

Exercise 3

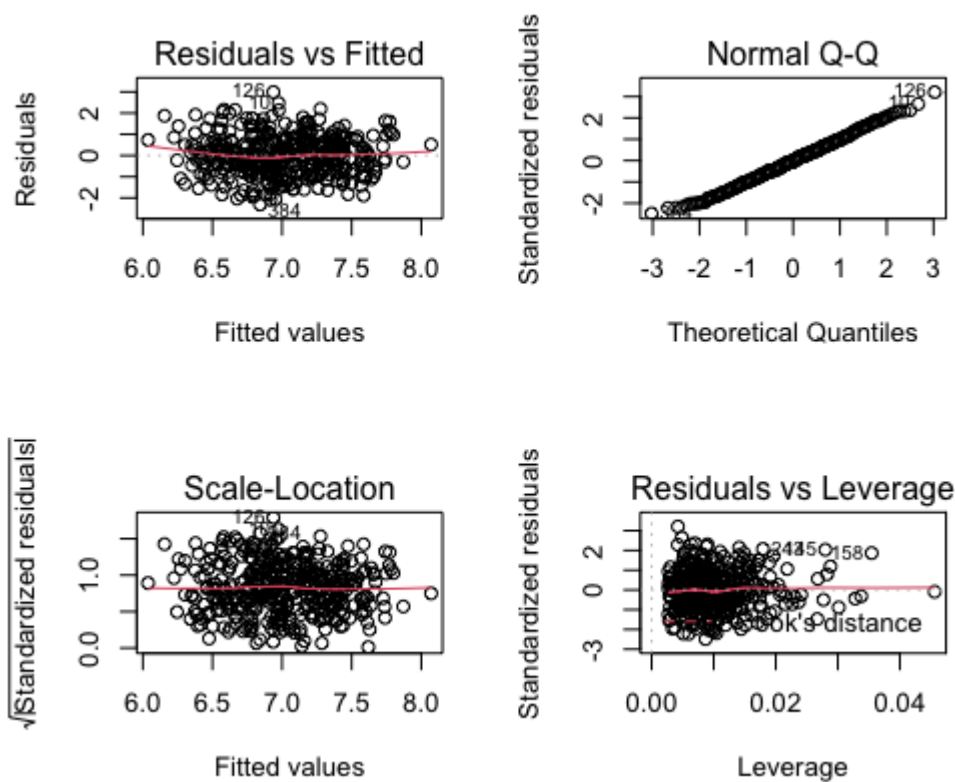
I fit a linear model.

I estimate from the model the coefficients and the standard deviation of the error, sigma

beta0	beta1	beta2	beta3	sigma
9.18676786	0.09930218	0.07258300	-0.00888967	0.9350057

The assumptions are $\text{Eps} \sim N(0, \sigma^2)$

I verify that the assumptions are fulfilled.



The diagnostic is good, I can see how the residuals are homoscedastic, and there seems not to be any significant leverage effect.

I test Gaussianity with a Shapiro test (H_0 : "data is Gaussian", H_1 : H_0^C). The p-value of the test is: 0.6653569, so I don't reject H_0 .

The model fulfills the assumptions.

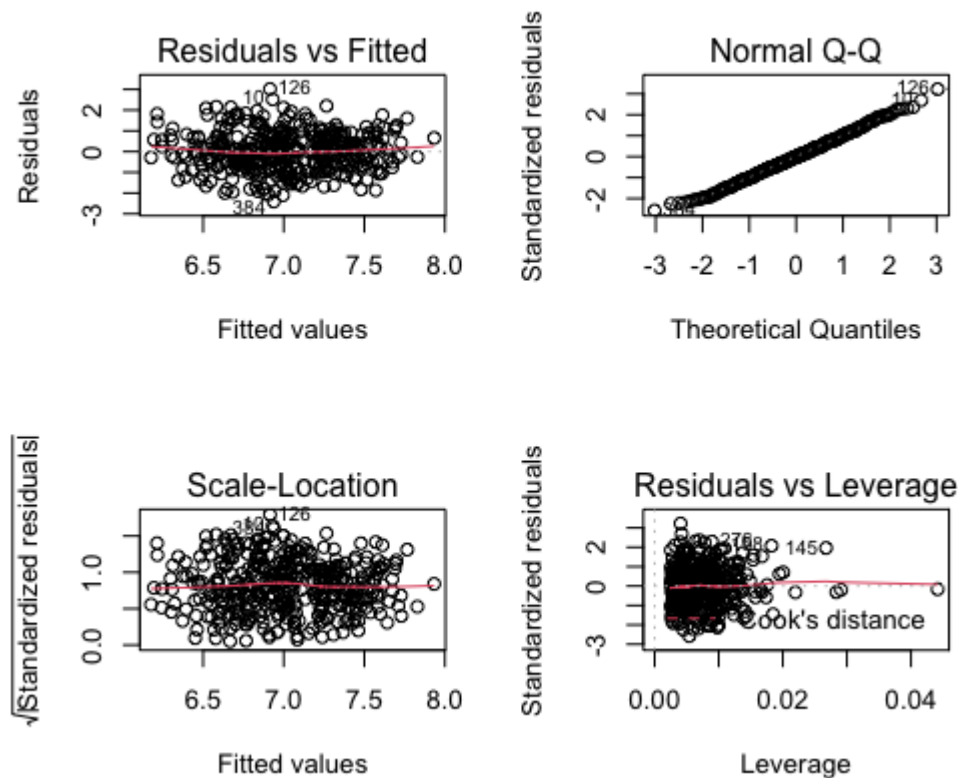
I perform a simultaneous test to see if loudness and energy are significant. The p-value I obtain is $2.112e-06$ which is low enough (at level 5%) to reject H_0 : "both coefficients are zero", so at least one is significant.

I try to remove just energy, since its one-at-the-time p-value for the test H_0 "coefficient is zero" is 0.0731, and obtain a nearly identical model with R^2 going from 0.1261 to 0.1212.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.182772	0.293896	31.245	< 2e-16 ***
loudness	0.170266	0.035008	4.864	1.66e-06 ***
tempo	-0.008962	0.001508	-5.944	6.09e-09 ***

and all coefficients significant one-at-the-time, this model is simpler and similar, so it's better. The new sigma estimated is 0.9376268. New $\beta_0 = 9.182772$, $\beta_1 = 0.170266$, $\beta_2 = 0$, $\beta_3 = -0.008962$.

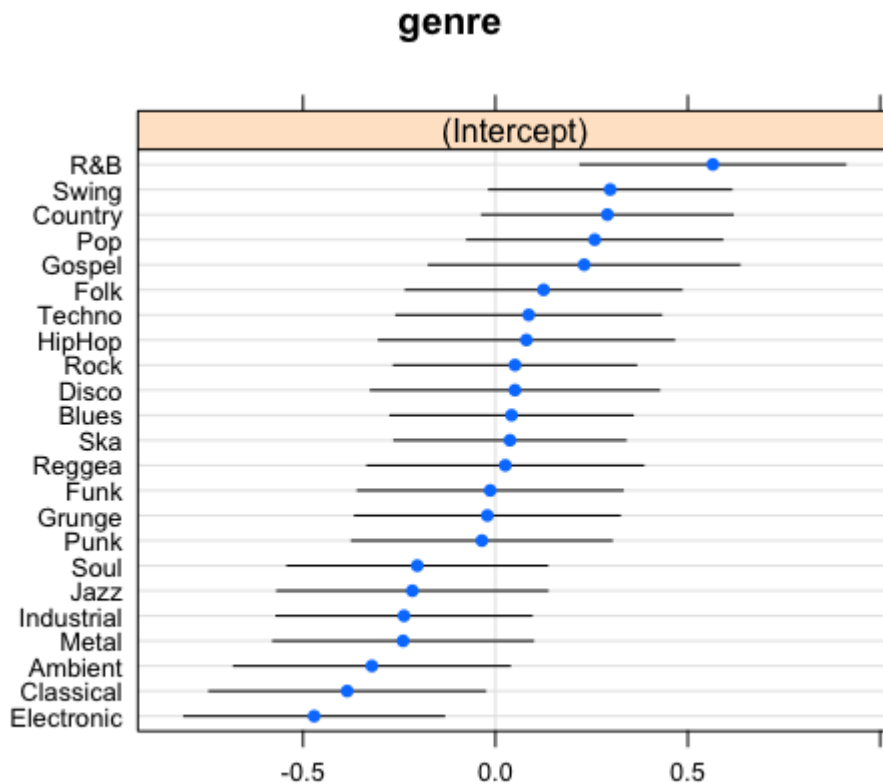
The residuals still fulfill the assumptions (p-value = 0.7943424)



I fit a mixed model by taking into account the grouper effect (random effect) on the intercept, starting from the model I have updated (the call is thus `lmer(danceability ~ loudness + tempo + (1|genre), data = d)`).

The PVRE (i.e. the variance explained by the grouping) is 0.1034443 which is pretty high.

The dotplot of the random effect of genre is the following, from which I can see that the genre associated with highest danceability is R&B with a positive random effect on the intercept of 0.56514396.



I assess the assumption on the within-group errors and the assumption on the random effects, which are both homoscedastic and from the plot I can say that they are both Gaussian.

