General Player Analysis

So far, we have analyzed the data for 50 individual players, but we would like to be able to draw broader conclusions about the overall effect the first free throw has on the probability of making the second free throw generally for all players.

To do this, we will define a logistic regression model with mixed effects.

This will allow us to model the probability of a player making the second free throw, given the result of the first free throw, while accounting for the baseline free throw probabilities of individual players.

We will then be able to look at the p-value for the shot1 coefficient to see if knowing the result of free throw 1 improves our ability to predict the probability of making free throw 2. If the p-value is statistically significant (< 0.05), it would indicate that the result of the first free throw (whether made or missed) has a significant effect on the likelihood of making the second free throw, beyond individual player baseline differences.

[9]:
```python
import pandas as pd
from pymer4.models import Lmer

#Create a DataFrame by reading CSV
df = pd.read_csv(r'free_throw_pairs.csv')

df.head(5)
```

[9]:

	player	shot1	shot2
0	Andrew Bynum	1	1
1	Andrew Bynum	1	0
2	Amare Stoudemire	1	1
3	Leandro Barbosa	0	1
4	Lamar Odom	1	1

[ ]:
```python
#Define the model that contains the fixed effect, shot1, and the random effect,
# the unique baseline free throw ability of each player
model = Lmer('shot2 ~ shot1 + (1 | player)', data=df, family='binomial')

# Fit the model
model.fit()

# View model summary
print(model.summary())
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: shot2~shot1+(1|player)

Family: binomial         Inference: parametric

Number of observations: 263300   Groups: {'player': 1091.0}
```

```
Log-likelihood: -134317.664     AIC: 268641.329

Random effects:

              Name    Var    Std
player  (Intercept)  0.271  0.521

No random effect correlations specified

Fixed effects:

Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: shot2~shot1+(1|player)

Family: binomial        Inference: parametric

Number of observations: 263300   Groups: {'player': 1091.0}

Log-likelihood: -134317.664     AIC: 268641.329

Random effects:

              Name    Var    Std
player  (Intercept)  0.271  0.521

No random effect correlations specified

Fixed effects:

             Estimate  2.5_ci  97.5_ci     SE     OR  OR_2.5_ci  OR_97.5_ci  \
(Intercept)      1.09   1.050    1.130  0.020  2.974      2.857       3.096
shot1            0.14   0.119    0.161  0.011  1.151      1.127       1.175

              Prob  Prob_2.5_ci  Prob_97.5_ci  Z-stat  P-val  Sig
(Intercept)  0.748        0.741         0.756  53.314    0.0  ***
shot1        0.535        0.530         0.540  13.093    0.0  ***
```

Looking at our model summary, we see that shot1 is highly statistically significant in the model.

Let's see how it effects our estimated free throw 2 probabilities to see if the difference is practically significant.

[ ]:
```python
def free_throw_compare(player_list):
    print("Estimated second free throw probabilities:\n")
    for player in player_list:
        missed_first = pd.DataFrame({
            'shot1': [0],
```

```python
            'player': [player]})
        made_first = pd.DataFrame({
            'shot1': [1],
            'player': [player]})
        prob_miss = model.predict(missed_first)[0]
        prob_make = model.predict(made_first)[0]
        difference = prob_make - prob_miss
        print(f"{player} missed first free throw: {prob_miss:.4f}")
        print(f"{player} made first free throw: {prob_make:.4f}")
        print(f"Difference: {difference:.4f}\n")

player_list = ["Ben Wallace", "Shaquille O'Neal",
               "LeBron James", "Stephen Curry"]

free_throw_compare(player_list)
```

```
Estimated second free throw probabilities:

Ben Wallace missed first free throw: 0.4507
Ben Wallace made first free throw: 0.4857
Difference: 0.0350

Shaquille O'Neal missed first free throw: 0.5590
Shaquille O'Neal made first free throw: 0.5933
Difference: 0.0343

LeBron James missed first free throw: 0.7611
LeBron James made first free throw: 0.7857
Difference: 0.0246

Stephen Curry missed first free throw: 0.9041
Stephen Curry made first free throw: 0.9156
Difference: 0.0115
```

From our model predictions, we can see that making the first free throw increases the probability of making the second free throw by roughly 1.2-3.5% depending on the player's average free throw percentage.

We can conclude that the result of the second free throw is not independent of the result of the first. However, while the statistical significance of the effect is clear, whether or not a difference of 1.2-3.5% is practically significant is less clear.