SS4853

Fang He

December 8, 2015

Notes from my time (2012)

Use the data file from the 2011 Youth Risk Behavior Survey. Calculate two new variables from the dataset:

- body mass index (BMI), which is a preson's weight in kilograms divided by their height in meters squared.
- age of the student (it is a coded value in the data file).

1. Using linear regression find the variables in the data file that best explain the variation in BMI. Make appropriate plots of the data.

In answering this question, you should research information about BMI and its correlates. To get you started, here is one article that will be of interest.

T.Ostbye, J.Pomerlau, M.Speechley, L.L. Pederson and K.L. Speechle (1995). Correlates of body mass index in the 1990 Ontario Health Survey. *Canadian Medical Association Journal* **152** (11): 1811 – 1817

2. Using logistic regression find the variables in the data file that best explain the proportion of students who have contemplated suicide. Make appropriate plots of the data.

In answering this question, you should research information about teen suicide and its correlates. To get you started, here is one article that will be of interest.

M.H. Swahn and R.M. Bossarte (2007). Gender, early alcohol use, and suicide ideation and attempts: findings from the 2005 Youth Risk Behavior Survey. *Journal of Adolescent Health* **41** (2): 175 – 181

First Class

What effects BMI? (Do test)

Test whether sex is significant.

Add random effect in the same model to see how p-value of sex change.

Second Class

Do logistic regression on suicide data.

Details

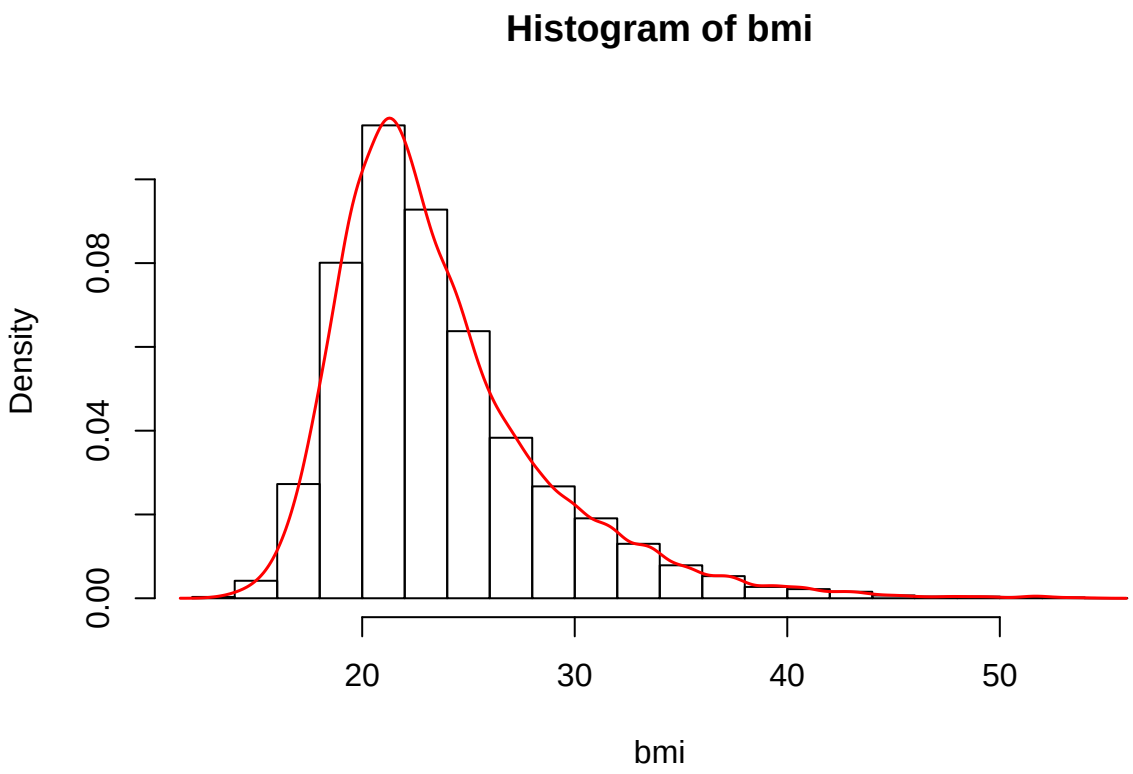
```
yrbs <- read.table("/Users/fhe7/Documents/School/Sampling/yrbs2013.txt",header=TRUE)
head(yrbs)
```

```
##   raceeth q1 q2 q3 q4 q5   q6    q7 q8 q9 q10 q11 q12 q13 q14 q15 q16 q17
## 1       3  3  1  1  2  C   NA    NA  2  5   5   2   2   1   1   1   1   1
## 2       3  3  1  1  2  C   NA    NA  2  2   1  NA  NA   1   1   1   1   1
## 3       7  4  1  1  1  D 1.73 84.37  2  1   5   6   8   5   5   2   5   8
## 4       3  4  1  1  2  C 1.60 55.79  2  3   1   2   2   1   1   1   1   1
## 5       3  4  1  1  2  C 1.50 46.72  1  5   1   1   1   1   1   1   3   2
## 6       3  4  1  1  2  C 1.57 67.13  1  5   2   1   1   1   1   1   1   3
##   q18 q19 q20 q21 q22 q23 q24 q25 q26 q27 q28 q29 q30 q31 q32 q33 q34 q35
## 1   2   1   2   2   2   2   2   2   2   2   2   1   1   2   1   1   1   1
## 2   5   1   1   2   2   2   2   2   2   2   2   1   1   2   1   1   1   1
## 3   8   1   8   2   3   2   2   1   1   1   1   2   3   1   2   7   3   5
## 4   3   1   3   2   2   2   1   2   2   2   2   1   1   1   1   1   1   1
## 5   1   1   1   2   2   3   1   2   1   2   2   1   1   2   1   1   1   1
## 6   1   1   1   1   4   3   1   1   1   2   2   1   1   2   1   1   1   1
##   q36 q37 q38 q39 q40 q41 q42 q43 q44 q45 q46 q47 q48 q49 q50 q51 q52 q53
## 1   1   2   1   1   1   4   2   1   1   1   1   3   5   1   1   1   1   1
## 2   1   2   1   1   1   7   5   4   3   6   6   1   1   1   1   1   1   1
## 3   2   1   3   1   7   7   2   7   7   8   6   7   2   6   6   6   1   1
## 4   1   2   1   1   1   3   5   2   1   2   5   5   5   3   1   1   1   1
## 5   1   2   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 6   1   2   1   1   1   2   6   1   1   1   1   1   1   1   1   1   1   1
##   q54 q55 q56 q57 q58 q59 q60 q61 q62 q63 q64 q65 q66 q67 q68 q69 q70 q71
## 1   1   1   1   1   2   2   1   1   1   1   1   1   4   1   2   2   2   1
## 2   1   1   1   1   2   2   1   1   1   1   1   1   4   1   2   2   2   3
## 3   6   1   6   1   1   1   5   3   4   2   3   2   4   1   1   2   2   7
## 4   1   1   1   1   2   1   4   2   3   3   2   4   3   3   2   2   2   2
## 5   1   1   1   1   2   2   1   1   1   1   1   1   3   2   2   2   2   4
## 6   1   1   1   1   1   1   6   2   2   2   2   4   3   1   2   2   2   2
##   q72 q73 q74 q75 q76 q77 q78 q79 q80 q81 q82 q83 q84 q85 q86 q87 q88 q89
## 1   4   2   1   1   1   5   1   1   5   7   6   1   2   1   1   1   1   3
## 2   3   2   1   1   2   3   2   4   3   7   7   1   1   3   2   1   1   2
## 3   7   4   7   1   7   7   4   8   8   7   7   1   2   1   1   1   1   2
## 4   2   1   1   1   1   3   1   3   1   4   5   6   1   2   2   1   1   2
## 5   2   1   1   1   3   2   2   5   3   5   7   6   1   1   2   1   2   2
## 6   3   2   1   1   3   1   1   8   2   7   4   1   1   1   2   1   1   3
##   q90 q91 q92 weight stratum   psu
## 1   1   1   5 0.8865     103 10970
## 2   1   1   3 0.8865     103 10970
## 3   1   1   1 0.8864     103 10970
## 4   1   1   3 0.8865     103 10970
## 5   1   1   6 0.8865     103 10970
## 6   1   1   5 0.8865     103 10970
```

```r
colnames(yrbs[,c(2,3,7,8)]) <- c("age","sex","height","weight")
attach(yrbs)

bmi <- q7/q6^2
```

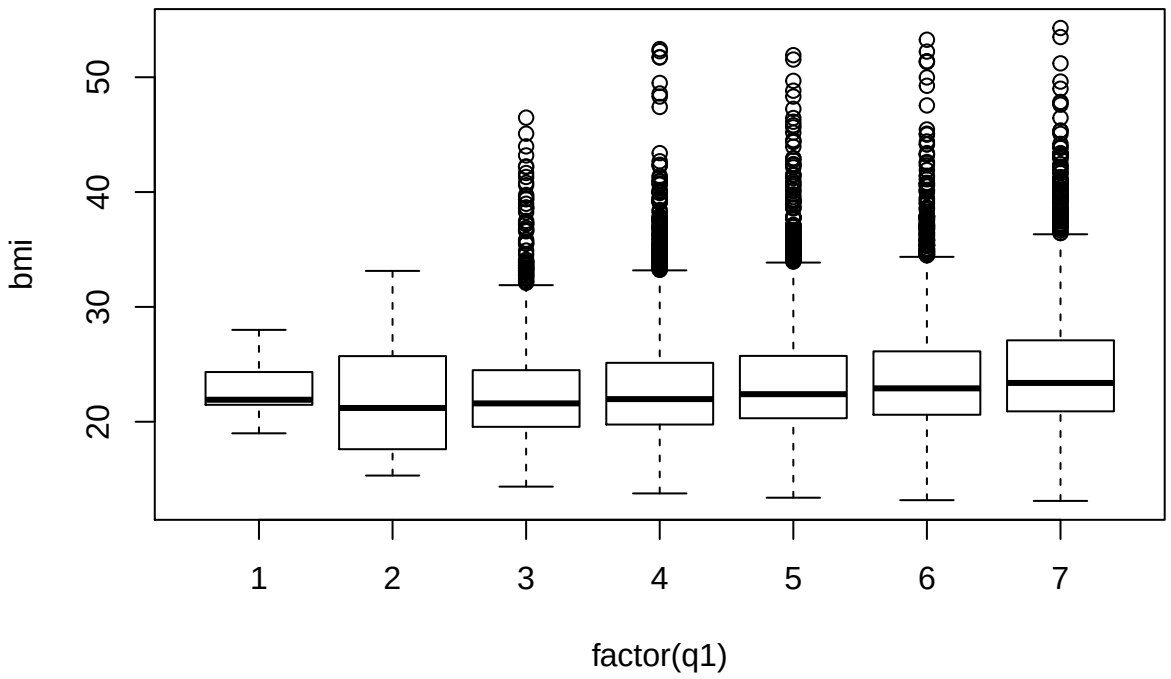
```r
hist(bmi,prob=TRUE)
lines(density(bmi,na.rm = T),col=2,lwd=1.5)
```



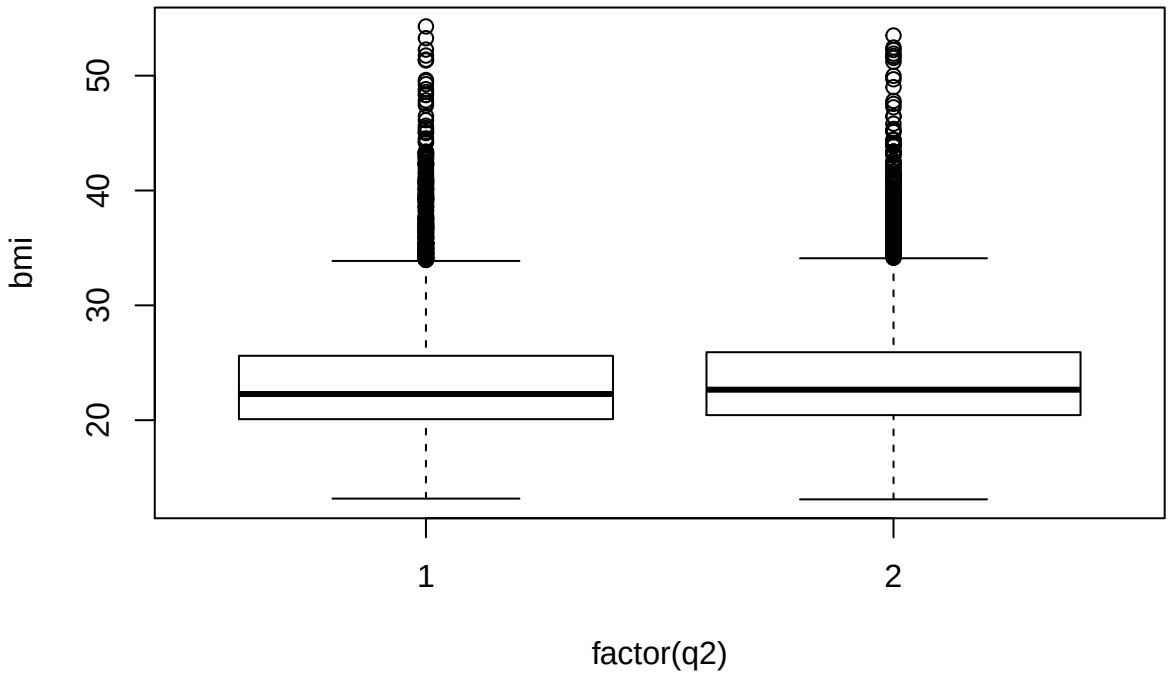
Linear Regression

Q: Why we use linear regression?

```
plot(bmi~factor(q1))
```



```r
plot(bmi~factor(q2))
```




```r
mod0 <- lm(bmi ~ factor(q1)+factor(q2),data=yrbs)
summary(mod0)
```

```
##
## Call:
## lm(formula = bmi ~ factor(q1) + factor(q2), data = yrbs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.713  -3.338  -1.151   2.139  29.692
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.62137    2.02719  11.159   <2e-16 ***
## factor(q1)2 -0.67892    2.44977  -0.277    0.782
## factor(q1)3 -0.13111    2.03125  -0.065    0.949
## factor(q1)4  0.28324    2.02851   0.140    0.889
## factor(q1)5  0.80665    2.02840   0.398    0.691
## factor(q1)6  1.23637    2.02824   0.610    0.542
## factor(q1)7  1.97249    2.02916   0.972    0.331
## factor(q2)2  0.22553    0.08865   2.544    0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.963 on 12571 degrees of freedom
##   (1004 observations deleted due to missingness)
## Multiple R-squared:  0.01772,    Adjusted R-squared:  0.01718
## F-statistic: 32.41 on 7 and 12571 DF,  p-value: < 2.2e-16
```

age - young did not affect much of BMI

gender - different

```
mod1 <- update(mod0,.~.-factor(q1))
## or mod1 <- lm(bmi ~ factor(q2),data=yrbs)
summary(mod1)
```

```
##
## Call:
## lm(formula = bmi ~ factor(q2), data = yrbs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.672  -3.373  -1.173   2.137  30.788
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.49766    0.06374 368.634  < 2e-16 ***
## factor(q2)2  0.28067    0.08927   3.144  0.00167 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.005 on 12577 degrees of freedom
##   (1004 observations deleted due to missingness)
## Multiple R-squared:  0.0007854,  Adjusted R-squared:  0.0007059
## F-statistic: 9.886 on 1 and 12577 DF,  p-value: 0.00167
```

```
anova(mod1,mod0)
```

```
## Analysis of Variance Table
##
## Model 1: bmi ~ factor(q2)
## Model 2: bmi ~ factor(q1) + factor(q2)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1  12577 315042
## 2  12571 309701  6    5340.8 36.131 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Consider random effect

```
library(lme4)
```

```
mod0_lmer <- lmer(bmi ~ q2 + factor(stratum) + (1|psu),data=yrbs)
summary(mod0_lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: bmi ~ q2 + factor(stratum) + (1 | psu)
##    Data: yrbs
```

```
##
## REML criterion at convergence: 76060.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1362 -0.6731 -0.2260  0.4357  6.0876
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  psu      (Intercept)  0.2243  0.4736
##  Residual             24.6403  4.9639
## Number of obs: 12579, groups:  psu, 54
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)        23.47843    0.24897   94.30
## q2                  0.28784    0.08875    3.24
## factor(stratum)102  0.48750    0.35227    1.38
## factor(stratum)103  0.32192    0.37639    0.86
## factor(stratum)111 -0.67212    0.34042   -1.97
## factor(stratum)112  0.23981    0.38581    0.62
## factor(stratum)113 -0.28259    0.38637   -0.73
## factor(stratum)201 -0.08045    0.30205   -0.27
## factor(stratum)202 -0.09193    0.47200   -0.19
## factor(stratum)203 -0.46678    0.43900   -1.06
## factor(stratum)211 -1.27210    0.33701   -3.77
## factor(stratum)212 -1.03897    0.36691   -2.83
## factor(stratum)213 -0.77853    0.35804   -2.17
## factor(stratum)214 -0.86282    0.45360   -1.90
##
## Correlation of Fixed Effects:
##             (Intr) q2     f()102 f()103 f()111 f()112 f()113 f()201 f()202
## q2          -0.541
## fctr(st)102 -0.501  0.002
## fctr(st)103 -0.471  0.006  0.331
## fctr(st)111 -0.517  0.000  0.365  0.342
## fctr(st)112 -0.459  0.005  0.322  0.302  0.334
## fctr(st)113 -0.459  0.006  0.322  0.301  0.333  0.294
## fctr(st)201 -0.581 -0.003  0.412  0.385  0.426  0.376  0.376
## fctr(st)202 -0.364 -0.016  0.264  0.247  0.273  0.241  0.240  0.307
## fctr(st)203 -0.403  0.003  0.283  0.265  0.293  0.259  0.258  0.331  0.211
## fctr(st)211 -0.524  0.003  0.369  0.346  0.382  0.337  0.337  0.431  0.275
## fctr(st)212 -0.484  0.008  0.339  0.317  0.351  0.310  0.309  0.395  0.253
## fctr(st)213 -0.494  0.004  0.347  0.325  0.360  0.317  0.317  0.405  0.259
## fctr(st)214 -0.390  0.004  0.274  0.257  0.284  0.250  0.250  0.320  0.205
##             f()203 f()211 f()212 f()213
## q2
## fctr(st)102
## fctr(st)103
## fctr(st)111
## fctr(st)112
## fctr(st)113
## fctr(st)201
## fctr(st)202
```

```
## fctr(st)203
## fctr(st)211  0.296
## fctr(st)212  0.272  0.354
## fctr(st)213  0.279  0.363  0.334
## fctr(st)214  0.220  0.287  0.263  0.270
```

Final Project - survey design - modeling with random effect - see the difference