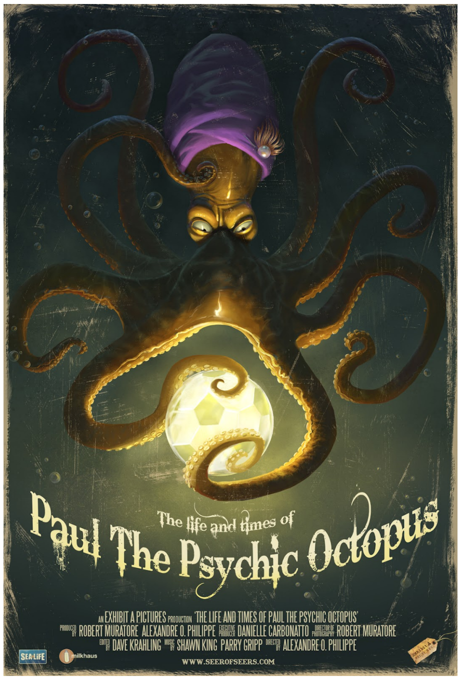
Hypothesis Tests about Proportions

1. Identify the
   - **population parameter**: a number summarizing the population.
     - What proportion of the "population" are in a certain category?
     - Denote by p (for proportion).
   - **sample statistic**: a number summarizing the sample.
     - In our sample, how many observational units are in that category?
     - Denote by X for a random variable, x for the observed value
2. State our hypotheses.
   - **Null hypothesis** (H_0): Nothing interesting is going on.
     - State in words what this means for the example we're looking at.
     - State as an equality involving the population parameter: $p =$ ____
   - **Alternative hypothesis** (H_A): Something interesting is going on.
     - State in words what this means for the example we're looking at.
     - State as an inequality involving the population parameter: $p <$ ____, or $p >$ ____, or $p \neq$ ____.
3. Find the sampling distribution for the sample statistic, assuming H_0 is true.
   - For now, our only choice is $\mathbf{X} \sim$ **Binomial**(____, ____) (fill in sample size and the value of p from the null hypothesis)
   - We need to **check conditions** before proceeding:
     - Were there any forms of bias in our sampling procedure?
     - Was a categorical variable recorded for each observational unit?
     - Is our sample statistic a count of how many observational units in our sample were in one of the categories?
     - Are different observational units in our sample independent?
4. Calculate a p-value: the probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true.
5. Draw a conclusion. Two options:
   - Make a statement about the strength of evidence against the null hypothesis provided by the sample data.
     - A p-value of .10 or less: some evidence against H_0
     - A p-value of .05 or less: fairly strong evidence against H_0
     - A p-value of .01 or less: very strong evidence against H_0
     - A p-value of .001 or less: extremely strong evidence against H_0
   - Make a decision by comparing the p-value to a specified significance level α.
     - Most commonly used values are $\alpha = 0.05$ and $\alpha = 0.01$, but there is no specific justification for these values.
     - "Reject" H_0 if the p-value is less than α
     - "Fail to Reject" H_0 if the p-value is greater than or equal to α

Example 1: Is Paul the Octopus Psychic?



(image credit: Paul J. Sullivan)

We observed 8 successful predictions out of 8 attempts.

Step 1: Identify population parameter and sample statistic.

- **Population parameter**:

The proportion of the time that Paul makes a prediction that he will be correct. (We have in mind here that the population is all predictions that Paul might ever make.) We will use the letter p to denote this proportion.

- **Sample statistic**:

The number of correct predictions in our sample. We observed a sample statistic of $x = 8$.

Step 2: State hypotheses

- **Null Hypothesis** (H_0):

Paul is not psychic, he is just guessing. $p = 0.5$

- **Alternative Hypothesis** (H_A):

Paul is psychic. $p > 0.5$

Step 3: Find sampling distribution of sample statistic, assuming H_0 is true

Check conditions for using a Binomial distribution:

- Were there any forms of bias in our sampling procedure?

It seems reasonable to think that these predictions would be representative of Paul's predictions more generally.

- Was a categorical variable recorded for each observational unit?

Yes: each prediction was either correct or incorrect (each observational unit is a prediction).

- Is our sample statistic a count of how many observational units in our sample were in one of the categories?

Yes: we counted how many predictions were correct.

- Are different observational units in our sample independent?

This seems reasonable. Knowing whether or not one of Paul's predictions was correct would not give me any information about whether or not another prediction was correct.

State the sampling distribution, assuming H_0 is true:

$X \sim \text{Binomial}(8, 0.5)$

Step 4: Calculate the p-value for the test

The p-value is the probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true.

In this example, that means $P(X \geq 8) = P(X > 7)$.

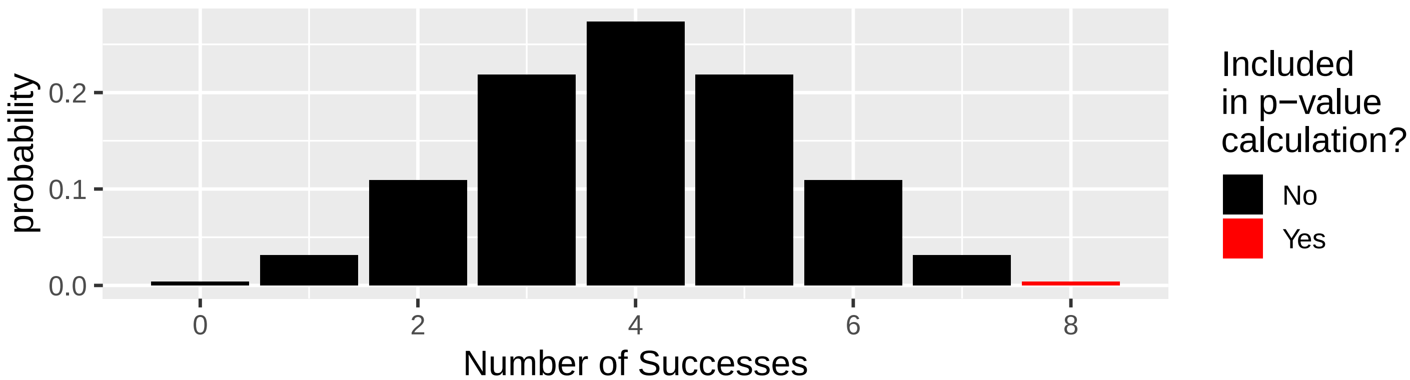
```
pbinom(q = 7, size = 8, prob = 0.5, lower.tail = FALSE)
```

```
## [1] 0.00390625
```

```
binom.test(x = 8,
  n = 8,
  p = 0.5,
  alternative = "greater")
```

```
##
##
##
## data:  8 out of 8
```

```
## number of successes = 8, number of trials = 8, p-value = 0.003906
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.687656 1.000000
## sample estimates:
## probability of success
##                      1
```



Step 5: Draw a conclusion

What is the strength of evidence against the null hypothesis provided by the data?

The p-value for this test was about 0.0039. This offers very strong evidence against the null hypothesis that Paul was just guessing.

Make a decision using a significance level of $\alpha = 0.01$.

Since the p-value is less than $\alpha = 0.01$, we reject the null hypothesis. The data provide statistically significant evidence that Paul is not just guessing.

Example 2: Blue M&M's

My favorite color of M&M's is blue. Mars Company (the makers of M&M's) last published the distribution of colors for M&M's in 2008, and at that time the proportion of M&M's that were blue was 0.16. However, I'm concerned that they may be putting fewer blue M&M's in now. I take a sample of 100 M&M's and count 12 blue M&M's in my sample. Does this provide evidence that the proportion of M&M's that are blue is lower now than it was in 2008?

Step 1: Identify population parameter and sample statistic.

- **Population parameter**:

$p =$ the proportion of all M&M's in current production that are blue.

- **Sample statistic**:

The number of M&M's in our sample that are blue. We observed $x = 12$.

Step 2: State hypotheses

- **Null Hypothesis** (H_0):

The proportion of M&M's that are blue is the same now as it was in 2008. $p = 0.16$

- **Alternative Hypothesis** (H_A):

The proportion of M&M's that are blue is lower now than it was in 2008. $p < 0.16$

Step 3: Find sampling distribution of sample statistic, assuming H_0 is true

Check conditions for using a Binomial distribution:

- Were there any forms of bias in our sampling procedure?

We might be concerned about taking a convenience sample, if we got all of our M&M's from a single bag. On the other hand, if we think that the contents of one bag are likely to be representative of the population of all M&M's more generally, this could be ok.

- Was a categorical variable recorded for each observational unit?

Yes: for each M&M, we recorded whether or not that M&M was blue.

- Is our sample statistic a count of how many observational units in our sample were in one of the categories?

Yes: we counted how many M&M's in our sample were blue.

- Are different observational units in our sample independent?

Yes: knowing that one M&M in my sample was blue doesn't change the probability that another M&M in my sample is blue.

State the sampling distribution, assuming H_0 is true:

$X \sim \text{Binomial}(100, 0.16)$

Step 4: Calculate the p-value for the test

The p-value is the probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true.

In this example, that means $P(X \leq 12)$

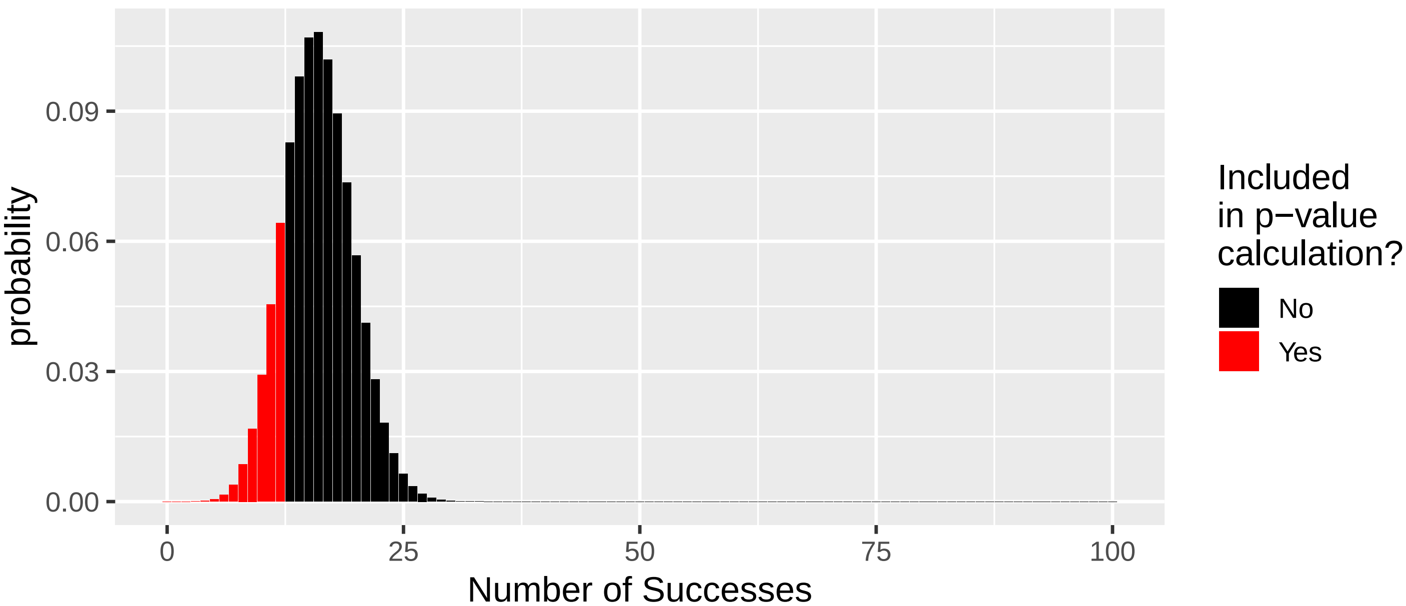
```r
pbinom(q = 12, size = 100, prob = 0.16)
```

```
## [1] 0.1703155
```

```r
binom.test(x = 12,
  n = 100,
  p = 0.16,
  alternative = "less")
```

```
##
##
##
## data:  12 out of 100
```

```
## number of successes = 12, number of trials = 100, p-value = 0.1703
## alternative hypothesis: true probability of success is less than 0.16
## 95 percent confidence interval:
##  0.0000000 0.1871661
## sample estimates:
## probability of success
##                   0.12
```



Step 5: Draw a conclusion

What is the strength of evidence against the null hypothesis provided by the data?

The p-value for the test is 0.1703. This is relatively large, so the data do not provide much evidence against the null hypothesis that the proportion of M&M's that are blue is the same as it was in 2008.

Make a decision using a significance level of $\alpha = 0.05$.

Since the p-value is greater than $\alpha = 0.05$, we fail to reject the null hypothesis. The sample data do not provide statistically significant evidence that the proportion of M&M's that are blue has changed since 2008.

Example 3: Absenteeism

According to the National Center for Education Statistics, in 1996, 66% of grade-school students in the U.S. had missed at least one day of school in the past month. A more recent survey of 8302 randomly selected students found that 5562 of them had missed at least one day of school in the past month. Has the rate of absenteeism changed since 1996?

Step 1: Identify population parameter and sample statistic.

- **Population parameter**:

The population parameter is the proportion of all current-day grade school students in the U.S. who have missed at least one day of school in the last month. We will use the symbol p for this parameter.

- **Sample statistic**:

The sample statistic is the number of students in our sample who have missed at least one day of school in the last month. We observed $x = 5562$ in our sample.

Step 2: State hypotheses

- **Null Hypothesis** (H_0):

The proportion of grade school students in the U.S. who have missed at least one day of school in the last month has not changed since 1996. $p = 0.66$.

- **Alternative Hypothesis** (H_A):

The proportion of grade school students in the U.S. who have missed at least one day of school in the last month is different than it was 1996. $p \neq 0.66$.

Step 3: Find sampling distribution of sample statistic, assuming H_0 is true

Check conditions for using a Binomial distribution:

- Were there any forms of bias in our sampling procedure?

We don't really have enough information to assess this. We are not told how the sample was collected.

- Was a categorical variable recorded for each observational unit?

Yes: each student either missed at least one day or did not.

- Is our sample statistic a count of how many observational units in our sample were in one of the categories?

Yes: we counted how many students in our sample had missed at least one day of school in the last month.

- Are different observational units in our sample independent?

If the students in our sample were randomly selected from across the US, this would be reasonable to assume.

State the sampling distribution, assuming H_0 is true:

$X \sim \text{Binomial}(8302, 0.66)$

Step 4: Calculate the p-value for the test (we won't do this manually for 2-sided tests about a proportion)

The p-value is the probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true.

In this example, that means "at least as far from the expected number of absentees if H_0 was true".

Expected number of absentees if H_0 is true: $8302 * 0.66 \approx 5479$

Our observed value was 5562.

How far above the expected value is this? $5562 - 5479 = 83$

"As extreme" in the other direction would be… $5479 - 83 = 5396$

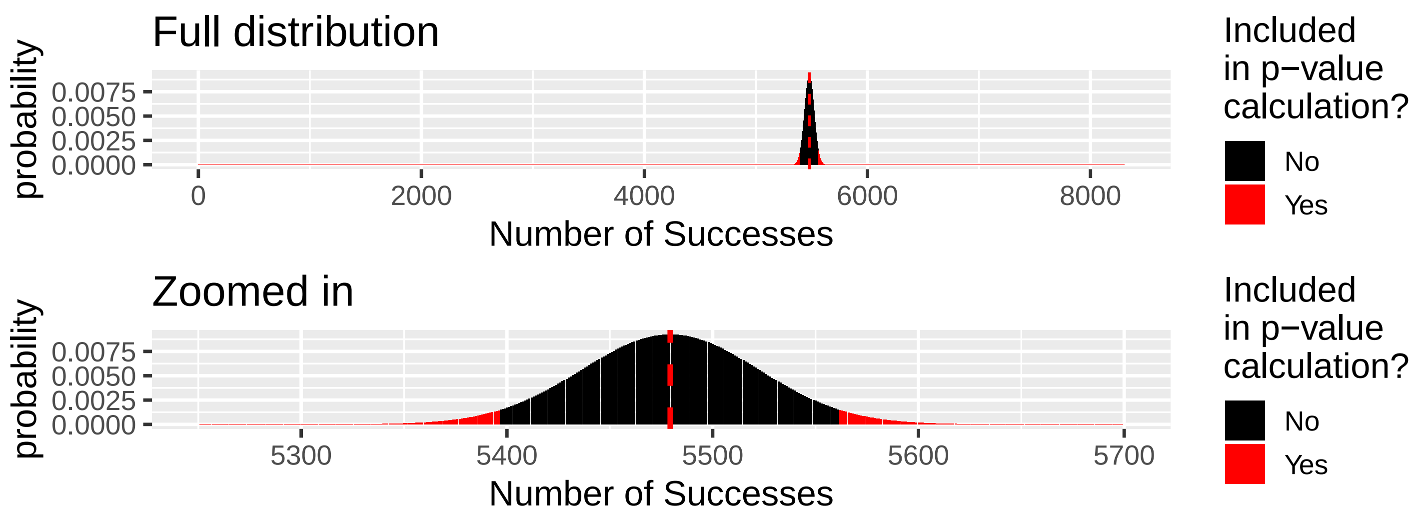
…so the p-value is: $P(X \leq 5396) + P(X \geq 5562)$

```
pbinom(5396, size = 8302, prob = 0.66) +
  pbinom(5562 - 1, size = 8302, prob = 0.66, lower.tail = FALSE)
```

```
## [1] 0.05595045
```

```r
binom.test(x = 5562,
  n = 8302,
  p = 0.66,
  alternative = "two.sided")
```

```
##
##
##
## data:  5562 out of 8302
## number of successes = 5562, number of trials = 8302, p-value =
## 0.05595
## alternative hypothesis: true probability of success is not equal to 0.66
## 95 percent confidence interval:
##  0.6597255 0.6800731
## sample estimates:
## probability of success
##               0.669959
```



Step 5: Draw a conclusion

What is the strength of evidence against the null hypothesis provided by the data?

The p-value was 0.05595. The data provide some evidence against the null hypothesis that the rate of absenteeism among grade school students is the same as it was in 1996.

Make a decision using a significance level of $\alpha = 0.05$.

The p-value is greater than $\alpha = 0.05$, so we fail to reject the null hypothesis. The data do not provide statistically significant evidence that the absenteeism rate among grade school students is different now than it was in 1996.