Homework 3

Problem 1- Insurance Claims! (IGLM 9.2)

The attached `insurance.csv` dataset contains information on the numbers of automobile insurance policies (`n`) and the number of claims (`y`), tabulated by the car's insurance category (`car`), the age category of the policy-holder (`age`), and the region where the policy-holder lives (`dist`, equal to 1 if the policy-holder lives in a major city, and 0 elsewhere). This data is derived from another dataset in Aitkin et. al. (1989).

a. Use *Poisson regression* to fit a model relating the number of claims `y`, against all tabulating variables (`car`, `age`, and `dist`), as well as all interaction terms between the categorical variables. Be aware that by default `R` will read the category labels as numbers! Use eg. `as.factor` to convert them to categorical.

```
library(tidyverse)
library(magrittr)
dat = read.csv('../../data/insurance.csv')

# largest possible interaction model:
mod1 = glm( y~ n + as.factor(car)*as.factor(age)*as.factor(dist), data=dat, family=poisson)
mod1 %>% summary
```

```
##
## Call:
## glm(formula = y ~ n + as.factor(car) * as.factor(age) * as.factor(dist),
##     family = poisson, data = dat)
##
## Deviance Residuals:
##  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [26]  0  0  0  0  0  0  0
##
## Coefficients: (1 not defined because of singularities)
##                                                 Estimate Std. Error z value
## (Intercept)                                    2.964e-01  7.532e+03   0.000
## n                                              1.223e-02  2.376e+01   0.001
## as.factor(car)2                               -1.657e+00  4.016e+03   0.000
## as.factor(car)3                                6.891e-01  2.234e+03   0.000
## as.factor(car)4                                1.612e+00  6.582e+03   0.000
## as.factor(age)2                               -1.945e+00  3.778e+03  -0.001
## as.factor(age)3                               -2.291e+00  4.016e+03  -0.001
## as.factor(age)4                               -3.443e+01  6.991e+04   0.000
## as.factor(dist)1                               1.520e-01  7.057e+03   0.000
## as.factor(car)2:as.factor(age)2               -3.908e+00  8.530e+03   0.000
## as.factor(car)3:as.factor(age)2               -8.720e-01  3.730e+03   0.000
## as.factor(car)4:as.factor(age)2                1.781e+00  1.212e+03   0.001
## as.factor(car)2:as.factor(age)3               -7.760e+00  1.663e+04   0.000
## as.factor(car)3:as.factor(age)3               -2.302e+00  7.247e+03   0.000
## as.factor(car)4:as.factor(age)3                1.428e+00  1.188e+02   0.012
## as.factor(car)2:as.factor(age)4               -5.114e+01  1.006e+05  -0.001
## as.factor(car)3:as.factor(age)4               -2.497e+00  6.582e+03   0.000
```

```
## as.factor(car)4:as.factor(age)4                    2.517e+01  4.664e+04   0.001
## as.factor(car)2:as.factor(dist)1                   2.775e+00  3.754e+03   0.001
## as.factor(car)3:as.factor(dist)1                   2.516e-01  2.186e+03   0.000
## as.factor(car)4:as.factor(dist)1                  -2.440e+01  4.843e+04  -0.001
## as.factor(age)2:as.factor(dist)1                   2.702e+00  3.469e+03   0.001
## as.factor(age)3:as.factor(dist)1                   2.739e+00  3.540e+03   0.001
## as.factor(age)4:as.factor(dist)1                   3.370e+01  6.287e+04   0.001
## as.factor(car)2:as.factor(age)2:as.factor(dist)1  2.896e+00  7.651e+03   0.000
## as.factor(car)3:as.factor(age)2:as.factor(dist)1  1.943e-01  3.540e+03   0.000
## as.factor(car)4:as.factor(age)2:as.factor(dist)1  2.140e+01  4.104e+04   0.001
## as.factor(car)2:as.factor(age)3:as.factor(dist)1  7.344e+00  1.495e+04   0.000
## as.factor(car)3:as.factor(age)3:as.factor(dist)1  2.405e+00  6.534e+03   0.000
## as.factor(car)4:as.factor(age)3:as.factor(dist)1  2.224e+01  4.232e+04   0.001
## as.factor(car)2:as.factor(age)4:as.factor(dist)1  4.607e+01  9.112e+04   0.001
## as.factor(car)3:as.factor(age)4:as.factor(dist)1  1.773e+00  5.869e+03   0.000
## as.factor(car)4:as.factor(age)4:as.factor(dist)1         NA         NA      NA
##                                                  Pr(>|z|)
## (Intercept)                                         1.000
## n                                                   1.000
## as.factor(car)2                                     1.000
## as.factor(car)3                                     1.000
## as.factor(car)4                                     1.000
## as.factor(age)2                                     1.000
## as.factor(age)3                                     1.000
## as.factor(age)4                                     1.000
## as.factor(dist)1                                    1.000
## as.factor(car)2:as.factor(age)2                     1.000
## as.factor(car)3:as.factor(age)2                     1.000
## as.factor(car)4:as.factor(age)2                     0.999
## as.factor(car)2:as.factor(age)3                     1.000
## as.factor(car)3:as.factor(age)3                     1.000
## as.factor(car)4:as.factor(age)3                     0.990
## as.factor(car)2:as.factor(age)4                     1.000
## as.factor(car)3:as.factor(age)4                     1.000
## as.factor(car)4:as.factor(age)4                     1.000
## as.factor(car)2:as.factor(dist)1                    0.999
## as.factor(car)3:as.factor(dist)1                    1.000
## as.factor(car)4:as.factor(dist)1                    1.000
## as.factor(age)2:as.factor(dist)1                    0.999
## as.factor(age)3:as.factor(dist)1                    0.999
## as.factor(age)4:as.factor(dist)1                    1.000
## as.factor(car)2:as.factor(age)2:as.factor(dist)1    1.000
## as.factor(car)3:as.factor(age)2:as.factor(dist)1    1.000
## as.factor(car)4:as.factor(age)2:as.factor(dist)1    1.000
## as.factor(car)2:as.factor(age)3:as.factor(dist)1    1.000
## as.factor(car)3:as.factor(age)3:as.factor(dist)1    1.000
## as.factor(car)4:as.factor(age)3:as.factor(dist)1    1.000
## as.factor(car)2:as.factor(age)4:as.factor(dist)1    1.000
## as.factor(car)3:as.factor(age)4:as.factor(dist)1    1.000
## as.factor(car)4:as.factor(age)4:as.factor(dist)1       NA
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5.6606e+03  on 31  degrees of freedom
```

```
## Residual deviance: 4.1224e-10  on  0  degrees of freedom
## AIC: 232.36
##
## Number of Fisher Scoring iterations: 20
```

b. Based on the modeling in (a) above, Aitkin et. al. determined that all the interaction terms were insignificant, and that both `car` and `age` could be treated as continuous variables rather than categories. Fit this new model, and compare to the model in (a) above. What conclusions do you reach

```
mod2 = glm( y~ n + car + age + as.factor(dist), data=dat, family=poisson)
mod2 %>% summary
```

```
##
## Call:
## glm(formula = y ~ n + car + age + as.factor(dist), family = poisson,
##     data = dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -7.505  -2.967  -1.253   1.614   8.008
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.395e+00  8.330e-02  40.757   <2e-16 ***
## n                 2.196e-04  8.308e-06  26.436   <2e-16 ***
## car              -4.363e-02  1.901e-02  -2.295   0.0218 *
## age               4.602e-01  2.466e-02  18.660   <2e-16 ***
## as.factor(dist)1 -1.596e+00  6.305e-02 -25.311   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5660.59  on 31  degrees of freedom
## Residual deviance:  501.79  on 27  degrees of freedom
## AIC: 680.15
##
## Number of Fisher Scoring iterations: 5
```

Compare between M_1 and M_2 using a deviance hypothesis test. Be careful here because M_2 is actually our "null model" in this case, despite the fact that we usually use M_1 to denote the null model:

```
D1 = 0
df1 = 0
D2 = 501.79
df2 = 27

delta.D = D2 - D1
df = df2-df1

p.val = pchisq(delta.D,df,lower.tail=FALSE)
p.val
```

```
## [1] 6.611538e-89
```

Since the p-value is less than any reasonable choice of α, we *reject* M_2 (the null model) in favor of M_1. Aitkin et. al. appear to have been mislead by the Wald test p-values: the interaction terms **do** matter in aggregate,

even if all together no single one matters.

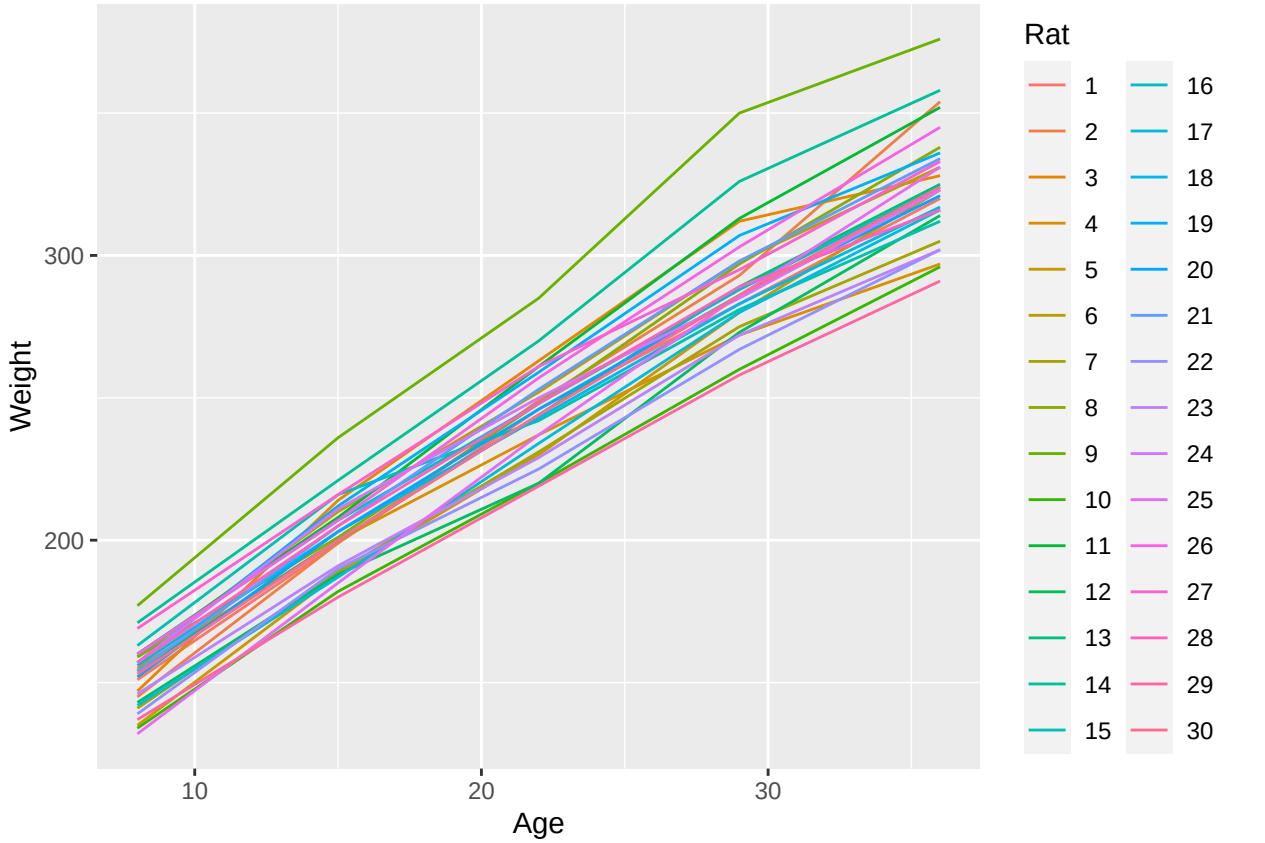
Note that here we could have also used the model AICs: the original (interaction) model has an AIC of ~232, far smaller than the reduced model's AIC of ~680. This indicates that M_1 substantially outperforms M_2 in terms of predictive power, even accounting for the large difference in model sizes.

Problem 2- Rats!

The attached `rats.csv` contains birthweights for 30 baby rats (in grams), measured every week for 5 weeks. For this problem we'll denote by Y_{jk} the weight of the j-th rat at age x_{jk} (in days) where $j = 1, ..., 30$ and $k = 1, .., 5$

a. Conduct an exploratory analysis of the data. Provide plots of rat-specific growth trajectories (piecewise-linear connections between observations for each rat, in a single figure). If you're using R this will probably be easiest in `ggplot2`, while if you're using Python you may want to checkout `plotnine` or `altair`.

```
dat = read.csv('../../data/rats.csv')

ggplot(dat,aes(x=age,y=weight,color=as.factor(rat))) + geom_line() + labs(x='Age',y='Weight',color='Rat
```



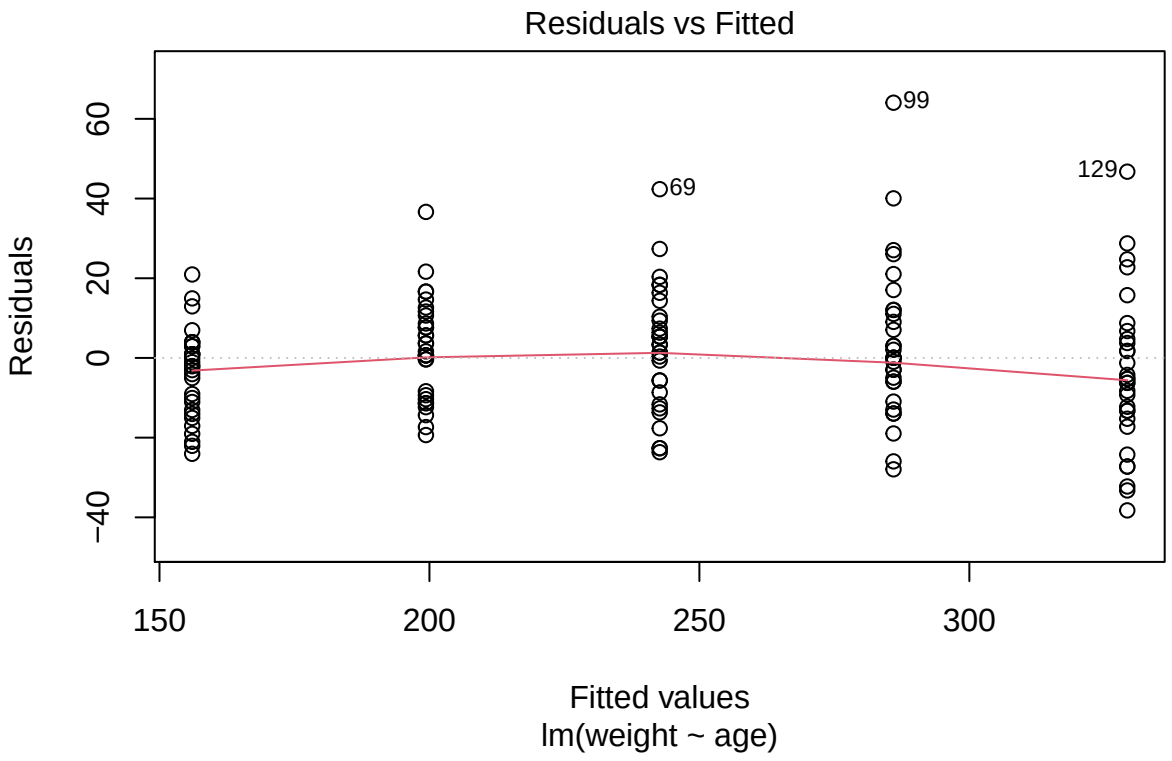
From this plot we see that while the *initial weight* appears to vary by rat, the actual growth rates in a given week do not appear to differ much. We should not expect a varying slope model to show a lot of differentiation across individuals.

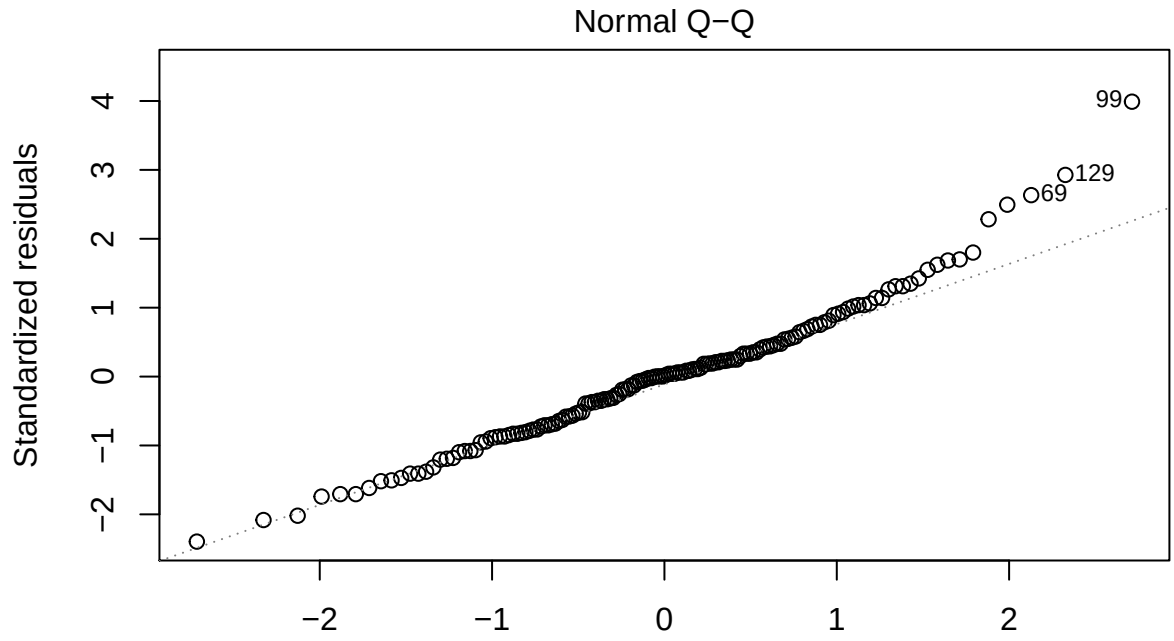
b. Fit the linear model $E[Y_{jk}] = \alpha + \beta x_{jk}$, assuming the variables Y_{jk} are all independent (ie. ignoring the fact that measurements corresponding to the same rat will have some correlation). Provide a summary of this model, check all diagnostics, and discuss the adequacy of the linear regression assumptions.

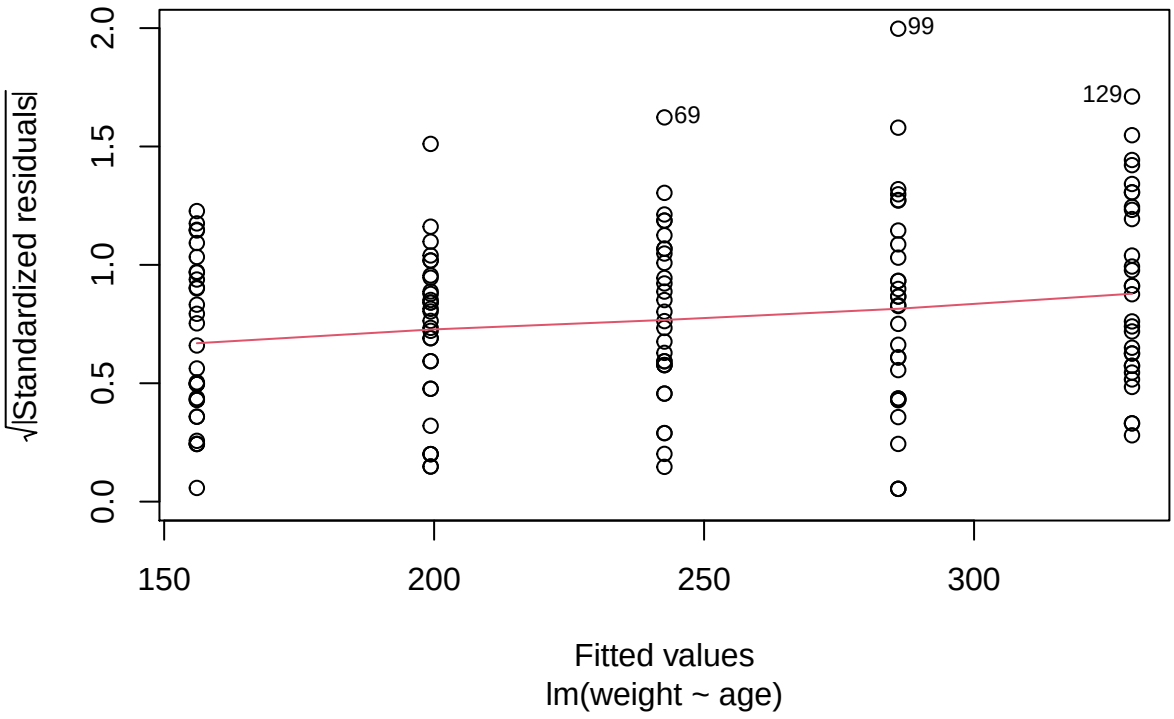
```r
mod.lm = lm(weight~age,data=dat)
summary(mod.lm)
```

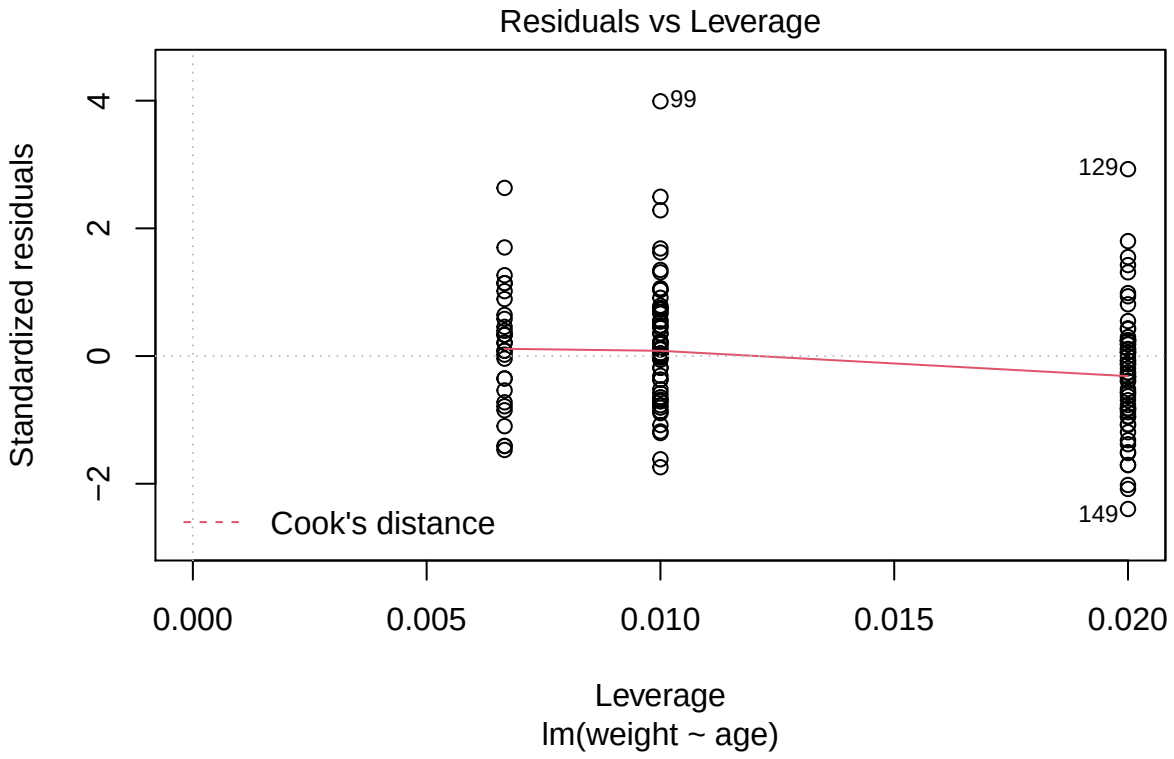
```
##
## Call:
## lm(formula = weight ~ age, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.253 -11.278   0.197   7.647  64.047
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 106.5676     3.2099   33.20   <2e-16 ***
## age           6.1857     0.1331   46.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.13 on 148 degrees of freedom
## Multiple R-squared:  0.9359, Adjusted R-squared:  0.9355
## F-statistic:  2161 on 1 and 148 DF,  p-value: < 2.2e-16
```

```r
plot(mod.lm)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(weight ~ age)
99
129
69


Scale−Location
√|Standardized residuals|
Fitted values
lm(weight ~ age)
99
129
69



Linearity is not very well satisified. Looking at the plot of the data this isn't surprising, we see a pretty clear "taper" in the growth rate in the last week or so. Heteroskedasticity also seems to be occurring here- could be caused by the non-linearity.

c. Fit the linear model $E[Y_{jk}] = \alpha_j + \beta_j x_{jk}$ (note that here the intercept and slope vary across individuals j). Provide a summary of this model, check all diagnostics, and discuss the adequacy of the linear regression assumptions.

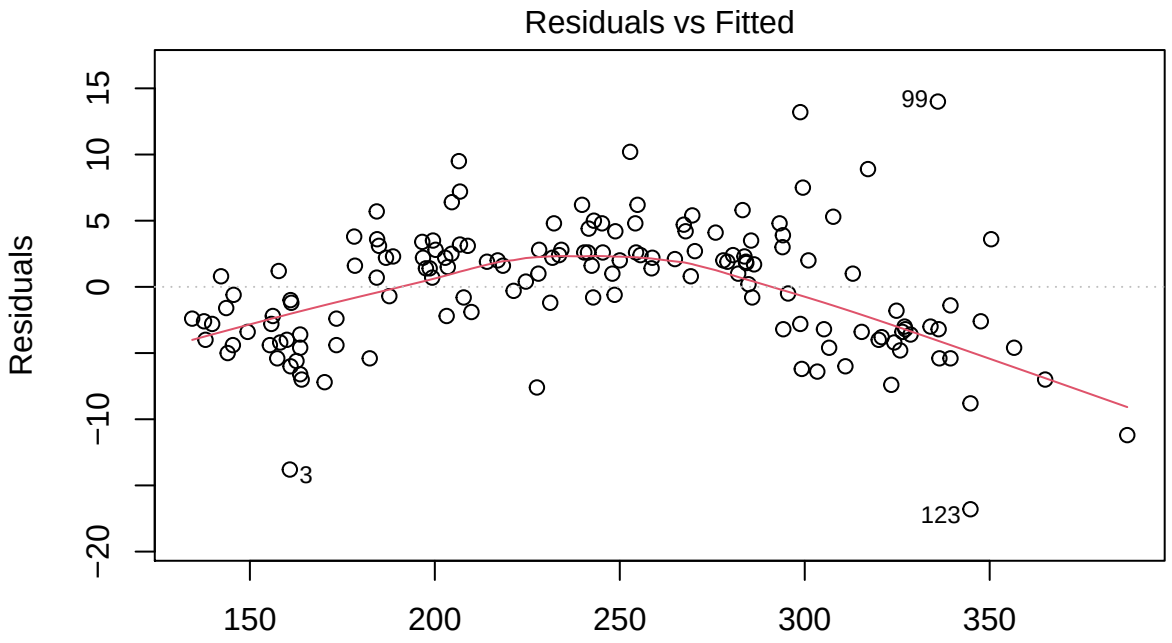
```
mod.lm.interaction = lm(weight~age*as.factor(rat),data=dat)
summary(mod.lm.interaction)
```

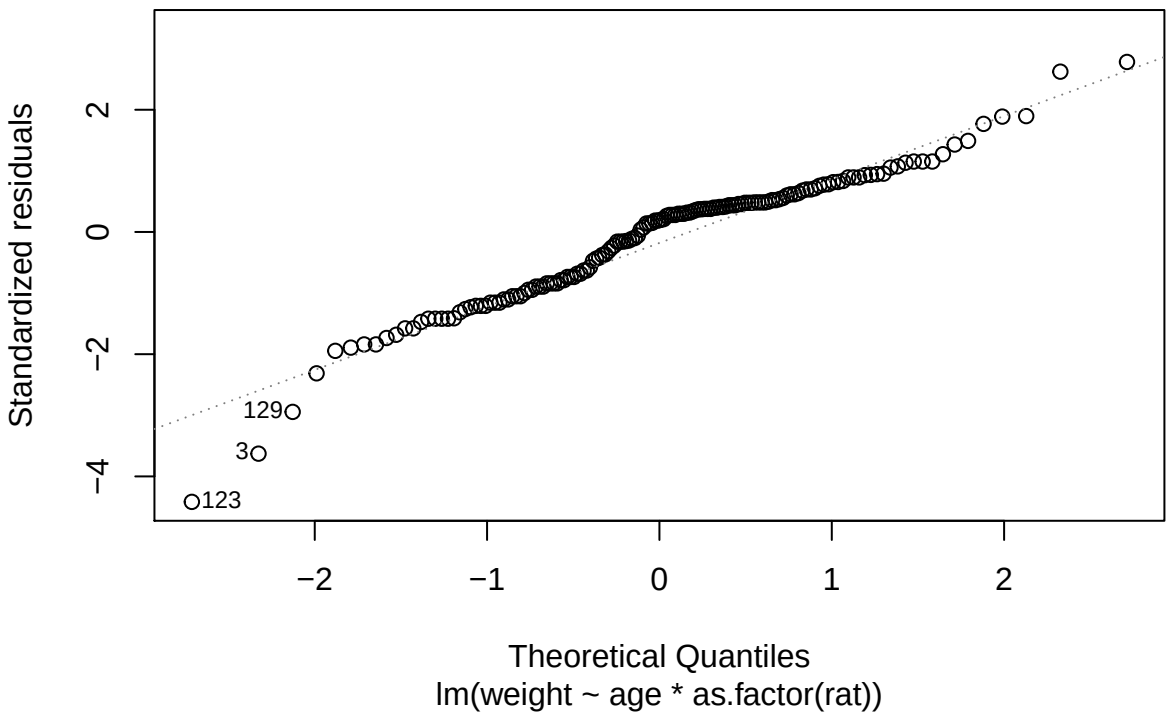
```
##
## Call:
## lm(formula = weight ~ age * as.factor(rat), data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.800  -3.350   1.000   2.775  14.000
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          107.17143    6.55498  16.350  < 2e-16 ***
## age                    6.02857    0.27171  22.187  < 2e-16 ***
## as.factor(rat)2      -20.08571    9.27013  -2.167  0.03290 *
## as.factor(rat)3        1.05714    9.27013   0.114  0.90946
## as.factor(rat)4       13.14286    9.27013   1.418  0.15971
## as.factor(rat)5      -23.05714    9.27013  -2.487  0.01472 *
## as.factor(rat)6        7.05714    9.27013   0.761  0.44848
## as.factor(rat)7       -9.08571    9.27013  -0.980  0.32966
## as.factor(rat)8       -1.25714    9.27013  -0.136  0.89243
## as.factor(rat)9       16.71429    9.27013   1.803  0.07473 .
```

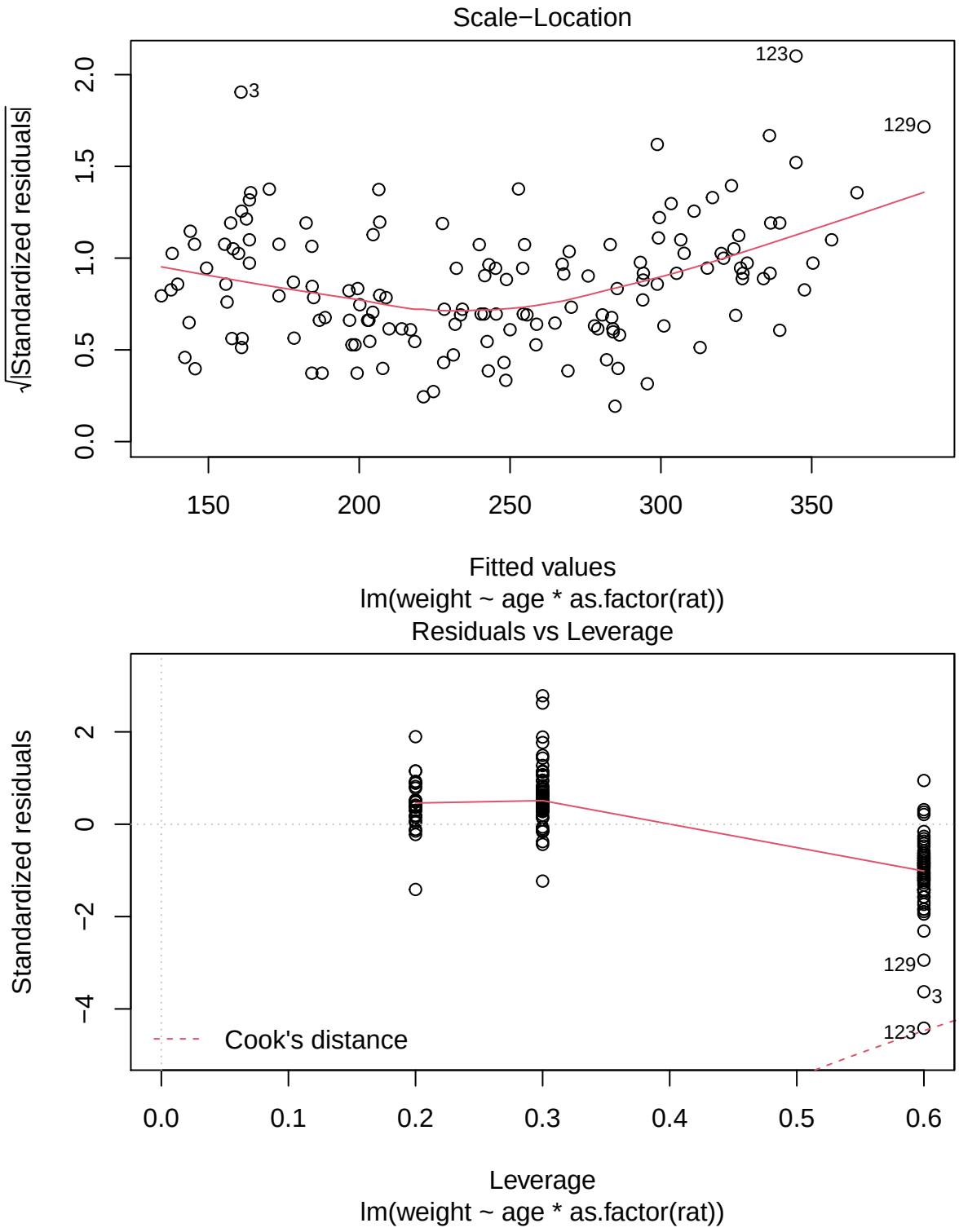
```
## as.factor(rat)10     -15.11429    9.27013  -1.630  0.10651
## as.factor(rat)11      -2.05714    9.27013  -0.222  0.82489
## as.factor(rat)12     -13.77143    9.27013  -1.486  0.14089
## as.factor(rat)13      -0.22857    9.27013  -0.025  0.98038
## as.factor(rat)14      11.48571    9.27013   1.239  0.21857
## as.factor(rat)15      21.54286    9.27013   2.324  0.02238 *
## as.factor(rat)16       9.68571    9.27013   1.045  0.29890
## as.factor(rat)17     -13.97143    9.27013  -1.507  0.13528
## as.factor(rat)18       6.88571    9.27013   0.743  0.45955
## as.factor(rat)19       4.65714    9.27013   0.502  0.61663
## as.factor(rat)20       2.11429    9.27013   0.228  0.82011
## as.factor(rat)21      -0.74286    9.27013  -0.080  0.93631
## as.factor(rat)22      -9.22857    9.27013  -0.996  0.32215
## as.factor(rat)23      -2.68571    9.27013  -0.290  0.77270
## as.factor(rat)24      10.42857    9.27013   1.125  0.26360
## as.factor(rat)25     -29.80000    9.27013  -3.215  0.00181 **
## as.factor(rat)26       0.77143    9.27013   0.083  0.93386
## as.factor(rat)27      19.71429    9.27013   2.127  0.03619 *
## as.factor(rat)28       9.48571    9.27013   1.023  0.30893
## as.factor(rat)29     -11.48571    9.27013  -1.239  0.21857
## as.factor(rat)30      -0.28571    9.27013  -0.031  0.97548
## age:as.factor(rat)2    1.28571    0.38426   3.346  0.00120 **
## age:as.factor(rat)3    0.54286    0.38426   1.413  0.16118
## age:as.factor(rat)4   -0.94286    0.38426  -2.454  0.01607 *
## age:as.factor(rat)5    0.65714    0.38426   1.710  0.09068 .
## age:as.factor(rat)6    0.14286    0.38426   0.372  0.71094
## age:as.factor(rat)7   -0.11429    0.38426  -0.297  0.76683
## age:as.factor(rat)8    0.45714    0.38426   1.190  0.23730
## age:as.factor(rat)9    1.28571    0.38426   3.346  0.00120 **
## age:as.factor(rat)10  -0.28571    0.38426  -0.744  0.45909
## age:as.factor(rat)11   0.95714    0.38426   2.491  0.01458 *
## age:as.factor(rat)12   0.07143    0.38426   0.186  0.85295
## age:as.factor(rat)13   0.12857    0.38426   0.335  0.73871
## age:as.factor(rat)14   0.81429    0.38426   2.119  0.03684 *
## age:as.factor(rat)15  -0.84286    0.38426  -2.193  0.03085 *
## age:as.factor(rat)16  -0.18571    0.38426  -0.483  0.63005
## age:as.factor(rat)17   0.27143    0.38426   0.706  0.48178
## age:as.factor(rat)18  -0.28571    0.38426  -0.744  0.45909
## age:as.factor(rat)19   0.44286    0.38426   1.152  0.25217
## age:as.factor(rat)20  -0.01429    0.38426  -0.037  0.97043
## age:as.factor(rat)21   0.44286    0.38426   1.152  0.25217
## age:as.factor(rat)22  -0.27143    0.38426  -0.706  0.48178
## age:as.factor(rat)23  -0.41429    0.38426  -1.078  0.28385
## age:as.factor(rat)24  -0.22857    0.38426  -0.595  0.55345
## age:as.factor(rat)25   1.10000    0.38426   2.863  0.00523 **
## age:as.factor(rat)26   0.62857    0.38426   1.636  0.10537
## age:as.factor(rat)27  -0.21429    0.38426  -0.558  0.57846
## age:as.factor(rat)28  -0.28571    0.38426  -0.744  0.45909
## age:as.factor(rat)29  -0.51429    0.38426  -1.338  0.18414
## age:as.factor(rat)30   0.08571    0.38426   0.223  0.82399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.015 on 90 degrees of freedom
```

```
## Multiple R-squared:  0.9946, Adjusted R-squared:  0.991
## F-statistic: 280.1 on 59 and 90 DF,  p-value: < 2.2e-16
```

```
plot(mod.lm.interaction)
```







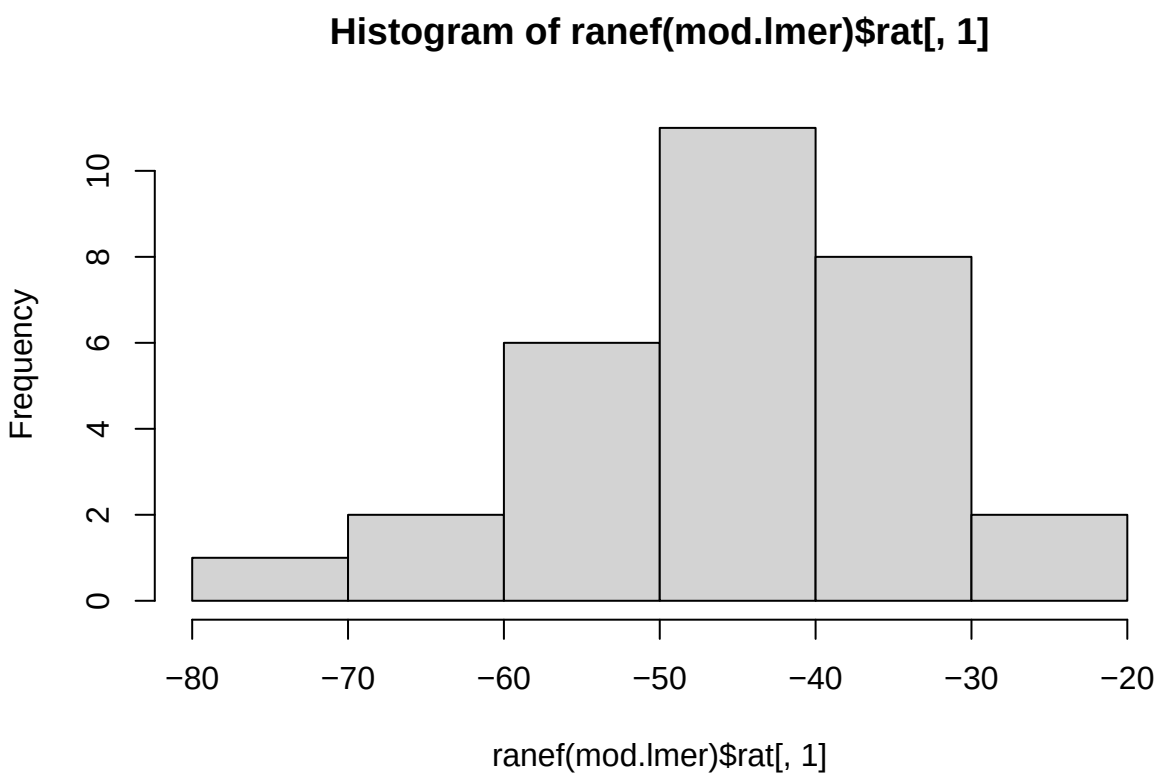
Linearity definitely not satisfied, and there appear to be some high-influence datapoints. Nevertheless, adjusted R^2 is higher than for the non-interaction model (maybe due to high influence points), but clearly some evidence that there are differences across individual rats.

d. Fit the hierarchical model $E[Y_{jk}] = \alpha_j + \beta_j x_{jk}$ where $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$ and $\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$. Provide a summary table.

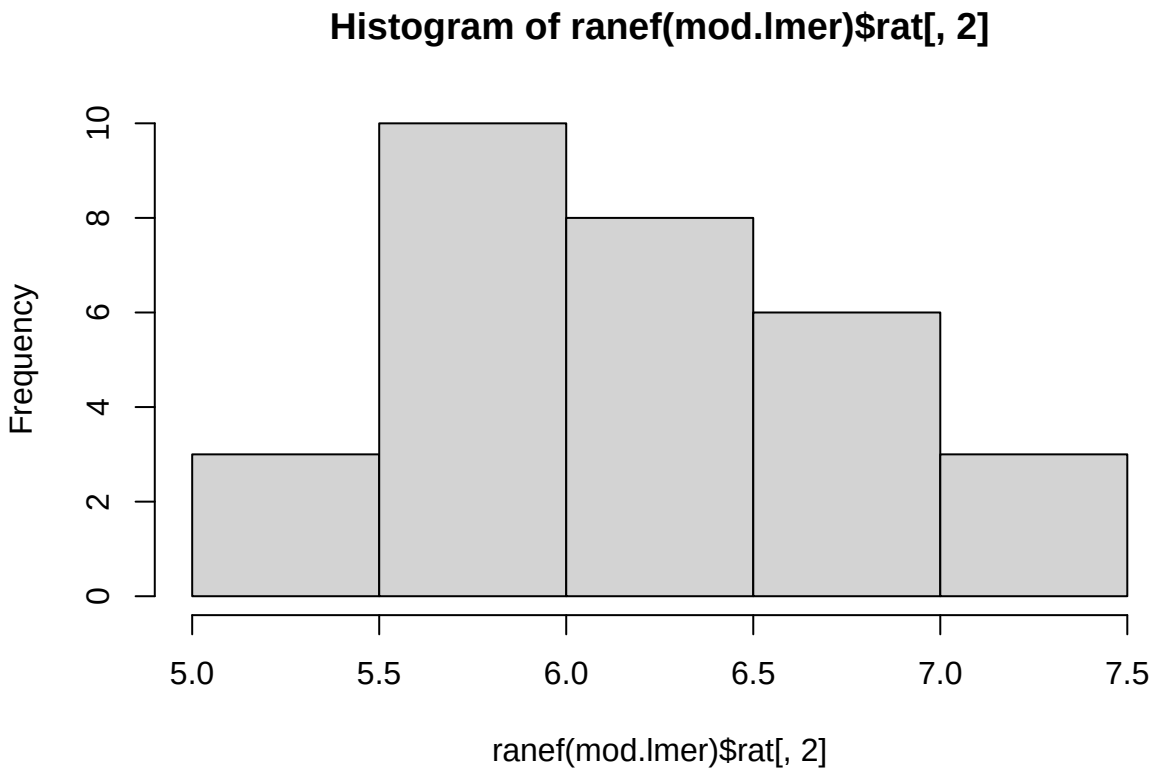
```
library(lme4)
mod.lmer = lmer(weight~(age|rat),data=dat)
summary(mod.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ (age | rat)
##    Data: dat
##
## REML criterion at convergence: 1236.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8060 -0.5793  0.1341  0.4738  2.3355
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  rat      (Intercept) 2247.74  47.410
##           age           38.51   6.206   -0.97
##  Residual               36.18   6.015
## Number of obs: 150, groups:  rat, 30
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  152.751      2.158   70.78
```

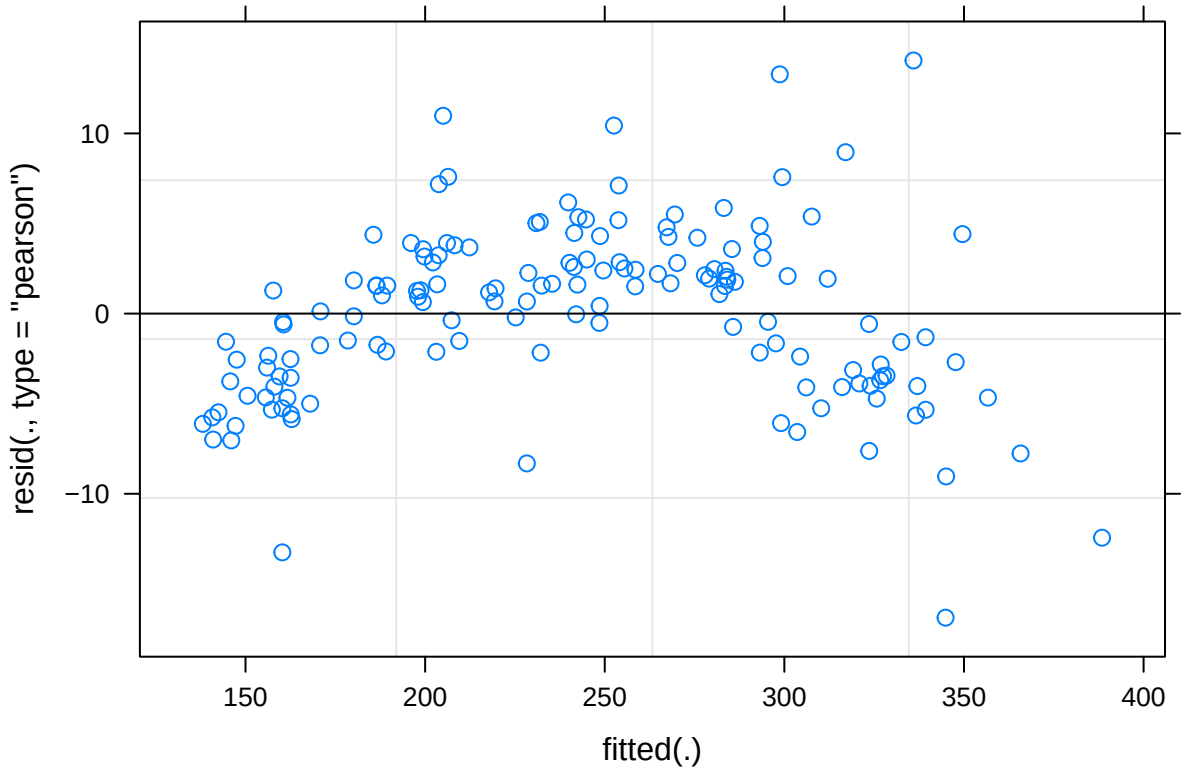
```
hist(ranef(mod.lmer)$rat[,1])
```



```
hist(ranef(mod.lmer)$rat[,2])
```




```
plot(mod.lmer)
```

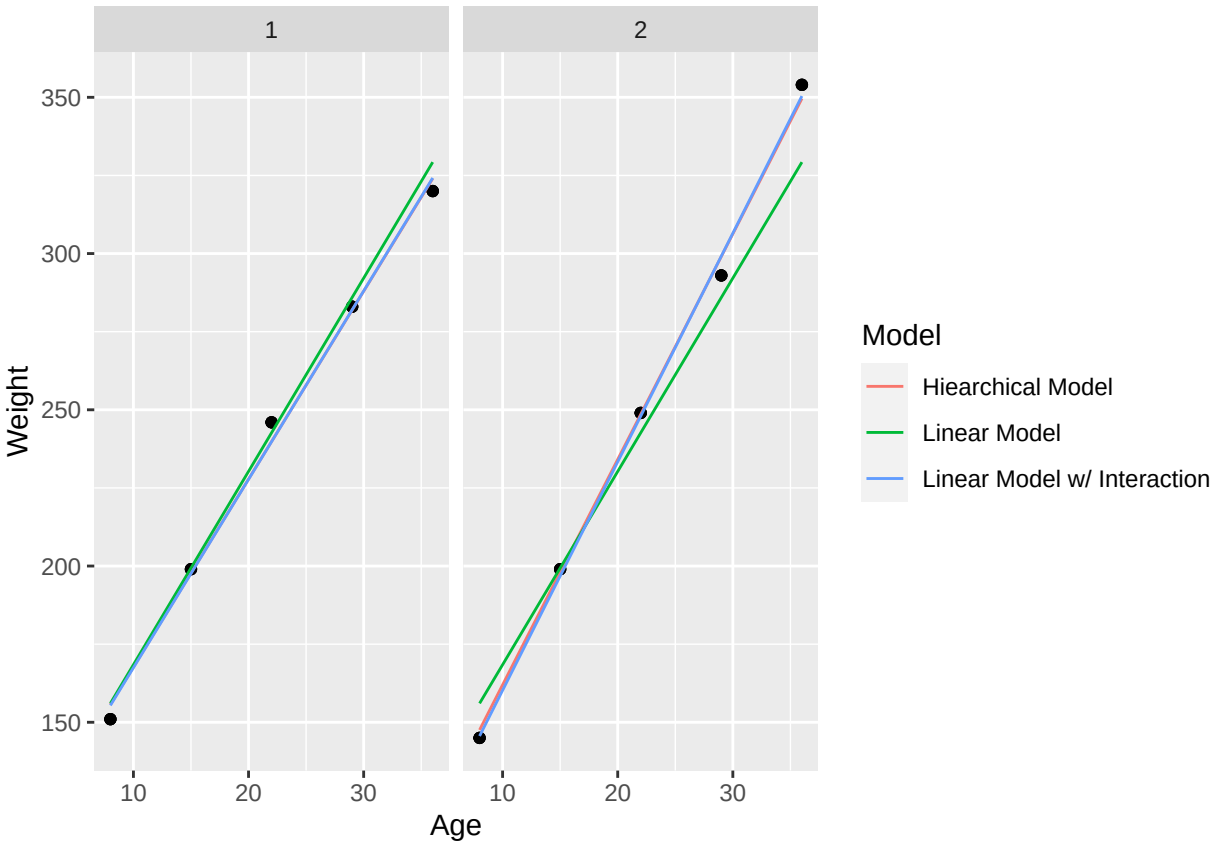


Again we see that linearity is not well satisified, so a transformation of the data may be needed. As expected,

there is far more variability in the rat intercepts than in the slopes. As a percentage of the mean, the standard deviation in the random intercepts is almost twice as high as the standard deviation in the slopes.

e. For at least two rats, plot the observed the data, as well as the predictions obtained from the hierarchical model in (d) and those from the two linear models fit in (b) and (c). Using these plots discuss which model you would prefer the most.

```
new.dat = data.frame('rat'=rep(1:2,each=5),
                     'age'=rep(unique(dat$age),2),
                     'weight'=dat %>%
                       filter(rat%in%c(1,2)) %>%
                       arrange(rat) %>%
                       dplyr::select(weight)
                     )

pred.lm = new.dat
pred.lm$pred = predict(mod.lm,newdata=new.dat)
pred.lm$model = 'Linear Model'

pred.lm.interaction = new.dat
pred.lm.interaction$pred = predict(mod.lm.interaction,newdata=new.dat)
pred.lm.interaction$model = 'Linear Model w/ Interaction'

pred.lmer = new.dat
pred.lmer$pred = predict(mod.lmer,newdata=new.dat)
pred.lmer$model = 'Hiearchical Model'

plt.df = rbind(pred.lm,pred.lm.interaction,pred.lmer)

ggplot(plt.df,aes(x=age)) +
         geom_point(aes(y=weight)) +
         geom_line(aes(y=pred,color=model)) +
         facet_wrap(~rat) +
         labs(x='Age',y='Weight',color='Model')
```



For both rats we see that all three models make *fairly* similar predictions. The linear model does appear to fit somewhat worse for rat 2 than either the hierarchical or interaction models, and this does fit with what we were seeing above. Here we shouldn't expect the hierarchical model to substantially outperform the interaction model, as all the rats have the same number of datapoints, so not much advantage to partial pooling, and moreover there isn't even much variability in the slopes.

Problem 3- Students!

In this problem we will work through some fundamental concepts in Bayesian analysis. This problem largely follows Richard McElreath's "Statistical Rethinking", section 4.7.

a. A class of n students is measured for height (in cm) each year for k years. Write down both a likelihood and a prior for a Bayesian simple linear regression of student height (as dependent variable) against year of measurement (as independent variable). Give a short interpretation of all components of the model (including likelihood and prior), and justify your choice of prior.

At a first pass our model might look something like:

$$\beta_0, \beta_1, \sigma^2 \sim \frac{1}{\sigma^2} \times \mathbb{1}_{\sigma>0}$$
$$\text{HEIGHT} \sim N(\beta_0 + \beta_1 \text{YEAR}, \sigma^2)$$

I started by choosing a maximally uninformative prior: uniform on the coefficients β_0 and β_1 and over $\log \sigma^2$ (which is equivalent to proportional to $\frac{1}{\sigma^2}$). This is probably not a great prior for this problem, since we know for example that it is likely $\beta_1 > 0$ since teenagers don't (usually) shrink with age. Nevertheless, in order to have something to talk about later we'll start from a bad prior.

The likelihood of this model has the usual interpretation as giving the distribution of HEIGHT in terms of YEAR of measurement. For every additional YEAR we expect the average HEIGHT to change by β_1, where

the average heights of students at the beginning of the study (YEAR=0) given by β_0. σ^2 gives the (constant) variance of student heights in a given year.

b. Now, pretend you are told (before collecting any data) that each student eats a diet plentiful in spinach, and thus it is guaranteed that each will get taller every year. Does this information change any of your priors, and if so how?

Yes, as mention above this implies that $\beta_1 > 0$, so I would adjust my priors to:

$$\beta_0, \beta_1, \sigma^2 \sim \frac{1}{\sigma^2} \times \mathbb{1}_{\sigma>0, \beta_1>0}$$

Ie. assign no prior probability mass to values of $\beta_1 < 0$. Notice that this will **guarantee** that my posterior will *also* not assign and probability mass to $\beta_1 < 0$, so this is a highly informative change (no amount of data will make me change this belief).

c. In addition to the information in (b) above (and before collecting any data) you learn that at outset of the study the average height of students is 120cm. Does this change any of your priors, and if so how?

Since β_0 is exactly equal to the average height at the start of the study, this implies that $\beta_0 = 120$ for sure. Therefore we no longer need to perform inference over this parameter: we know its value exactly and thus can just drop it from our prior. The model now becomes:

$$\beta_1, \sigma^2 \sim \frac{1}{\sigma^2} \times \mathbb{1}_{\sigma^2>0, \beta_1>0}$$
$$\text{HEIGHT} \sim N(120 + \beta_1 \text{YEAR}, \sigma^2)$$

d. In addition to (b) and (c) above (and still before collecting any data) you learn that the variance of heights among students in the same year cannot be more than 64cm. Does this information change any of your priors, and if so how?

As with learning that $\beta_1 > 0$, this information causes us to adjust our priors **support** (the parameter values for which the prior probability density is nonzero). Our final priors are:

$$\beta_1, \sigma^2 \sim \frac{1}{\sigma^2} \times \mathbb{1}_{0<\sigma<64, \beta_1>0}$$

e. Finally, using the priors you selected at the end of (d) above, describe the prior distribution of the students' heights at the end of the first year of the study. With the aid of a software tool such as R or Python (or a physical tool such as a coin, a bag of dice, or a FERMIAC), draw 100 samples from this prior distribution and plot them in a histogram. Does this prior distribution accurately reflect your beliefs about the students' heights at the end of the first year? Does it accurately reflect what is physiologically possible? If not, what would you change in your priors to resolve the discrepancy?

We are being asked to draw samples of HEIGHT from the model:

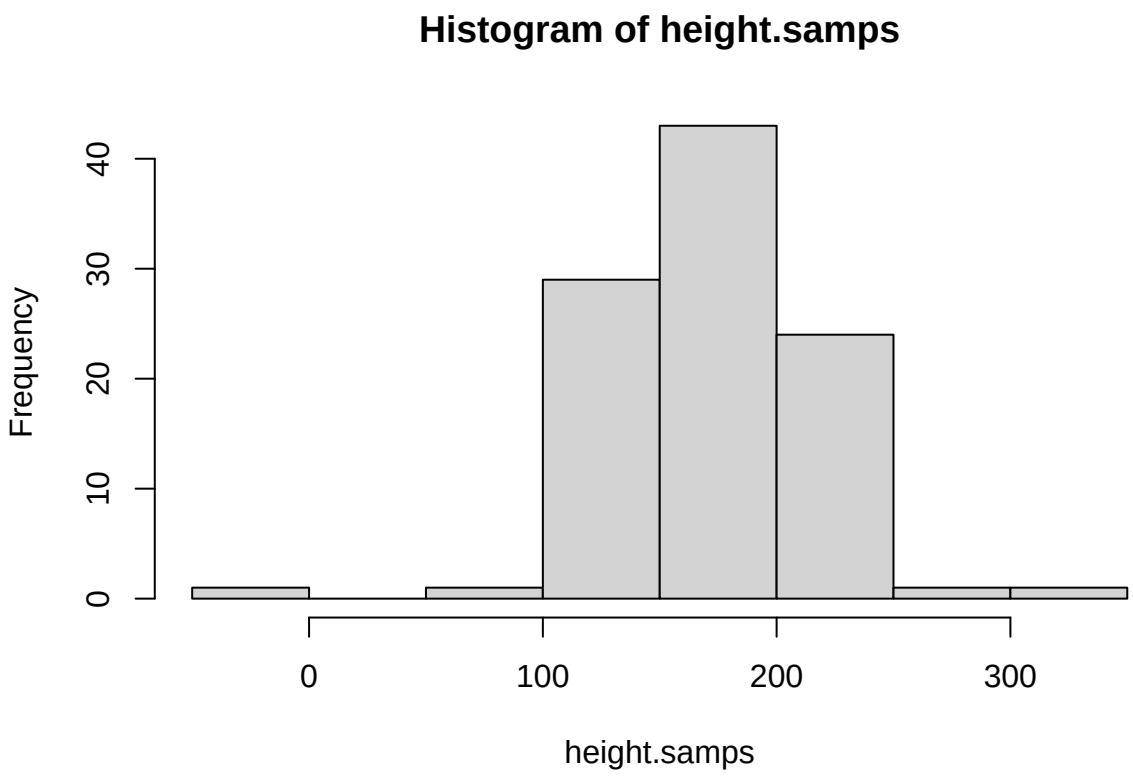
$$\beta_1, \sigma^2 \sim \frac{1}{\sigma^2} \times \mathbb{1}_{0<\sigma^2<64, \beta_1>0}$$
$$\text{HEIGHT} \sim N(120 + \beta_1 \times 1, \sigma^2)$$

To make things a little easier on myself, I will assume that it is impossible for a student to grow more than 1 meters in a single measurement period (since this would imply the existence of ~15 foot tall students at the end of the experiment):

$$\beta_1, \sigma^2 \sim \frac{1}{\sigma^2} \times \mathbb{1}_{0<\sigma^2<64, 0<\beta_1<100}$$
$$\text{HEIGHT} \sim N(120 + \beta_1 \times 1, \sigma^2)$$

Moreover, I will exploit the fact that $\sigma^2 \sim \frac{1}{\sigma^2}$ is equivalent to a uniform distribution over $\log \sigma^2$:

```
n = 100
sigma2.samps = exp(runif(n,0,log(64)))
beta1.samps = runif(n,0,100)
height.samps = rnorm(n,120+beta1.samps,sigma2.samps)

hist(height.samps)
```



This seems like a fairly reasonable predictive prior distribution. Although it *does* seem to allow for students to be a *little* too tall at the end of a month (notice that the range goes as high as 250cm, which is like 6'6"), the majority of the mass lies around plausible human heights. This seems like a reasonably informative set of priors, with enough uncertainty to allow us to be"surprised" by our data.