Elementary statistics

UW CSE 160

Winter 2017

A dice-rolling game

- Two players each roll a die
- The higher roll wins
  - Goal: roll as high as you can!
- Repeat the game 6 times

Hypotheses regarding the outcome

- Luck
- Fraud
  - loaded die
  - inaccurate reporting

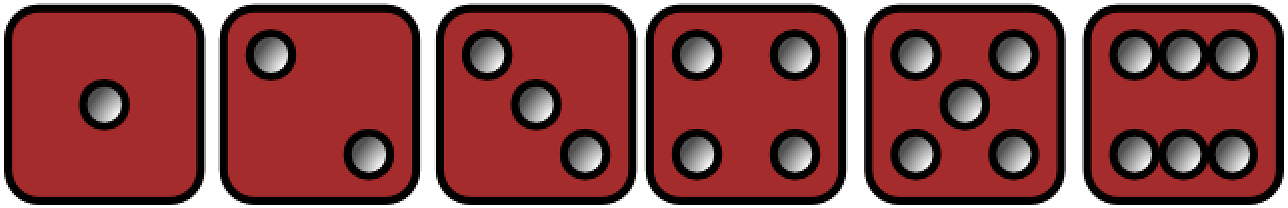
- How likely is luck?
- How do we decide?

Questions that statistics can answer

- I am flipping a coin. Is it fair?
  How confident am I in my answer?
- I have two bags of beans, each containing some black and some white beans. I have a handful of beans. Which bag did the handful come from?
- I have a handful of beans, and a single bag. Did the handful come from that bag?

- Does this drug improve patient outcomes?
- Which website design yields greater revenue?
- Which baseball player should my team draft?
- What premium should an insurer charge?
- Which chemical process leads to the best-tasting beer?

What can happen when you roll a die?

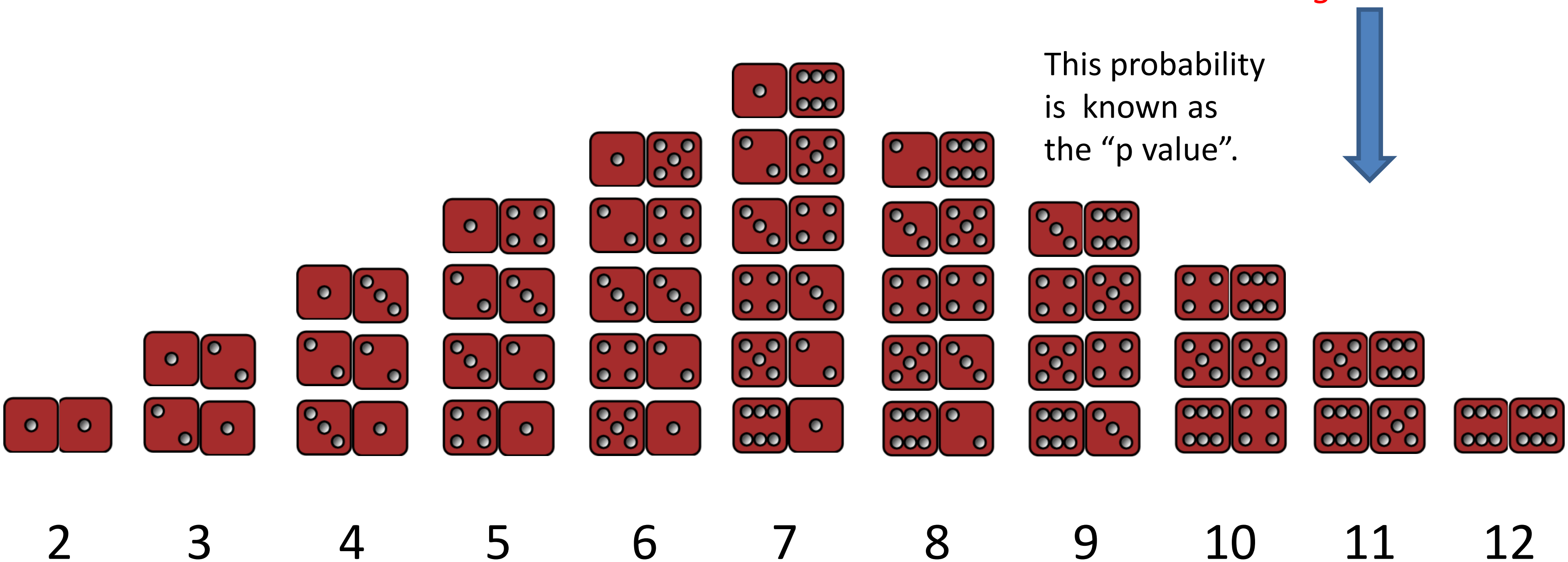


What is the likelihood of each?

What can happen when you roll two dice?

How likely are you to roll *11 or higher*?

This probability is known as the "p value".



How to compute p values

- Via a statistical formula
  - Requires you to make assumptions and know which formula to use
- Computationally (simulation)
  - Run many experiments
  - Count the fraction with a better result
    - Requires a metric/measurement for “better”
  - Requires you to be able to run the experiments
  - We will use this approach exclusively

Analogy between hypothesis testing and mathematical proofs

"The underlying logic [of hypothesis testing] is similar to a proof by contradiction. To prove a mathematical statement, A, you assume temporarily that A is false. If that assumption leads to a contradiction, you conclude that A must actually be true."

From the book *Think Statistics* by Allen Downey

Interpreting p values

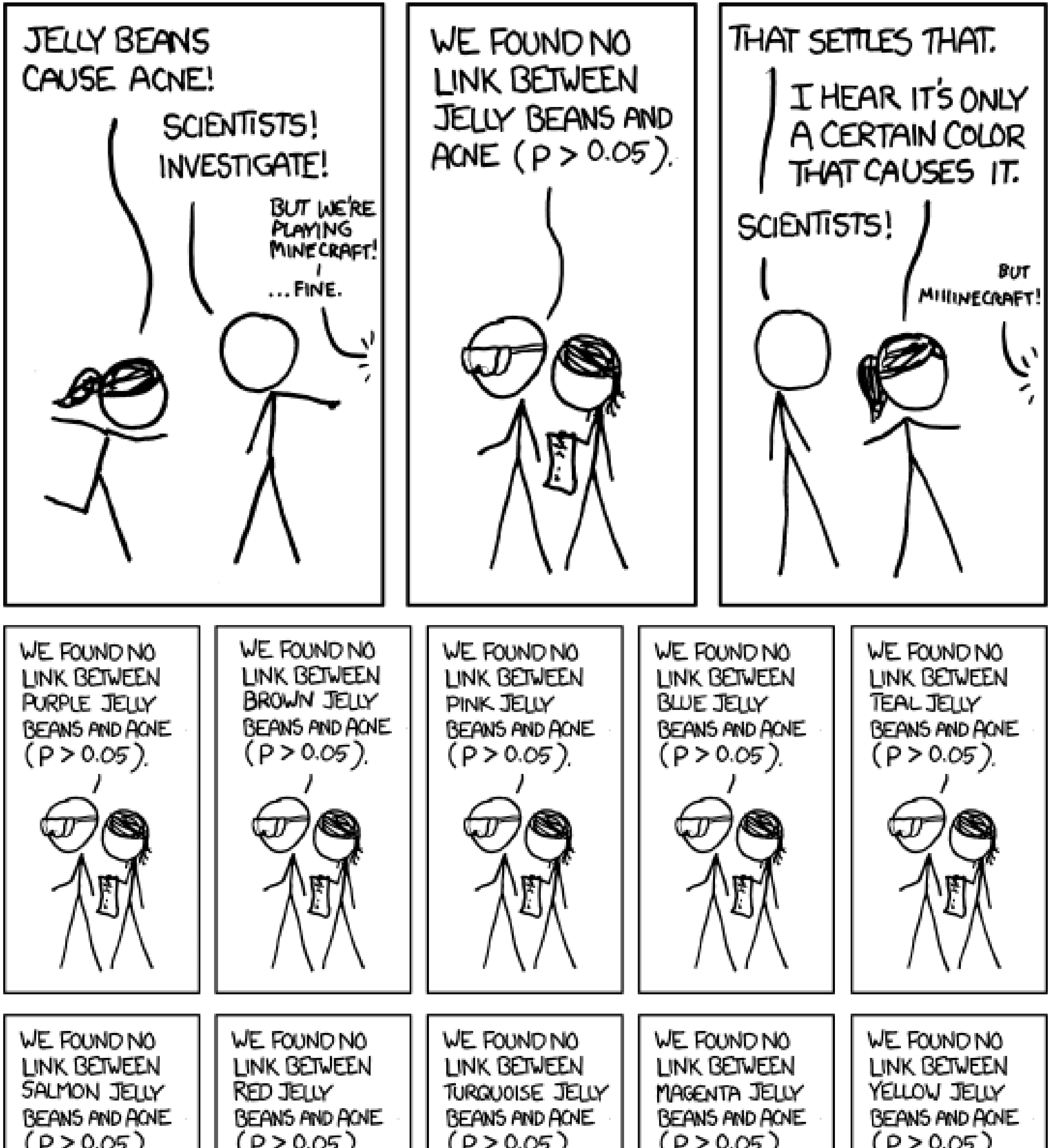
p value of 5% or less = statistically significant

- This is a *convention*; there is nothing magical about 5%

Two types of errors may occur in statistical tests:

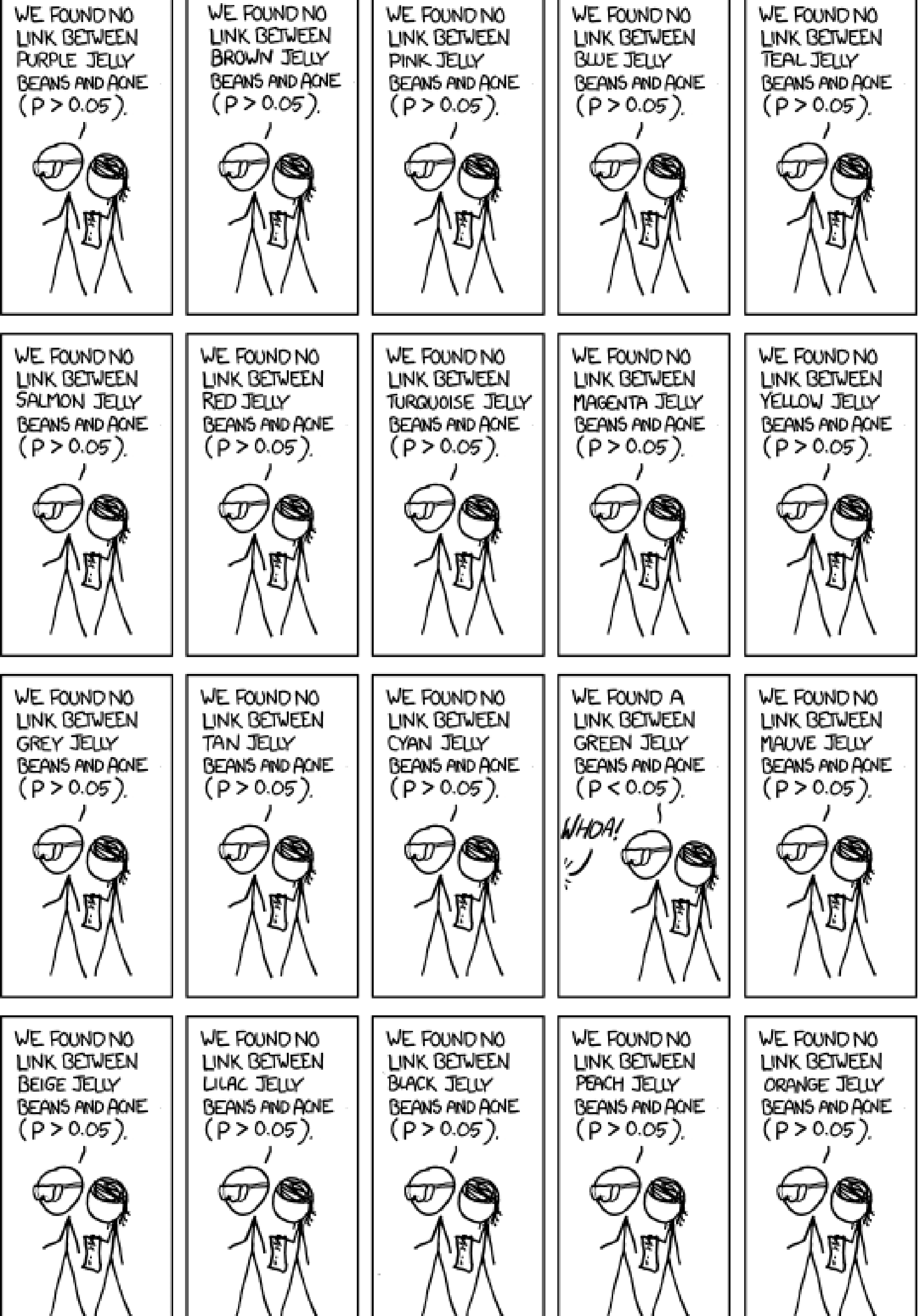
- false positive (or false alarm or Type I error): no real effect, but report an effect (through good/bad luck or coincidence)
  - If no real effect, a false positive occurs about 1 time in 20
- false negative (or miss or Type II error): real effect, but report no effect (through good/bad luck or coincidence)

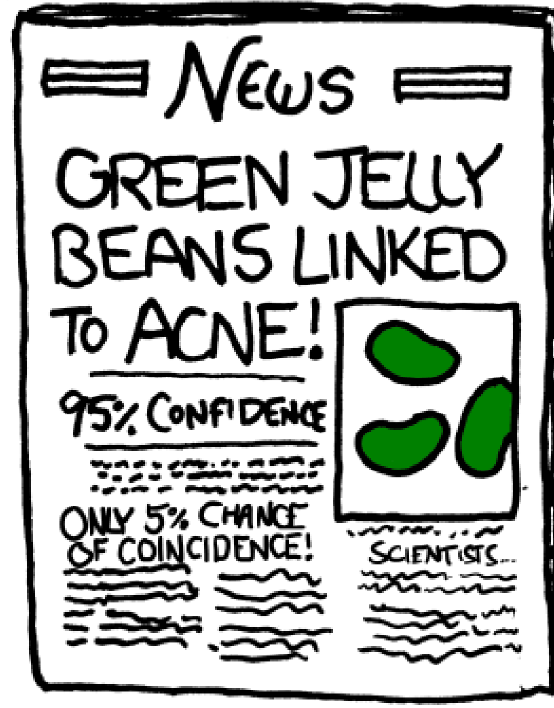
The *larger* the sample, the *less the likelihood* of a false positive or negative



A false positive

http://xkcd.com/882/





http://xkcd.com/882/

Summary of statistical methodology

1. Decide on a metric (bigger value = better)
2. Observe what you see in the real world
3. Hypothesize that what you saw is normal/typical

   This is the "null hypothesis"
4. Simulate the real world many times
5. How different is what you observed from the simulations?

   What percent of the simulation values are the real world values bigger than?
6. If the percentage is 95% or more, reject the null hypothesis

A common error

1. Observe what you see in the real world
2. Decide on a metric (bigger value = better)

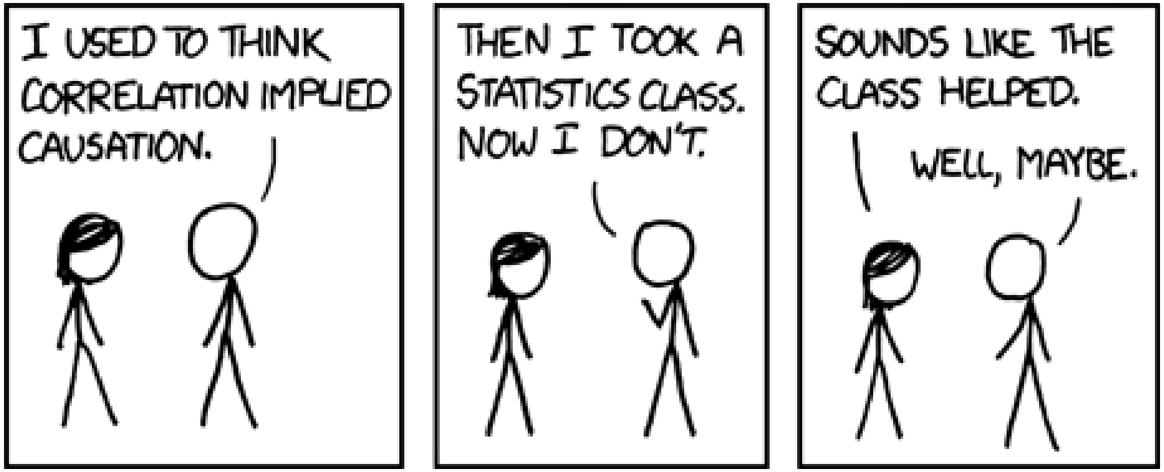
This is *backwards*

For any observation, there is something unique about it.

Example: Roll dice, then be amazed because what are the odds you would get exactly that combination of rolls?

Correlation ≠ causation

Ice cream sales and rate of drowning deaths are correlated



http://xkcd.com/552/

Statistical significance
≠ practical importance

Don’t trust your intuition

- People have very bad statistical intuition
- It’s much better to follow the methodology and do the experiments