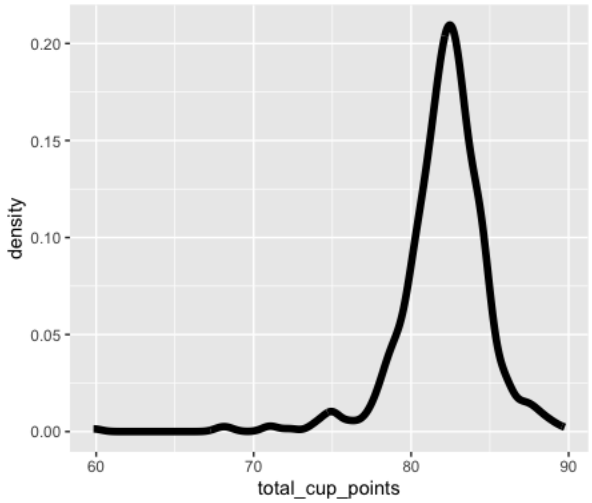
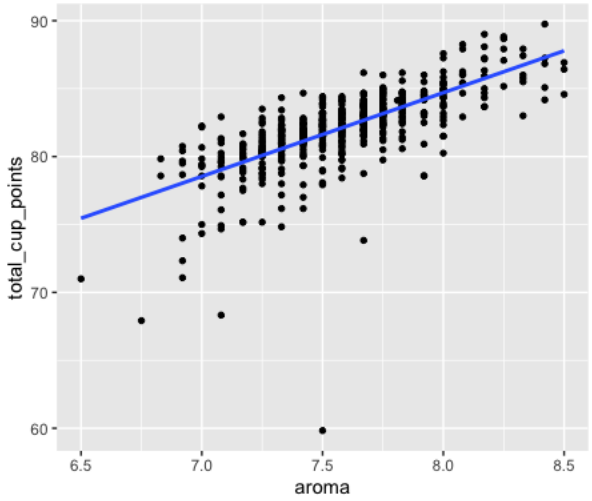
1a)



There are some outliers as can be seen from 60 to 75. But after that we see that the data is normally distributed. There is long tail towards left, but we still go ahead with normal assumptions.



Y is linearly associated with X.

1b) Assuming $\in_i \sim N(0, \sigma^2)$,

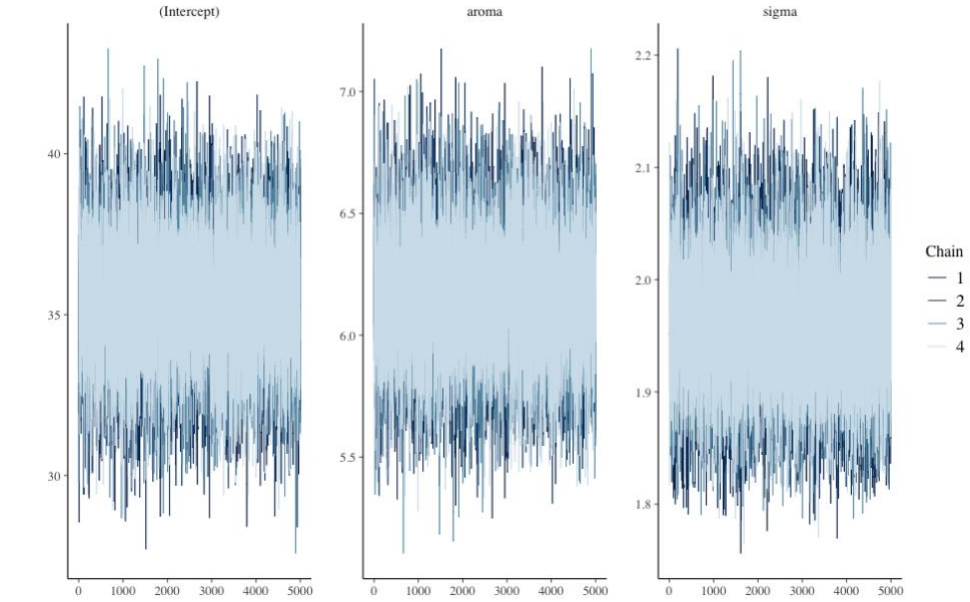
$$L(\beta_0, \beta_1, \sigma^2 | (y_1, x_1), \ldots, (y_n, x_n) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu_i)^2$$

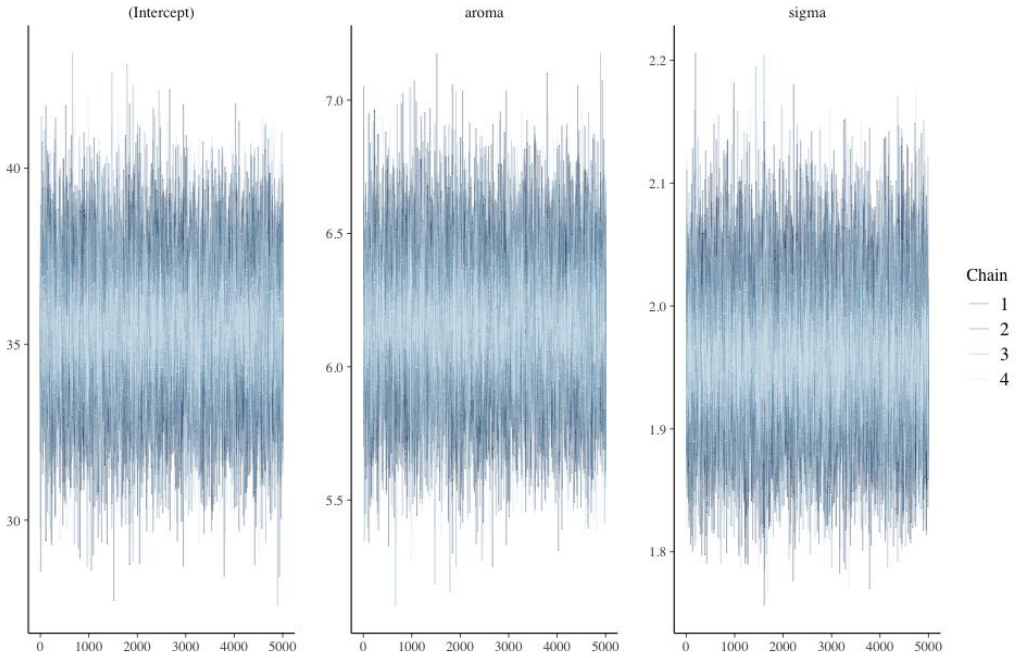
1c) coffee_model = stan_glm(total_cup_points ~ aroma, data = CoffeeData, family = gaussian,
prior_intercept = normal(80, 30), prior = normal(4, 8),
prior_aux = exponential(0.25),
chains = 4, iter = 5000*2, seed = 84735)

1d). neff_ratio(coffee_model)

(Intercept) aroma sigma

0.97025 0.96560 0.97135
rhat(coffee_model)(Gelman Rubin Statistics)
(Intercept) aroma sigma
1.0000030 1.0000002 0.9998541



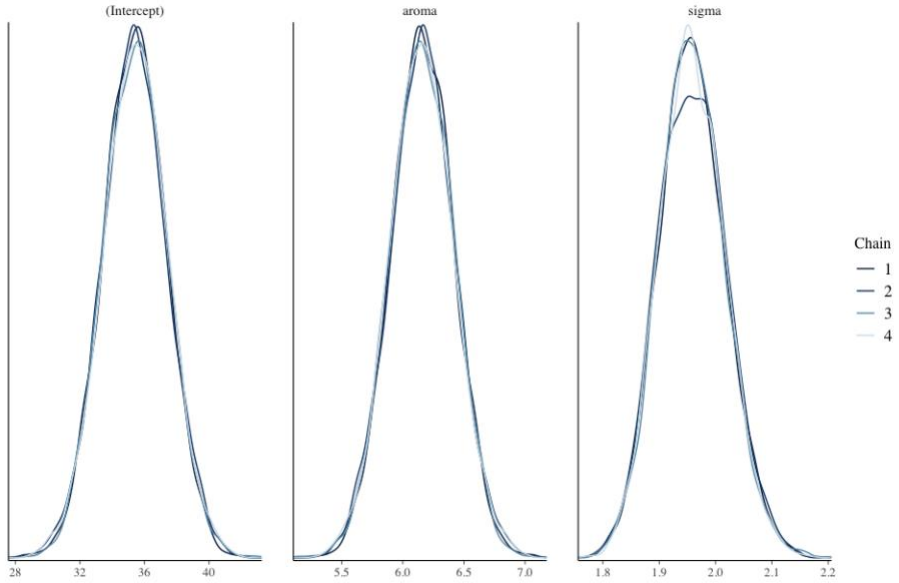


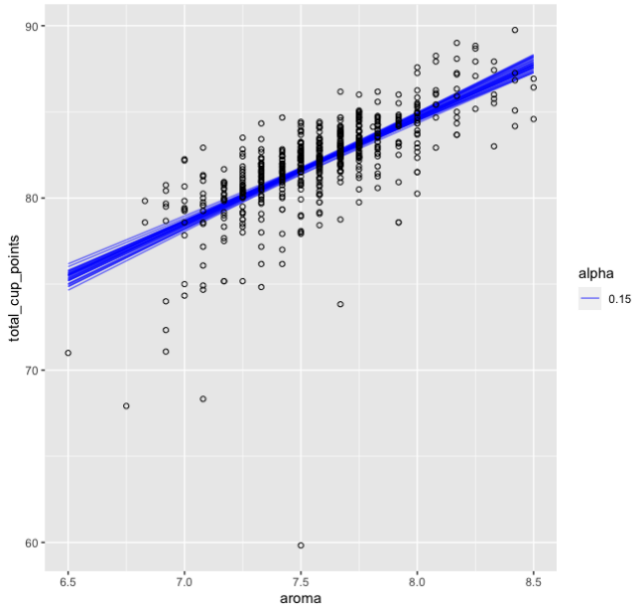
The model appears to have converged as seen from the plots above.

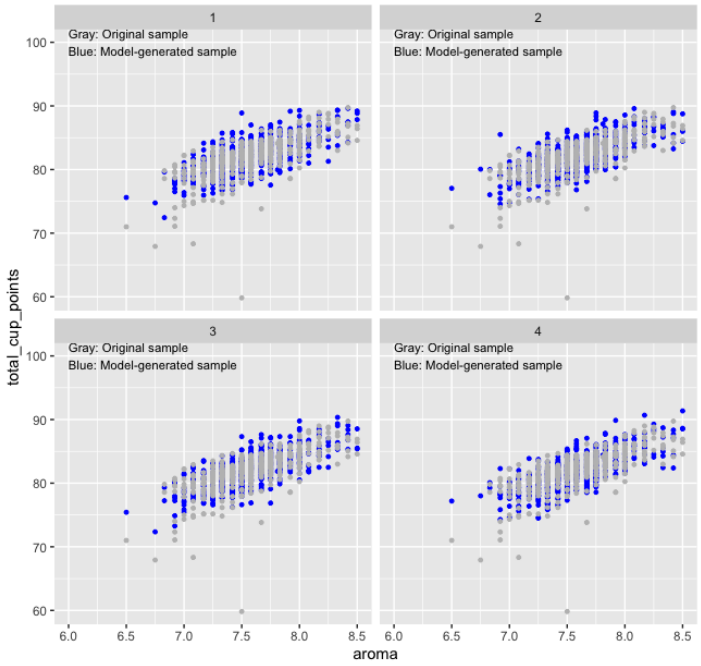
1e).

term	estimate	std.error	conf.low	conf.high
1 (Intercept)	35.4	1.97	32.2	38.6
2 aroma	6.16	0.260	5.73	6.59
3 sigma	1.96	0.0588	1.87	2.06
4 mean_PPD	82.1	0.116	81.9	82.3

This is the summary of the posterior distribution of the beta parameters and the sigma.





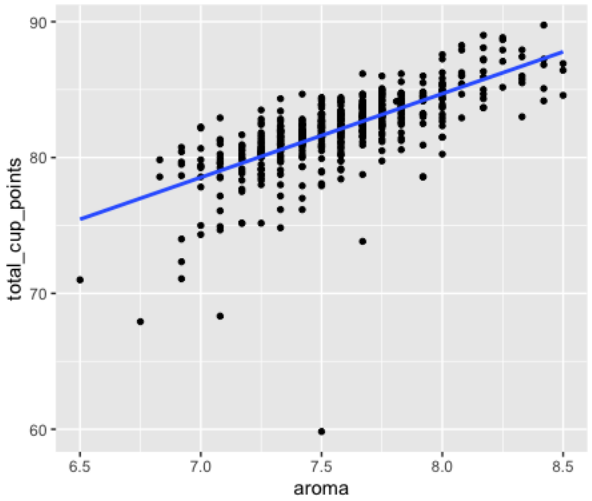


1f) Model Assumptions :

a) The model is independent

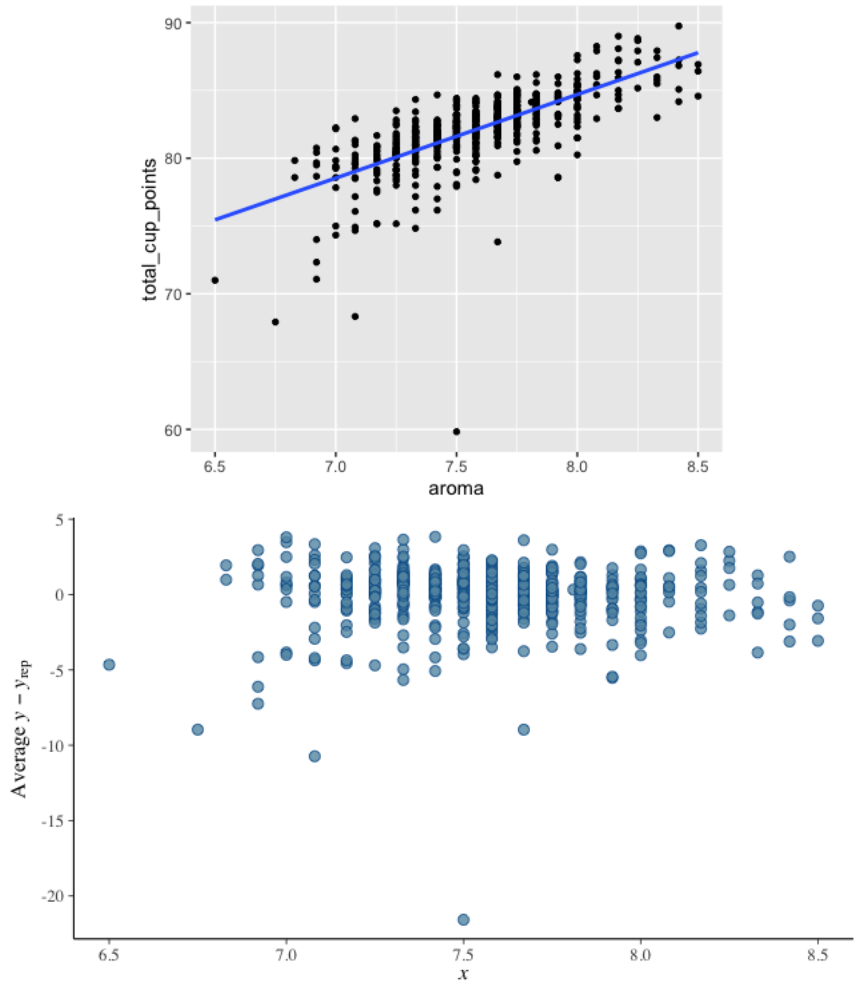
Since data is time independent, we can assume this is satisfied

b) The expected value of Y can be written as linear function of X



c) Normality and constant variance – At any x_i, the observations y_{ij} vary normally around mean((μ_i) and the variability σ is constant.

The plot below shows that there is constant variability around each value of x_i. Moreover, as discussed before y_{ij} is also normally distributed.



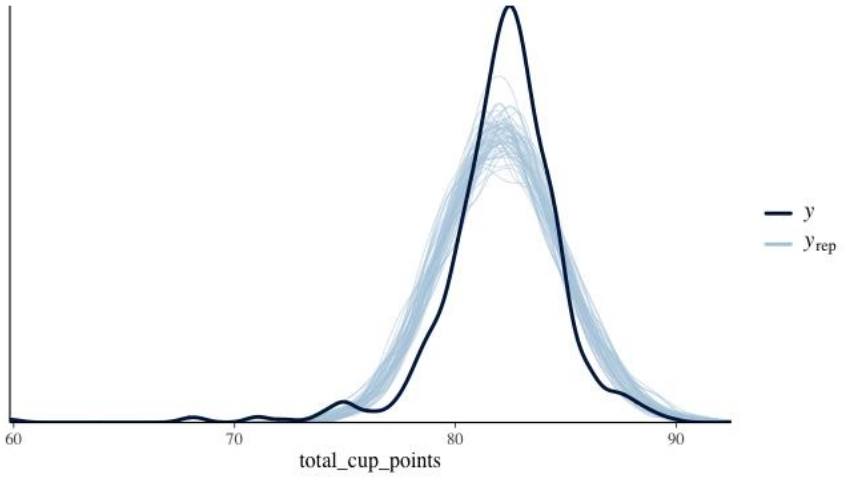
The plot above shows no evidence of unequal variance or lack of independence.

1g) 35.4 is the average rating of coffee, when the aroma value is zero. The 90% confidence interval is (32.2, 38.6)

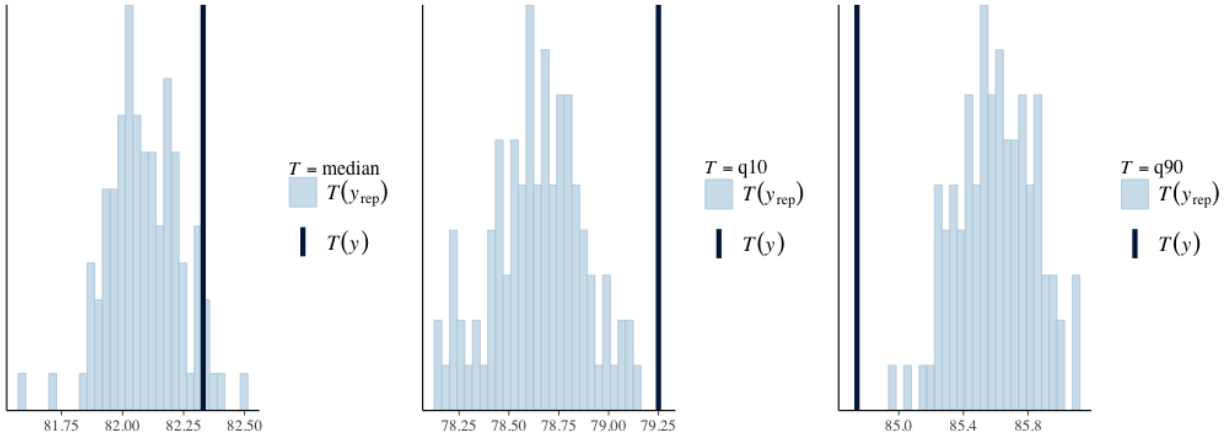
For increase of aroma = 1, the average rating increases by 6.16 and the 90% confidence interval of this increase is (5.73,6.59)

For each value of x_i, the standard deviation of y_{ij} is 1.96 and the 90% confidence interval is (1.87,2.06)

1h) The dimension of yrep will be 50x572

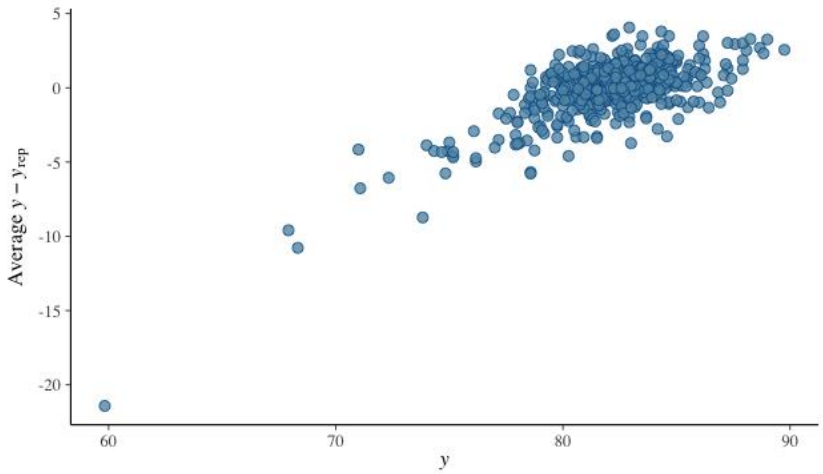


1i)



The median , the 10th quantile and the 90th quantile should be distributed around T(y), but we don't see that in the above figure. They are the ends. So , we can say that the model does not represent the test statistics very well.

1j)



We can see a linear relationship between the residuals and y = total_cup_points/ coffee rating. This indicates that there is strong evidence of missing predictor(s).