Introduction to Statistical Investigations

Second Edition

Tintle, Chance, Cobb, Rossman, Roy, Swanson, VanderStoep

Chapter 1.3

Significance: How Strong is the Evidence?

Review: Interpreting p-values

- P-values are conditional probabilities:

The probability of observing a statistic as extreme or more extreme than $\hat{p}$, assuming the null hypothesis is true.

- We know how to interpret probability in context:

If <random process> is repeated a very large number of times, then, in the long run, the proportion of repetitions in which <event of interest> occurs will approach <probability>.

Interpreting p-values

We can put these together.

- **Repeated random process**: the null-hypothesis version of the study activity
  - Harley guesses 10 times at random
  - 20 students choose R-P-S at random
- **Event of interest**: an observed statistic as extreme or more extreme than $\hat{p}$
  - Harley gets at least 9/10 correct
  - 2/20 or fewer students choose scissors
- **Probability**: p-value

Interpreting p-values

What does it mean to say "the p-value for the Harley study was 0.01"?

If *Harley guesses 10 times at random*, and this process is repeated a very large number of times, then, in the long run, the proportion of repetitions in which *Harley gets at least 9/10 correct* will approach *0.01*.

Now you try for "the p-value for the R-P-S study was 0.149."

Alternative Measure of Strength of Evidence

Section 1.3

One vs. Two-sided tests

Two-Sided Tests

- In a two-sided test, the alternative hypothesis is the logical negation of the null.

- **Null hypothesis:** How people decide which face is Tim's is no different from random guessing. (H_0: $\pi = 0.50$)

- **Alternative hypothesis:** How people decide which face is Tim's is different from random guessing. (H_a: $\pi \neq 0.50$)

Let's guess people's faces

- Given 2 photos and a name (Tim), can we guess who is named Tim?

Is the face on the left Bob or Tim?

Two-Sided Tests and p-values

- Changing the alternative hypothesis affects how we compute the p-value.
- The p-value is the probability 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.

Strength of Evidence

- P-values are one measure for the strength of evidence.

- Another approach is to *standardize* our observed statistic.

- *Standardizing* gives a quick, informal way to evaluate strength of evidence.
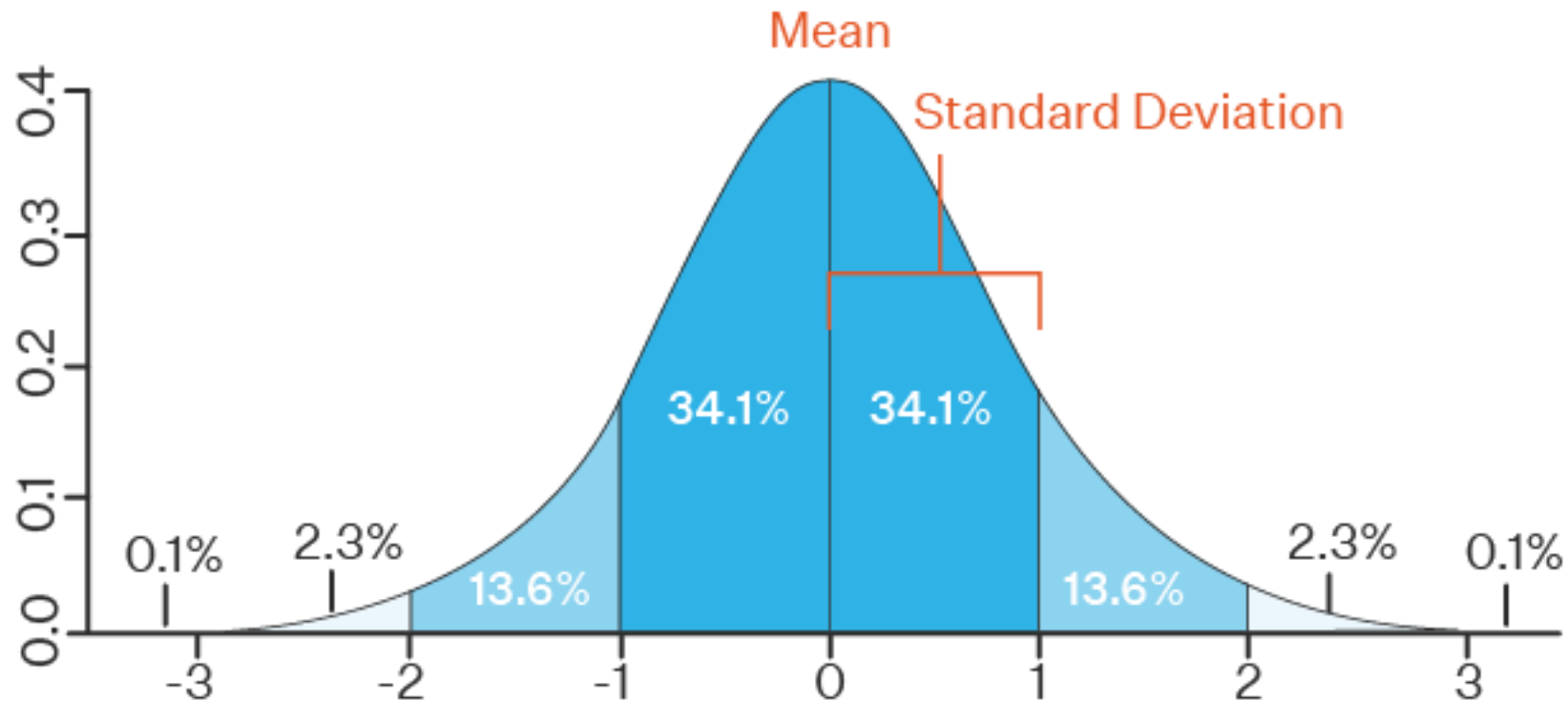
Describing a Distribution

- Some aspects to look for in a distribution:
    - ***Shape***
    - ***Center***
    - ***Variability*** (how spread out is the data?)
    - ***Unusual observations***

Describing a Distribution

- For our null distribution:
  - *Shape:* bell-curve ("normal")
  - *Center:* mean (average)
  - *Variability:* standard deviation (SD)
  - *Unusual observations:* none

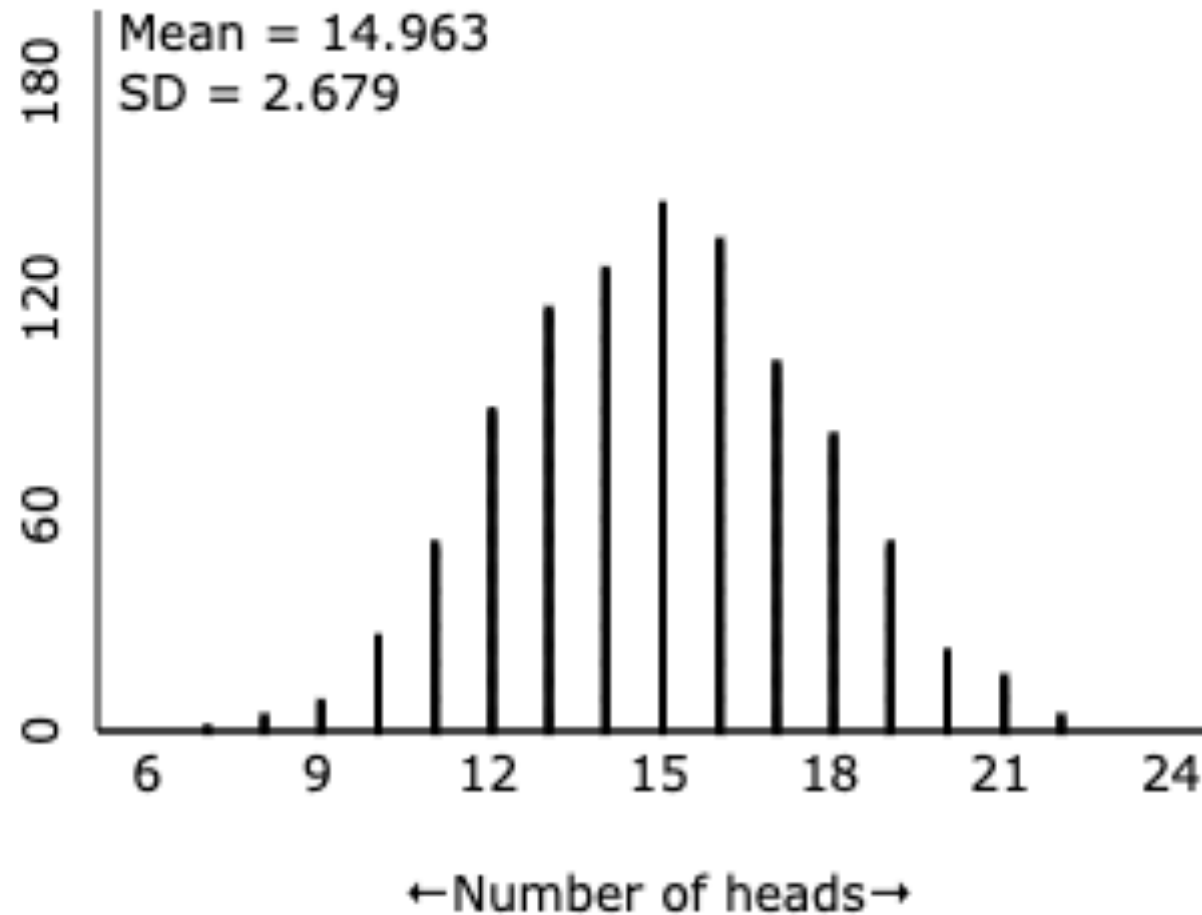# What's a summary statistic?

Describing a Distribution

- There are two main ways to find the mean and SD of a null distribution:
  - Have the applet compute them. (Click "Summary Stats" box in **One Proportion** applet.)
  - Use an appropriate theory-based approach (Section 1.5).

Describing a Distribution

☑ Summary Statistics

Mean = 14.963
SD = 2.679

←Number of heads→

The Standardized Statistic

- The ***standardized statistic*** (denoted *z*) is the number of standard deviations an observed statistic falls above or below the mean of a null distribution.

$$z = \frac{(\text{observed statistic}) - (\text{mean of null distribution})}{(\text{SD of null distribution})}$$

The Standardized Statistic

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
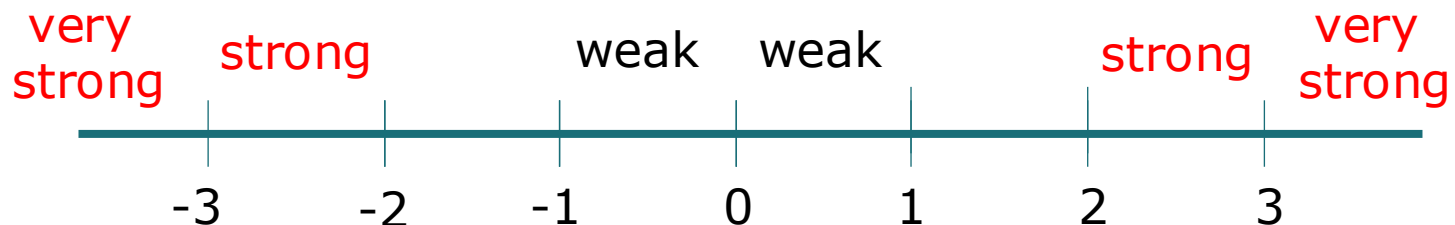
Strength of evidence against H_0 for various values of z

Strength of Evidence

- Two-sided tests increase the p-value compared to one-sided tests (it about doubles).
- So more evidence is needed to conclude the alternative in a two-sided test.
- The guidelines given for interpreting z-values correspond to two-sided p-values.
- But two-sided p-values don't convey whether observed statistic is more or less than what the null predicts, whereas z-values do.