# Binomial Tests

*TLScofield*

*09/22/2015*

## Binomial Test

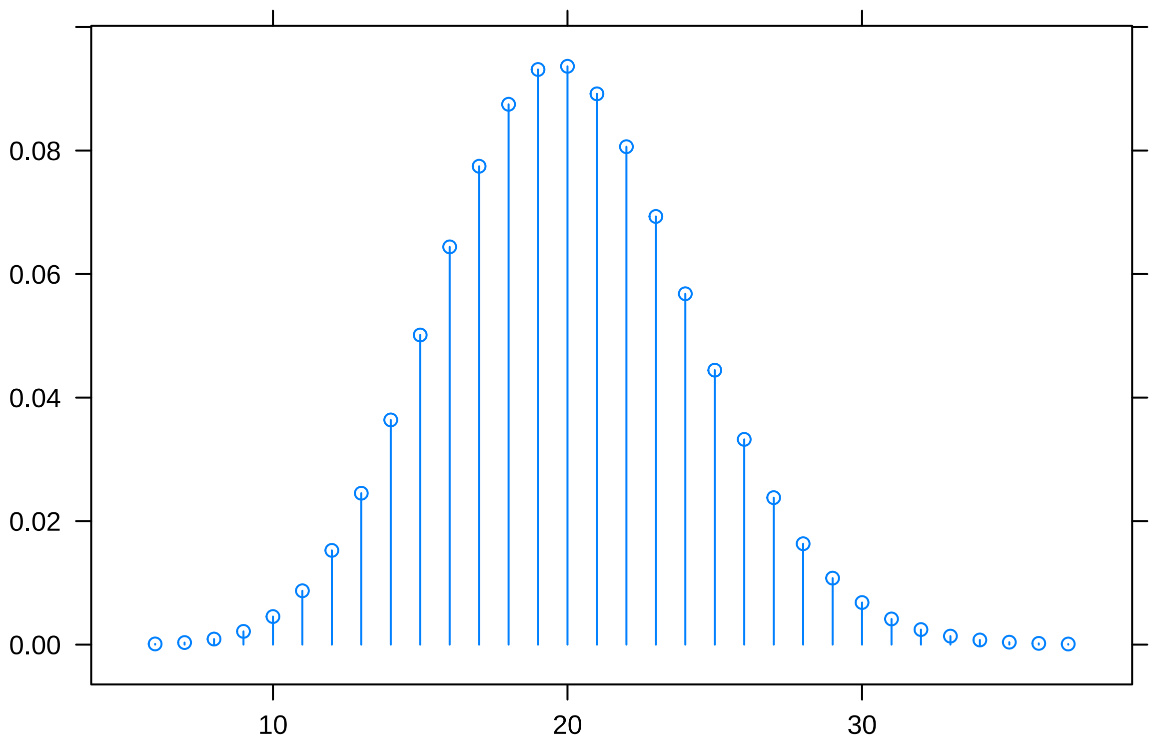
**Example**: Instances of sevens as digits from a random digit generator

Step 1: Formulate hypotheses. $H_0 : \pi = \frac{1}{10}$, $H_a : \pi \neq \frac{1}{10}$

Step 2: Obtain a test statistic. In this case, I indicated a sample of 200 random digits yielded 35 sevens.

Step 3: Calculate the $P$-value. The $P$-value is the conditional probability of obtaining a test statistic at least as extreme as ours given that the null hypothesis is true. It may be helpful to view the **null distribution** $X \sim \text{Binom}(200, 1/10)$, which is the pmf for $X$ under the null hypothesis.

```
plotDist("binom", params=c(200,1/10))
```




If you locate $X = 35$, you see that this value, along with ones more extreme such as $X = 36$, $X = 37$, etc., do not occur very often in the null distribution. We may evaluate the pmf to see just how often values in this range occur.

```
1 - pbinom(34,200,1/10)
```

```
## [1] 0.0007843467
```

This is not yet the $P$-value, as it gives the probability only of obtaining a result of 35 or *higher*. To get that we should add in the probabilities of occurrences in the opposite *tail*. One can often get quite close to this **two-tailed** $P$-value by doubling the area in the one tail (what we found already). The times this does not work out are generally those in which the null distribution is not symmetric, as is somewhat the case here.

As described in the text, we include additional probabilities precisely for those values of $X$ with probabilities (under the null hypothesis) not exceeding the probability that $X = 35$. Here is some code to do this.

```
nullPMFvals = dbinom(0:200, 200, 1/10)
```

The result of the calculation above is that the vector `nullPMFvals` contains the probilities of $X = 0, 1, 2, \ldots, 200$. We could compare each of these to the probability that $X = 35$:

```
referenceProb = dbinom(35, 200, 1/10)
head( nullPMFvals <= referenceProb )
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE
```

The actual comparison resulted in 201 boolean (TRUE or FALSE) values, but by using the `head()`, I suppressed all but the first 6 of them. We obtain a $P$-value by adding up all the numbers in `nullPMFvals` that were smaller than the reference probability with the command

```
sum( nullPMFvals[ nullPMFvals <= referenceProb ] )
```

```
## [1] 0.001269343
```

There ought to be an easier way, you say. Perhaps, by reading the text, you have discovered one already, the `binom.test()` command. It provides more than we ask or understand at the moment, but it will give us the $P$-value, among other things.

```
binom.test(35, 200, 1/10)
```

```
##
##
##
## data:  35 out of 200
## number of successes = 35, number of trials = 200, p-value =
## 0.001269
## alternative hypothesis: true probability of success is not equal to 0.1
## 95 percent confidence interval:
##  0.1250244 0.2348840
## sample estimates:
## probability of success
##                  0.175
```

Step 4: Draw a conclusion. At this point we may want to make a call, such as a jury might. Is there reasonable doubt here to reject the null hypothesis in favor of the alternative? The null hypothesis led to the null distribution $X \sim \text{Binom}(200, 1/10)$ which, if it reflects the reality of the situation, would mean we were present for an event that occurs only rarely, just over once in 1000 attempts. One might hold a fairly high threshold for *reasonable doubt*, and fail to reject the null hypothesis despite the rarity of the result. (That is the language we use. Avoid saying, "I accept the null hypothesis as true," for the data never proves this; rather, data often comes in which is consistent with a true null hypothesis.) On the other hand, you may feel justified in dismissing (or "rejecting") the null hypothesis in favor of the alternative, believing that $\pi$ is, in fact, different from 1/10.

**Example**: Field goal kicker

Suppose your professional football team has dismissed its kicker on the grounds that empirical data suggests he only makes kicks at a distance 40-45 yards 30% of the time. You are "interviewing" a kicker as a possible replacement. The interview consists of having the interviewee attempt 10 kicks in the range 40-45 yards. How well would this kicker need to do to convince you, say, with a $P$-value not exceeding 5%, that he is better than a 30% kicker at this distance?

Step 1: $H_0 \colon \pi = 0.3$, $H_a \colon \pi > 0.3$. It makes sense to have the alternative be 1-sided, given that we only seek evidence of this kicker being better than the one we let go.

Step 2: We will forego gathering data, because of the nature of the question being asked. We are charged with the task of forming, in advance, a **decision rule** for how good "good enough" should be.

Step 3: In the absence of a test statistic, one cannot compute a $P$-value, per se. Still, we can use the null distribution $X \sim \text{Binom}(10, .3)$ to determine a threshold value for $X$ so that instances in which a kicker who truly is only able to make 30% of tries would rarely (not more often than 5% of the time) surpass this threshold. To get this threshold, we use the `qbinom()` command.

```
qbinom(.95, 10, .3)
```

```
## [1] 5
```

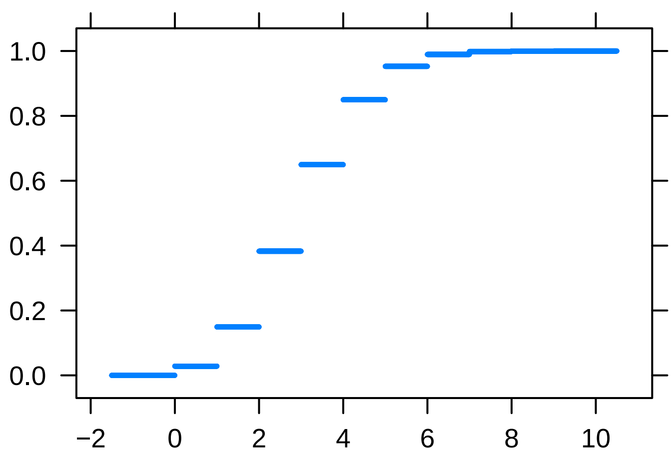
This command asks where to place a value for $X$ in the null distribution so that 95% of values lie to its left. A closer look at the pmf for the null distribution reveals that the probability of making 5 of 10 for a 30% kicker is rather high, about 0.1:

```
dbinom(4:6,10,.3)
```

```
## [1] 0.20012095 0.10291935 0.03675691
```

Because of this, the CDF makes a significant jump at $X = 5$, from around 0.85 to around 0.95:

```
plotDist("binom", params=list(10, .3), kind="cdf", pch=19, cex=.2)
```



```r
pbinom(4:6,10,.3)
```

```
## [1] 0.8497317 0.9526510 0.9894079
```

It is really the results of $X = 6$ or more successful attempts that are among the rarest (cumulatively less than 5%) favorable showings for a kicker who is truly a 30% kicker.

We can use several different commands to calculate just how rare an event of $X \geq 6$ is from a 30% field goal kicker. Here are three possibilities. Note the use of the `alternative` switch for `binom.test()`, explained in the text.

```r
sum(dbinom(6:10, 10, .3))
```

```
## [1] 0.04734899
```

```r
1 - pbinom(5, 10, .3)
```

```
## [1] 0.04734899
```

```r
binom.test(6, 10, .3, alternative="greater")
```

```
##
##
##
## data:  6 out of 10
## number of successes = 6, number of trials = 10, p-value = 0.04735
## alternative hypothesis: true probability of success is greater than 0.3
## 95 percent confidence interval:
##  0.3035372 1.0000000
## sample estimates:
## probability of success
##                    0.6
```