Does Size Matter? (Estimation of Banana Weight with a regression modeling appraoch)

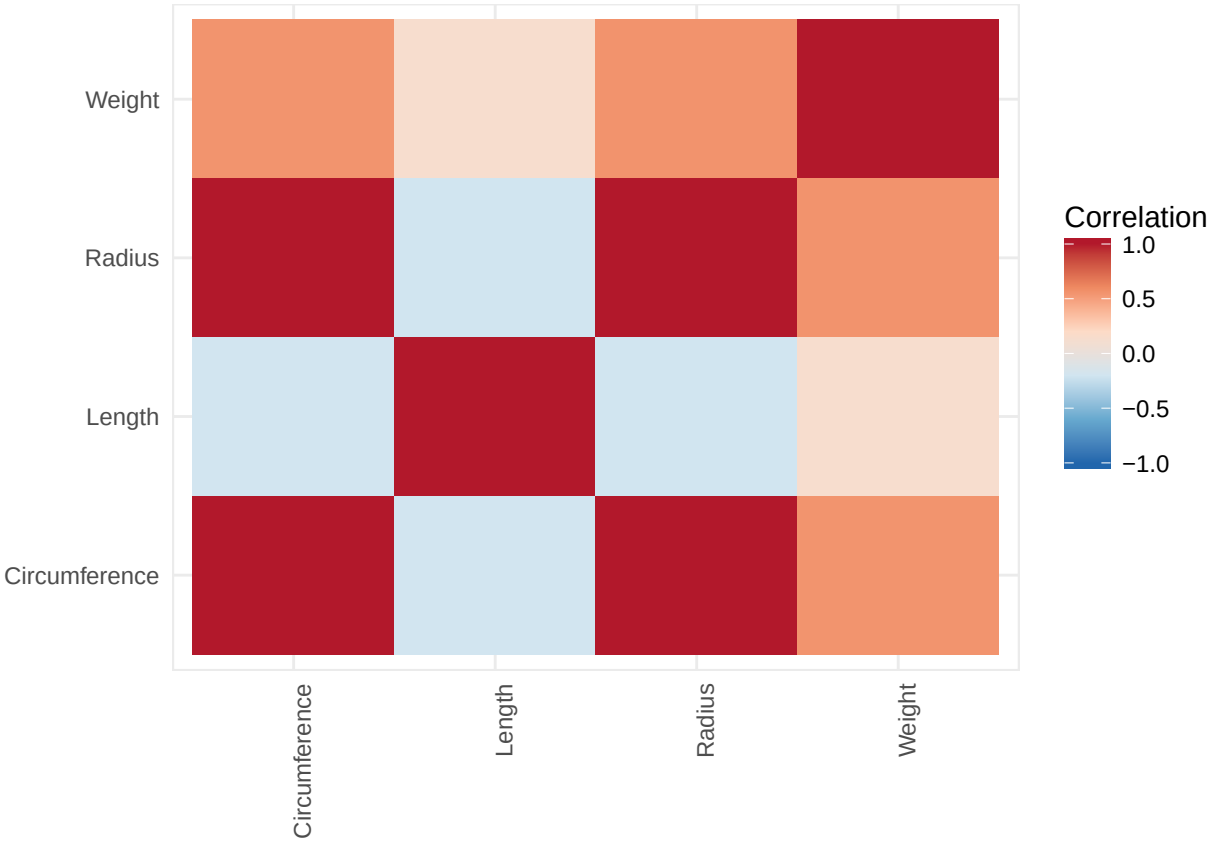
Scott Graham, Kaisa Roggeveen

February 13, 2018

Summary

Introduction

The purpose of this study was to determine the most effective regression model to predict the weight of a banana using external measurements. This study also demonstrated multiple techniques for developing regression models. These models were then examined to demonstrate their effectiveness at creating regression models.

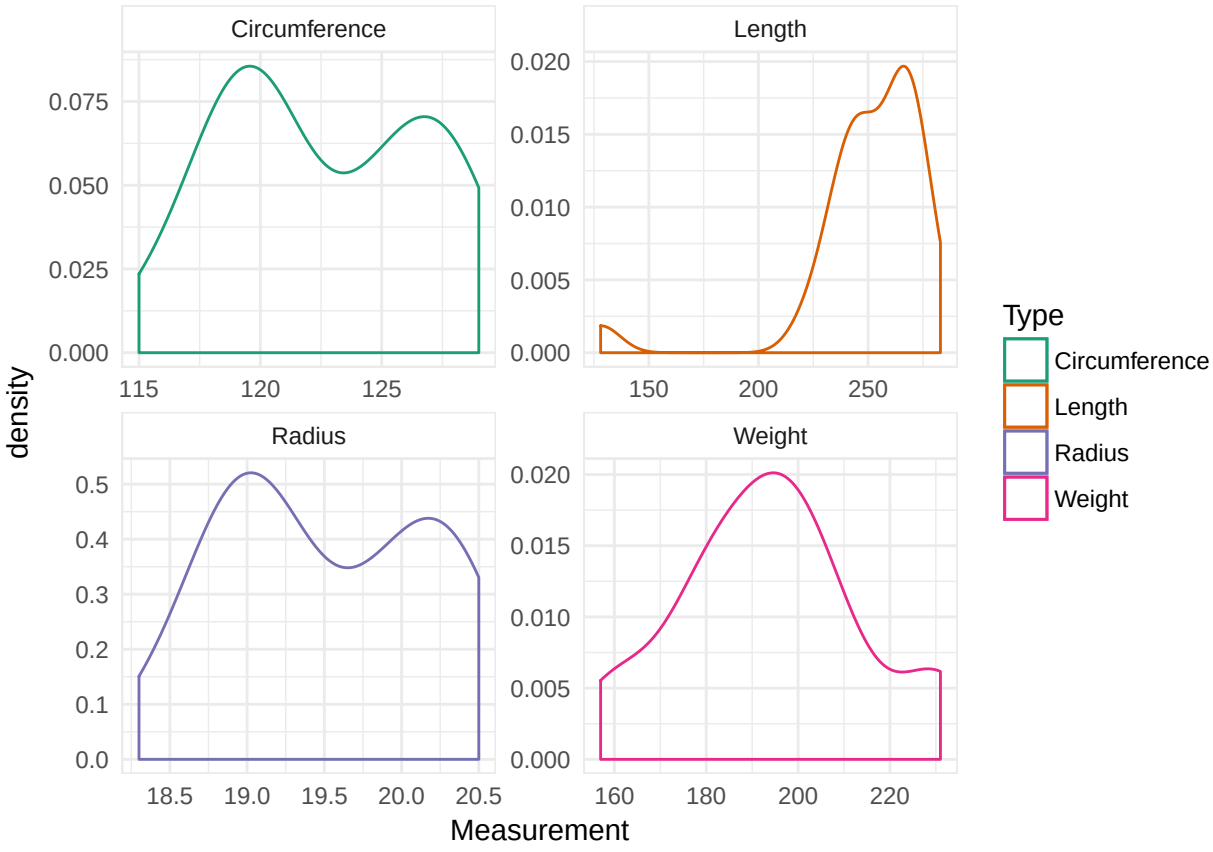


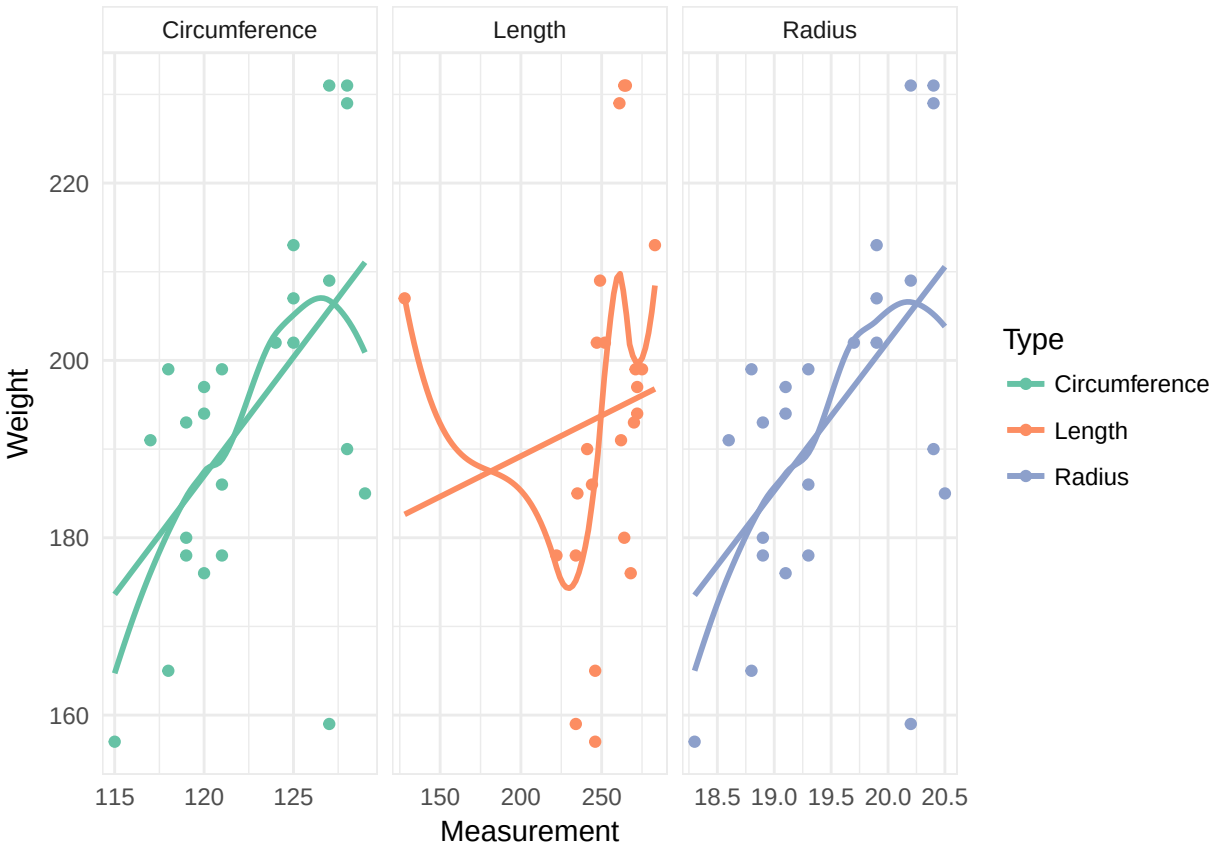

```
## Warning: Computation failed in `stat_bin()`:
## is.numeric(width) is not TRUE

## Warning: Computation failed in `stat_bin()`:
## is.numeric(width) is not TRUE

## Warning: Computation failed in `stat_bin()`:
## is.numeric(width) is not TRUE
```

```
## Warning: Computation failed in `stat_bin()`:
## is.numeric(width) is not TRUE
```






```
##
## Call:
## lm(formula = Weight_log ~ Length_log + Radius_log + Circumference_log,
##     data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24351 -0.06228  0.02400  0.06062  0.11528
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        10.0594    26.3199   0.382    0.706
## Length_log          0.1230     0.1275   0.965    0.346
## Radius_log          7.5264    14.0928   0.534    0.599
## Circumference_log  -5.7885    14.1601  -0.409    0.687
##
## Residual standard error: 0.09248 on 20 degrees of freedom
## Multiple R-squared:  0.3318, Adjusted R-squared:  0.2316
## F-statistic:  3.31 on 3 and 20 DF,  p-value: 0.04104
```

```
##
## Call:
## lm(formula = Weight_log ~ Radius_log + Length_log, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24861 -0.05259  0.02164  0.05202  0.11544
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.6702     1.9132  -0.350  0.72961
## Radius_log    1.7702     0.5596   3.163  0.00468 **
## Length_log    0.1223     0.1249   0.979  0.33888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09062 on 21 degrees of freedom
## Multiple R-squared:  0.3262, Adjusted R-squared:  0.2621
## F-statistic: 5.084 on 2 and 21 DF,  p-value: 0.01583
```

```
##
## Call:
## lm(formula = Weight_log ~ Radius_log, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25201 -0.05851  0.02530  0.05814  0.12150
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3046     1.6319   0.187  0.85363
## Radius_log    1.6689     0.5494   3.038  0.00604 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09054 on 22 degrees of freedom
## Multiple R-squared:  0.2955, Adjusted R-squared:  0.2635
## F-statistic: 9.227 on 1 and 22 DF,  p-value: 0.006043
```

```
##
## Call:
## lm(formula = Weight_log ~ Length_log, data = .)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.204865 -0.072290 -0.000595  0.055375  0.177835
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.99043    0.80371   6.209    3e-06 ***
## Length_log   0.04917    0.14574   0.337    0.739
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1076 on 22 degrees of freedom
## Multiple R-squared:  0.005146,   Adjusted R-squared:  -0.04007
## F-statistic: 0.1138 on 1 and 22 DF,  p-value: 0.739
```

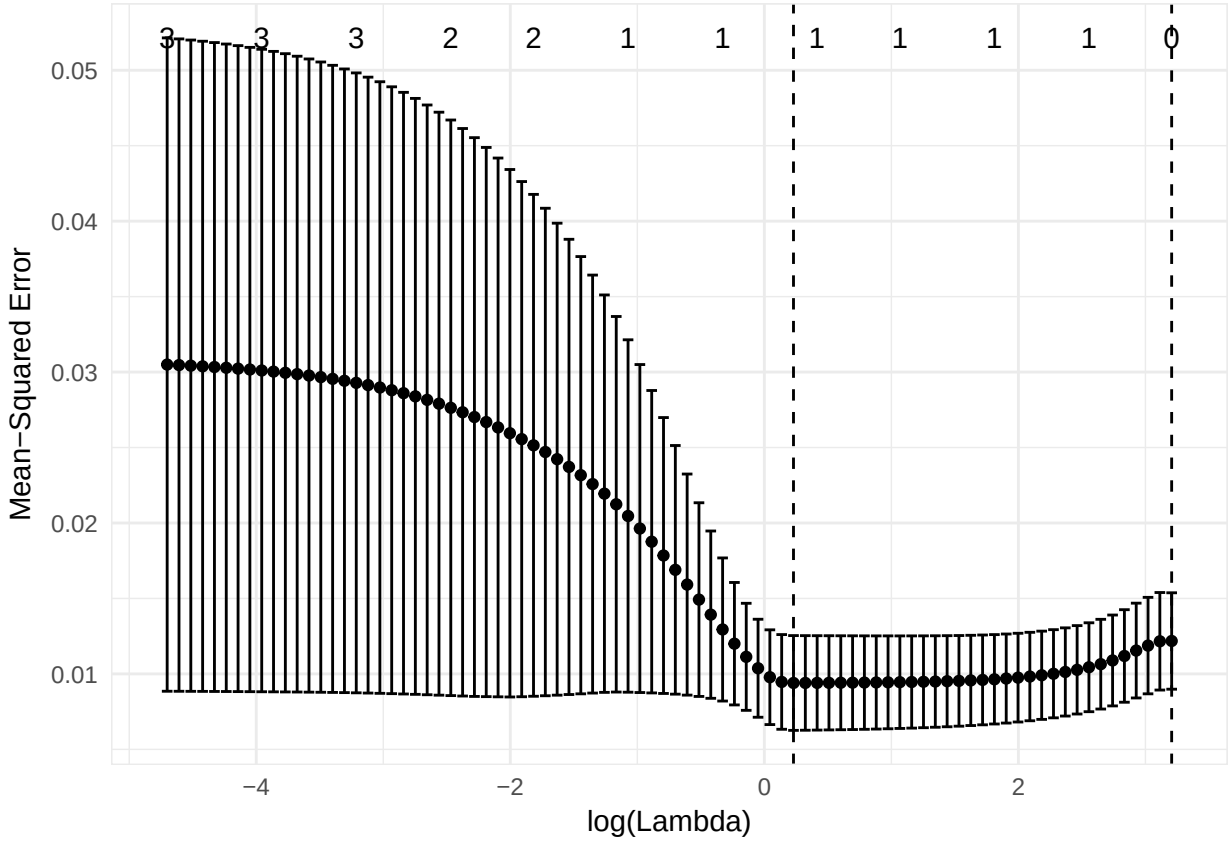
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
21	0.1724705	NA	NA	NA	NA
20	0.1710414	1	0.0014291	0.1671057	0.687041

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
22	0.1803370	NA	NA	NA	NA
20	0.1710414	2	0.0092956	0.5434713	0.5890658

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
22	0.1803370	NA	NA	NA	NA
21	0.1724705	1	0.0078665	0.9578258	0.3388762

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations
## per fold
```

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations
## per fold
```

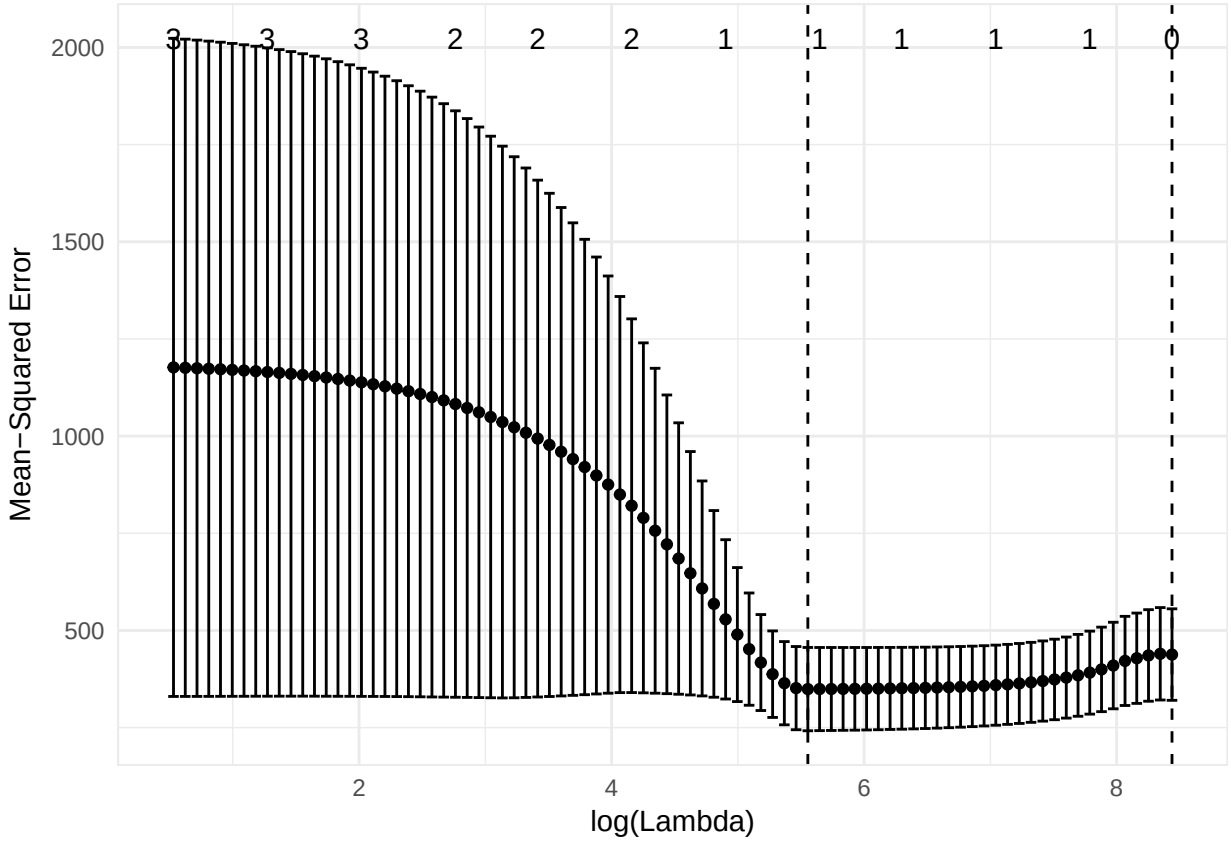


	1
(Intercept)	0.55713
Length	0.00000
Radius	0.00000
Circumference	0.00000
Length_log	0.00000
Radius_log	1.58392
Circumference_log	0.00000

	1
(Intercept)	5.261471
Length	0.000000
Radius	0.000000
Circumference	0.000000
Length_log	0.000000
Radius_log	0.000000
Circumference_log	0.000000

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations
## per fold

## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations
## per fold
```



	1
(Intercept)	-730.8216
Length	0.0000
Radius	0.0000
Circumference	0.0000
Length_log	0.0000
Radius_log	311.3111
Circumference_log	0.0000

	1
(Intercept)	193.7917
Length	0.0000
Radius	0.0000
Circumference	0.0000
Length_log	0.0000
Radius_log	0.0000
Circumference_log	0.0000

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.002597 0.469236 0.673413 0.613353 0.802255 0.984631
```