1_transformed,sacksum1_transfor
med,kolhsum1_transformed,FA_transformed,protsum1_transformed,fibesum1_transformed,NATRsum1_transformed),type="pearson"
)
```

	bmi	POLYsum1	MONOsum1	mfetsum1	fettsum1	sacksum1	kolhsum1	FA	protsum1	fibesum1	NATRsum1
bmi	1										
POLYsum1	0.07	1									
MONOsum1	0.09	0.80	1								
mfetsum1	0.05	0.63	0.91	1							
fettsum1	0.07	0.80	0.97	0.96	1						
sacksum1	-0.02	0.23	0.29	0.30	0.32	1					
kolhsum1	0.02**	0.43	0.46	0.45	0.49	0.70	1				
FA	0.06	0.98	0.74	0.57	0.75	0.24	0.43	1			
protsum1	0.10	0.60	0.73	0.69	0.74	0.34	0.69	0.56	1		
fibesum1	0.00 n.s.	0.35	0.24	0.20	0.26	0.41	0.78	0.37	0.52	1	
NATRsum1	0.10	0.68	0.81	0.71	0.79	0.32	0.63	0.64	0.88	0.48	1

all p-values < 2e-16, except where marked

Pairwise Pearson correlation coefficient for all variables, expressed as % of TEI, included in the model:

	bmi	POLYsum1	MONOsum1	mfetsum1	fettsum1	sacksum1	kolhsum1	FA	protsum1	fibesum1	NATRsum1
bmi	1										
POLYsum1	0.04	1									
MONOsum1	0.07	0.56	1								
mfetsum1	0.00 n.s.	0.16	0.74	1							
fettsum1	0.04	0.53	0.90	0.89	1						
sacksum1	-0.07	-0.29	-0.36	-0.30	-0.36	1					
kolhsum1	-0.07	-0.48	-0.82	-0.77	-0.88	0.50	1				
FA	0.02	0.97	0.45	0.57	0.44	-0.24	-0.40	1			
protsum1	0.09	-0.05	0.05	-0.02	-0.02**	-0.38	-0.23	-0.08	1		
fibesum1	-0.06	-0.14	-0.58	-0.62	-0.62	0.08	0.63	-0.08	-0.01 n.s.	1	
NATRsum1	0.09	0.21	0.37	0.08	0.22	-0.37	-0.34	0.17	0.54	-0.10	1

all p-values < 2e-16, except where marked

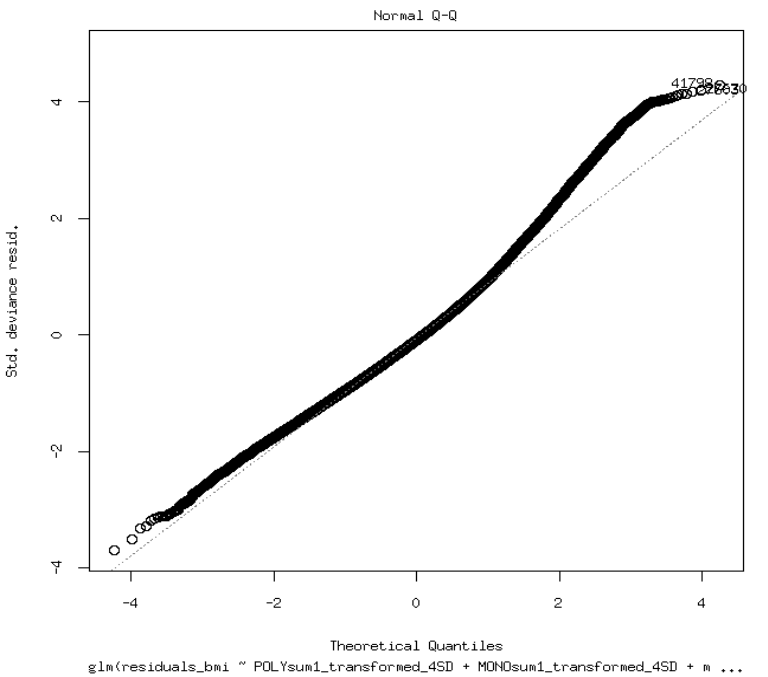
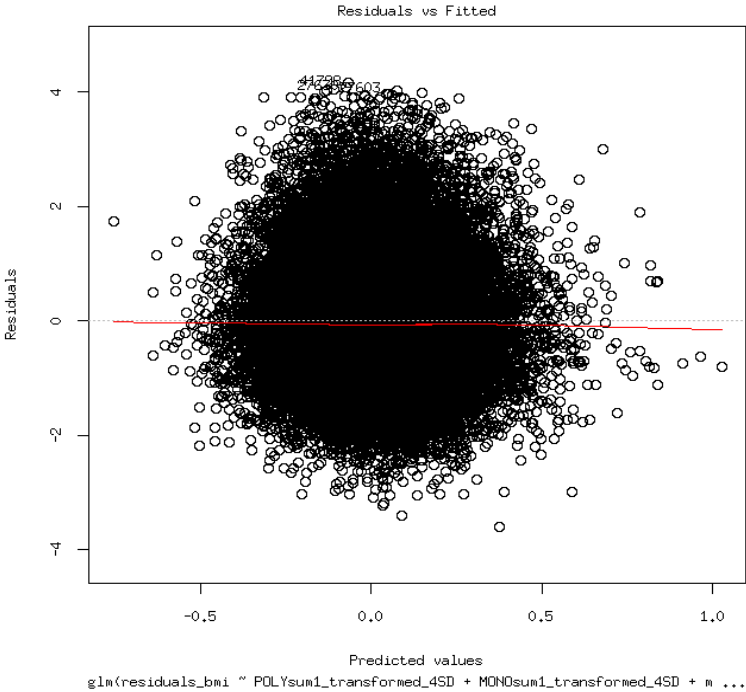
Pairwise Pearson correlation coefficient for all variables, expressed as % of TEI and categorized:

	bmi	POLYsum1	MONOsum1	mfetsum1	fetsum1	sacksum1	kolhsum1	FA	protsum1	fibesum1	NATRsum1
bmi	1										
POLYsum1	0.03	1									
MONOsum1	0.04	0.32	1								
mfetsum1	0.00 n.s.	0.07	0.45	1							
fetsum1	0.04	0.38	0.44	0.42	1						
sacksum1	-0.02**	-0.15	-0.13	-0.05	-0.12	1					
kolhsum1	-0.05	-0.36	-0.46	-0.36	-0.70	0.21	1				
FA	0.00 n.s.	-0.24	0.27	-0.09	-0.18	0.12	0.18	1			
protsum1	0.05	0.01 n.s.	0.04	0.03	0.03	-0.12	-0.12	-0.01 n.s.	1		
fibesum1	0.00 n.s.	0.04	0.20	0.18	0.21	0.02**	-0.25	-0.02	0.02	1	
NATRsum1	0.09	0.08	0.13	0.07	0.11	-0.05	-0.13	-0.06	0.07	-0.32	1

all p-values < 2e-16, except where marked

Raw continuous variables, fitted together by: `glm(bmi_4SD_sd~age + agesq + gender_factor + year + ffq_factor + POLYsum1_transformed_4SD + MONOsum1_transformed_4SD + mfetsum1_transformed_4SD + fettsum1_transformed_4SD + sacksum1_transformed_4SD + kolhsum1_transformed_4SD + FA_transformed_4SD + protsum1_transformed_4SD + fibesum1_transformed_4SD + NATRsum1_transformed_4SD, family = gaussian(link = "identity"))`

	Regression coefficient	Variance explained	p-value	vif
POLYsum1	0.1407375	0.000384	2.91e-05	54.419921
MONOsum1	0.1295815	0.000557	4.79e-07	31.765487
mfetsum1	-0.3515174	0.002782	< 2e-16	46.706314
fettsum1	0.1981662	0.000344	7.57e-05	120.048480
acids	-0.2087798	0.001177	2.52e-13	39.097753
kolhsum1	-0.0948393	0.001179	2.37e-13	8.041505
sacksum1	-0.0054483	1.2e-05	0.46002	2.604278
protsum1	0.1417485	0.003634	< 2e-16	5.813655
NATRsum1	0.0697070	0.000674	3.09e-08	7.601622
fibesum1	-0.0442983	0.000539	7.41e-07	3.830762

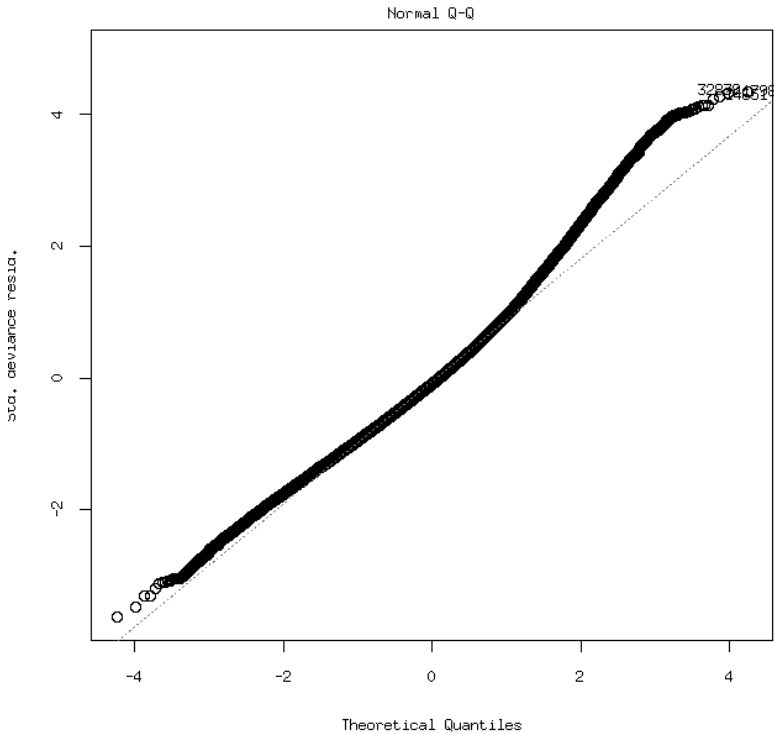
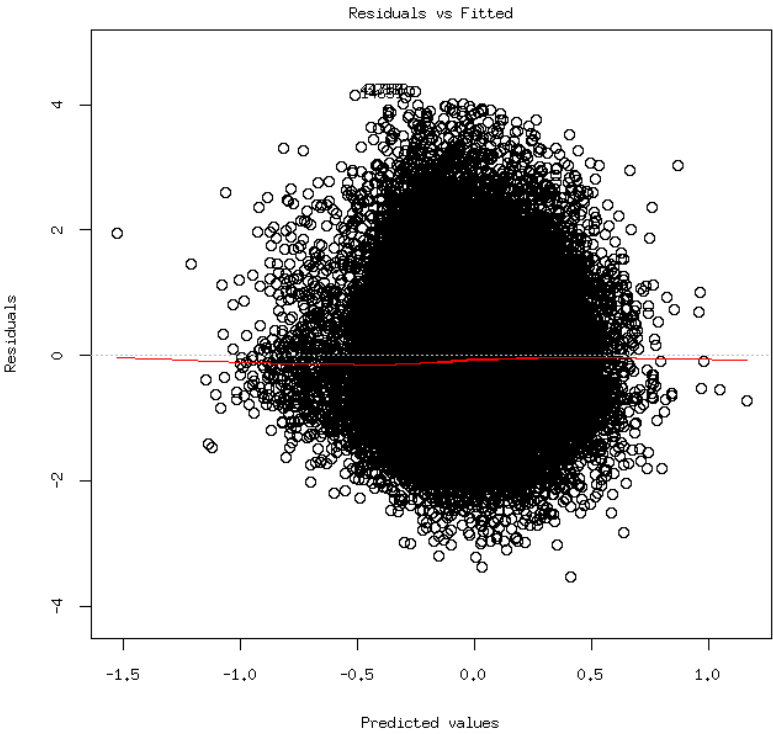


Raw continuous, fitted separately by: `glm(bmi~age + agesq + gender_factor + year + ffq_factor +
VIP_data_independant[,c(variable)], family = gaussian(link = "identity"))`

	Regression coefficient	Variance explained	p-value
POLYsum1	0.0398984	0.0013	< 2e-16
MONOsum1	0.0592131	0.002464	< 2e-16
mfetsum1	0.0036171	1e-05	0.5031204
fettsum1	0.0322182	0.000754	< 2e-16
acids	0.024749	0.000516	1.2e-06
kolhsum1	-0.0191932	0.000322	0.0001264
sacksum1	-0.0394297	0.001431	< 2e-16
protsum1	0.077037	0.005269	< 2e-16
NATRsum1	0.0874177	0.006282	< 2e-16
fibesum1	-0.016959	0.000282	0.0003361

Continuous expressed in % of TEI, fitted together by: `glm(bmi~age + agesq + gender_factor + year + ffq_factor + POLYsum1_ofTEI_transformed + MONOsum1_ofTEI_transformed + mfetsum1_ofTEI_transformed + fetsum1_ofTEI_transformed + sacksum1_ofTEI_transformed + kolhsum1_ofTEI_transformed + FA_ofTEI_transformed + protsum1_ofTEI_transformed + fibesum1_ofTEI_transformed + NATRsum1_ofTEI_transformed, family = gaussian(link = "identity"))`

	Regression coefficient	Variance explained	p-value	vif
POLYsum1	0.1098	0.000464	4.36e-06	27.429570
MONOsum1	0.06735	0.00046	4.82e-06	10.369828
mfetsum1	-0.2018	0.002506	< 2e-16	16.993811
fetsum1	0.1464	0.000584	2.55e-07	38.347210
acids	-0.1565	0.001216	1.01e-13	21.253716
kolhsum1	0.003378	2e-06	0.793610	7.944828
sacksum1	-0.001805	2e-06	0.766830	1.772334
protsum1	0.08669	0.00375	< 2e-16	2.097602
NATRsum1	0.03824	0.000731	8.16e-09	2.456449
fibesum1	-0.03414	0.000498	1.95e-06	2.103118

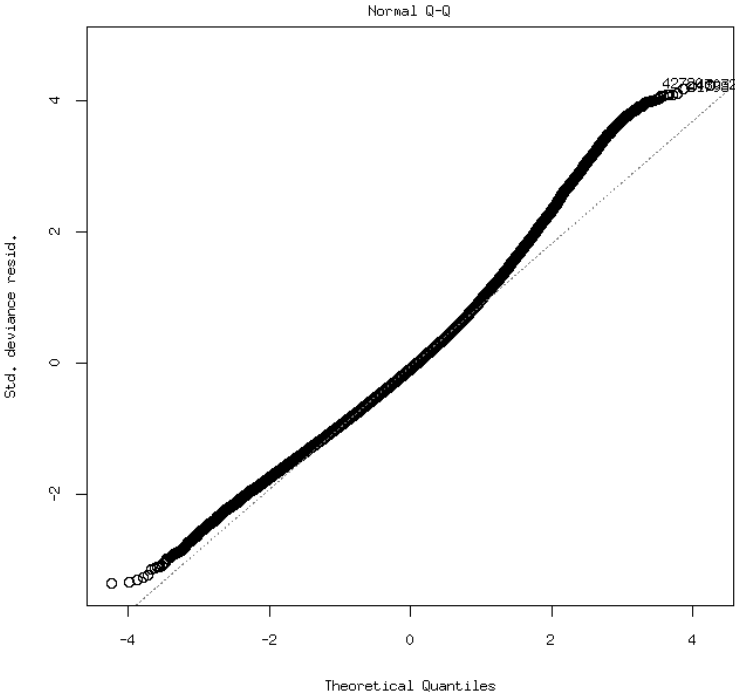
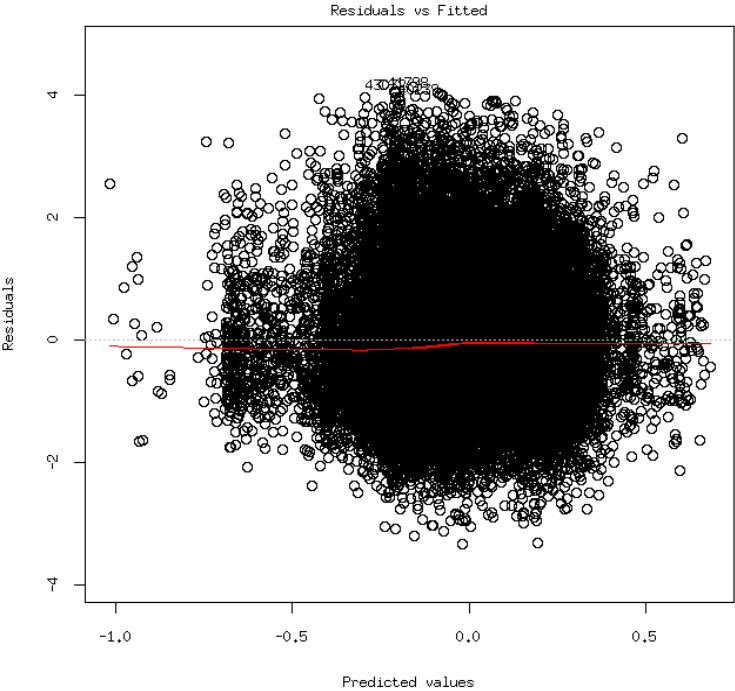


Continuous expressed in % of TEI, fitted separately by: `glm(bmi~age + agesq + gender_factor + year + ffq_factor + VIP_data_independant[,c(variable)], family = gaussian(link = "identity"))`

	Regression coefficient	Variance explained	p-value
POLYsum1	0.0375058	0.001273	< 2e-16
MONOsum1	0.0742868	0.004165	< 2e-16
mfetsum1	-0.0115707	0.000117	0.0208154
fetsum1	0.0309661	0.000789	< 2e-16
acids	0.018312	0.000314	0.0001535
kolhsum1	-0.0690082	0.003707	< 2e-16
sacksum1	-0.0550085	0.002863	< 2e-16
protsum1	0.1181138	0.013484	< 2e-16
NATRsum1	0.1211507	0.014003	< 2e-16
fibesum1	-0.0407374	0.001319	< 2e-16

Categorized according to the recommended cut points, fitted together by: `glm(bmi~age + agesq + gender_factor + year + ffq_factor + POLYsum1_ofTEI_categorized_g + MONOsum1_ofTEI_categorized_g + mfetsum1_ofTEI_categorized_g + fettsum1_ofTEI_categorized_g + sacksum1_ofTEI_categorized_g + kolhsum1_ofTEI_categorized_g + FA_ofTEI_categorized_g + protsum1_ofTEI_categorized_g + fibesum1_categorized_g + NATRsum1_categorized_g, family = gaussian(link = "identity"))`

	Regression coefficient	Variance explained	p-value	vif
POLYsum1	0.02643	0.000172	0.00514	1.330790
MONOsum1	0.06633	0.000438	7.76e-06	1.675835
mfetsum1	-0.1431	0.001016	9.77e-12	1.423775
fettsum1	0.02481	7.8e-05	0.05933	2.212601
acids	0.009544	4e-06	0.67994	1.129397
kolhsum1	-0.03141	0.00013	0.01488	2.353182
sacksum1	-0.001314	0	0.94279	1.099673
protsum1	0.2787	0.003151	< 2e-16	1.050402
NATRsum1	0.09898	0.001449	4.25e-16	1.321859
fibesum1	0.06554	0.000446	6.53e-06	1.291556



Categorized according to the recommended cut points, fitted separately by: `glm(bmi~age + agesq + gender_factor + year + ffq_factor + VIP_data_independant[,c(variable)], family = gaussian(link = "identity"))`

	Regression coefficient	Variance explained	p-value
POLYsum1	0.0525053	0.000806	< 2e-16
MONOsum1	0.0813032	0.000935	< 2e-16
mfetsum1	-0.0399634	0.000108	0.0267953
fetsum1	0.0578997	0.000801	< 2e-16
acids	-0.0487452	0.000106	0.0278558
kolhsum1	-0.0775953	0.001484	< 2e-16
sacksum1	-0.0561989	0.000218	0.0016114
protsum1	0.3064736	0.003932	< 2e-16
NATRsum1	0.0991399	0.001695	< 2e-16
fibesum1	0.0413634	0.000223	0.0014196

Focusing on sugar and interaction with other variables when fitting against bmi , few examples:
(in all models, basic covariates are also included)

	regression coeff	p-value	vif
sacksum1	-0.03943	6.63e-16	1.120491
kolhsum1	-0.01919	0.000126	1.176650
fibesum1	-0.01696	0.000336	1.051876
mfetsum1	0.003617	0.5031	1.370631
fettsum1	0.03222	4.49e-09	1.418434
sacksum1+kolhsum1	-0.04816 ; 0.01363	2.04e-13 ; 0.0421	2.0079 ; 2.1103
sacksum1+kolhsum1+sacksum1*kolhsum1	-0.048 ; 0.012 ; -0.014	4.14e-13 ; 0.067144 ; 0.000345	2.009346 ; 2.116882 ; 1.012116
sacksum1+fibesum1	-0.038 ; -0.003	5.07e-13 ; 0.587	1.316 ; 1.236
sacksum1+fibesum1+sacksum1*fibesum1	-0.043 ; -0.002 ; -0.022	1.99e-15 ; 0.683 ; 1.46e-07	1.347924 ; 1.236961 ; 1.027558
sacksum1+mfetsum1	-0.044 ; 0.017	< 2e-16 ; 0.00194	1.2203 ; 1.4927
sacksum1+mfetsum1+sacksum1*mfetsum1	-0.042 ; 0.017 ; -0.017	4.22e-16 ; 0.002344 ; 0.000256	1.2355 ; 1.4931 ; 1.0286
sacksum1+fettsum1	-0.054 ; 0.051	< 2e-16 ; < 2e-16	1.2428 ; 1.5733
sacksum1+fettsum1+sacksum1*fettsum1	-0.051 ; 0.051 ; -0.021	< 2e-16 ; < 2e-16 ; 4.24e-06	1.2540 ; 1.5734 ; 1.023
sacksum1+ mfetsum1 + fibesum1 + kolhsum1	-0.056 ; 0.012 ; -0.027 ; 0.037	1.48e-15 ; 0.05368 ; 0.00163 ; 0.00164	2.2856 ; 1.7394 ; 3.4855 ; 6.5437
sacksum1+ fettsum1 + fibesum1 + kolhsum1	-0.054 ; 0.053 ; -0.024 ; 0.014	9.42e-15 ; < 2e-16 ; 0.00562 ; 0.22857	2.2823 ; 1.8837 ; 3.4118 ; 6.4475
sacksum1+ mfetsum1 + fettsum1 + fibesum1 + kolhsum1	-0.059 ; -0.291 ; 0.337 ; -0.059 ; 0.038	< 2e-16 ; < 2e-16 ; < 2e-16 ; 1.86e-11 ; 0.0013	2.2872 ; 14.2251 ; 15.4102 ; 3.6215 ; 6.5438
sacksum1 + fibesum1 + kolhsum1	-0.057 ; -0.03 ; 0.045	4.57e-16 ; 0.000376 ; 4.91e-05	2.2775 ; 3.3812 ;5.7946

Fitting other variables against sugar :
(in all models, basic covariates are also included)

	regression coeff	p-value	vif
kolhsum1	0.68	< 2e-16	1.177203
fibesum1	0.374	< 2e-16	1.051820
mfetsum1	0.317	< 2e-16	1.370057
fetsum1	0.353	< 2e-16	1.417309
fibesum1+kolhsum1	1.037 ; -0.419	< 2e-16 ; < 2e-16	3.342284 ; 2.984543
fetsum1+fibesum1	0.23 ; 0.305	< 2e-16 ; < 2e-16	1.6149 ; 1.1983
kolhsum1+fetsum1	0.687 ; -0.014	< 2e-16 ; 0.0018	1.5498 ; 1.8655
fetsum1 + fibesum1 + kolhsum1	-0.042 ; -0.423 ; 1.059	< 2e-16 ; < 2e-16	1.879721 ; 3.005663 ; 3.884361

Partial correlation with bmi:
(in all models, basic covariates are also included)

	regression coeff	p-value
kolhsum1	-0.01794978	0.0001281271
fibesum1	-0.01671706	0.0003611425
sacksum1	-0.03764379	9.400732e-16
fibesum1, (kolhsum1)	-0.003829232	0.4139341
fibesum1, (kolhsum1, sacksum1)	-0.01666753	0.000376094
fibesum1, (sacksum1)	-0.002482752	0.5963136
sacksum1, (kolhsum1)	-0.03443119	2.008087e-13
sacksum1, (kolhsum1, fibesum1)	-0.03805709	4.565394e-16
sacksum1, (fibesum1)	-0.0338241	5.251835e-13
kolhsum1, (sacksum1)	0.00950748	0.04250747
kolhsum1, (sacksum1, fibesum1)	0.01902661	4.912507e-05
kolhsum1, (fibesum1)	-0.007576845	0.1059682

Sugar fitted against bmi within each category of carbs(adjusted for basic cov):

	regression coeff	p-value
sacksum1(carbs score==0)	-1.864e-02	0.016853
sacksum1(carbs score==1)	-3.244e-02	5.88e-06
sacksum1(carbs score==2)	-0.0164567	0.596274