1 Executive Summary

Shaquille O'Neal (refereed to as Shaq), has long been criticized for his poor free throw shooting. During the 2000 NBA playoffs, commentators remarked that his free-throw shooting varied dramatically from game to game. However, upon looking at the data there is no strong evidence that the variability in success probability is not due to chance.

2 Introduction

We want to find out if the commentators claim is valid: that Shaq's success probability for free-throws varies from game to game. Or, contrary to the commentators claim that Shaq's success probability for free-throws is constant for all 23 games. This condenses down to testing if the success probability (making a free throw) is the same for the 23 games (i.e. constant.)

In this analysis we will be using the model $\pi_i = \alpha$ where π_i is a probability of success for game i. To analyze this model, we will look at the following hypotheses:

$H_0 : \pi_i = \alpha \ \forall i = 1, \cdots, 23$ (constant probability model hold)
$H_1 : \pi_i \neq \alpha$ for at least one $i = 1, \cdots, 23$ (constant probability model does not hold)

Data used in the analysis is taken from the NBA website (www.nba.com). The data is from 23 games in the 2000 NBA playoffs. The raw data gives the number of free-throws made (success) and the number of attempts for each of the 23 games.

3 Results

3.1 Constant Probability Model

This model assumes that each game has the same free-throw conversion rate, i.e. : $\pi_i = \alpha$.

3.1.1 Estimate α

We run a regression of the free throws (binary data, 0 or 1) versus a constant to get an estimation for α (the intercept).

Table 1: Constant Probability Model - α

	Statistic	Values
1	$\hat{\alpha}$	0.456
2	Standard Error	0.029
3	Confidence Interval	(0.400,0.513)

Shaq's estimated probability of making a free throw is 0.456 [95% CI (0.400, 0.513)] for each of the 23 games.

3.1.2 Global Lack of Fit - Chi-Squared Approximation

One way to check the global goodness of fit of a model is with chi-squared approximation. However, prior to using the chi-squared approximation, we have to check if the assumption is met, i.e. we have to check if $n^* \gg \dim(H_0 \cup H_1)$. In this case, $n^* = 296$, which is the total number of attempts. Also, $\dim(H_0 \cup H_1) = 23$, the number of games or observations. $296 \gg 23$ so we can proceed with the chi-squared approximation.

Pearson's X-squared test statistic is 35.511, which corresponds to a p-value of 0.034 given 22 degrees of freedom. Note that degrees fo freedom is $\dim(H_0 \cup H_1) - \dim(H_0) = 23 - 1 - 22$ degrees of freedom. Given that the p-value is small, we reject the null hypothesis that the model of constant probability across the 23 games hold. In other words, we have evidence that there is a lack of fit.

3.1.3 Local Lack of Fit - Standardized Residuals

Another way to look at goodness of fit, is to check standardized residuals for local lack of fit. Generally, if a standardized residual is larger than 2 in absolute value, that is considered an indication of local lack of fit.

We have one standardized residual that is larger than 2, which demonstrates a local lack of fit for the 14th Game (residual=3.327). Besides the residual for the 14th game, there are no other outliers. The significant lack of fit for the 14th game is due to having such a different proportion than the other games. In the 14th game Shaq shot 9/9 (100%), all of the other proportions are between 0.167 and 0.800.

3.1.4 Game 14 - Outlier causing lack of fit?

If a chi-squared approximation is used to analysis goodness of fit on the data *without Game 14*, the result is a $X^2 = 24.618$ on 21 degrees of freedom with a p-value of 0.264. We would fail to reject the null hypothesis that the model of constant probability holds for the 23 games. In addition, there are no standardized residuals that have an absolute value greater than 1.693, suggest there is no local lack of fit.

3.2 Constant Probability Model with Over-dispersion

If our previous assumption of independent trials (one free throw is independent of another) is incorrect, then our previous goodness of fit tests and our estimates of α and its standard error have questionable validity since those values were computed under the assumption that there was independence between trials.

Is it feasible that the trials (i.e. free throws) are not independent? It is possible that a player within a game will be doing well, having successful free throws, and will build off that momentum and confidence to continue playing well. In sports terms, this means getting a "hot hand" and means a players is able to have successful shots multiple times in a row. If this situation happens, there is positive within-game correlation. When there is a positive within-game correlation, this will lead to over-dispersion to what is expected in the binomial model that assumes independence.

In our example, look at Game 14 where Shaq shot 9/9 on the free throw line. His "hot hand" in this game leads to a more success with likely dependent trials than expected under our constant probability model.

The constant probability model rejection (using the independence assumption) has two possible reasons: (1) the constant Probability assumption is not true or (2) the constant probability assumption could be true, but the interdependent assumption is not.

We will look at the second reason, keeping in mind Game 14 when Shaq went 9/9, much higher than any other game. We expect over-dispersion if we assume our "hot hand" theory is true which leads to positive within-game correlation.

We need to assume the model includes an unknown dispersion parameter, this will update our results.

3.2.1 Estimate Dispersion and Scale

For a binomial model, we estimate the dispersion $\hat{\phi} = X^2/df$ where X^2 is the Pearson X-squared statistic and df is the degrees of freedom used in the chi-squared approximation. We have $\hat{\phi} = {}^{35.511}/_{22} = 1.614$.

So our estimate for dispersion, $\hat{\phi}$, is 1.614 and our estimate for the scale, $\sqrt{\hat{\phi}}$, is 1.270. This suggest that there is over-dispersion.

3.2.2 Estimate α, assuming $\phi \neq 1$

We are able to use the dispersion and scale to adjust the approximate standard error for α and thus the confidence interval. The constant probability model, assuming $\phi \neq 1$, approximate standard error is related to the constant probability model, assuming $\phi = 1$, approximate standard error in the flowing manner:

$$\text{s.e.}\,(\alpha)_{\phi\neq 1} = \sqrt{\hat{\phi}} \cdot \text{s.e.}\,(\alpha)_{\phi=1}$$

Table 2: Constant Probability Model assuming $\phi \neq 1$ - α

	Statistic	Values
1	$\hat{\alpha}$	0.456
2	Standard Error of α	0.037
3	Confidence Interval for α	(0.384,0.528)

From the results above, given the over-dispersion, the standard error increases and the confidence interval widens.

4 Conclusion

The constant probability model, that chance of success in free throws is the same for all 23 games, reasonably hold. With the exception of the large outlier from the 14th game, the observed proportions are in line with those predicted using the constant probability model. This suggests that Shaq's shooting percentage variability can be explained by randomness. There is no strong case that the success probability varies from game to game. There is some over-dispersion, likely due to within-game correlation among the trials. Taking into account over-dispersion, Shaq's estimated probability of making a free throw is 0.456 [95% CI (0.384, 0.528)] for each of the 23 games.