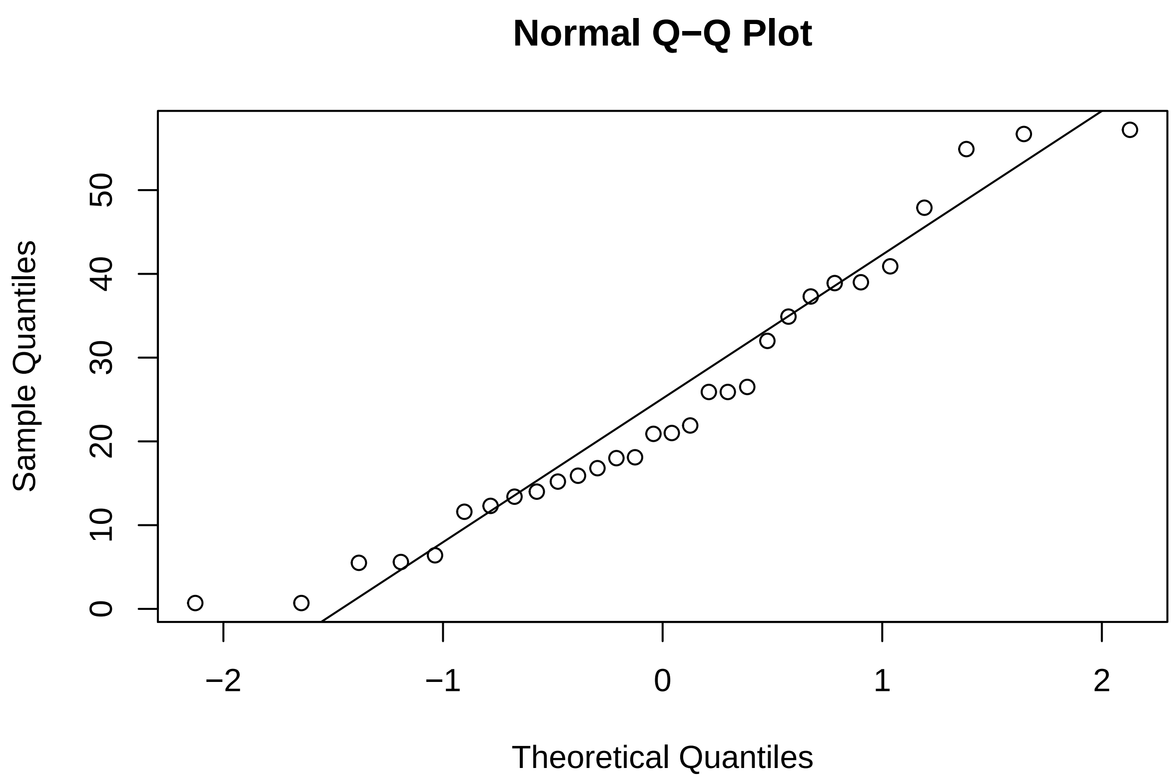
Zhang_exam1_Cutler

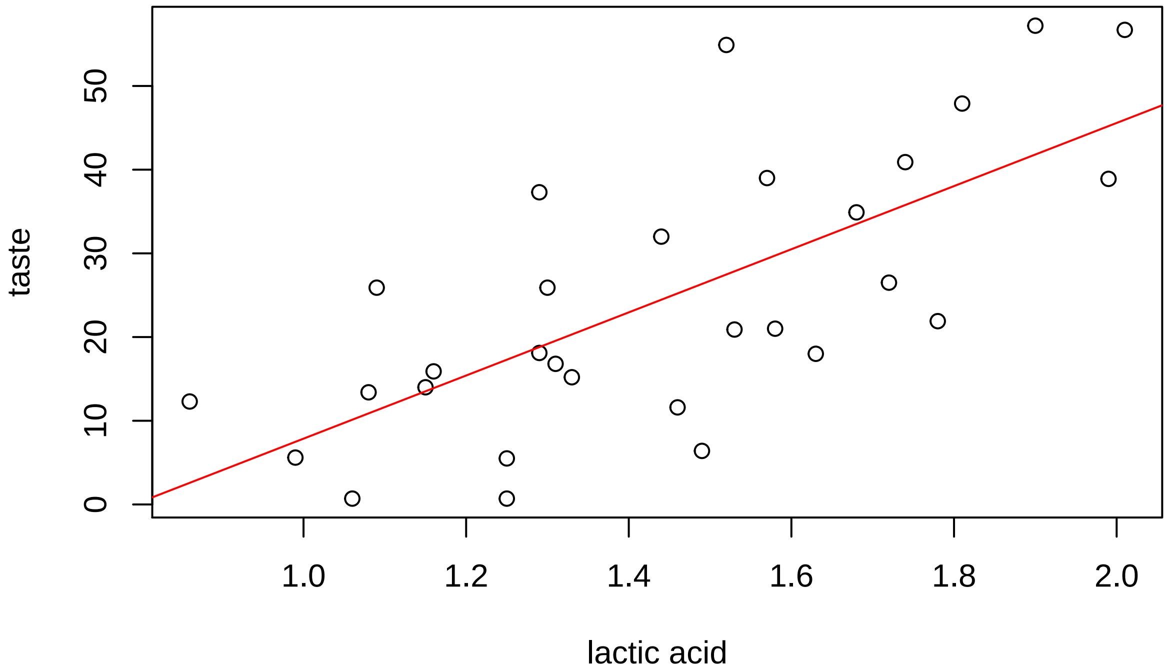
```
## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.
## TMB was built with Matrix version 1.2.15
## Current Matrix version is 1.2.14
## Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRAN for a
```

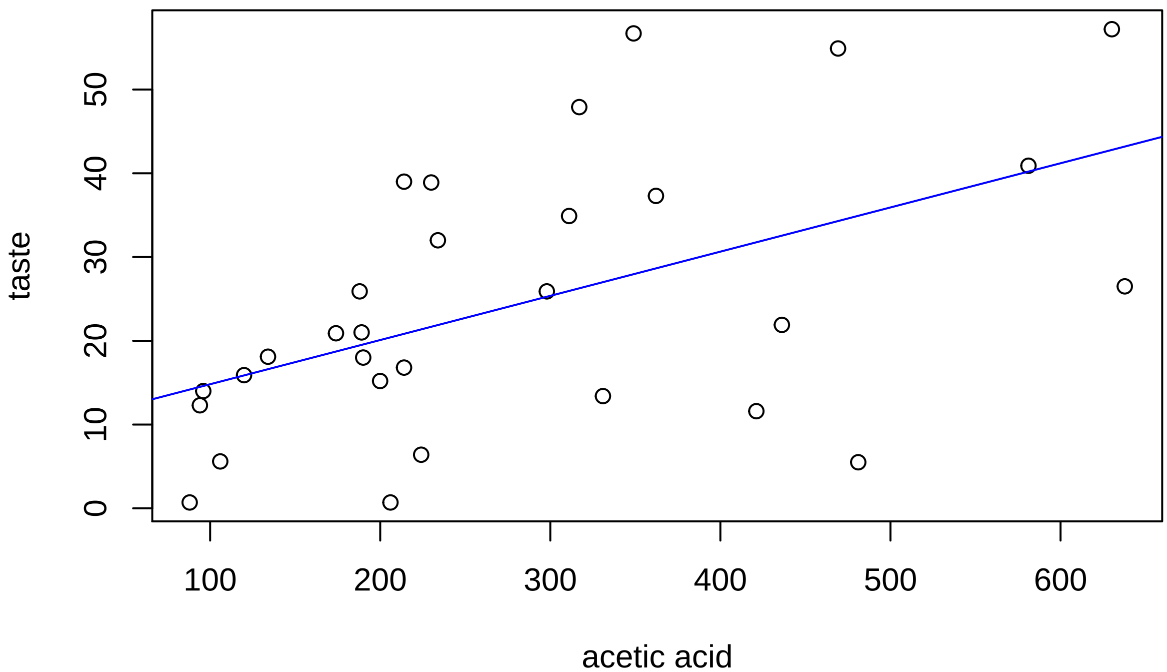
1. Why is violation of normality in the outcome variable not a problem:

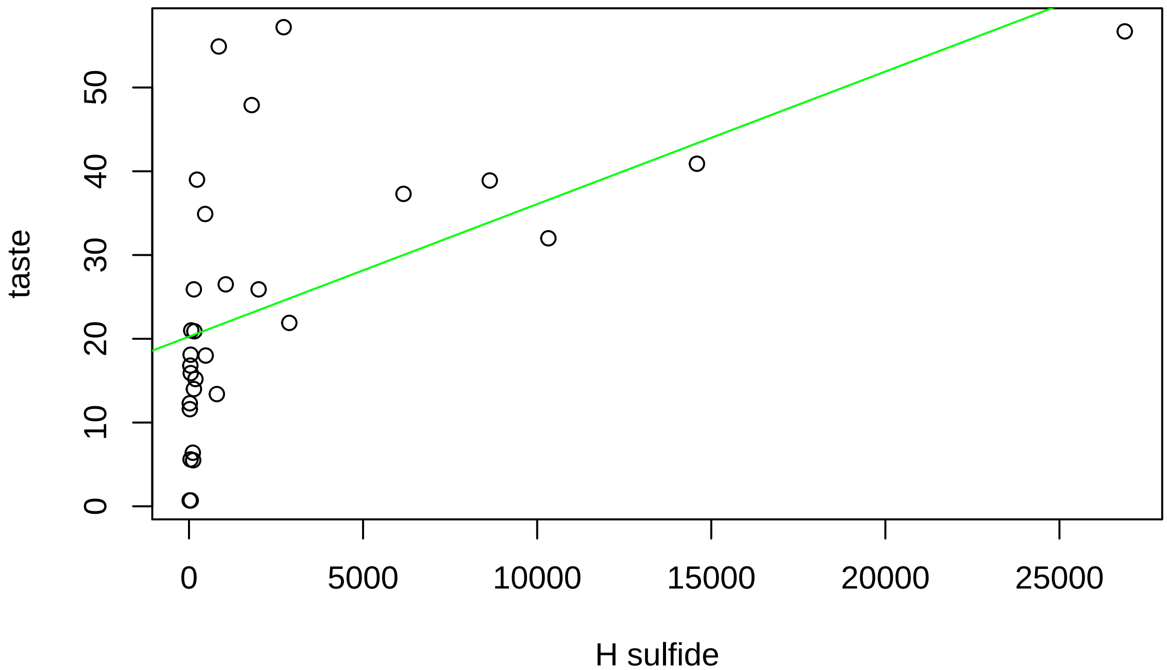
Because in this case it's close enough to being normal:



2. Simple Linear Regression and LOESS:







taste

Predictors

Estimates

CI

p

(Intercept)

-29.86

-50.60 – -9.12

0.009

Lacid

37.72

23.63 – 51.81

<0.001

Observations

30

R2 / adjusted R2

0.496 / 0.478

taste

Predictors

Estimates

CI

p

(Intercept)

9.54

-1.05 – 20.12

0.088

Aacid

0.05

0.02 – 0.09

0.004

Observations

30

R2 / adjusted R2

0.263 / 0.237

taste

Predictors

Estimates

CI

p

(Intercept)

20.26

14.82 – 25.69

<0.001

Hsulf

0.00

0.00 – 0.00

0.001

Observations

30

R2 / adjusted R2

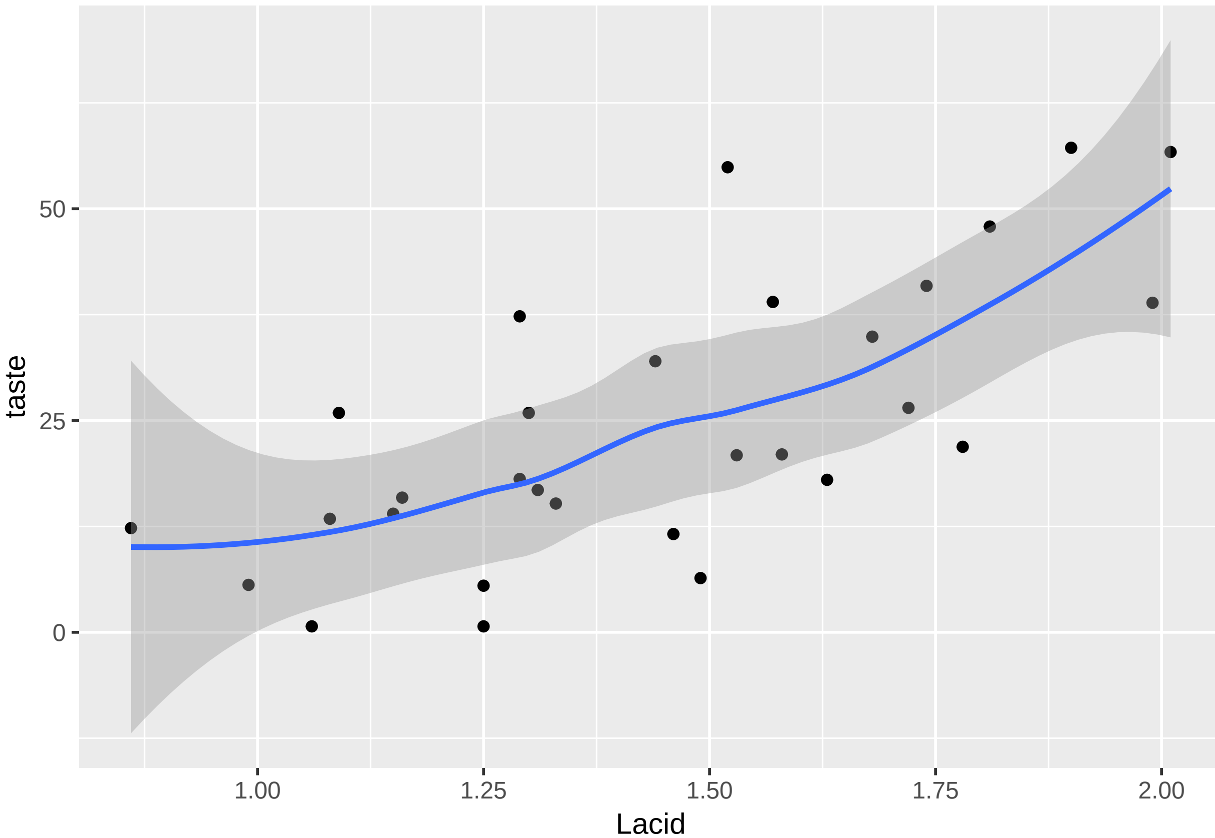
0.314 / 0.290

```
##                  2.5 %    97.5 %
## (Intercept) -51.53573 -8.181935
## Lacid        22.99928 52.440613
```

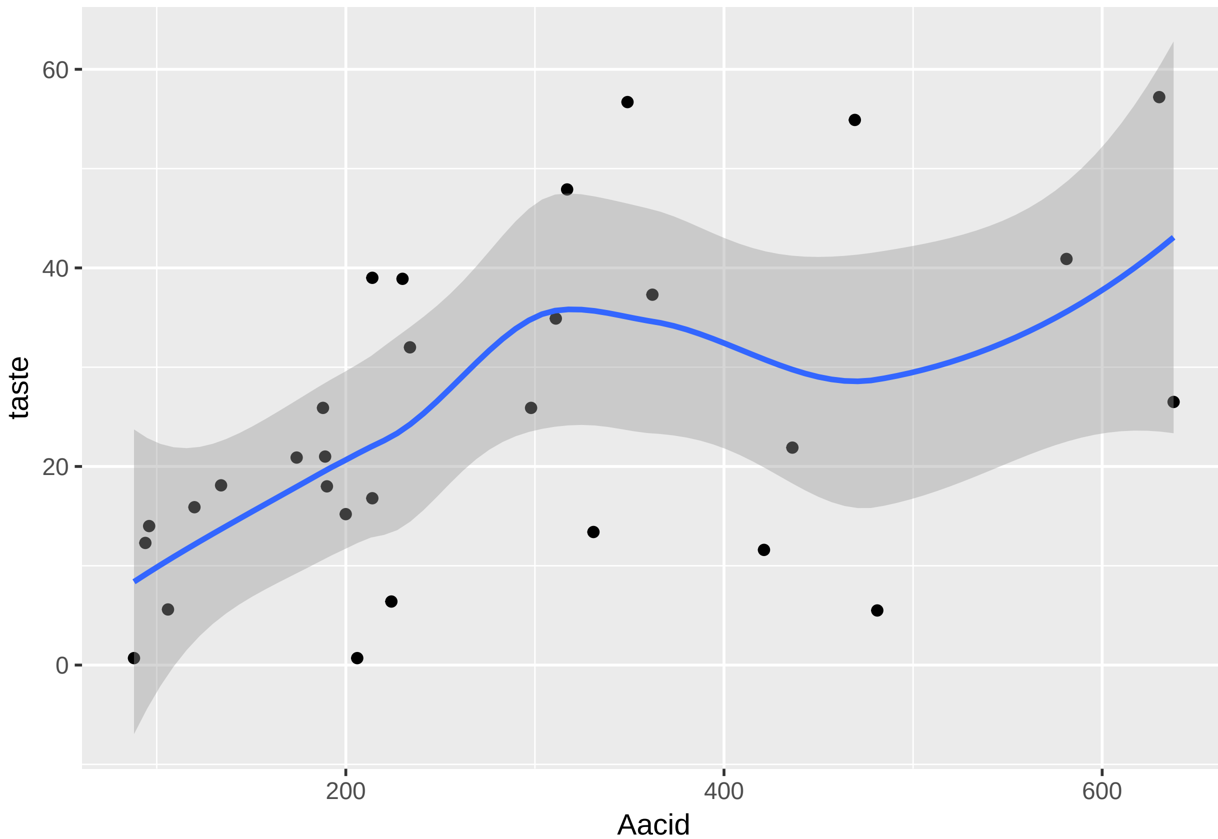
```
##                   2.5 %      97.5 %
## (Intercept) -1.52478617 20.60252313
## Aacid        0.01860862  0.08691771
```

```
##                   2.5 %       97.5 %
## (Intercept) 14.57774505 25.933841789
## Hsulf        0.00067731  0.002487749
```

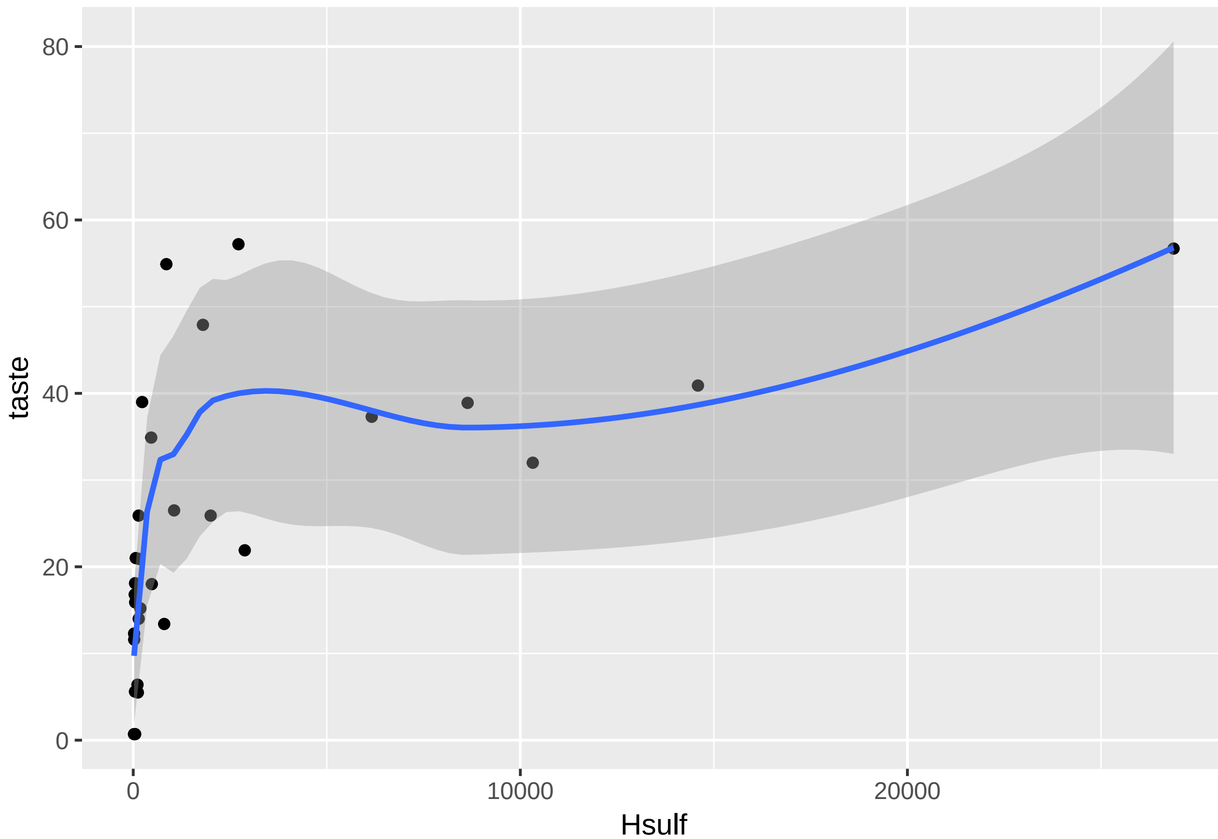
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```


80
60
40
20
0
taste
0
10000
20000
Hsulf

3. Equations of models:

Association between taset and acetic acid:

Taste and lactic acid:

Taste and H sulfide:

4. Sentences quantifying the associations:

Taste and acetic acid:

Taste and lactic acid:

Taste and H sulfide:

5. Because unless we use CIs, our point estimate of the population parameter in question will always be guaranteed to be wrong (not equal to the true parameter), at least when dealing with continuous data. This is because the probability of a random variable taking on a specific value is essentially zero, due to the laws of probability. A CI will at least give us a good chance of including the true parameter within the sweep of the CI.

6. 'Statistic' refers to a function of a sample (e.g. X-bar is defined as the sum of the Xi's divided by the sample size), and is meant to approximate a true population parameter. Parameter estimates can change depending on what methods are used to calculate them, while the sample statistics shouldn't vary for the same data. An example illustrating this will come up below, in question 24.

7. A least squares estimate is an estimate that minimizes the amount of squared error, or distance between Y-hat (predicted Y values) and the observed Y values.

8. Comparing linear and quadratic models for lactic acid:

taste

Predictors

Estimates

CI

p

(Intercept)

22.22

-69.09 – 113.54

0.637

Lacid

-36.76

-164.73 – 91.21

0.578

Lacid 2

25.51

-18.06 – 69.08

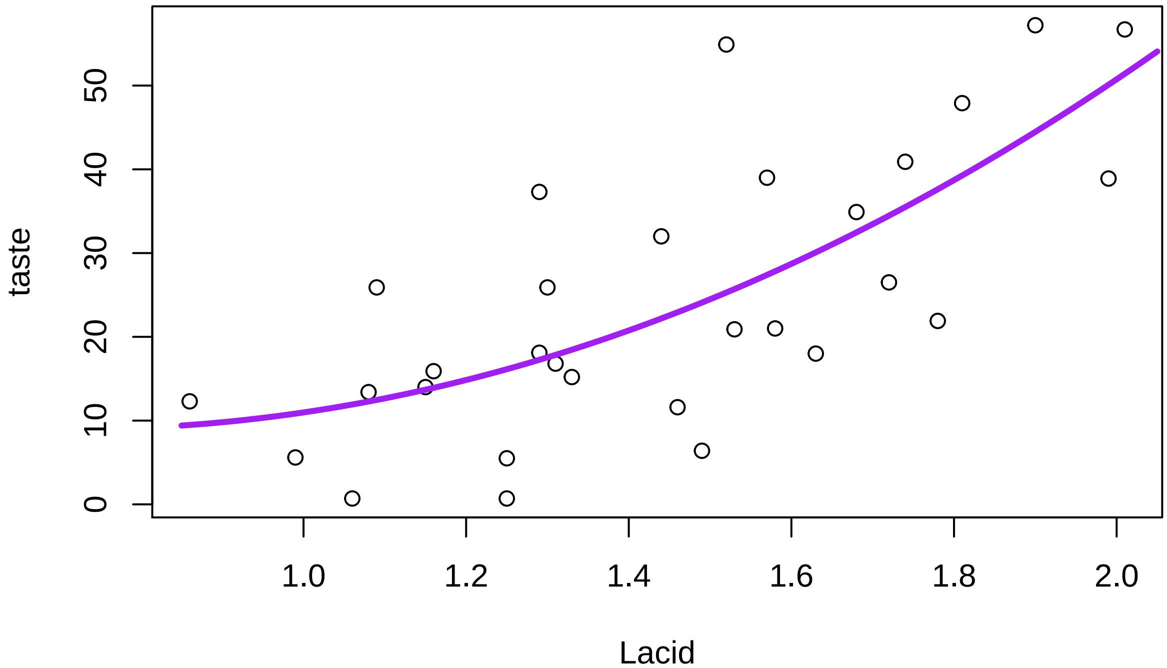
0.261

Observations

30

R2 / adjusted R2

0.519 / 0.484



MSE and R^2 for lactic acid linear model:

```
## [1] 128.7496
```

```
## [1] 0.4959486
```

MSE and R^2 for lactic acid quadratic model:

```
mse.Lq = mean(Lquad$residuals^2); mse.Lq
```

```
## [1] 122.7613
```

```
# Get r^2 from Lquad?
```

The quadratic model looks better than the linear, based on the higher R^2 value and the lower MSE.

9. Why must the quadratic model also include an unsquared term?

10. Comparing linear and quadratic models for acetic acid:

taste

Predictors

Estimates

CI

p

(Intercept)

-0.46

-21.90 – 20.99

0.967

Aacid

0.13

-0.02 – 0.27

0.093

Aacid 2

-0.00

-0.00 – 0.00

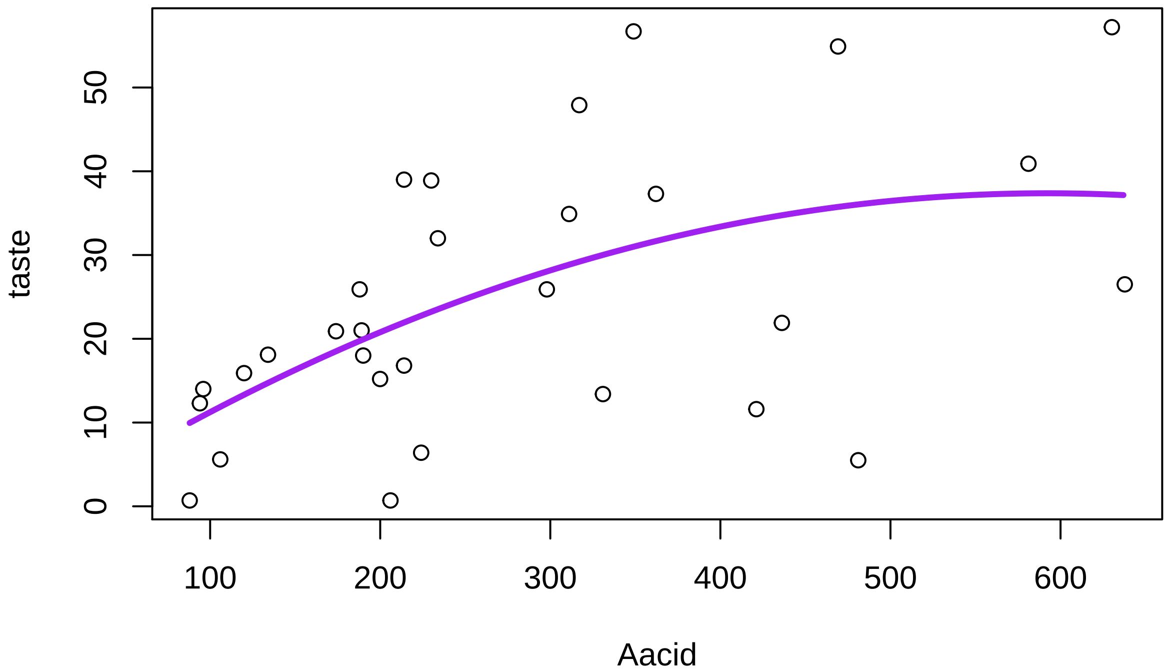
0.303

Observations

30

R2 / adjusted R2

0.292 / 0.240



MSE and R^2 for acetic acid linear model:

```
## [1] 188.1431
## [1] 0.2634246
```

MSE and R^2 for acetic acid quadratic model:

```
## [1] 180.7652
```

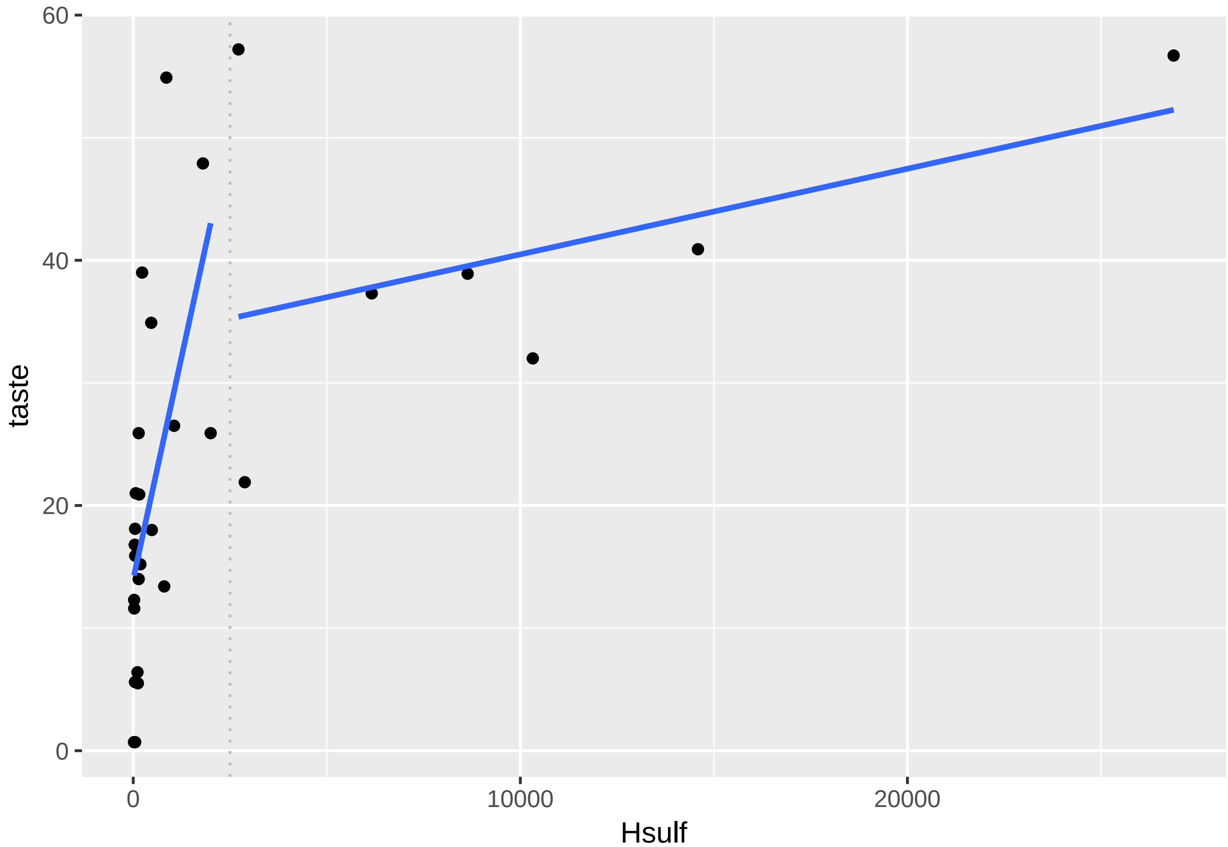
11. Define the type III SS for the acetic acid quadratic model:

```
## Anova Table (Type III tests)
##
## Response: taste
##              Sum Sq Df F value  Pr(>F)
## (Intercept)    0.3  1  0.0017 0.96705
## Aacid        609.2  1  3.0331 0.09296 .
## Aacid2       221.3  1  1.1020 0.30313
## Residuals   5423.0 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The type III sum of squares for the acetic acid quadratic model is:

12. Run a segmented regression model on hydrogen sulfide with a knot at concentration of 2500 mg/L:

```
## Warning: Ignoring unknown parameters: method, se
```



13. Hydrogen sulfide segmented regression model's coefficients and hypothesis tests:

14. MSE and R^2 for the hydrogen sulfide segmented regression model:

15. Simple linear regression of log-transformed hydrogen sulfide:

taste

Predictors

Estimates

CI

p

(Intercept)

-9.79

-21.46 – 1.89

0.112

log HS

5.78

3.92 – 7.63

<0.001

Observations

30

R2 / adjusted R2

0.571 / 0.556

MSE and R^2 for the logHS SLR model and untransformed SLR model:

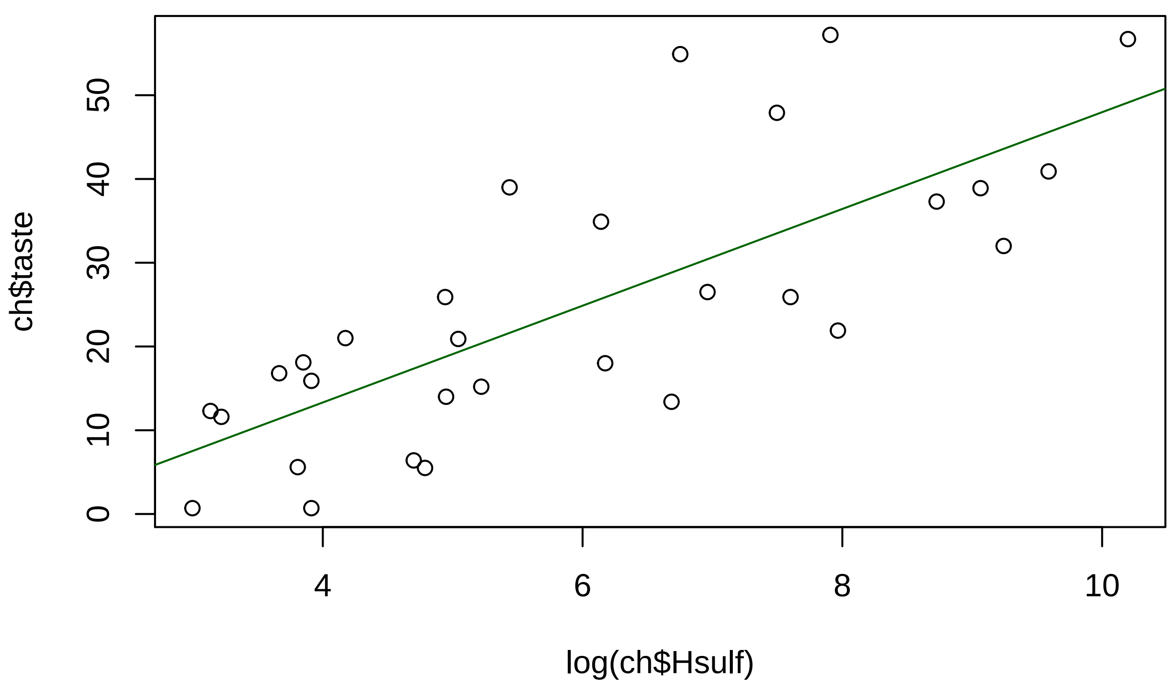
logHS:

```
## [1] 109.5392
```

Untransformed:

```
## [1] 175.1911
```

Plot of the log-transformed data:



16. Comparing the hydrogen sulfide log-transformed SLR model to the segmented model (including definitions of the MSE and R^2 statistics):

17. Residual plots for the hydrogen sulfide log-transformed SLR model and segmented model (which is preferrable?):

18. What kind of residuals did I use to perform the regression diagnostics, and why?

19. What are the consequences for a statistical model whose underlying assumptions are violated? Why do we care?

20. What is the meaning of the intercept in a regression model?

21. We do not always interpret the intercept when we report results from a regression model. Under what circumstances would we be able to interpret the intercept?

22. Which multivariable regression model best fits the observations on taste? How did I arrive at my final model?

taste

Predictors

Estimates

CI

p

(Intercept)

11.93

-123.02 – 146.89

0.864

Lacid

-5.39

-101.09 – 90.31

0.913

Aacid

-0.29

-0.92 – 0.35

0.388

log HS

0.00

-29.63 – 29.64

1.000

Lacid:Aacid

0.18

-0.24 – 0.61

0.403

Lacid:logHS

2.26

-15.69 – 20.21

0.807

Aacid:logHS

0.03

-0.08 – 0.15

0.578

Lacid:Aacid:logHS

-0.02

-0.09 – 0.05

0.572

Observations

30

R2 / adjusted R2

0.680 / 0.578

The full model above shows no significant interactions. This indicates that the reduced or main effects model is worth examining:

taste

Predictors

Estimates

CI

p

(Intercept)

-27.14

-45.33 – -8.96

0.007

Lacid

19.20

2.62 – 35.78

0.032

Aacid

0.00

-0.03 – 0.03

0.781

log HS

3.84

1.45 – 6.23

0.004

Observations

30

R2 / adjusted R2

0.653 / 0.613

```
##                    2.5 %      97.5 %
## (Intercept) -46.21346358 -8.07066654
## Lacid         1.81641427 36.58636137
## Aacid        -0.02645859  0.03484837
## logHS         1.32904794  6.34370461
```

The reduced model is my choice for the final model, given that no significant interactions exist.

23. Describe the association of taste with the chemicals using the coefficients and CIs from the final model. Recall in your description that the model's coefficients represent adjusted estimates.

24. Comparison between the coefficient and CI for the log-transformed hydrogen sulfide SLR model and the coefficient and CI for the log-transformed hydrogen sulfide in the MLR model:

SLR:

```
## (Intercept)       logHS
##   -9.786909    5.776095
```

```
##                  2.5 %   97.5 %
## (Intercept) -21.991270 2.417453
## logHS         3.838588 7.713602
```

As can be seen, the coefficients from the SLR model for log-transformed hydrogen sulfide and those of the same log-transformed chemical in the MLR model are different. Adjustment affects parameter estimates in MLR models . . .