Assume that pairs of free throws are independent of each other and that our probabilities of success within each group (made/missed first shot) remain constant for all trials.

Let the random variable Y_1 represent the number of successful 2nd free throw shots in a sample of n_1 free throw pairs, where the 1st free throw shot was unsuccessful, drawn from a large population.

Based on our assumptions, Y_1 can be closely modelled by a binomial distribution with the following probability function:

$$P(Y_1 = y_1; \theta_1) = \binom{n_1}{y_1} \theta_1^{y_1} (1 - \theta_1)^{n_1 - y_1}$$
$$\text{for } y_1 = 0, 1, ..., n_1 \text{ and } 0 \leq \theta_1 \leq 1$$

where θ_1 represents the probability of making the 2nd free throw shot after missing the 1st free throw shot.

Suppose we observed y_1 successful 2nd free throws after missing the 1st free throw from a sample of size n_1. The likelihood function for θ_1 based on this data is

$$\begin{aligned}
L(\theta_1) &= P(y_1 \text{ 2nd free throws were made; } \theta_1) \\
&= \binom{n_1}{y_1} \theta_1^{y_1} (1 - \theta_1)^{n_1 - y_1} \\
&\propto \theta_1^{y_1} (1 - \theta_1)^{n_1 - y_1} \\
&\propto \ln[\theta_1^{y_1} (1 - \theta_1)^{n_1 - y_1}] \\
&= \ln(\theta_1^{y_1}) + ln[(1 - \theta_1)^{n_1 - y_1}] \\
&= y_1 \ln(\theta_1) + (n_1 - y_1) \ln(1 - \theta_1)
\end{aligned}$$

$$\begin{aligned}\frac{d}{d\theta_1}(y_1 \ln(\theta_1) + (n_1 - y_1)\ln(1-\theta_1)) &= y_1 \frac{d}{d\theta_1}(\ln(\theta_1)) + (n_1 - y_1)\frac{d}{d\theta_1}(\ln(1-\theta_1)) \\ &= \frac{y_1}{\theta_1} - \frac{n_1 - y_1}{1-\theta_1} \\ &= \frac{y_1 - n_1\theta_1}{\theta_1(1-\theta_1)}\end{aligned}$$

We maximize our likelihood function by setting the derivative, $\frac{y_1 - n_1\theta_1}{\theta_1(1-\theta_1)}$, equal to zero and solving for θ_1:

$$0 = \frac{y_1 - n_1\theta_1}{\theta_1(1-\theta_1)} \implies \frac{y_1}{\theta_1(1-\theta_1)} = \frac{n_1\theta_1}{\theta_1(1-\theta_1)} \implies y_1 = n_1\theta_1 \implies \theta_1 = \frac{y_1}{n_1}$$

Therefore, the maximum likelihood estimate for θ_1 is $\hat{\theta}_1 = \frac{y_1}{n_1}$.

As an example, our observed data for LeBron James is $y_1 = 730$ and $n_1 = 951$, giving us $\hat{\theta}_1 = \frac{730}{951} \approx 0.7676$.

Defining Y_2 as the number of successfull 2nd free throw shots in a sample of n_2 free throw pairs where the 1st free throw shot was successful, and applying the same process, we get $\hat{\theta}_2 = \frac{1871}{2389} \approx 0.7832$, where θ_2 is defined as the probability of making the 2nd free throw shot after making the 1st free throw shot.

We would like to see if the difference between θ_1 and θ_2 is statistically significant at the 0.05 level.

Our null hypothesis is $H_O : \theta_1 = \theta_2$, and the alternative hypothesis is $H_A : \theta_1 \neq \theta_2$.

From the National Institute of Standards and Technology (www.itl.nist.gov/div898/handbook/prc/section3/prc33.htm), we have that

$$Z = \frac{\tilde{\theta}_1 - \tilde{\theta}_2}{\sqrt{\tilde{\theta}(1-\tilde{\theta})(\frac{1}{n_1} + \frac{1}{n_2})}} \sim G(0,1), \text{ where } \tilde{\theta} = \frac{n_1\tilde{\theta}_1 + n_2\tilde{\theta}_2}{n_1 + n_2}$$

.

Our test statistic is $D = |Z|$, and our observed value of D is $d \approx 0.9777$.

Our approximate p-value based on the Gaussian approximation is

$$
\begin{aligned}
\text{p-value} &= P(D \geq d; H_O) \\
&= P\left(\frac{|\tilde{\theta}_1 - \tilde{\theta}_2|}{\sqrt{\tilde{\theta}(1-\tilde{\theta})(\frac{1}{n_1} + \frac{1}{n_2})}} \geq 0.9777\right) \\
&\approx P(|Z| \geq 0.9777) \\
&= 2[1 - P(Z \leq 0.9777)] \\
&\approx 2(1 - 0.8359) \\
&= 0.3282
\end{aligned}
$$

Since our p-value is greater than 0.05, we conclude that there is insufficient evidence to reject the null hypothesis for LeBron James.

We can repeat this process for each player with a large enough sample size using python.