Hypothesis Tests for Population Proportions

Evan L. Ray

November 6, 2017

Is Paul the Octopus Psychic?

Recall our procedure for hypothesis testing:

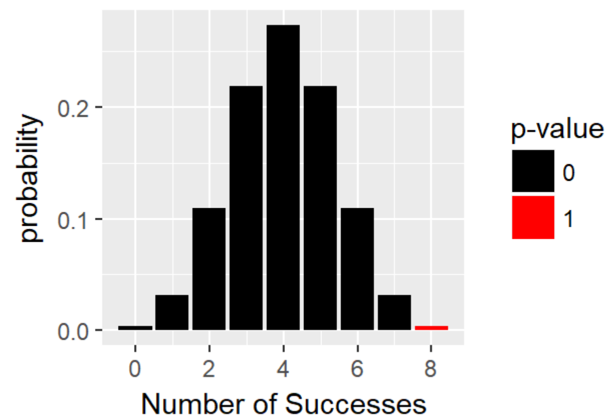
1. Collect **data:** for each of 8 trials, was the prediction correct?
2. Calculate a **sample statistic** (called the test statistic):
   - $x =$ total number correct (8 in our case)
3. Obtain the **sampling distribution** of the test statistic, assuming a **null hypothesis** of no effect (in this case, assuming Paul is just guessing)
4. Calculate the **p-value**: probability of getting a test statistic "at least as extreme" as what we observed in step 2
5. If the p-value is low, reject the null hypothesis and conclude that Paul is psychic!

More Carefully…

- **Test Statistic**: $x = 8$ (observed number correct)
- **Null Hypothesis**: Paul was just guessing: $p = 0.5$
- **Alternative Hypothesis**: Paul is psychic: $p > 0.5$
- **Sampling Distribution**, assuming null hypothesis is true:

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

(check assumptions!)

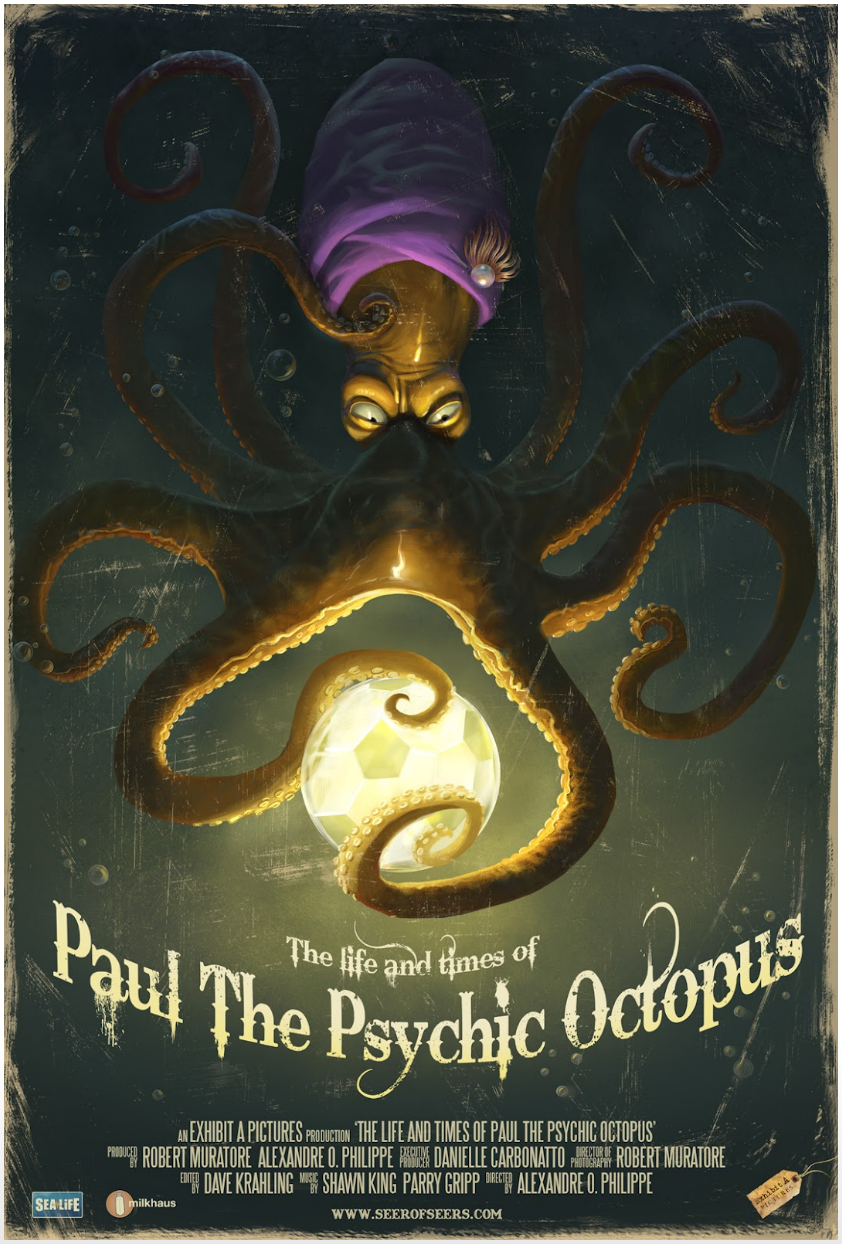


- **p-value**:

$P(X \geq 8) = 0.0039$

- **Conclusion**: It's unlikely that Paul would get 8/8 right if he was just guessing, so we reject the null hypothesis and conclude that he is psychic!

IMPORTANT FACT!!!



- Hypothesis tests are **guaranteed** tell you the wrong thing sometimes!!!!!!!!!!!!!!!!!
- This is similar to the fact that confidence intervals are **guaranteed** to miss the population parameter sometimes.
- More on this in a few days.

(image credit: Paul J. Sullivan)

More on Hypotheses

- **Null Hypothesis:** (Short Hand: H_0)
  - Nothing has changed since the past…
  - People are just guessing…
  - Nothing interesting is going on…
  - $p =$ (proportion from the past)/(chance of being right if just guessing)/(etc…)
- **Alternative Hypothesis:** (Short Hand: H_A)
  - Times have changed!
  - People know what they're doing!
  - The world is fascinating!
  - $p \neq$ (value from null hypothesis)
  - $p >$ (value from null hypothesis)
  - $p <$ (value from null hypothesis)

Examples of Hypotheses

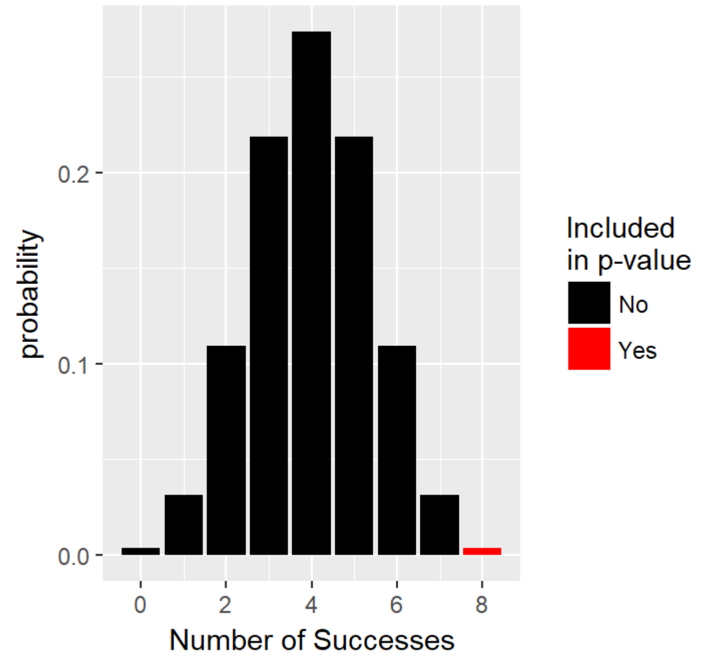
- Paul the Octopus, 8 right out of 8
  - Null Hypothesis (H_0): $p = 0.5$
  - Alternative Hypothesis (H_A): $p > 0.5$
- Proportion of M and M's that are blue (concerned it's lower now!); 12 blue out of 100
  - Null Hypothesis (H_0): $p = 0.16$
  - Alternative Hypothesis (H_A): $p < 0.16$
- The National Center for Education Statistics released a report in 1996 saying that 66% of students had missed at least one day of school in the past month. A more recent survey of 8302 students found that 5562 of them had missed at least one day of school. Has the rate of absenteeism changed?
  - Null Hypothesis (H_0): $p = 0.66$
  - Alternative Hypothesis (H_A): $p \neq 0.66$

More on P-Values

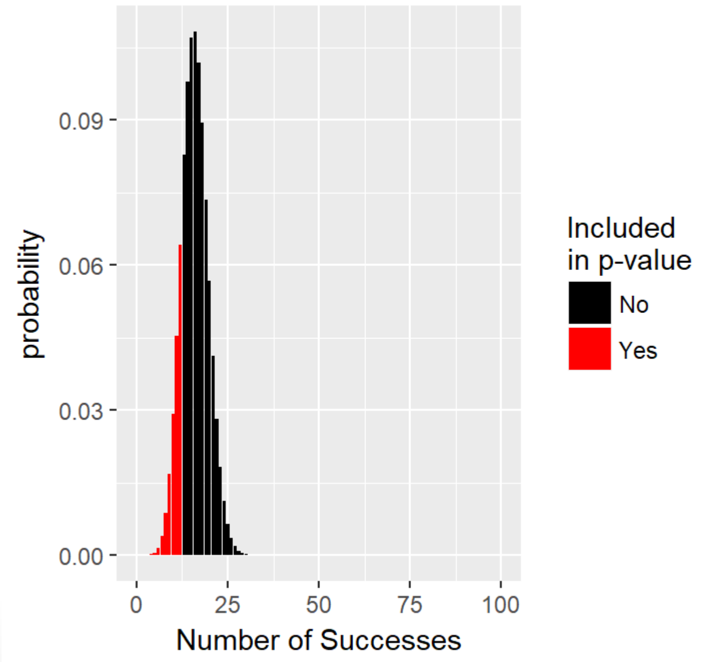
- **p-value**: probability of getting a test statistic "at least as extreme" as what we observed, assuming H_0 is true
- What counts as "at least as extreme" depends on the form of the alternative hypothesis

P-Values for One-Sided Tests

- Paul predicts 8 of 8 correctly
- $H_0: p = 0.5$
- $H_A: p > 0.5$
- **p-value:** $P(X \geq 8) = 0.0039$ if $X \sim \text{Binomial}(8, 0.5)$

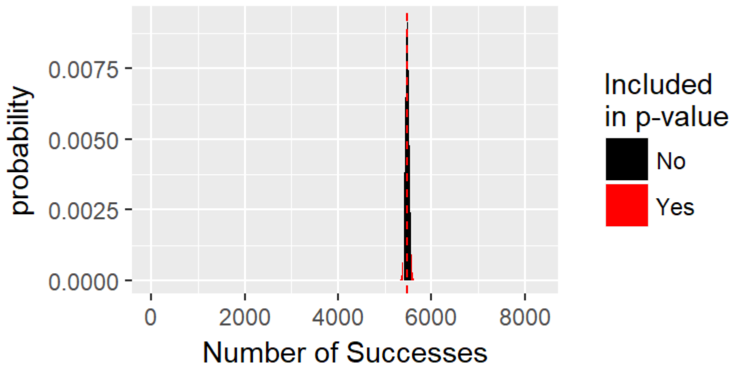


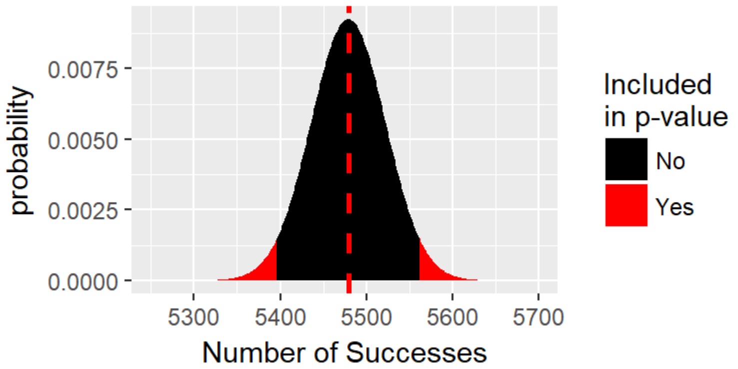
- 12 Blue M&M's out of 100
- $H_0: p = 0.16$
- $H_A: p < 0.16$
- **p-value:** $P(X \leq 12) = 0.1703$ if $X \sim \text{Binomial}(100, 0.16)$



P-Values for Two-Sided Tests

- 5562 out of 8302 students missed at least one day of school.
- $H_0: p = 0.66$, $H_A: p \neq 0.66$
- If H_0 is true, $X \sim \text{Binomial}(8302, 0.66)$
- "At least as extreme": at least as far from the expected value
- $E(X) = np = 8302 * 0.66 = 5479.32$





- R actually does something slightly different, but the results will usually be the same as what's described here.

Calculation of P-Values in R

- Suppose we have a data frame with a variable indicating success/failure:

```
paul_guesses
```

```
##    result
## 1 correct
## 2 correct
## 3 correct
## 4 correct
## 5 correct
## 6 correct
## 7 correct
## 8 correct
```

Calculation of P-Values in R (Cont'd)

- One-sided: $H_A: p > 0.5$

```
binom.test(paul_guesses$result,
  success = "correct",
  p = 0.5,
  alternative = "greater")
```

```
##
##
##
## data:  paul_guesses$result  [with success = correct]
## 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
```

Calculation of P-Values in R (Cont'd)

- One-sided: H_A: $p < 0.16$
- Suppose we know the number of trials ($n = 100$ M&M's) and number of successes ($x = 12$ blue)

```
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
```

Calculation of P-Values in R (Cont'd)

- Two-sided: H_A: $p \neq 0.66$
- Suppose we know the number of trials ($n = 8302$ students) and number of successes ($x = 5562$ missed school)

```
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:
```

Drawing Conclusions

- **p-value**: probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true
  - e.g., probability of getting at least 8 predictions right if Paul is just guessing
- If the p-value is **small**, that is evidence that the null hypothesis may not be true

Drawing Conclusions

- **p-value**: probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true
  - e.g., probability of getting at least 8 predictions right if Paul is just guessing
- If the p-value is **small**, that is evidence that the null hypothesis may not be true
- If we need to make a decision about whether or not the null hypothesis is true, we can see whether the p-value is smaller than a cutoff of our choosing
- The cutoff is the **significance level** of the test
- Denote the significance level by α (alpha)
- A common significance level: $\alpha = 0.05$
- But this choice is arbitrary

Drawing Conclusions

- If the p-value $< \alpha$, we "reject" H_0: the data offer enough evidence to conclude that H_0 is not true at the significance level α.
- If the p-value $\geq \alpha$, we "fail to reject"" H_0: the data **don't** offer enough evidence to conclude that H_0 is not true at the significance level α.

Note About the Book

- The procedure described in these slides is different from what's in the book.
- Our method uses
  - **sample statistic** = number of successes in the sample
  - **sampling distribution** = modeled with a Binomial
- The book's method uses
  - **sample statistic** = proportion of successes in the sample
  - **sampling distribution** = modeled with a Normal
- Everything else is the same (hypotheses, p-values, conclusions), and both methods are valid.
- Our procedure requires less work:
  - fewer assumptions to check (more broadly applicable)
  - fewer computations (e.g. no need to calculate $\sqrt{p(1-p)/n}$)