V. Statistical analysis in R (solution)

Data Science Lab, University of Copenhagen

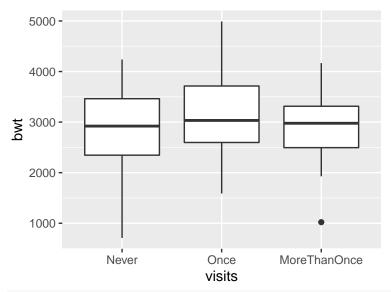
14 March, 2022

Data

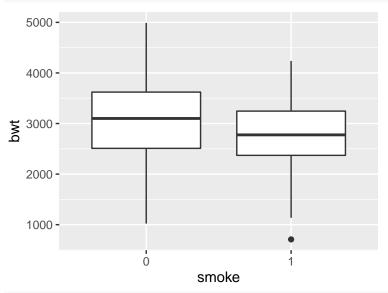
```
### Open help page for birthwt
?birthwt
### Make a tibble with the data
birthData <- as_tibble(birthwt)</pre>
birthData
## # A tibble: 189 x 10
##
        low
                      lwt
               age
                           race smoke
                                          ptl
                                                  ht
                                                        ui
                                                              ftv
                                                                     bwt
##
      <int> <int>
                                                                  <int>
    1
           0
                               2
                                     0
                                            0
                                                                    2523
##
                19
                      182
                                                   0
                                                          1
##
    2
           0
                33
                      155
                               3
                                     0
                                            0
                                                   0
                                                         0
                                                                3
                                                                   2551
##
    3
                20
                      105
                               1
                                                         0
                                                                   2557
                      108
                                                                   2594
##
    4
           0
                21
                               1
                                     1
                                            0
                                                   0
                                                         1
                                                                2
##
    5
           0
                18
                      107
                               1
                                     1
                                            0
                                                   0
                                                                   2600
##
    6
           0
                21
                      124
                               3
                                     0
                                            0
                                                   0
                                                         0
                                                                   2622
    7
##
           0
                22
                      118
                                     0
                                            0
                                                   0
                                                         0
                                                                   2637
                17
                      103
                               3
                                     0
                                            0
##
    8
           0
                                                   0
                                                         0
                                                                   2637
                                                                1
##
    9
                29
                      123
                                                                1
                                                                   2663
## 10
           0
                26
                      113
                                                                   2665
## # ... with 179 more rows
### Make smoke into a factor
# Check if `smoke` is numerical
is.numeric(birthData$smoke)
## [1] TRUE
# Alternatively, review the summary of the whole dataset
summary(birthData)
```

```
##
         low
                                              lwt
                                                               race
                            age
##
                                                : 80.0
    Min.
            :0.0000
                       \mathtt{Min}.
                             :14.00
                                        \mathtt{Min}.
                                                          Min.
                                                                  :1.000
##
    1st Qu.:0.0000
                       1st Qu.:19.00
                                        1st Qu.:110.0
                                                          1st Qu.:1.000
                       Median :23.00
##
    Median :0.0000
                                        Median :121.0
                                                          Median :1.000
                                                :129.8
    Mean
            :0.3122
                       Mean
                              :23.24
                                        Mean
                                                          Mean
                                                                  :1.847
                                        3rd Qu.:140.0
##
    3rd Qu.:1.0000
                       3rd Qu.:26.00
                                                          3rd Qu.:3.000
                                                :250.0
##
    Max.
            :1.0000
                       Max.
                              :45.00
                                        Max.
                                                          Max.
                                                                  :3.000
##
        smoke
                            ptl
                                                ht
                                                                    ui
    Min.
            :0.0000
                              :0.0000
                                         Min.
                                                 :0.00000
                                                             Min.
                                                                     :0.0000
                       Min.
##
    1st Qu.:0.0000
                       1st Qu.:0.0000
                                         1st Qu.:0.00000
                                                             1st Qu.:0.0000
                       Median :0.0000
                                                             Median :0.0000
    Median :0.0000
                                         Median :0.00000
```

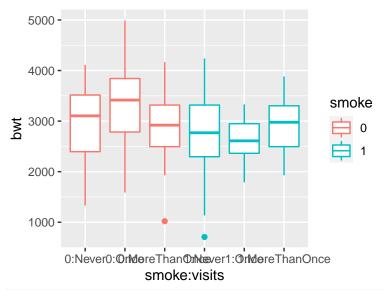
```
## Mean
          :0.3915
                  Mean :0.1958
                                  Mean :0.06349
                                                   Mean :0.1481
   3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.00000
                                                   3rd Qu.:0.0000
##
  Max.
        :1.0000
                  Max. :3.0000 Max. :1.00000 Max. :1.0000
##
                       bwt
        ftv
## Min. :0.0000
                  Min. : 709
## 1st Qu.:0.0000
                  1st Qu.:2414
## Median :0.0000
                  Median:2977
                  Mean :2945
## Mean :0.7937
## 3rd Qu.:1.0000
                   3rd Qu.:3487
## Max. :6.0000
                  Max. :4990
# Finally, make smoke into a factor
birthData <- mutate(birthData, smoke = factor(smoke))</pre>
# Check the ftv variable, make it a factor and collapse some of the levels
table(birthData$ftv)
##
   0 1
##
           2
               3
                      6
## 100 47 30
               7
birthData <- mutate(birthData, ftvFac = factor(ftv))</pre>
birthData <- mutate(birthData, visits = fct_collapse(ftvFac, Never="0", Once="1", other_level="MoreThan
summary(birthData)
##
        low
                                      lwt
                                                               smoke
                       age
                                                     race
                                 Min. : 80.0
## Min. :0.0000
                  Min. :14.00
                                                     :1.000
                                               Min.
                                                               0:115
  1st Qu.:0.0000
                  1st Qu.:19.00
                                 1st Qu.:110.0
                                               1st Qu.:1.000
                                                               1: 74
## Median :0.0000
                  Median :23.00
                                Median :121.0 Median :1.000
## Mean :0.3122
                  Mean :23.24
                                 Mean :129.8
                                                Mean :1.847
##
   3rd Qu.:1.0000 3rd Qu.:26.00
                                 3rd Qu.:140.0 3rd Qu.:3.000
## Max. :1.0000
                  Max. :45.00 Max. :250.0 Max. :3.000
##
       ptl
                        ht
                                         ui
                                                       ftv
## Min. :0.0000
                         :0.00000 Min.
                                        :0.0000 Min.
                                                         :0.0000
                  Min.
##
  1st Qu.:0.0000
                  1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000
                  Median :0.00000 Median :0.0000 Median :0.0000
## Mean :0.1958
                        :0.06349
                                                 Mean :0.7937
                  Mean
                                   Mean :0.1481
##
   3rd Qu.:0.0000
                   3rd Qu.:0.00000
                                   3rd Qu.:0.0000
                                                  3rd Qu.:1.0000
## Max.
         :3.0000
                  Max. :1.00000
                                   Max.
                                         :1.0000 Max. :6.0000
##
       bwt
                 ftvFac
                                visits
## Min. : 709
                 0:100
                                   :100
                        Never
## 1st Qu.:2414
                 1: 47
                                   : 47
                        Once
## Median :2977
                 2: 30
                        MoreThanOnce: 42
## Mean :2945
                 3: 7
## 3rd Qu.:3487
                 4: 4
        :4990
## Max.
                 6: 1
### Parallell boxplots
ggplot(birthData, aes(x=visits, y=bwt)) + geom_boxplot()
```



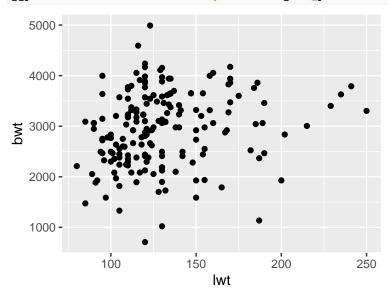
ggplot(birthData, aes(x=smoke, y=bwt)) + geom_boxplot()



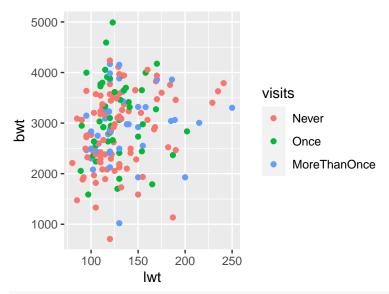
Groupwise boxplots
ggplot(birthData, aes(x=smoke:visits, y=bwt, col=smoke)) + geom_boxplot()



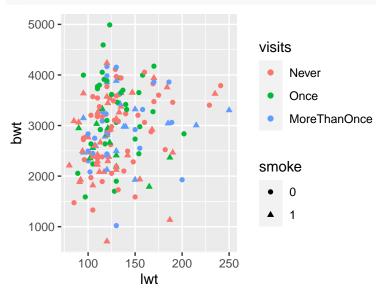
Scatterplot
ggplot(birthData, aes(x=lwt, y=bwt)) + geom_point()



Scatter plot with points coloured after visits, and point types after smoke status
ggplot(birthData, aes(x=lwt, y=bwt, col=visits)) + geom_point()



ggplot(birthData, aes(x=lwt, y=bwt, col=visits, pch=smoke)) + geom_point()



Regression

##

```
### Simple linear regression
reg1 <- lm(bwt ~ lwt, data=birthData)
summary(reg1)
##
## Call:
## lm(formula = bwt ~ lwt, data = birthData)</pre>
```

Residuals:
Min 1Q Median 3Q Max
-2192.12 -497.97 -3.84 508.32 2075.60
##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

```
## (Intercept) 2369.624
                              228.493
                                        10.371
                                                   <2e-16 ***
                                                   0.0105 *
## lwt
                    4.429
                                 1.713
                                          2.585
##
                     0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Signif. codes:
## Residual standard error: 718.4 on 187 degrees of freedom
## Multiple R-squared: 0.0345, Adjusted R-squared: 0.02933
## F-statistic: 6.681 on 1 and 187 DF, p-value: 0.0105
confint(reg1)
##
                        2.5 %
                                   97.5 %
## (Intercept) 1918.867879 2820.37916
                    1.048845
                                  7.80937
### Model validation
par(mfrow=c(2,2))
plot(reg1)
                                                   Standardized residuals
                                                                       Normal Q-Q
                 Residuals vs Fitted
     2000
Residuals
                                                         \alpha
                                                         0
     -2000
                                                                                          2
            2800
                     3000
                              3200
                                                                  -2
                                                                                                3
                                      3400
                                                            -3
                                                                     Theoretical Quantiles
                     Fitted values
/Standardized residuals
                                                   Standardized residuals
                                                                  Residuals vs Leverage
                   Scale-Location
                                                         \alpha
                                                         0
                               8
                                                                      Cooks distance
                                                         ကု
                                       0
            2800
                     3000
                              3200
                                      3400
                                                            0.00
                                                                    0.02
                                                                           0.04
                                                                                   0.06
                                                                                          0.08
                     Fitted values
                                                                          Leverage
### Include age and ftv as covariate
reg2 <- lm(bwt ~ lwt + age + ftv, data=birthData)</pre>
summary(reg2)
##
## Call:
## lm(formula = bwt ~ lwt + age + ftv, data = birthData)
## Residuals:
##
         Min
                    1Q
                          Median
                                         3Q
   -2218.60 -488.73
                           11.56
                                    516.63
                                            1907.42
##
```

```
## Coefficients:
##
              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2223.506
                          301.567
                                   7.373 5.4e-12 ***
                                    2.344 0.0201 *
## lwt
                 4.121
                           1.758
## age
                 7.486
                           10.285
                                    0.728
                                          0.4676
## ftv
                15.364
                           51.114
                                    0.301 0.7641
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 720.9 on 185 degrees of freedom
## Multiple R-squared: 0.03831, Adjusted R-squared: 0.02271
## F-statistic: 2.457 on 3 and 185 DF, p-value: 0.06446
# Just to view the coefficients
summary(reg2)$coefficients
##
                 Estimate Std. Error t value
                                                   Pr(>|t|)
## (Intercept) 2223.505618 301.567140 7.3731694 5.402210e-12
                 4.120744
                           1.757787 2.3442798 2.012498e-02
## age
                 7.485748 10.285178 0.7278189 4.676446e-01
## ftv
                15.363944 51.113921 0.3005824 7.640705e-01
### Prediction
newData <- data.frame(lwt=100, age=25, ftv=0)</pre>
newData
##
    lwt age ftv
## 1 100 25
predict(reg2, newData)
##
## 2822.724
predict(reg2, newData, interval="prediction")
##
         fit
                  lwr
                           upr
## 1 2822.724 1390.148 4255.299
### Use only data from mothers with weight below 160
reg3 <- lm(bwt ~ lwt + age + ftv, data=filter(birthData, lwt<160))
summary(reg3)
##
## Call:
## lm(formula = bwt ~ lwt + age + ftv, data = filter(birthData,
##
      lwt < 160))
##
## Residuals:
               1Q Median
                               3Q
                                      Max
      Min
## -2233.9 -487.0
                     25.7 476.2 1852.5
##
## Coefficients:
              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1824.063
                          444.508
                                   4.104 6.52e-05 ***
                 7.182
                                    2.149 0.0332 *
                            3.342
## lwt
## age
                 9.179
                           11.624
                                    0.790 0.4310
                16.986
                           60.012
                                   0.283 0.7775
## ftv
```

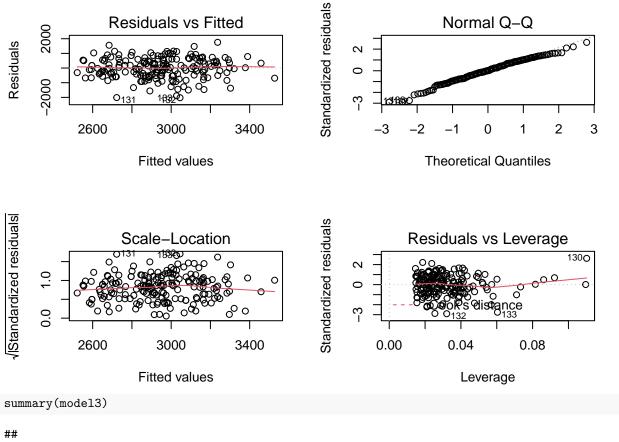
```
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 716.3 on 157 degrees of freedom
## Multiple R-squared: 0.03956,
                                   Adjusted R-squared:
## F-statistic: 2.155 on 3 and 157 DF, p-value: 0.09549
# Just to view the coefficients
summary(reg3)$coefficients
                 Estimate Std. Error t value
                                                   Pr(>|t|)
## (Intercept) 1824.063262 444.508301 4.1035528 0.0000651941
## lwt
                 7.182184
                           3.341787 2.1492047 0.0331503757
## age
                 9.178577 11.624301 0.7896024 0.4309508695
## ftv
              16.986307 60.011783 0.2830495 0.7775117320
ANOVA
### Oneway ANOVA against smoke
oneway1 <- lm(bwt ~ smoke, data=birthData)</pre>
summary(oneway1)
## Call:
## lm(formula = bwt ~ smoke, data = birthData)
## Residuals:
##
      Min
               1Q Median
                               3Q
                                      Max
## -2062.9 -475.9
                     34.3 545.1 1934.3
##
## Coefficients:
              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3055.70
                           66.93 45.653 < 2e-16 ***
                           106.97 -2.653 0.00867 **
## smoke1
               -283.78
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 717.8 on 187 degrees of freedom
## Multiple R-squared: 0.03627,
                                   Adjusted R-squared: 0.03112
## F-statistic: 7.038 on 1 and 187 DF, p-value: 0.008667
summary(oneway1)$coefficients
##
               Estimate Std. Error
                                     t value
                                                  Pr(>|t|)
                          66.93326 45.652875 2.463035e-103
## (Intercept) 3055.6957
## smoke1
              -283.7767 106.96877 -2.652893 8.666726e-03
emmeans(oneway1,~smoke)
  smoke emmean
                  SE df lower.CL upper.CL
##
           3056 66.9 187
                             2924
                                      3188
           2772 83.4 187
                                      2937
## 1
                             2607
## Confidence level used: 0.95
pairs(emmeans(oneway1,~smoke))
```

```
## contrast estimate SE df t.ratio p.value
## 0 - 1
                 284 107 187
                               2.653 0.0087
### Oneway ANOVA against visits
oneway2 <- lm(bwt ~ visits, data=birthData)</pre>
summary(oneway2)
##
## Call:
## lm(formula = bwt ~ visits, data = birthData)
## Residuals:
##
       Min
                 1Q
                     Median
                                   3Q
                                            Max
## -2156.14 -484.88
                       26.12
                               578.86 1882.00
##
## Coefficients:
##
                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                      2865.14
                                   72.62 39.456
                                                   <2e-16 ***
## visitsOnce
                       242.86
                                  128.42
                                          1.891
                                                   0.0602 .
## visitsMoreThanOnce
                        85.74
                                  133.52
                                          0.642
                                                  0.5216
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 726.2 on 186 degrees of freedom
## Multiple R-squared: 0.01888,
                                   Adjusted R-squared:
## F-statistic: 1.79 on 2 and 186 DF, p-value: 0.1698
emmeans(oneway2,~visits)
## visits
                          SE df lower.CL upper.CL
                emmean
## Never
                 2865 72.6 186
                                     2722
                                              3008
## Once
                  3108 105.9 186
                                     2899
                                              3317
## MoreThanOnce
                  2951 112.1 186
                                     2730
                                              3172
## Confidence level used: 0.95
pairs(emmeans(oneway2,~visits))
## contrast
                        estimate SE df t.ratio p.value
## Never - Once
                          -242.9 128 186 -1.891 0.1441
## Never - MoreThanOnce
                           -85.7 134 186 -0.642 0.7970
## Once - MoreThanOnce
                           157.1 154 186
                                          1.019 0.5658
##
## P value adjustment: tukey method for comparing a family of 3 estimates
drop1(oneway2,test="F")
## Single term deletions
##
## Model:
## bwt ~ visits
         Df Sum of Sq
                           RSS
                                  AIC F value Pr(>F)
## <none>
                      98081730 2493.2
## visits 2 1887925 99969656 2492.8 1.7901 0.1698
### Twoway ANOVA without interaction
twoway1 <- lm(bwt ~ visits + smoke, data=birthData)</pre>
```

```
summary(twoway1)
##
## Call:
## lm(formula = bwt ~ visits + smoke, data = birthData)
## Residuals:
##
       Min
                 1Q
                      Median
                                   3Q
                                           Max
## -2034.02 -504.02
                       34.98 536.36 1816.31
##
## Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
##
                      2980.92
                                   86.74 34.367
## (Intercept)
                                                   <2e-16 ***
## visitsOnce
                       192.77
                                  128.60
                                           1.499
                                                   0.1356
## visitsMoreThanOnce
                       74.10
                                  131.98
                                          0.561
                                                   0.5752
                      -257.28
                                  108.37 -2.374
                                                   0.0186 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 717.3 on 185 degrees of freedom
## Multiple R-squared: 0.04789, Adjusted R-squared: 0.03245
## F-statistic: 3.102 on 3 and 185 DF, p-value: 0.02795
summary(twoway1)$coefficients
                       Estimate Std. Error
                                              t value
## (Intercept)
                     2980.91672
                                86.73734 34.3671676 3.021284e-82
## visitsOnce
                      192.77220 128.59583 1.4990548 1.355639e-01
## visitsMoreThanOnce
                      74.10202 131.98092 0.5614601 5.751634e-01
                     -257.28159 108.37493 -2.3739954 1.862096e-02
## smoke1
### Twoway ANOVA with interaction, test for interaction in two ways
twoway2 <- lm(bwt ~ visits * smoke, data=birthData)</pre>
summary(twoway2)
##
## Call:
## lm(formula = bwt ~ visits * smoke, data = birthData)
## Residuals:
       Min
                 1Q
                      Median
                                   3Q
                                           Max
## -2045.82 -460.29
                       50.71
                               505.18 1713.77
##
## Coefficients:
##
                            Estimate Std. Error t value Pr(>|t|)
                                          96.22 30.714 <2e-16 ***
## (Intercept)
                             2955.40
                                         154.30
## visitsOnce
                                                  2.079
                              320.83
                                                           0.039 *
## visitsMoreThanOnce
                               12.20
                                         172.13
                                                  0.071
                                                           0.944
## smoke1
                             -200.58
                                         143.44 -1.398
                                                           0.164
## visitsOnce:smoke1
                             -458.32
                                         278.50 -1.646
                                                           0.102
## visitsMoreThanOnce:smoke1
                              159.27
                                         266.27
                                                  0.598
                                                           0.550
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 713.6 on 183 degrees of freedom
```

```
## Multiple R-squared: 0.06783,
                                  Adjusted R-squared: 0.04236
## F-statistic: 2.663 on 5 and 183 DF, p-value: 0.02378
anova(twoway2, twoway1)
## Analysis of Variance Table
## Model 1: bwt ~ visits * smoke
## Model 2: bwt ~ visits + smoke
   Res.Df
              RSS Df Sum of Sq
                                    F Pr(>F)
## 1
       183 93189162
       185 95182096 -2 -1992934 1.9568 0.1443
drop1(twoway2,test="F")
## Single term deletions
##
## Model:
## bwt ~ visits * smoke
              Df Sum of Sq
                                 RSS
                                     AIC F value Pr(>F)
## <none>
                            93189162 2489.5
## visits:smoke 2 1992934 95182096 2489.5 1.9568 0.1443
# Compute the expected birth weight of infants for smokers and non-smokers, respectively, on average ov
emmeans(twoway2, ~smoke)
## NOTE: Results may be misleading due to involvement in interactions
   smoke emmean
                  SE df lower.CL upper.CL
## 0
           3066 70.1 183
                             2928
## 1
           2766 96.4 183
                             2576
                                      2956
## Results are averaged over the levels of: visits
## Confidence level used: 0.95
Models with numerical as well as categorical predictors
### Model with linear (lwt,bwt) association.
### Intercept differ between smokers and non-smokers, one common slope.
model1 <- lm(bwt ~ lwt + smoke, data=birthData)</pre>
summary(model1)
##
## Call:
## lm(formula = bwt ~ lwt + smoke, data = birthData)
## Residuals:
                 1Q
                     Median
                                   3Q
                       29.16 521.76 1967.76
## -2030.90 -445.69
##
## Coefficients:
              Estimate Std. Error t value Pr(>|t|)
                          230.836 10.835
## (Intercept) 2501.125
                                           <2e-16 ***
                 4.237
                          1.690
                                  2.507
                                            0.0130 *
## smoke1
             -272.081
                          105.591 -2.577 0.0107 *
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 707.8 on 186 degrees of freedom
## Multiple R-squared: 0.06777,
                                   Adjusted R-squared: 0.05775
## F-statistic: 6.761 on 2 and 186 DF, p-value: 0.001464
### Model with linear (lwt,bwt) association. Intercept and slope both differ between visit groups
model2 <- lm(bwt ~ lwt * smoke, data=birthData)</pre>
summary(model2)
##
## Call:
## lm(formula = bwt ~ lwt * smoke, data = birthData)
## Residuals:
       \mathtt{Min}
                 1Q
                     Median
                                   ЗQ
                                            Max
## -2038.80 -454.76
                       28.36 530.84 1976.84
## Coefficients:
              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2350.578
                          312.733 7.516 2.35e-12 ***
                 5.387
                            2.335
                                   2.307 0.0222 *
                41.384
                                   0.092
## smoke1
                          451.187
                                            0.9270
## lwt:smoke1
                -2.422
                            3.388 -0.715 0.4757
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 708.8 on 185 degrees of freedom
## Multiple R-squared: 0.07034,
                                  Adjusted R-squared: 0.05527
## F-statistic: 4.666 on 3 and 185 DF, p-value: 0.003621
### Test if slopes differ between visit groups
anova(model2, model1)
## Analysis of Variance Table
##
## Model 1: bwt ~ lwt * smoke
## Model 2: bwt ~ lwt + smoke
   Res.Df
                RSS Df Sum of Sq
                                      F Pr(>F)
## 1
       185 92937722
       186 93194298 -1 -256576 0.5107 0.4757
### Model with many effects (no interactions)
model3 <- lm(bwt ~ lwt + smoke + age + visits, data=birthData)</pre>
par(mfrow=c(2,2))
plot(model3)
```



```
## Call:
## lm(formula = bwt ~ lwt + smoke + age + visits, data = birthData)
##
## Residuals:
##
        Min
                  1Q
                       Median
                                     3Q
                                             Max
   -2024.13 -491.65
                         6.56
                                 507.56
                                         1753.25
##
## Coefficients:
##
                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                      2346.442
                                   301.420
                                             7.785
                                                       5e-13 ***
                                     1.727
                                             2.413
                                                     0.0168 *
## lwt
                          4.167
## smoke1
                       -244.402
                                   107.187
                                            -2.280
                                                     0.0238 *
                                                     0.6754
## age
                         4.300
                                    10.252
                                             0.419
## visitsOnce
                        184.314
                                   130.194
                                             1.416
                                                     0.1586
                                             0.242
                                                     0.8089
## visitsMoreThanOnce
                         32.318
                                   133.410
##
                  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Signif. codes:
##
## Residual standard error: 708.6 on 183 degrees of freedom
## Multiple R-squared: 0.08076,
                                     Adjusted R-squared: 0.05564
## F-statistic: 3.215 on 5 and 183 DF, p-value: 0.008306
```

Logistic regression

```
### Logistic regression with many predictors (no interactions)
logreg1 <- glm(low ~ lwt + smoke + age + visits, data=birthData, family="binomial")</pre>
```

```
summary(logreg1)
##
## Call:
## glm(formula = low ~ lwt + smoke + age + visits, family = "binomial",
      data = birthData)
##
## Deviance Residuals:
      Min 10 Median
                                  3Q
                                          Max
## -1.3128 -0.8665 -0.6732 1.2390
                                       2.0197
##
## Coefficients:
##
                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)
                      1.370311 1.016623 1.348 0.1777
                     -0.012276 0.006138 -2.000
                                                  0.0455 *
## lwt
## smoke1
                      0.619077 0.330428 1.874
                                                  0.0610 .
## age
                     -0.031764 0.033933 -0.936
                                                   0.3492
## visitsOnce
                     -0.413043
                                0.424527 -0.973
                                                   0.3306
## visitsMoreThanOnce -0.148285
                                0.420965 -0.352 0.7247
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## (Dispersion parameter for binomial family taken to be 1)
##
      Null deviance: 234.67 on 188 degrees of freedom
##
## Residual deviance: 221.91 on 183 degrees of freedom
## AIC: 233.91
##
## Number of Fisher Scoring iterations: 4
Linear mixed models
### Make artificial center variable
set.seed(123)
center <- sample(rep(1:19, each=10)[1:189])</pre>
birthData <- mutate(birthData,center=factor(center))</pre>
### Remember to install lme4 before this can run
# install.packages("lme4")
library(lme4)
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following objects are masked from 'package:tidyr':
##
      expand, pack, unpack
### Linear mixed model with random effect of center
lmm1 <- lmer(bwt ~ lwt + smoke + age + visits + (1|center), data=birthData)</pre>
```

boundary (singular) fit: see ?isSingular

summary(lmm1)

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: bwt ~ lwt + smoke + age + visits + (1 | center)
##
     Data: birthData
##
## REML criterion at convergence: 2958
##
## Scaled residuals:
##
       Min
            1Q
                    Median
## -2.85637 -0.69380 0.00926 0.71625 2.47412
##
## Random effects:
## Groups Name
                      Variance Std.Dev.
## center (Intercept) 9.668e-12 3.109e-06
## Residual
                       5.022e+05 7.086e+02
## Number of obs: 189, groups: center, 19
## Fixed effects:
                    Estimate Std. Error t value
## (Intercept)
                    2346.442 301.420
                                       7.785
## lwt
                       4.167
                               1.727
                                        2.413
## smoke1
                    -244.402 107.187 -2.280
                      4.300
                               10.252 0.419
## age
                                       1.416
## visitsOnce
                    184.314
                               130.194
## visitsMoreThanOnce 32.318 133.410
                                       0.242
## Correlation of Fixed Effects:
##
             (Intr) lwt
                           smoke1 age
                                      vsts0n
## lwt
             -0.613
            -0.196 0.046
## smoke1
             -0.620 -0.173 0.003
## age
## visitsOnce -0.026 0.048 0.160 -0.218
## vstsMrThnOn 0.055 -0.058 0.031 -0.190 0.335
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```