Parametric Bootstrapping

CJ Carani, Dylan Hunt and Matt Kaznikov

02-21-2024

Background:

W.A. Nelson, renowned American statistician, from the Encyclopedia of Ecology states, "Bootstrapping methods are a numerical approach to generating confidence intervals that use either resampled data or simulated data to estimate the sampling distribution of the maximum likelihood parameter estimates". In addition to generating confidence intervals, bootstrapping is used often for data where there is a lack of observations or concerns about model assumptions (like normal distribution or skewed data). Bootstrapping solves this issue by using a technique called sampling with replacement. Sampling with replacement is taking the original data and randomly sampling observations while putting the data back into the original dataset. These steps are then repeated thousands of times to generate a simulated dataset. The idea is that this simulated data will provide us with better reasoning on how well our models perform and which model to use by creating a comparison with the use of the new data. While non-parametric bootstrapping bases the simulated data on the original data, parametric bootstrapping simulates data based on a model's estimates. The steps to use parametric bootstrapping are as follows:

1) Fit a model to the original dataset and record that model's parameters using their estimated values.
2) Next using the recorded estimated parameter values simulate many datasets.
3) For each newly created dataset, record each of the estimated parameters.
4) Then calculate statistics that are important for your conclusions. (p-values, F-statistic, t-statistics, etc.)
5) Create confidence intervals based off of the simulated values.

There are multiple reasons that we would want to use simulated data based on a model, but the most common one is when we have confidence in a model but doubt a model's standard error terms. While if we had confidence in both the standard error and the model we would use the theory based R estimates. And, if we had faith in the standard error but not in the model we would use non-parametric bootstrapping. We must have faith in the model for parametric bootstrapping because parametric bootstrapping operates on the assumption that the estimates follow a normal distribution. This assumption comes from the fact that the simulated values are generated based on the fitted model. This simulation allows us to generate confidence intervals that are more precise and trustworthy than the original data.

Methods:

For each of the three models we use parametric bootstrapping to improve our datasets or to improve our models. The steps used to complete this follow the steps highlighted within our background section. First we create a model using the original dataset using the typical functions (lm and glmer). Following this step, we record the parameter estimates and standard deviation. Next we create vectors of a certain size, x (10,000) that will be used to hold the parameter estimates and standard deviations for each of the simulated models. Then we create a duplicate dataset using the original dataset. We are now able to begin our for loop that will run for x numbers of times. Now that we are within our for loop we want to begin to create our new simulated data. We use a mutate command to change the duplicated dataset to hold simulated data instead of the original data. Following, our model is run on the new simulated data then those estimates are collected within the vectors we created earlier. After the for loop is done we use the collected simulated data estimates to run tests to understand how well our models are and their estimated parameters.

Data/Results:

Linear Model Comparison:

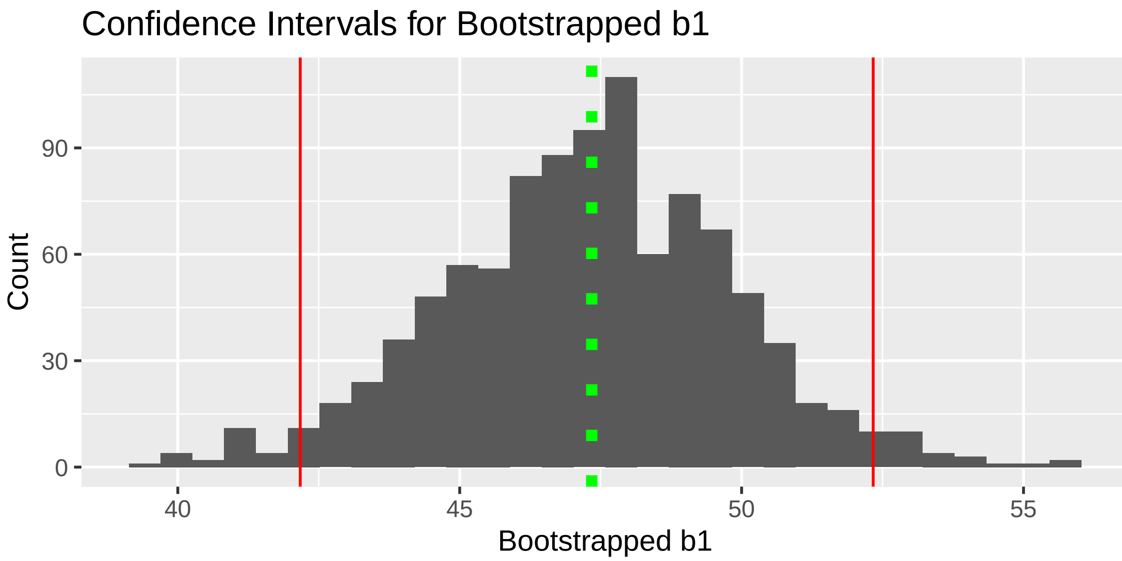
Model 1: Reduced Model

```
##
## Call:
## lm(formula = price ~ bedrooms + room_type, data = Airbnb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -213.86  -27.47   -8.47   16.53  930.50
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             69.496      4.982  13.948  < 2e-16 ***
## bedrooms                47.342      2.506  18.892  < 2e-16 ***
## room_typePrivate room  -53.372      3.998 -13.350  < 2e-16 ***
## room_typeShared room   -82.138     10.130  -8.108 1.03e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68.36 on 1557 degrees of freedom
## Multiple R-squared:  0.4027, Adjusted R-squared:  0.4015
## F-statistic: 349.9 on 3 and 1557 DF,  p-value: < 2.2e-16
```

Model 1: Bootstrapping

```
##   bootstrap_b0     bootstrap_b1    bootstrap_b2       bootstrap_b3
##  Min.   :-82.78   Min.   :39.46   Min.   :-11.80524   Min.   :-35.72647
##  1st Qu.:-69.23   1st Qu.:45.60   1st Qu.: -2.47724   1st Qu.: -6.16005
##  Median :-65.97   Median :47.36   Median :  0.06112   Median :  0.08577
##  Mean   :-65.99   Mean   :47.29   Mean   :  0.06336   Mean   :  0.28372
##  3rd Qu.:-62.84   3rd Qu.:49.03   3rd Qu.:  2.69659   3rd Qu.:  7.21172
##  Max.   :-48.87   Max.   :55.79   Max.   : 15.38318   Max.   : 28.54360
##  bootstrap_sigma
##  Min.   :64.74
##  1st Qu.:67.47
##  Median :68.33
##  Mean   :68.29
##  3rd Qu.:69.10
##  Max.   :72.44
```

Model 1: Confidence Interval for Bootstrapped b0




```
##      2.5%    97.5%
## 42.17177 52.33334
```

Model 2: Full Model

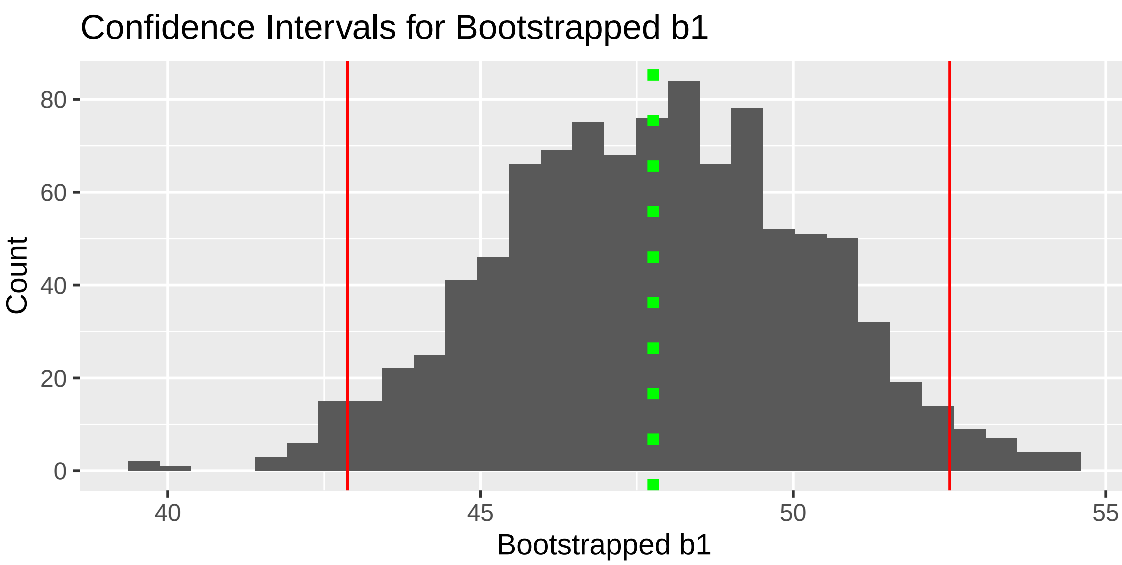
```
##
## Call:
## lm(formula = price ~ bedrooms + room_type + overall_satisfaction +
##     reviews, data = Airbnb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -225.14  -28.17   -7.71   17.81  920.90
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -72.21935   24.62148  -2.933   0.0034 **
## bedrooms                 47.76056    2.47712  19.281  < 2e-16 ***
## room_typePrivate room -54.19290    3.95588 -13.699  < 2e-16 ***
## room_typeShared room  -81.64921   10.02502  -8.145 7.73e-16 ***
## overall_satisfaction   30.35449    5.06886   5.988 2.63e-09 ***
## reviews                -0.11371    0.04913  -2.314   0.0208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.55 on 1555 degrees of freedom
## Multiple R-squared:  0.4176, Adjusted R-squared:  0.4157
## F-statistic:   223 on 5 and 1555 DF,  p-value: < 2.2e-16
```

Model 2: Bootstrapping

```
##   bootstrap_b0     bootstrap_b1    bootstrap_b2        bootstrap_b3
##  Min.   :-280.0   Min.   :39.71   Min.   :-14.22436   Min.   :-29.1728
##  1st Qu.:-225.0   1st Qu.:46.03   1st Qu.: -2.85747   1st Qu.: -6.2459
##  Median :-209.6   Median :47.84   Median : -0.05566   Median :  0.3115
```

```
##  Mean    :-208.5   Mean   :47.76   Mean    : -0.02701   Mean    :  0.3794
##  3rd Qu.:-192.4   3rd Qu.:49.45   3rd Qu.:  2.71135   3rd Qu.:  7.4302
##  Max.    :-126.6   Max.   :54.43   Max.    : 14.40938   Max.    : 31.4701
##   bootstrap_b4     bootstrap_b5    bootstrap_sigma
##  Min.    :14.16   Min.    :14.16   Min.    :64.04
##  1st Qu.:26.95   1st Qu.:26.95   1st Qu.:66.80
##  Median :30.66   Median :30.66   Median :67.55
##  Mean    :30.43   Mean    :30.43   Mean    :67.59
##  3rd Qu.:33.67   3rd Qu.:33.67   3rd Qu.:68.37
##  Max.    :44.85   Max.    :44.85   Max.    :71.66
```

Model 2: Confidence Intervals for Bootstrapped b1

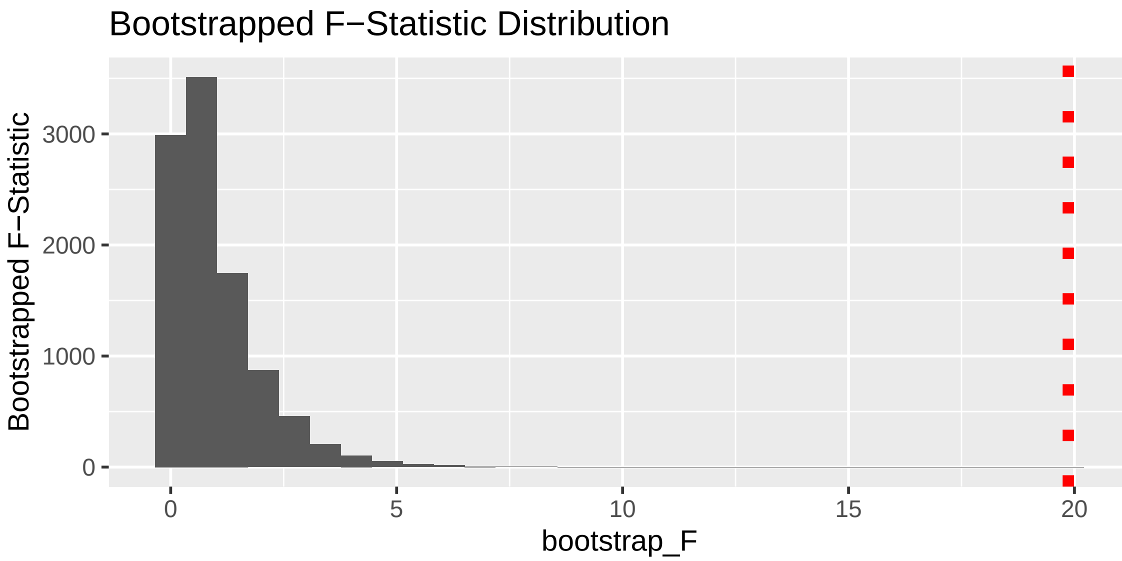



```
##      2.5%    97.5%
## 42.87524 52.50396
```

ANOVA Test

```
## Analysis of Variance Table
##
## Model 1: price ~ bedrooms + room_type + overall_satisfaction + reviews
## Model 2: price ~ bedrooms + room_type
##   Res.Df     RSS Df Sum of Sq      F   Pr(>F)
## 1   1555 7094695
## 2   1557 7275958 -2   -181263 19.864 3.03e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bootstrapping F-Statistic




```
## [1] 0
```

Linear Model Comparison Results:

For our first model we used the Airbnb data to test the overall Airbnb price as seen in class. We had a reduced model where we only had bedrooms and room_type and a full model where we also added two extra explanatory variables of overall_satisfaction and reviews. After our models were made, we wanted to understand the reliability of the relationship between Airbnb pricing and the number of bedrooms. To conduct this, we tested parametric bootstrapping of our b1 variable among both models and received confidence intervals in return. For the reduced model we received a 95% confidence interval ranging from 42.67896 to 51.96590, and for the full model it ranged from 42.63542 to 52.69015. We also added a green dotted line in each of the bootstrapping b1 plots for our estimated b1 from the original models. As expected, this estimate is in the center of the distribution which reaffirms that our data was generated from the models. This displays how bootstrapping simulates data based on our model estimates and how we would expect a bell curve with our estimate in the center. After this, we conducted bootstrapping on the F statistic to compare the goodness of fit between the two models for 10,000 resamples. The observed F-statistic value proved to be statistically significant due to the right distribution of the F-statistics. What this tells us is that the full model (with two extra explanatory variables) is very unlikely to have occurred by chance. This conclusion is also consistent with what we saw in our investigation where variables like overall_satisfaction which was added in the second model is highly correlated with price.

Model without Large Dataset and Comparison:

Summary of Model 1 & Model 2

```
##
## Call:
## lm(formula = MATINGS ~ AGE, data = elephants)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1158 -1.3087 -0.1082  0.8892  4.8842
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

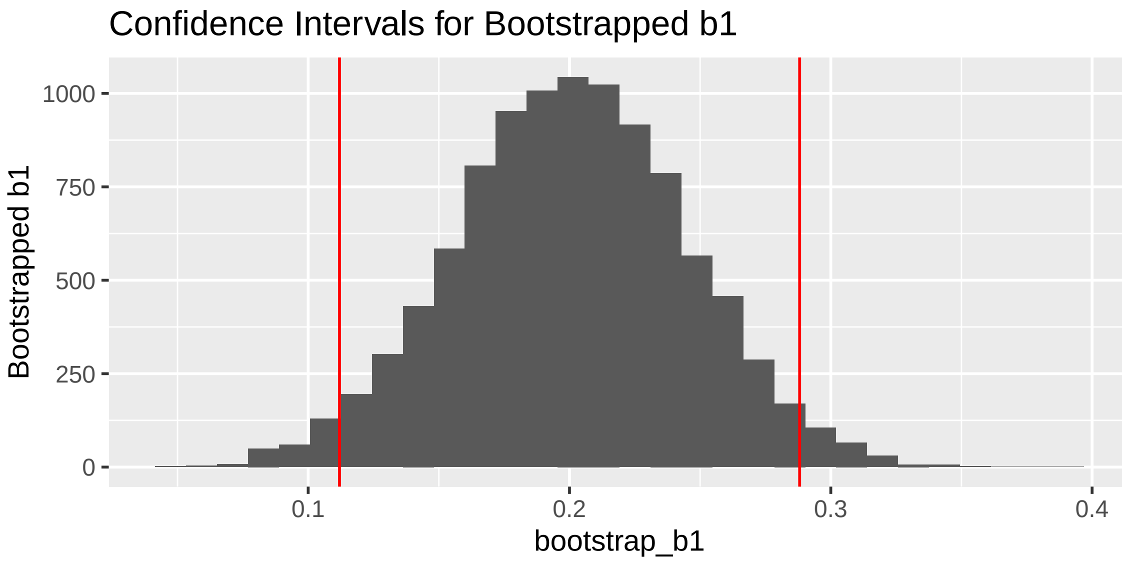
```
## (Intercept) -4.50589    1.61899  -2.783  0.00826 **
## AGE          0.20050    0.04443   4.513 5.75e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.849 on 39 degrees of freedom
## Multiple R-squared:  0.343,  Adjusted R-squared:  0.3262
## F-statistic: 20.36 on 1 and 39 DF,  p-value: 5.749e-05
```

```
##
## Call:
## lm(formula = MATINGS ~ AGE + age2, data = elephants)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0461 -1.2021 -0.1683  0.9962  4.9539
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.526838   9.464669   0.161    0.873
## AGE         -0.131194   0.514562  -0.255    0.800
## age2         0.004414   0.006821   0.647    0.521
##
## Residual standard error: 1.863 on 38 degrees of freedom
## Multiple R-squared:  0.3502, Adjusted R-squared:  0.316
## F-statistic: 10.24 on 2 and 38 DF,  p-value: 0.0002772
```

Model 1: Bootstrapping

```
##   bootstrap_b0      bootstrap_b1     bootstrap_sigma
##  Min.   :-11.446   Min.   :0.04248   Min.   :1.139
##  1st Qu.: -5.622   1st Qu.:0.17082   1st Qu.:1.694
##  Median : -4.531   Median :0.20079   Median :1.836
##  Mean   : -4.519   Mean   :0.20069   Mean   :1.838
##  3rd Qu.: -3.434   3rd Qu.:0.23078   3rd Qu.:1.978
##  Max.   :  1.325   Max.   :0.38599   Max.   :2.661
```

Model 2: Confidence Intervals for Bootstrapped b1

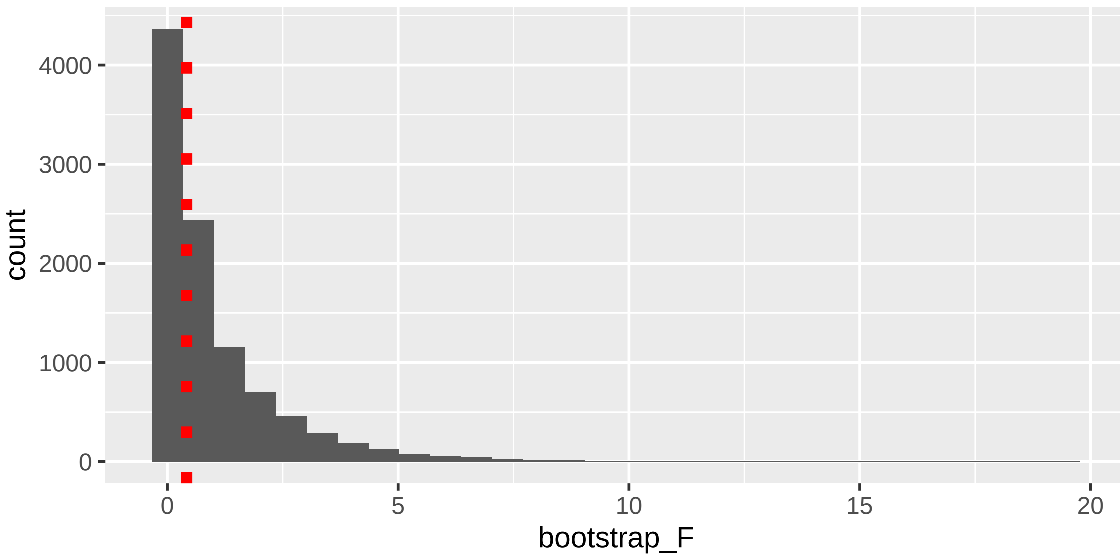



```
##      2.5%     97.5%
## 0.1119849 0.2880765
```

ANOVA Test

```
## Analysis of Variance Table
##
## Model 1: MATINGS ~ AGE
## Model 2: MATINGS ~ AGE + age2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     39 133.28
## 2     38 131.83  1    1.4526 0.4187 0.5215
```

Bootstrapping F-Statistic



###

Model without Large Dataset and Comparison Results:

To showcase a dataset that is not significantly large, we used the elephants dataset by Poole(1989) with matings as a response variable, as a dataset without a large amount of data. The major reason for using this dataset is because of its lack of data which makes it a prime example to use parametric bootstrapping. We are concerned about the estimates for the parameters as well as its F-statistic because of the low number of observations. In E_M1, for every additional year of age for an elephant, on average we expect to see the number of matings increase by 0.2. In E_M2, for every additional year of age for an elephant, we expect on average the number of matings to quadratically increase by a factor of 0.0058. The bootstrap confidence interval in the first histogram shows that the estimated parameter based on the real data is significant when compared against the bootstrapped data. The second histogram created is used to compare the F-statistic of the full model to the bootstrapping data. The F-statistic is very common which does not give us evidence that the full model is better than the reduced model. Our results are consistent with the results from class because in both examples, a quadratic term age^2 is not enough to create a significant model.

Model Comparison for Multilevel GLM:

Summary of GLM1 & GLM2:

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: foul.home ~ foul.diff + (1 | game)
##    Data: refdata
##
##      AIC      BIC   logLik deviance df.resid
##   6792.5   6812.1  -3393.3   6786.5     4969
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.6995 -0.9055 -0.6518  0.9655  1.6849
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  game   (Intercept) 0.273    0.5225
## Number of obs: 4972, groups:  game, 340
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.18886    0.04434  -4.259 2.05e-05 ***
## foul.diff   -0.26821    0.03895  -6.887 5.71e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## foul.diff 0.368
```

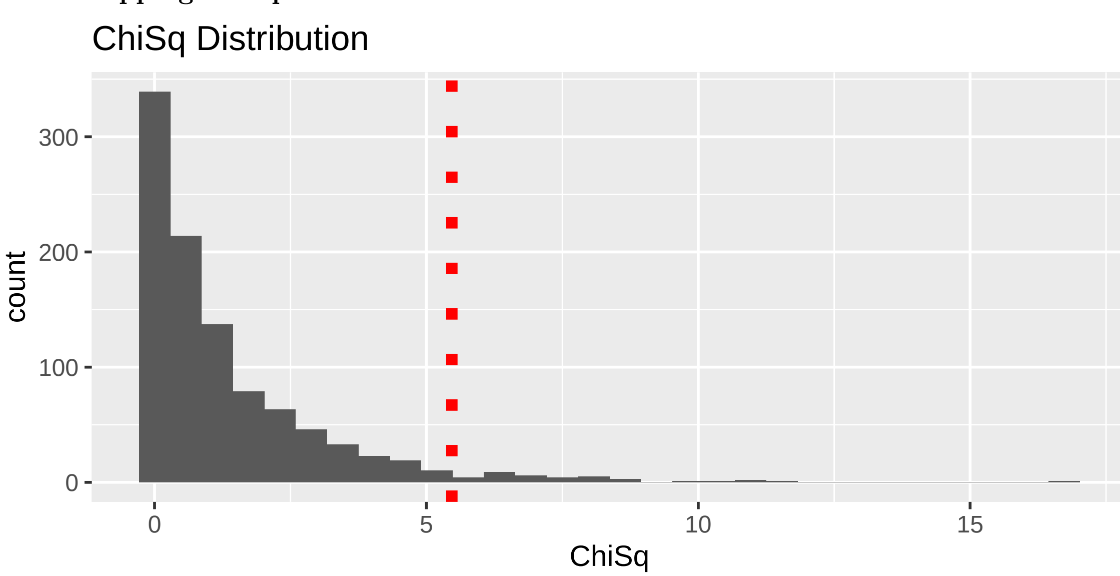
```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: foul.home ~ foul.diff + (foul.diff | game)
##    Data: refdata
##
##      AIC      BIC   logLik deviance df.resid
##   6791.1   6823.6  -3390.5   6781.1     4967
##
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q     Max
## -1.6399 -0.9087 -0.6349  0.9528  1.7687
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  game   (Intercept) 0.294141 0.54235
##         foul.diff   0.001235 0.03514  -1.00
## Number of obs: 4972, groups:  game, 340
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.15684    0.04637  -3.382 0.000719 ***
## foul.diff   -0.28533    0.03835  -7.440    1e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## foul.diff 0.192
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

ANOVA Test

```
## Data: refdata
## Models:
## glm1: foul.home ~ foul.diff + (1 | game)
## glm2: foul.home ~ foul.diff + (foul.diff | game)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## glm1    3 6792.5 6812.1 -3393.3   6786.5
## glm2    5 6791.1 6823.6 -3390.5   6781.1 5.4682  2    0.06495 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bootstrapping Chisq



Simulation Based P-Value

```
## [1] 0.037
```

Model Comparison for Multilevel GLM Results:

To display model comparison for a multilevel generalized linear model we used the basketball referees dataset with the goal of finding the odds of a home team getting called for a foul. From using parametric bootstrapping we get a simulated P-value of 0.037. This P-Value is smaller than the previously observed P-value. This result is consistent with the idea that when boundary constraints are an issue in estimating variance and correlation of random effects, theory based P-value tests are too conservative leading to P-values that are estimated higher than they actually occur. Our simulated P-value supports this hypothesis that our above P-value was likely too large due to a conservative approach to calculating the P-values in the ANOVA test. The simulated P-value also tells us that we do have enough evidence to say that our full model is better than our reduced model because the p-value is lower than 0.05. Thus, we can say that the addition of a random effect for foul differential is proven to be beneficial in modeling the odds a foul is called on a home team. This is consistent with our notes and discussion in class that it is clear that the addition of the random slope term of the foul differential makes a difference and is warranted.

Conclusions:

Parametric Bootstrapping is useful for generating simulated data from model estimates to solve problems that occur due to dataset size, issues with assumptions, or comparing models. This method of bootstrapping can be used on various different types of models. One of the more useful ways to use parametric bootstrapping is by using it to generate additional data to test various models to see which one is more useful for explaining a given dataset. We used this in all three of our examples and then used parametric bootstrapping to estimate the f-statistic and chi squared values. Parametric bootstrapping is very useful when we believe in the model, but do not believe in the standard error formulas. This is because parametric bootstrapping assumes that our simulated values follow a normal distribution because the data is based on a model. Non-parametric bootstrapping is useful when we believe in the standard error formulas, but not the model. This does not assume a normal distribution because we are generating our data by sampling with replacement. Overall, bootstrapping can be very useful when we doubt our confidence intervals, do not have enough data, or have doubts about model assumptions. But, parametric bootstrapping provides another way to fix these problems when we believe in our model and want to have our simulated data follow a normal distribution.

References:

https://www.sciencedirect.com/topics/earth-and-planetary-sciences/bootstrapping#:~:text=Parametric%20bootstrapping%20
https://stat455-w22.github.io/stat455-w22-notes/multilevel-generalized-linear-models.html#parametric-bootstrapping https://bookdown.org/roback/bookdown-BeyondMLR/ch-GLMM.html#cs:refs