**Overview:**

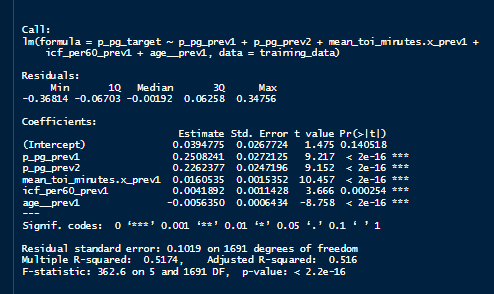
The purpose of the is research is to create a model that will allow us to predict players scoring production at 5v5 into the future and leverage that against their market value to find play driving players at superb value. This will mostly impact middle of the lineup players as top-end talent is signed for long periods of time and are outliers in the data. Additionally, these predictions can help internally justify an overpayment against market value.

**Data & Contents:**

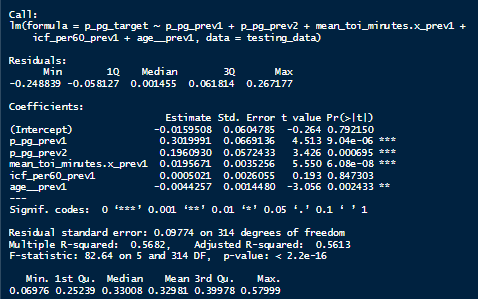
* Data from HockeyReference.com queried via HockeyR package
* Filtered to forwards only, 5v5 stats, > 10 GP in each season, and < 0.80 5v5 points per game (outlier remove). Players must have two seasons of data to be included in model.

**Model & Model Eval:**

* Stepwise regression to model the first season into the future.
* Training: 17-18, 19-20 to 22-23 season



* Test: 18-19 season



* Assumptions made on model features when predicting 2+ seasons into the future. Leveraged research on the age curve and point production to create age buckets of ‘26u’, ‘27-29’, ‘30+’ and rank each player into a quartile based on their point per game production from the most recent season and used the average change from each model feature in each quartile to create the updated model feature. See code below: A computer screen shot of a code

  Description automatically generated
* **Evaluations:**

A graph with black dots

Description automatically generated

A graph of a normal q-q plot

Description automatically generated

A graph of a number of dots

Description automatically generated

A graph with blue lines

Description automatically generated

**Outcome:**

* Although the r2 scores weren’t spectacular and only account for ~50% of the variability the models did generalize well from train to test. And as evidenced by the residuals of each season the model does a pretty good job predicting both +1 season ahead using strictly regression techniques while +2 seasons and beyond predict well using both regression techniques and assumptions. Also as we hoped the model does a better job predicting middle of the lineup player than others.

A graph of a graph showing the number of projections

Description automatically generated with medium confidence

A graph with numbers and a black background

Description automatically generated

A screen shot of a graph

Description automatically generated

A graph of a number of points

Description automatically generated with medium confidence

A graph of different colored squares

Description automatically generated

A graph of a graph

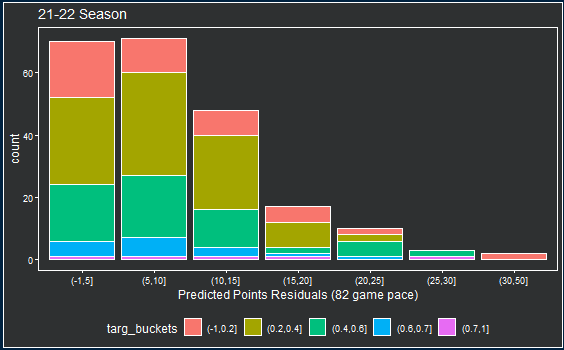
Description automatically generated with medium confidence

A screen shot of a graph

Description automatically generated

A graph with numbers and a black background

Description automatically generated



**Next Steps:**

With access to better data I would like to include both team level and career to date features into the model to see if it would help it perform more effectively. As I progress in my modelling capabilities, possibly testing out this data using a quadratic model, considering the Age Curve in professional sports, or a Lasso Regression to draw out important features rather than just using a Stepwise technique.