

Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Textbook versions and hw.
2. Simulating null distributions and SEs.
3. p-values.
4. Heart transplant example.

Read chapter 1.

Hw1 is due Tue 10/9 now. 1.3.16, 1.4.26, and the names and emails of 2 students.

<http://www.stat.ucla.edu/~frederic/13/F18> .

Switching sections is proving difficult. A lot of people want A to B,C, or D, but not the other way around.

1. Textbook and hw.

If you have a different edition of the textbook than the 2016 edition, then make sure you are doing the correct hw problems.

Hw1 is due Tue 10/9. 1.3.16, 1.4.26, and the names and emails of 2 students.

1.3.16 is on p84 and is about Rhesus monkeys, exercise 1.2.18, which is on p80. It starts "For this study: a. State the null hypothesis and the alternative hypothesis in the context of this study. b. Determine the standardized statistic from the data. (Hint: you will need to get the standard deviation of the simulated statistics from the null distribution in an applet." But you don't need an applet. You can use the theoretical value of $\sqrt{[\pi(1-\pi)/n]}$, where π is the probability of the monkey getting it right under the null hypothesis, or do simulations in *R*. For instance, in *R* you could do:

```
pi2 = ## insert your answer to the null hypothesis part of question a here.
```

```
a = rep(0,10000)
```

```
for(i in 1:10000){
```

```
  b = runif(40)
```

```
  c = (b < pi2)
```

```
  a[i] = mean(c)
```

```
}
```

```
sd(a)
```

```
## compare with
```

```
sqrt(pi2 * (1-pi2) / n)
```

Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

Hw1 is due Thu 10/4. 1.3.16, 1.4.26, and the names and emails of 2 students.

Also, on 1.3.16, it is a little unclear whether the alternative hypothesis should be 1-sided or 2-sided. Here they state they want to test whether the monkeys do "better" than you'd expect just by chance with random guessing. Not "better or worse". So a 1-sided hypothesis is appropriate here. Similarly with 1.4.26.

1.4.26 is on p90. It starts "Researchers wanted to test the hypothesis that living in the country is better for your lungs than living in a city."

Be careful in part c. The table gives you $P(\# \text{ of heads} = i)$, not $P(\# \text{ of heads} \geq i)$, for $i = 0, 1, 2, 3, \dots, 7$.

2. Simulating null distributions and Standard Errors.

Continuing the HG example from last time, we observe $p = 15.34\%$ in our sample, and under H_0 , the population percentage $\pi = 10\%$.

So we see a difference of 5.34% . This is our quantity of interest, and it is usually a difference like this. We want to see if that quantity of interest, 5.34% , is bigger than what we'd expect by chance under the null hypothesis.

The Standard Error (SE) is the standard deviation of the quantity of interest under the null hypothesis. Your book calls this the *standard deviation of the null distribution*.

Many stat books just tell you the formulas to get the SE. Your book is different. They want to emphasize that in many cases you can estimate the SE by simulations.

In this example, under H_0 , women with HG are just like the rest in terms of probability of delivering preterm. We have a SRS of size 254 from a population with $\pi = 10\%$ having preterm delivery. We can simulate 254 draws on the computer, where each draw is independent of the others and has a 10% chance of being preterm, and then see what results we get. In R, I did

```
x = runif(254)
```

```
y = (x < 0.1)
```

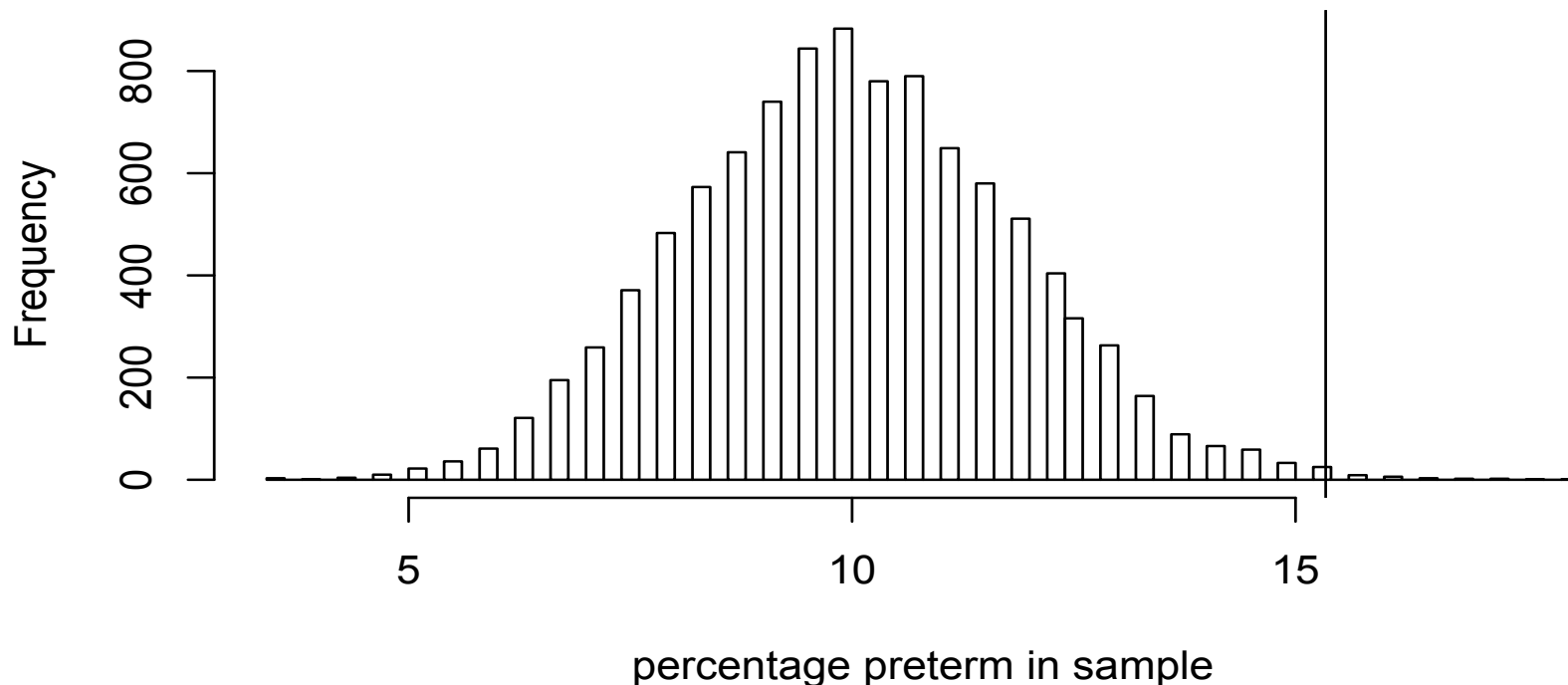
```
p = mean(y)
```

The first time, I got $p = 0.1259843$. 12.60%.

I tried it many times, and here is what I got.

```
a = rep(0,10000)
for(i in 1:10000){ x = runif(254); a[i] = mean(x<.1)}
hist(a*100,main="simulated preterm percentages", nclass=100,
      xlab="percentage preterm in sample")
abline(v=15.34)
mean(abs(a-.10)>.0534)      ## 0.0051
sd(a-.10)                   ## 0.01885409
sqrt(.10 * .90 / 254) ## 0.01882367
```

simulated preterm percentages



3. p-values.

The p-value is the probability, assuming H_0 is true, that the test statistic will be at least as extreme as that observed.

"What are the chances of that?"

The key idea is that the convention is to compute the probability of getting something as extreme as you observed or more extreme.

e.g. $n = 5$, $\pi_0 = 50\%$, $p = 4/5$. The probability that $p = 4/5$ is 15.625%.

However, what if $n = 400$, $\pi_0 = 50\%$, and $p = 201/400$? Now the probability of getting 201/400 is 3.97%, but obviously the data are consistent with the null hypothesis that $\pi = 50\%$.

Typically, one does a two-sided test, which means that by "extreme", we mean extreme in either direction. We want to see how in line our observed value of $p=15.34\%$ is with our null hypothesis of a population percentage of 10%. Could our sample of 15.34% preterm have come from a population of 10% preterm? A simulation with $p>15.34\%$ would be more extreme than what we observed, and also a simulation with $p<4.66\%$ would be more extreme than what we observed.

Guidelines for evaluating strength of evidence from p-values

- p-value > 0.10 , not much evidence against null hypothesis
- $0.05 < \text{p-value} \leq 0.10$, moderate evidence against the null hypothesis
- $0.01 < \text{p-value} \leq 0.05$, strong evidence against the null hypothesis
- $\text{p-value} \leq 0.01$, very strong evidence against the null hypothesis

Continuing the HG example from last time, using simulations of H_0 we obtained samples of 254 values, and just by chance, in 0.51% of these samples, at least 15.34% or more were preterm or less than 4.66% were preterm.

So we'd say the p-value is 0.51% for this two-sided test.

The observed difference is highly significant, and we have strong evidence against the null hypothesis of HG pregnancies having a 10% chance of being preterm like other pregnancies.

4. Heart Transplant Example.

Example 1.3

Heart Transplants

- The *British Medical Journal* (2004) reported that heart transplants at St. George's Hospital in London had been suspended after a spike in the mortality rate
- Of the last 10 heart transplants, 80% had resulted in deaths within 30 days
- This mortality rate was over five times the national average.
- The researchers used 15% as a reasonable value for comparison.

Heart Transplants

- Does a heart transplant patient at St. George's have a higher probability of dying than the national rate of 0.15?
- Observational units
 - The last 10 heart transplantations
- Variable
 - If the patient died or not
- Parameter
 - The actual probability of a death after a heart transplant operation at St. George's

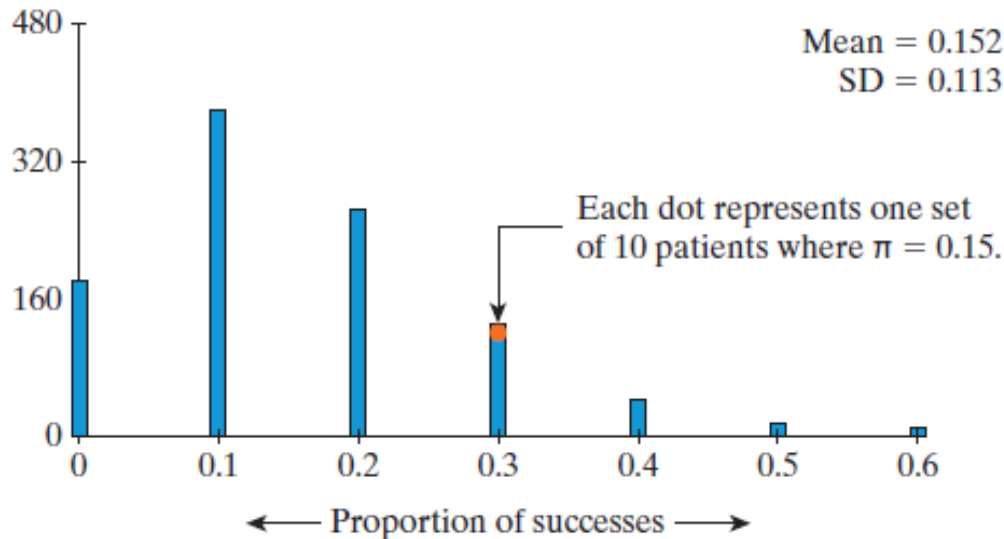
Heart Transplants

- **Null hypothesis:** Death rate at St. George's is the same as the national rate (0.15).
- **Alternative hypothesis:** Death rate at St. George's is higher than the national rate.
- $H_0: \pi = 0.15$ $H_a: \pi > 0.15$
- Our **statistic** is 8 out of 10 ($\hat{p} = 0.8$)

Heart Transplants

Simulation

- Null distribution of 1000 repetitions of drawing samples of 10 “patients” where the probability of death is equal to 0.15.



What is the p-value?

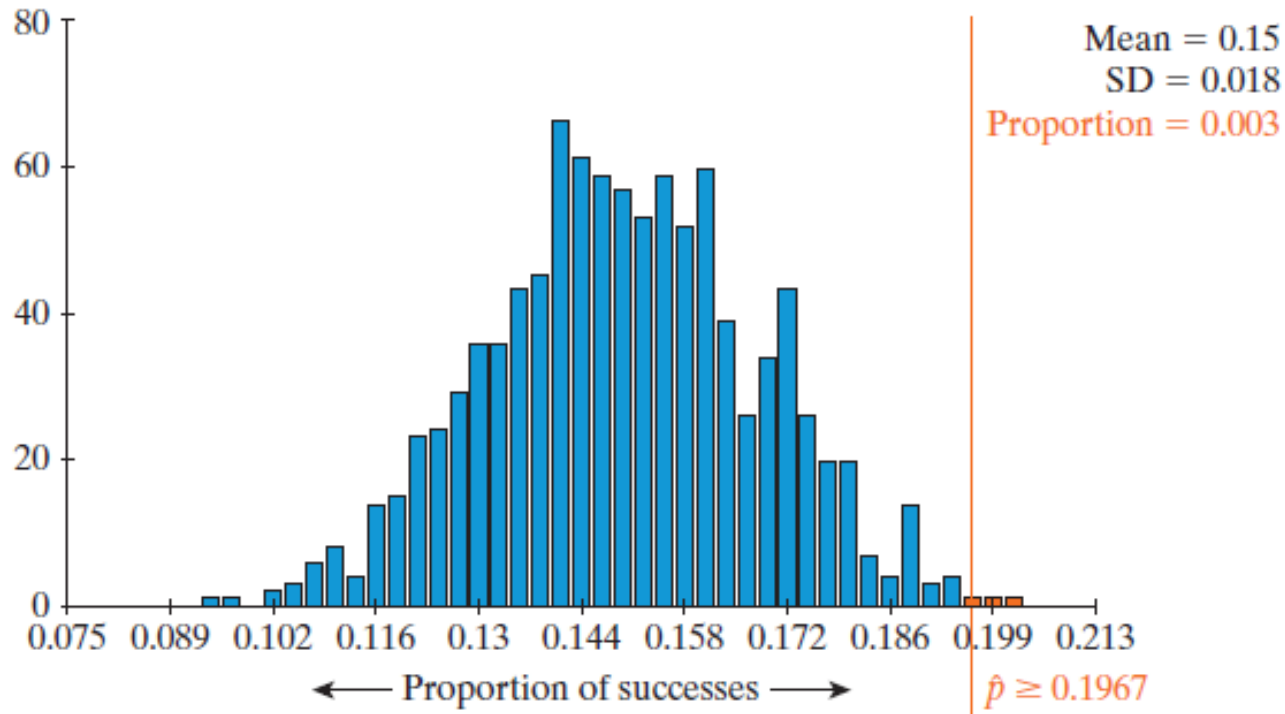
Heart Transplants

Strength of Evidence

- Our p-value is 0, so we have very strong evidence against the null hypothesis.
- Even with this strong evidence, it would be nice to have more data.
- Researchers examined the previous 361 heart transplantations at St. George's and found that 71 died within 30 days.
- Our new statistic, \hat{p} , is $71/361 \approx 0.1967$

Heart Transplants

- Here is a null distribution and p-value based on the new statistic.



Heart Transplants

- The p-value was about 0.003
- We still have very strong evidence against the null hypothesis, but not quite as strong as the first case
- Another way to measure strength of evidence is to ***standardize*** the observed statistic

The Standardized Statistic

- The ***standardized statistic*** is the number of standard errors our sample statistic is above the mean of the null distribution (or below the mean if it is negative).
- $$z = \frac{\text{statistic} - \text{mean of null distribution}}{\text{standard error}}$$
- The book uses the phrase *standard deviation of the null distribution* in place of standard error.
- For a single proportion, we will use the symbol z for standardized statistic.
- Note: In the formula above, we can either use the mean of the actual null distribution or (better yet) the long-term proportion (probability) given in the null hypothesis.

The Standardized Statistic

- Here are the standardized statistics for our two studies.

$$z = \frac{0.80 - 0.15}{0.113} = 5.75 \quad z = \frac{0.197 - 0.15}{0.018} = 2.61$$

- In the first, our observed statistic was 5.75 standard errors above the mean.
- In the second, our observed statistic was 2.61 standard errors above the mean.
- Both of these are very strong, but we have stronger evidence against the null in the first.