1a)

Chart, line chart

Description automatically generated

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.

Chart, scatter chart

Description automatically generated

Y is linearly associated with X.

1b) Assuming

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

Histogram

Description automatically generated with low confidence

A picture containing application

Description automatically generated

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.

Chart, histogram

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

1f) Model Assumptions :

1. The model is independent

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

1. The expected value of Y can be written as linear function of X

Chart, scatter chart

Description automatically generated

1. Normality and constant variance – At any , the observations vary normally around mean( and the variability is constant.

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

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

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 , the standard deviation of is 1.96 and the 90% confidence interval is (1.87,2.06)

1h) The dimension of yrep will be 50x572

Chart

Description automatically generated

1i)

Chart, histogram

Description automatically generated

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)

Chart, scatter chart

Description automatically generated

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).