

MODEL SELECTION AND CROSS VALIDATION

Licks to the Center of a Tootsie Pop



MARISSA JENSEN

960:486 SPRING 2020

Jack Mardekian

ABSTRACT:

“How many licks does it take to get to the center of a tootsie pop?” A previous 2015 study by Purdue University revealed an average of 252 licks (tootsie.com). While answering the age-old question, the study neglected to propose contributing factors to this number. Why does it take ~250 licks to get to the center of a tootsie pop? Rather than asking Mr. Owl, we took matters into our own hands. Here, model selection and cross validation simulations dependent on p-values suggest that among Candyland, USA* residents, the number of licks is mostly dependent on stimulated saliva flow, followed by tongue length, and finally, pop diameter.

INTRODUCTION:

“How many licks does it take to get to the center of a tootsie pop?” The question first posed in a tootsie pop commercial in 1970. The commercial depicted a young boy asking a number of animals the notorious question but was adapted to focus on one animal in particular: Mr. Owl. In the commercial, Mr. Owl counts his licks aloud: “One... two... three...” only to be finished with a crunch, the stick returned to the boy empty, and the question remaining unanswered. The commercial and its outstanding question became a staple of American popular culture, and even the basis of student studies searching for an answer. A 2015 Purdue University study averaged 252 licks to reach the center (tootsie.com). While Mr. Owl was a ways off, it begs the question – Why? Why does it take ~250 licks to reach the center of a tootsie pop? Data collected from 16 Candyland, USA* residents was used to help answer this question. Factors like pop diameter, tongue size, gender, stimulated saliva flow, and number of licks were all recorded per individual. Data on the average pop size, tongue size, and stimulated saliva flow were researched and recorded as follows:

Average pop size (cm)	2.70	(Borelli, 2015)
Average tongue size (cm)	Men: 8.5 Women: 7.9	(onhealth.com, 2016)
Average stimulated saliva flow	4-5 mL/min	(Iorgulescu, 2009)

More specifically, the contributing factors to average licks were analyzed via model selection and cross validation. Model selection, best described as selecting a statistical model among several models given data, will assist in deciding which factors contribute to the result: number of licks. These factors will be investigated in a number of ways: [stepwise] forward selection, backward selection, and “both” selection. The factors will be weighed via P-values. 5-fold cross validation is also conducted, splitting the data into 5 groups to be resampled. This is particularly convenient for limited data samples, as our set is only 16 individuals (Brownlee, 2018).

MATERIALS AND METHODS:

The data collected from 16 participating Candyland, USA* residents was recorded as follows:

Tongue size (cm)	Pop diameter (cm)	Sex (1 M, 2 F)	Stimulated saliva flow rate (ml/min)	Licks (30 sec interval)
8.5	2.67	1	4.3	250
7.8	2.72	1	4.7	241

8.5	2.70	1	5	210
7.8	2.69	2	4.3	251
8.3	2.73	2	4.8	220
8.7	2.7	2	5.1	199
7.7	2.7	2	3.9	271
8.3	2.67	1	4.6	250
8.8	2.72	1	5.3	179
7.6	2.69	2	3.6	300
7.8	2.7	2	4	250
8.5	2.72	1	4	240
8.1	2.73	1	3.8	265
7.7	2.68	2	4.6	254
8.4	2.7	1	4	260
7.8	2.7	2	4.8	239

Here, licks are measured in 30-second intervals. This data was used to generate a model for the factors contributing to average licks in R Studio. The packages used in R Studio were faraway, olssr, and caret. The code written is specifically as follows **:

```

1 # Candyland data
2 library(faraway)
3 tongue <- c(8.5,7.8,8.5,7.8,8.3,8.7,7.7,8.3,8.8,7.6,7.8,8.5,8.1,7.7,8.4,7.8)
4 pop <- c(2.67,2.72,2.70,2.69,2.73,2.70,2.70,2.67,2.72,2.69,2.70,2.72,2.73,2.68,2.70,2.70)
5 gender <- c(1,1,1,2,2,2,1,1,2,2,1,1,2,1,2)
6 saliva <- c(4.3,4.7,5.0,4.3,4.8,5.1,3.9,4.6,5.3,3.6,4.0,4.0,3.8,4.6,4.0,4.8)
7 licks <- c(250,241,210,251,220,199,271,250,179,300,250,240,265,254,260,239)
8 # This is used to generate vectors of data
9 tootsie <- data.frame (tongue,pop,gender,saliva,licks)
10 # This combines the data into a set, called tootsie
11 tootsie_mod <- lm(licks~ tongue + pop + gender + saliva, data=tootsie)
12 summary(tootsie_mod)
13 # This step models a linear regression of the tootsie data based upon all the factors recorded
14 require(olsrr)
15 #P value is how we will be conducting the selection. Now we conduct our model selection via stepwise, forward, both and backward
16 forward <- ols_step_forward_p(tootsie_mod,details=FALSE)
17 forward
18 both <- ols_step_both_p(tootsie_mod,details=FALSE)
19 both
20 back <- ols_step_backward_p(tootsie_mod,details=FALSE)
21 back
22 forward <-lm(licks~ saliva + tongue + pop + gender, data=tootsie)
23 summary(forward)
24 both <- lm(licks~ saliva + tongue + pop, data=tootsie)
25 summary(both)
26 backward <- lm(licks~ saliva + tongue + pop + gender, data=tootsie)
27 summary(backward)
28 # The following package will allow us to conduct a 5-fold cross validation for the models created
29 require(caret)
30 set.seed(12345)
31 train.control <-trainControl(method="cv", number=5)
32 model_forward <- train(licks~ saliva + tongue + pop + gender, data= tootsie, method="lm", trControl=train.control)
33 print(model_forward)
34 set.seed(12345)
35 train.control <-trainControl(method="cv", number=5)
36 model_both <- train(licks~ saliva + tongue + pop, data=tootsie, method="lm", trControl=train.control)
37 print(model_both)
38 set.seed(12345)
39 train.control <-trainControl(method="cv", number=5)
40 model_backward <-train(licks~ saliva + tongue + pop + gender, data=tootsie, method="lm", trControl= train.control)
41 print(model_backward)
42 summary(model_forward)
43 summary(model_both)
44 summary(model_backward)

```

Faraway:

The faraway package contains the information for linear models.

Olssr:

The olssr package contains the information/simulations for stepwise selection. The simulations used specifically are explained below:

Stepwise_forward_p— This is stepwise forward selection dependent on p-values. In this, the model begins with no variables; also known as the “null model.” The first step adds the most

significant variable, and continues to add the next most significant variable until there are no variables left that meet the criteria; in this case, significant p-values ($p < 0.05$) (Flom, 2018).

Stepwise_both_p— This is a “both” selection, dependent on p-values. In this, the selection alternates between forward and backward, bringing in or taking out variables until there are no variables left that meet the criteria; in this case, significant p-values ($p < 0.05$) (Flom, 2018).

Stepwise_backward_p— This is a stepwise backward elimination dependent on p-values. In this, the model begins with all the variables selected, and removes the least significant one at each step until there are no variables left that meet the criteria; in this case, significant p-values ($p < 0.05$) (Flom, 2018).

Caret Package:

The caret package contains the information/simulations for cross validation. Here, “trainControl” and “train” from the caret package are employed to train the model for cross validation. Cross validation algorithms are best described below:

- “1. Reserve a small sample of the data set
2. Build (or train) the model using the remaining part of the data set
3. Test the effectiveness of the model on the the reserved sample of the data set. If the model works well on the test data set, then it’s good.”

- Kassambara, 2018

In this case, a 5-fold cross validation is conducted for all three models. Set.seed is employed so that data is replicable. Finally, summaries of each model are generated for comparison.

RESULTS:

The stepwise selection code ran as follows **:

```
> require(olsrr)
> forward <- ols_step_forward_p(tootsie_mod,details=FALSE)
> forward
```

Step	Variable Entered	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	saliva	0.7904	0.7754	32.3708	133.3738	13.8496
2	tongue	0.8953	0.8791	12.1731	124.2746	10.1600
3	pop	0.9386	0.9232	4.9993	117.7308	8.0970
4	gender	0.9480	0.9291	5.0000	117.0589	7.7795

```
> both <- ols_step_both_p(tootsie_mod,details=FALSE)
> both
```

Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	saliva	addition	0.790	0.775	32.3710	133.3738	13.8496
2	tongue	addition	0.895	0.879	12.1730	124.2746	10.1600
3	pop	addition	0.939	0.923	4.9990	117.7308	8.0970

```
> back <- ols_step_backward_p(tootsie_mod,details=FALSE)
> back
[1] "No variables have been removed from the model."
```

The forward selection added all variables, saliva, tongue, pop and gender to the model. The order in which variables were added is present in this summary; saliva, then tongue, then pop, then gender. This order is consistent with the algorithm's prioritization of variable contribution; in other words, the order in which the variables were added is also their ranking from most to least significant.

The summary of the forward selection model reads **:

```
> summary(forward)
```

Call:
lm(formula = licks ~ saliva + tongue + pop + gender, data = tootsie)

Residuals:

Min	1Q	Median	3Q	Max
-16.323	-3.667	1.081	3.828	9.877

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1576.823	291.483	5.410	0.000214	***
saliva	-38.369	4.765	-8.052	6.14e-06	***
tongue	-30.977	7.360	-4.209	0.001463	**
pop	-333.871	107.793	-3.097	0.010152	*
gender	-6.975	4.933	-1.414	0.185049	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.78 on 11 degrees of freedom
Multiple R-squared: 0.948, Adjusted R-squared: 0.9291
F-statistic: 50.17 on 4 and 11 DF, p-value: 5.366e-07

Note that p-value for gender is > 0.05

As the code suggests, the p-value for saliva is < 0.001 , the p-value for tongue is < 0.01 , the p-value for pop is < 0.05 , but the p-value for gender is > 0.05 .

The “both” selection added only three variables: saliva, tongue, and pop. This is consistent with gender being added last to the forward selection; both has deemed the least significant variable from the forward selection an insignificant variable in the both selection.

The summary of the “both” selection model reads **:

```
> both <- lm(licks~ saliva + tongue + pop, data=tootsie)
> summary(both)
```

Call:
lm(formula = licks ~ saliva + tongue + pop, data = tootsie)

Residuals:

Min	1Q	Median	3Q	Max
-18.4752	-4.1491	0.9309	6.0256	8.2900

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1502.817	298.446	5.035	0.000291	***
saliva	-40.326	4.746	-8.498	2.02e-06	***
tongue	-24.702	6.111	-4.042	0.001632	**
pop	-326.058	112.044	-2.910	0.013079	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.097 on 12 degrees of freedom
Multiple R-squared: 0.9386, Adjusted R-squared: 0.9232
F-statistic: 61.14 on 3 and 12 DF, p-value: 1.53e-07

Here, gender is eliminated. The p-value for saliva is < 0.001 , the p-value for tongue is < 0.01 , and the p-value for pop is < 0.05 .

The backward selection removed no variables from the model. It includes saliva, tongue, pop and gender, as does the forward selection model.

The summary of the backward model selection reads **:

```
> backward <- lm(licks~ saliva + tongue + pop + gender, data=tootsie)
> summary(backward)
```

Call:
lm(formula = licks ~ saliva + tongue + pop + gender, data = tootsie)

Residuals:

Min	1Q	Median	3Q	Max
-16.323	-3.667	1.081	3.828	9.877

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1576.823	291.483	5.410	0.000214	***
saliva	-38.369	4.765	-8.052	6.14e-06	***
tongue	-30.977	7.360	-4.209	0.001463	**
pop	-333.871	107.793	-3.097	0.010152	*
gender	-6.975	4.933	-1.414	0.185049	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.78 on 11 degrees of freedom
Multiple R-squared: 0.948, Adjusted R-squared: 0.9291
F-statistic: 50.17 on 4 and 11 DF, p-value: 5.366e-07

As the code suggests, the p-value for saliva is < 0.001 , the p-value for tongue is < 0.01 , the p-value for pop is < 0.05 , but the p-value for gender is < 1 .

The forward selection cross-validation code ran as follows **:

```
> require(caret)
> set.seed(12345)
> train.control <- trainControl(method="cv", number=5)
> model_forward <- train(licks~ saliva + tongue + pop + gender, data= tootsie, method="lm", trControl=train.control)
> print(model_forward)
Linear Regression

16 samples
4 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 12, 13, 13, 13, 13
Resampling results:

    RMSE      Rsquared   MAE
9.558873  0.9292211  8.37598

Tuning parameter 'intercept' was held constant at a value of TRUE
```

The “both” selection cross-validation code ran as follows **:

```
> set.seed(12345)
> train.control <- trainControl(method="cv", number=5)
> model_both <- train(licks~ saliva + tongue + pop, data=tootsie, method="lm", trControl=train.control)
> print(model_both)
Linear Regression

16 samples
3 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 12, 13, 13, 13, 13
Resampling results:

    RMSE      Rsquared   MAE
8.557208  0.937969  7.516689

Tuning parameter 'intercept' was held constant at a value of TRUE
```

The backward selection cross-validation code ran as follows **:

```
> set.seed(12345)
> train.control <- trainControl(method="cv", number=5)
> model_backward <- train(licks~ saliva + tongue + pop + gender, data=tootsie, method="lm", trControl= train.control)
> print(model_backward)
Linear Regression

16 samples
 4 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 12, 13, 13, 13, 13
Resampling results:

    RMSE      Rsquared   MAE
9.558873  0.9292211  8.37598

Tuning parameter 'intercept' was held constant at a value of TRUE
```

For clarity, cross-validation values are displayed in a tabular format with their respective models below:

Cross-validation model	RMSE value	Rsquared value	MAE value
forward	9.559	0.929	8.376
both	8.557	0.938	7.517
backward	9.559	0.929	8.376

The “both” model produced the lowest RMSE, the lowest MAE and highest Rsquared. Forward and backward models have higher RMSE values than “both”, lower Rsquared values and higher MAE values.

DISCUSSION:

It is interesting to note that saliva ultimately had the most contribution to number of licks. Saliva is followed by tongue, then pop. Stimulated saliva flow and tongue length did not display any correlation. The initial model selection simulations were inconsistent; forward and backward

selection yielded one answer, while “both” yielded another. Cross-validation results with the tabled values above bring clarity to the results. The RMSE value, indicative of how close the model points are to the real data points, is lowest in “both,” indicating the best fit. Additionally, the Rsquared value indicating a relative measure of fit (whereas RMSE is actual) is highest in “both,” indicating better fit. It is important to note that trusting the Rsquared value alone can be misleading; even the best Rsquared values can be products of overfitting. However, the RMSE and MAE values are consistent with “both” being the best model. The MAE value, or mean absolute error value, is lowest in “both,” indicating a better fit.

CONCLUSION:

“How many licks does it take to get to the center of a tootsie pop?” The famous fifty-year-old question put to the test by Purdue University revealed it takes approximately 252 licks. While answering the question “how?” the study did not answer the question “why?”. Data collected from Candyland, USA* has allowed us to investigate the answer to that very question. Investigating number of licks on the basis of pop diameter, stimulated saliva flow, tongue length and gender have led to this conclusion: the number of licks is mostly dependent on stimulated saliva flow, followed by tongue length, and finally, pop diameter. All of the values from cross validation, the gender variable p-value of ~ 0.18 , and the previous knowledge of men and women having different tongue length averages, suggests that the best model includes saliva, pop, and tongue while excluding gender.

Literature Cited

Borrelli, L. (2015, February 11). How Many Licks It Takes To Get To The Center Of A Tootsie Pop, Based On Science. Retrieved from <https://www.medicaldaily.com/tootsie-pop-mystery-solved-how-many-licks-it-takes-get-center-based-science-321802>

Flom, P. (2018, December 11). Stopping stepwise: Why stepwise selection is bad and what you should use instead. Retrieved from <https://towardsdatascience.com/stopping-stepwise-why-stepwise-selection-is-bad-and-what-you-should-use-instead-90818b3f52df>

Iorgulescu, G. (2009). Saliva between normal and pathological. Important factors in determining systemic and oral health. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052503/>

Kassambara, A. (2018). *Machine learning essentials: practical guide in R*. Scotts Valley, Kalifornien: CreateSpace Independent Publishing Platform.

Mardekian, J. (2020). Ex_Cross_Validation_Output.txt

Mardekian, J. (2020). FatModelSelection_out.txt.

Tootsie Roll Inc. (n.d.). Retrieved from <https://tootsie.com/howmanylick-experiments>

10 Fun Facts About Your Tongue and Taste Buds. (2016, June 21). Retrieved from https://www.onhealth.com/content/1/tongue_facts

APPENDICES:

*All data from Candyland, USA was fictionally generated. No software was used. The 16 residents' data was completely made up by Marissa Jensen, using the previous research of pop diameter, tongue sizes (and the gender correlation), stimulated saliva flow, and the 2015 Purdue University study.

**All code, R knowledge, etc. is directly from Professor Mardekian and the material from 960:486 Spring semester of 2020. Files FatModelSelection_out.txt and Ex_Cross_Validation_Output.txt contain code and output for example simulations done in class. This code was used as a model for the model selection and cross validation simulations in this project. These are cited in Literature Cited (Jensen, 12).

This project in its entirety was completed by Marissa Jensen. The idea, research, code, writing, etc. was produced by Marissa Jensen only. She is the sole contributor of this project. The code models aforementioned are acknowledged and cited appropriately. The research is acknowledged and cited appropriately.