Lebron James free throws with Confidence Intervals

Patrick Kelly

July 11, 2016

setwd("~/Desktop/PK_Portfolio/Khan_Academy/Probability")

Khan Academy Video - English

Ten free throws in a row "LeBron James"

If LeBron James has a lifetime successful free throw percentage of 75%, what is the probability that he will make 10 in a row, assuming that each attempt is an indepedent event?

Theoretical P(10 in a row) = (75%)^10 = (75/100)^10 = 0.75^10 = 0.056 = 5.6 %

Use the binomial theorem for the same result

```
round(dbinom(10, size=10, prob=0.75),3)
```

```
## [1] 0.056
```

Simulate shooting 10 free throws 1000 times

Then repeat 500 times

```
n <- 1000
P_ten_500 <- NULL
for(i in 1:500){
trial <- NULL
for (j in 1:n) {
       trial = rbind(trial,t(as.matrix(rbinom(10,1,0.75))))
    }
trial <- data.frame(trial)
Successes <- apply(trial,1,sum)
Successes <- data.frame(Successes)
Ten_out_of_ten <- ifelse(Successes$Successes == 10, 1,0)
P_Ten_out_of_ten <- sum(Ten_out_of_ten)
options(digits=4)
P_Ten_out_of_ten <- P_Ten_out_of_ten / n
P_Ten_out_of_ten
P_ten_500 <- rbind(P_ten_500,P_Ten_out_of_ten)
}
```

```r
row.names(P_ten_500) <- NULL
P_ten_500 <- data.frame(P_ten_500)
```

```r
sim_Mean <- mean(P_ten_500$P_ten_500)
sim_Mean
```

```
## [1] 0.05581
```

```r
sim_sd <- sd(P_ten_500$P_ten_500)
sim_sd
```

```
## [1] 0.007236
```

```r
sim_se <- sim_sd / sqrt(500)
sim_se
```

```
## [1] 0.0003236
```

```r
qnorm(0.975)
```

```
## [1] 1.96
```

```r
margin_of_error_95 <- qnorm(0.975) * sim_se
margin_of_error_95
```

```
## [1] 0.0006343
```

```r
Lower_limit = sim_Mean - margin_of_error_95
Lower_limit
```

```
## [1] 0.05517
```

```r
Upper_limit = sim_Mean + margin_of_error_95
Upper_limit
```

```
## [1] 0.05644
```

```r
CI <- t.test(P_ten_500$P_ten_500, conf.level = 0.95)
CI[[4]][1] # Lower Limit
```

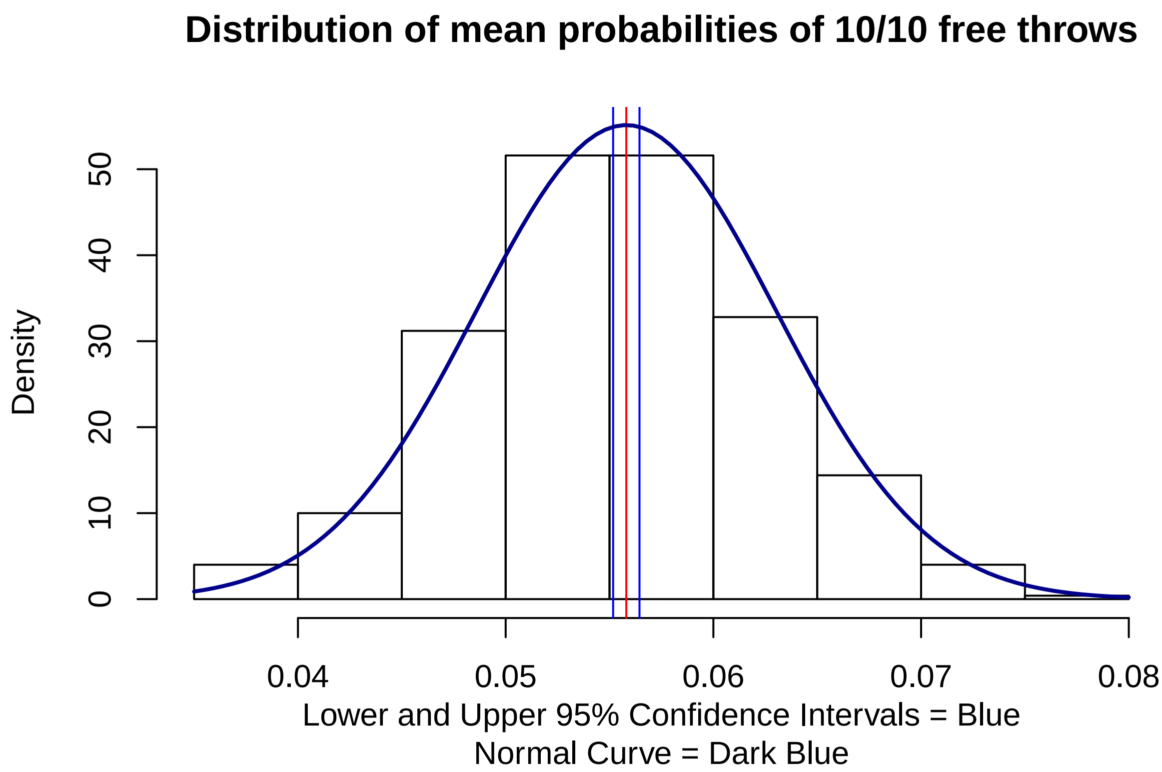
```
## [1] 0.05517
```

```r
CI[[4]][2] # Upper Limit
```

```
## [1] 0.05644
```

Histogram plot of the experimental mean probabilities

```r
hist(P_ten_500$P_ten_500, freq = FALSE,xlab = "Lower and Upper 95% Confidence Intervals = Blue\nNormal C
    main = "Distribution of mean probabilities of 10/10 free throws",
    ylim = c(0,55))
abline(v=sim_Mean, col="red")
abline(v=Lower_limit , col = "blue")
abline(v=Upper_limit, col = "blue")
curve(dnorm(x, mean=sim_Mean, sd=sim_sd),
          col="darkblue", lwd=2, add=TRUE, yaxt="n")
```

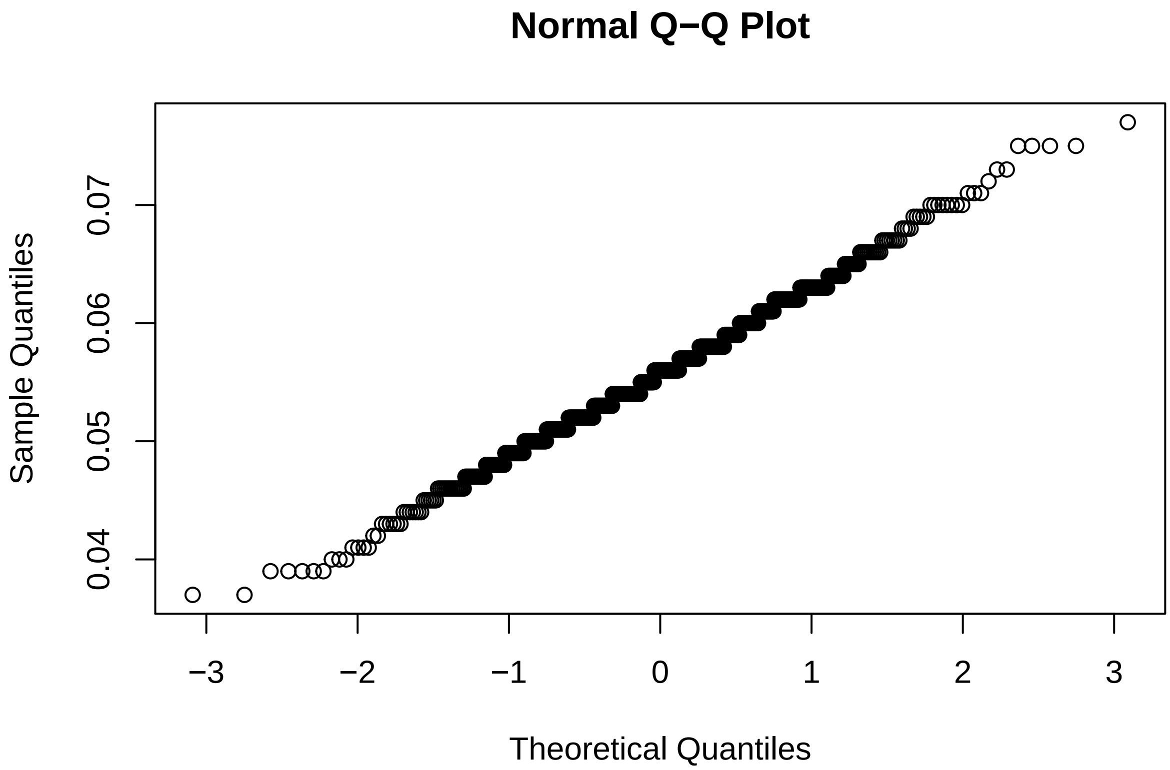


Tests for normality

```r
shapiro.test(P_ten_500$P_ten_500)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  P_ten_500$P_ten_500
## W = 1, p-value = 0.2
```

```r
## If P > 0.05 then the null hypothesis that the distribution is normal, is not rejected.
## Quantile-Quantile plot
qqnorm(P_ten_500$P_ten_500)
```



Important statistical concepts are used in this example.

1. The Law of Large Numbers: As the sample size increases the sample mean will be closer to the theoretical value. Large sample size = narrow confidence limits.

2. The Central Limit Theorem: The central limit theorem (CLT) states that the means of random samples drawn from any distribution with mean m and variance s2 will have an approximately normal distribution with a mean equal to m and a variance equal to s2 / n.

3. Standard Error of the Sample Mean = sqrt(s2 / n) = s / sqrt(n).

4. Margin of Error of the Sample Mean = Critical Value * Standard Error of the Mean. Critical values for the standard normal distribution of z scores are 1.64 for 90% confidence, 1.96 for 95% confidence and 2.54 for 99% confidence.

5. The Confidence Interval (2-tailed) about the Sample Mean = Sample Mean plus or minus the Margin of Error.

6. Interpretation of 95% confidence interval: Based on our sample data, we are 95% confident that the "theoretical" mean of 0.056 is between the Lower Limit and the Upper Limit. In other words, we expect that once in 20 times, the experimental mean may fall outside of these limits.