

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

1. p-values.
2. Heart transplant example.
3. Standardized statistic
4. What impacts p-values and strength of evidence

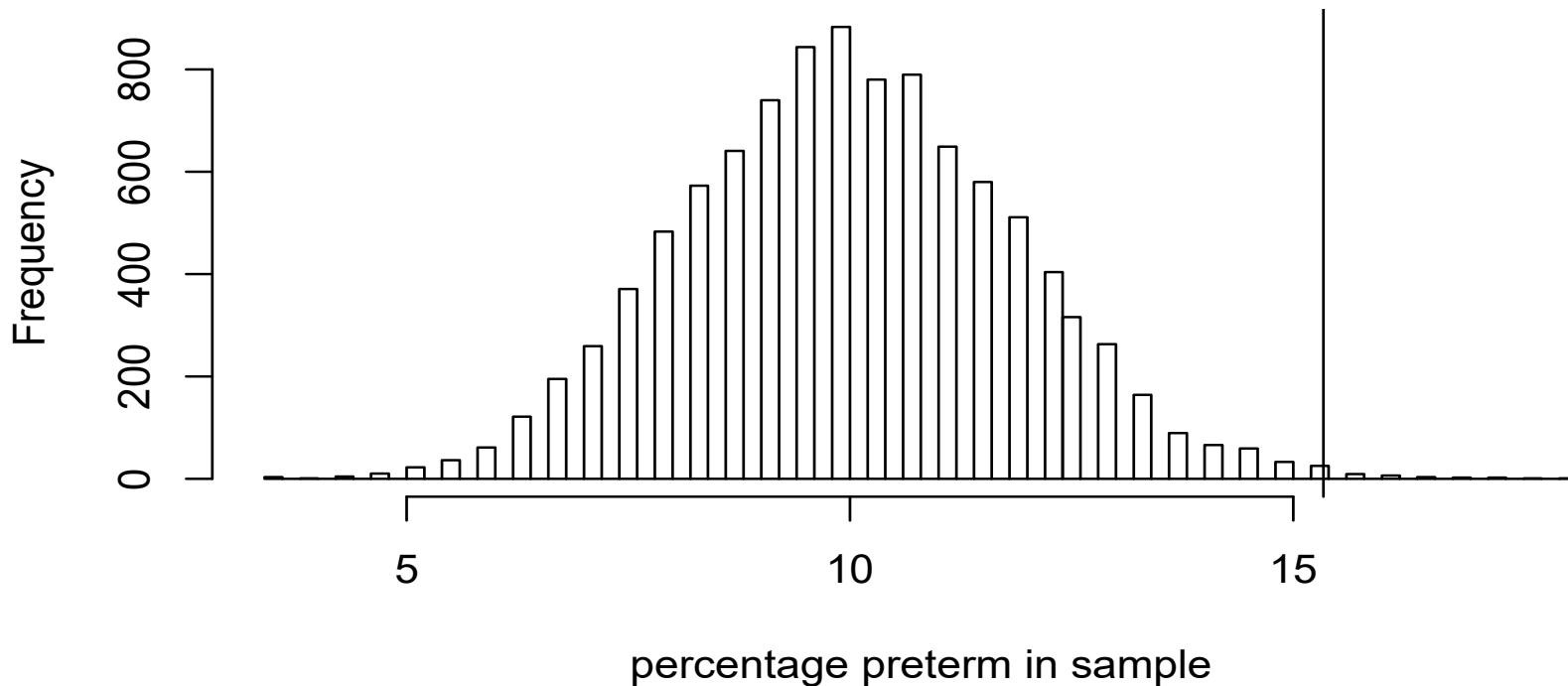
Read through chapter 1.

Hw1 is due Fri Apr17, 1159pm, and should be submitted by EMAIL to STATGRADER@STAT.UCLA.EDU .

The course website is <http://www.stat.ucla.edu/~frederic/13/S26>

```
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)
sd(a)          ## 0.01885409
sqrt(.10 * .90 / 254) ## 0.01882367
```

simulated preterm percentages



1. 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. $n = 5$, $\pi_0 = 50\%$, $\hat{p} = 4/5$. The probability that $\hat{p} = 4/5$ is 15.625%.

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

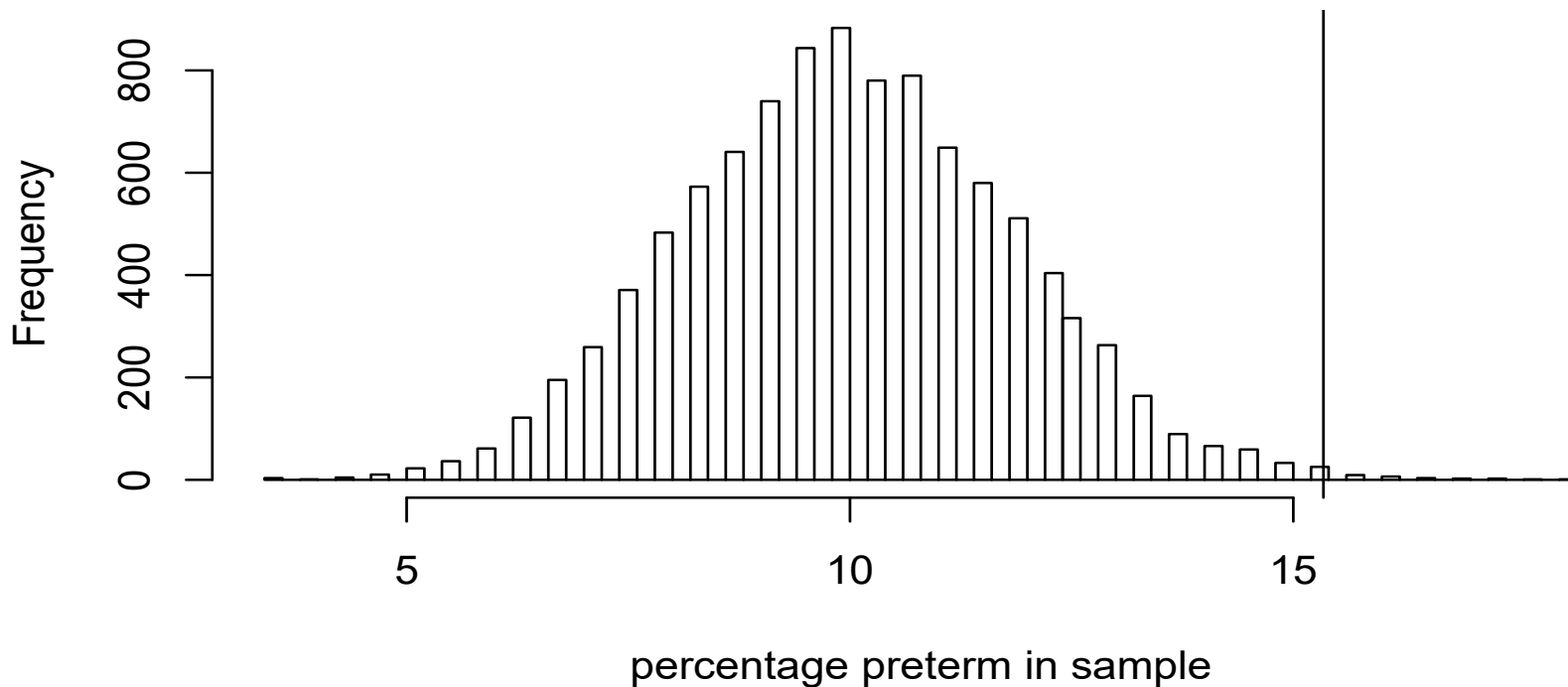
Typically, one does a two-sided test, which means that by "extreme", we mean extreme in either direction. We want to see how in line our observed value of $\hat{p} = 15.34\%$ is with our null hypothesis of a population percentage of 10%. Could our sample of 15.34% preterm have come from a population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

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

simulated preterm percentages



Continuing the HG example, using simulations of H_0 we obtained samples of 254 values, and 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.

2. 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

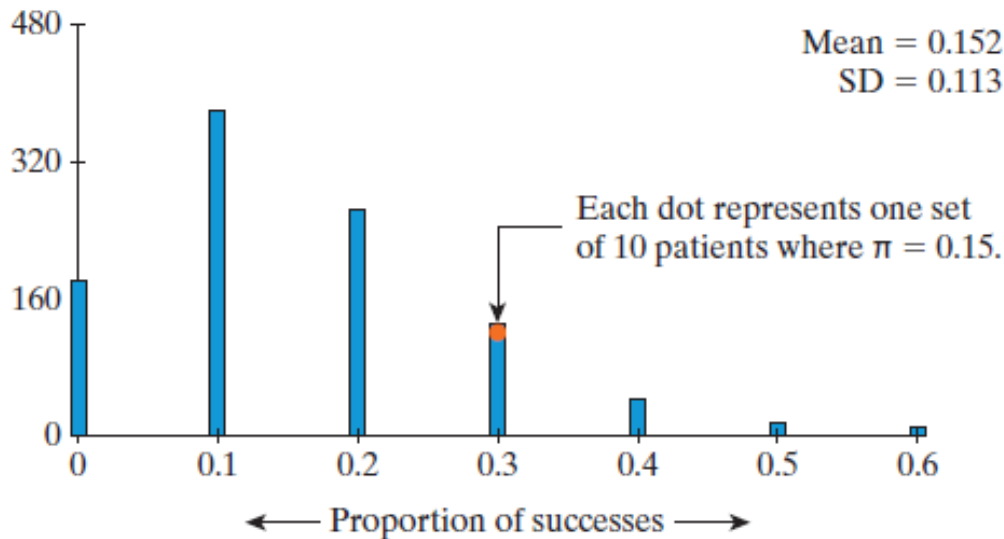
Heart Transplants

- **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?

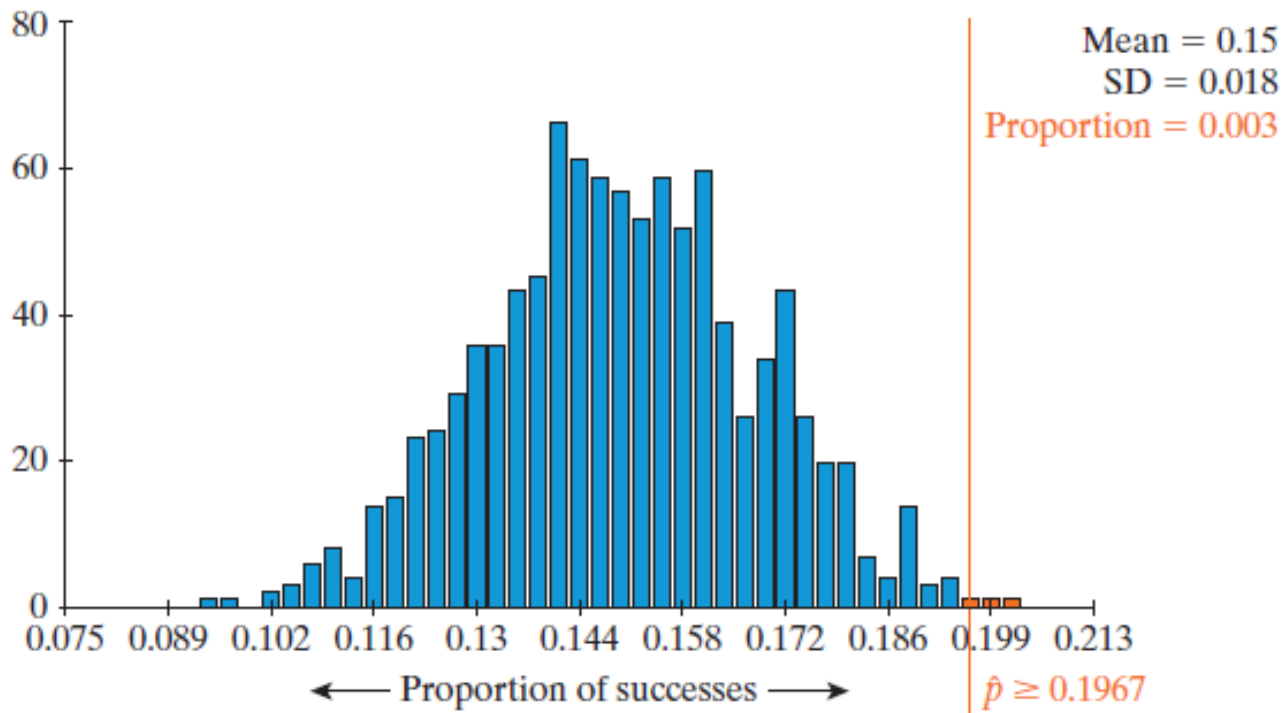
Heart Transplants

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

3. The Standardized Statistic

- The *standardized statistic* is the number of standard deviations our sample statistic is above the mean of the null distribution (or below the mean if it is negative).
- $$z = \frac{\textit{statistic} - \textit{mean of null distribution}}{\textit{standard deviation of null distribution}}$$
- The sd of the null distribution is the *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 deviations above the mean.
- In the second, our observed statistic was 2.61 standard deviations above the mean.
- Both of these are very strong, but we have stronger evidence against the null in the first.

Guidelines for strength of evidence

- If a standardized statistic is below -2 or above 2, we have strong evidence against the null.

Standardized Statistic	Evidence Against Null
between -1.5 and 1.5	not much
below -1.5 or above 1.5	moderate
below -2 or above 2	strong
below -3 or above 3	very strong