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

1. Note on 1-sided and 2-sided tests.
2. Predicting faces example.
3. 2-sided tests.
4. Predicting house elections.
5. Normal distribution, CLT, and Halloween candy example.

Read chapters 2 and 3.
Hw1 is due Wed Jan22. 1.3.16 and 1.4.26. Also, on the bottom of your hw, print the names and emails of two other students in the class.
HW should be submitted BY EMAIL to STATGRADER@STAT.UCLA.EDU for sections a and b, and to STATGRADER2@STAT.UCLA.EDU for c and d.

The course website is http://www.stat.ucla.edu/~frederic/13/W25

Two-Sided Tests

- The change to the alternative hypothesis affects how we compute the p-value.
- Remember that the p-value is the probability (assuming the null hypothesis is true) of obtaining a proportion that is equal to or **more extreme** than the observed statistic
- In a *two-sided test*, **more extreme** goes in both directions.

Example 1.4

Predicting Elections from Faces

Predicting Elections

- Do voters make judgments about candidates based on facial appearances?
- More specifically, can you predict an election by choosing the candidate whose face is more competent-looking?
- Participants were shown two candidates and asked who has the more competent-looking face.

Who has the more competent looking face?

- 2004 Senate Candidates from Wisconsin

Winner

Loser

Bonus: One is named Tim and the other is Russ. Which name is the one on the left?

- 2004 Senate Candidates from Wisconsin

Russ Tim

Predicting Elections

- They determined which face was the more competent for the 32 Senate races in 2004.
- What are the observational units?
  - The 32 Senate races
- What is the variable measured?
  - If the method predicted the winner correctly

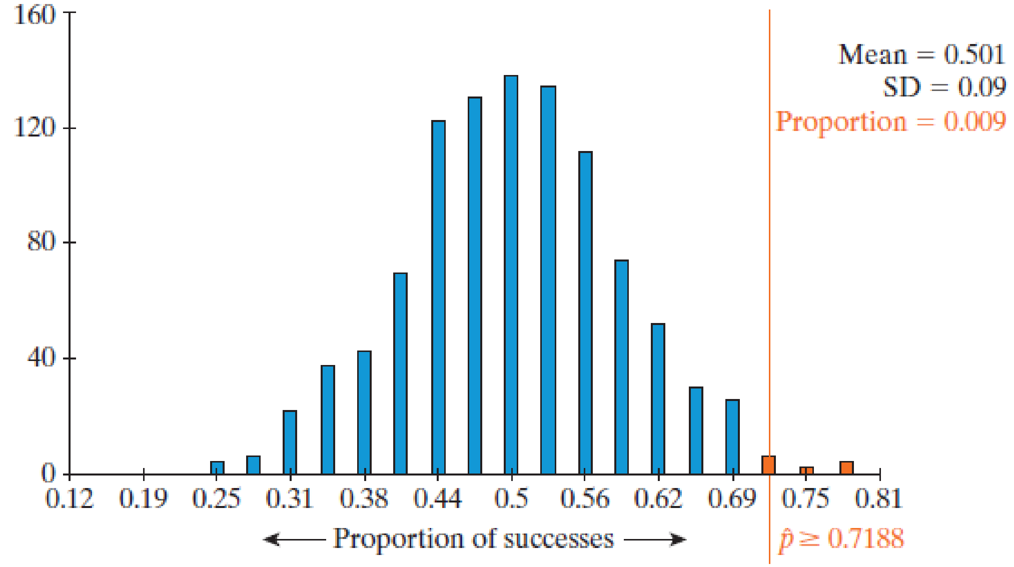
Predicting Elections

- Null hypothesis: The probability this method predicts the winner equals 0.5. (H_0: $\pi = 0.5$)
- Alternative hypothesis: The probability this method predicts the winner is greater than 0.5. (H_a: $\pi > 0.5$)

- This method predicted 23 of 32 races, hence $\hat{p} = 23/32 \approx 0.719$, or 71.9%.

Predicting Elections

1000 simulated sets of 32 races



Predicting Elections

- With a p-value of 0.009 we have strong evidence against the null hypothesis.
- When we calculate the standardized statistic we again show strong evidence against the null.

$$z = \frac{0.7188 - 0.5}{0.09} = 2.43.$$

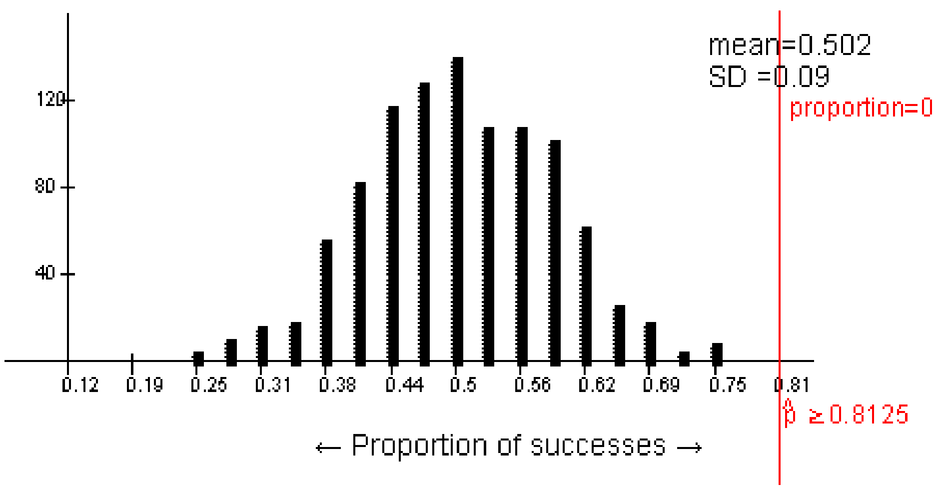
What do the p-value and standardized statistic mean?

What affects the strength of evidence?

1. The effect size, which is the difference between the observed statistic ($\hat{p}$) and null hypothesis parameter (π_0).
2. Sample size.
3. If we do a one or two-sided test.

Effect size, i.e. the difference between $\hat{p}$ and π_o

- What if researchers predicted 26 elections instead of 23?
  - 26/32 = 0.8125 never occurs just by chance



Difference between $\hat{p}$ and the null parameter

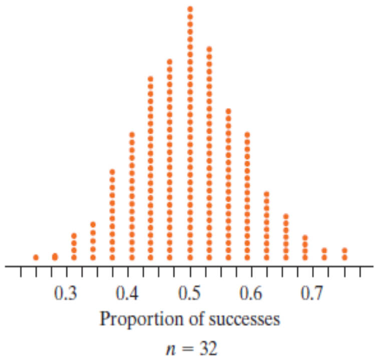
- The farther away the observed statistic is from the average value of the null distribution (or π_0), the more evidence there is against the null hypothesis.

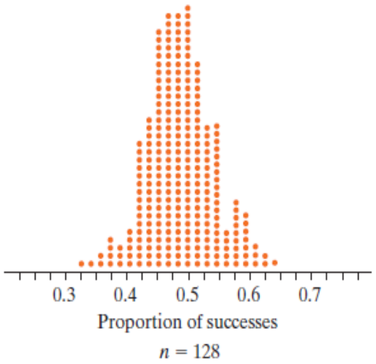
Sample Size

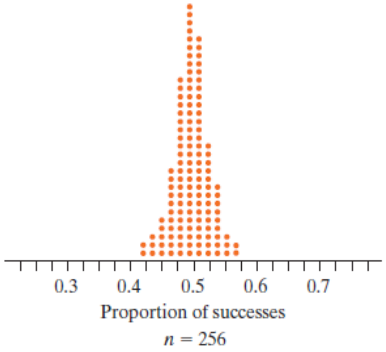
Suppose the sample proportion stays the same, do you think increasing sample size will increase, decrease, or have no impact on the strength of evidence against the null hypothesis?

Sample Size

- The null distribution changes as we increase the sample size from 32 senate races to 128 races to 256 races.
- As the sample size increases, the variability (standard error) decreases.







Sample Size

- What does decreasing variability mean for statistical significance (with same sample proportion)?
- 32 elections
  - p-value = 0.009 and z = 2.43
- 128 elections
  - p-value = 0 and z =5.07
- 256 elections
  - Even stronger evidence
  - p-value = 0 and z = 9.52

Sample Size

- As the sample size increases, the variability decreases.
- Therefore, as the sample size increases, the evidence against the null hypothesis increases (as long as the sample proportion stays the same and is in the direction of the alternative hypothesis).

Two-Sided Tests

- What if researchers were wrong; instead of the person with the more competent face being elected more frequently, it was actually less frequently?

  $H_0: \pi = 0.5$

  $H_a: \pi > 0.5$

- With this alternative, if we get a sample proportion less than 0.5, we would get a p-value greater than 50%.
- This is a ***one-sided*** test.
- Often one-sided is too narrow
- In fact most research uses two-sided tests.

Two-Sided Tests

- In a two-sided test the null can be rejected when sample proportions are in either tail of the null distribution.

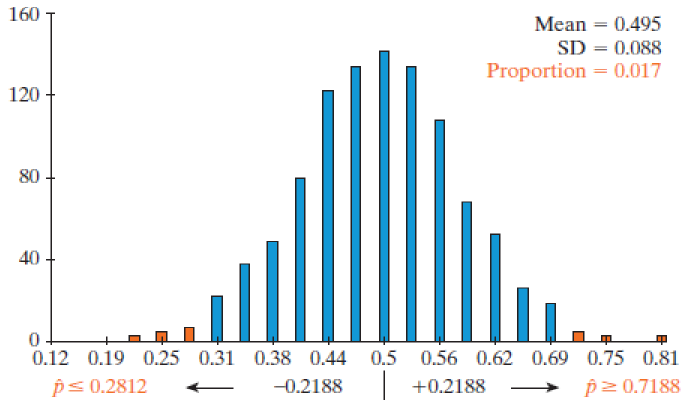
Null hypothesis: The probability this method predicts the winner equals 0.50. (H_0: $\pi = 0.50$)

Alternative hypothesis: The probability this method predicts the winner **is not** 0.50.

(H_a: $\pi \neq 0.50$)

Two-Sided Tests

- Continuing with the example of predicting elections based on faces, since our sample proportion was 0.7188 and 0.7188 is 0.2188 *above* 0.5, we also need to look at 0.2188 *below* 0.5.
- The p-value will include all simulated proportions 0.7188 and above as well as those 0.2812 and below.



Two-Sided Tests

- 0.7188 or greater was obtained 9 times
- 0.2812 or less was obtained 8 times
- The p-value is (8 + 9 = 17)/1000 = 0.017.
- Two-sided tests increase the p-value (it about doubles) and hence decrease the strength of evidence.
- Two-sided tests are said to be more conservative. More evidence is needed to reject the null hypothesis.

Predicting House Elections

- Researchers also predicted the 279 races for the House of Representatives in 2004.
- They correctly predicted the winner in 189/279 ≈ 0.677, or 67.7% of the races.
- The House's sample percentage (67.7%) is a bit smaller than the Senate (71.9%), but the sample size is larger (279) than for the senate races (32).
- Do you expect the strength of evidence to be stronger, weaker, or essentially the same for the House compared to the Senate?

Predicting House Elections

Distance of the observed statistic to the null hypothesis value

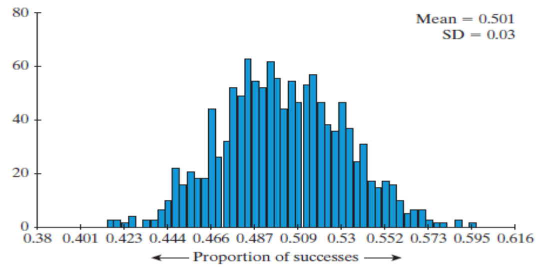
- The statistic in the House is 0.677 compared to 0.719 in the Senate
- Slight decrease in the effect size.

Sample size

- The sample size is almost 10 times as large (279 vs. 32)
- This will increase the strength of evidence.

Predicting House Elections

Null distribution of 279 sample House races



Simulated statistics ≥0.677 didn't occur at all so the estimated p-value is 0

Predicting House Elections

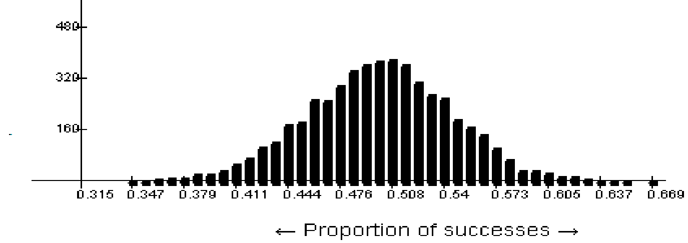
- What about the standardized statistics?
  - For the Senate it was 2.43
  - For the House is 5.90.
- The larger sample size for the House outweighed the smaller effect size in this particular case. We have stronger evidence against the null using the data from the House.

Predicting Elections

- Do voters make judgments about candidates based on facial appearances?
- More specifically, can you predict an election by choosing the candidate whose face is more competent-looking?
- Participants were shown two candidates and asked who has the more competent-looking face.

Normal distribution, CLT, and halloween candy example.

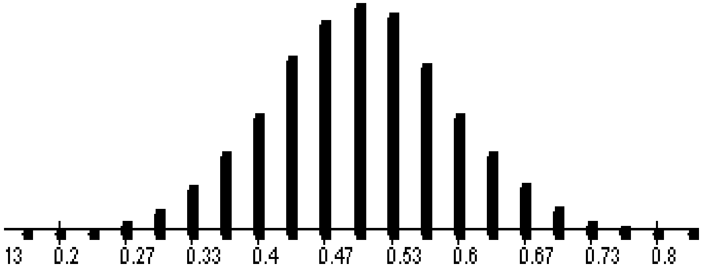
Section 1.5

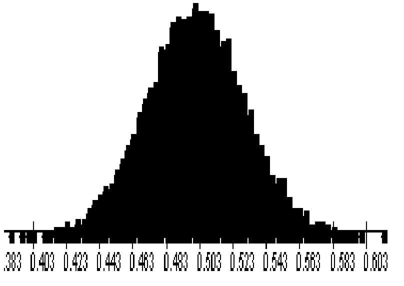


- The shape of most of our simulated null distributions always seems to be bell shaped. This shape is called the normal distribution.
- The Central Limit Theorem (CLT) dictates that, as *n* gets large, the sample mean or proportion becomes approximately normally distributed.
- When we do a test of significance using theory-based methods, only how our p-values are found will change. Everything else will stay the same.

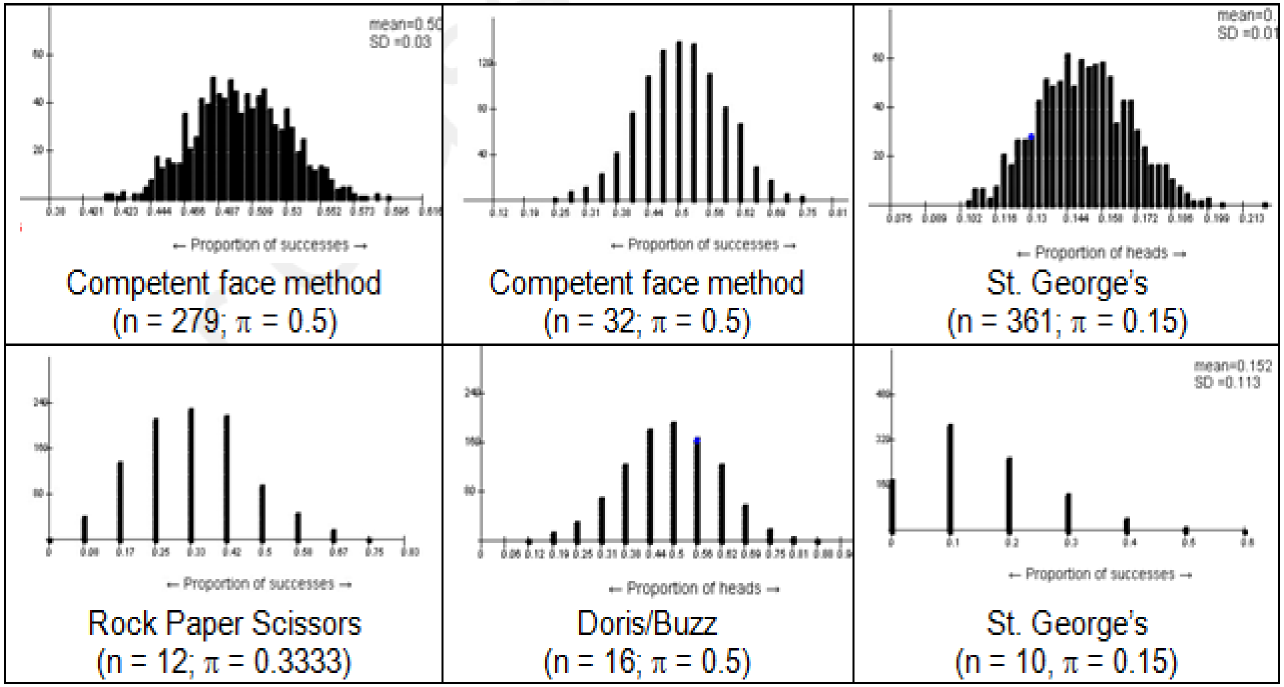
The Normal Distribution

- Both of these are centered at 0.5.
  - The one on the left represents samples of size 30.
  - The one on the right represents samples of size 300.
  - Both could be described as normal distributions.





- Which ones will normal distributions fit?



When can I use a theory-based test that uses the normal distribution?

- The shape of the randomized null distribution is affected by the sample size and the proportion under the null hypothesis.
- The larger the sample size the better.
- The closer the null proportion is to 0.5 the better.
- For testing proportions, you should have at least 10 successes and 10 failures in your sample to be confident that a normal distribution will fit the simulated null distribution nicely.