

# STA 130 (Winter 2016): An Introduction to Statistical Reasoning and Data Science

Mondays 2:10 – 4:00 (GB 220)  
and Wednesdays 2:10 – 4:00 (various)

Jeffrey Rosenthal

Professor of Statistics, University of Toronto

[www.probability.ca](http://www.probability.ca)

[j.rosenthal@math.toronto.edu](mailto:j.rosenthal@math.toronto.edu)

sta130-1

## Introduction

- What is “statistics”? Is it ...
  - A mathematical theory of probability and randomness.
  - A method of analysing data that arise in medicine, economics, opinion polls, market research, social studies, scientific experiments, and more?
    - Software packages for testing and manipulating data?
    - A way of providing clients with consulting reports about trends and analysis?
- Answer: All of the above!

sta130-2

## Statistical Skills

- So, what skills are needed to be a good statistician?
  - Mathematical abilities: equations, sums, integrals, ...?
  - Computer abilities: software, programming, ...?
  - Communication abilities: oral presentations, written reports, group discussions, ...?
- Answer: All of the above!
- STA 130 is a brand new course.
  - Normally, statistics studies begin in second year, with STA 220 or 247 or 257.
  - So why a new first-year course?
  - Is it for ...

sta130-3

- A general introduction to the field of statistics?
- A demonstration of the importance of statistics, to recruit strong science students to study more statistics?
- To teach students to use statistical computer software?
- To help students express their statistical ideas well in spoken English, to help them meet clients, actively participate in class discussions, ask questions in class, etc.?
- To teach students to describe their statistical ideas well in written English, to help them write professional reports, clearly explain their solutions on homework assignments, etc.?
- To mentor students so they can meet other statistics students, participate more in university activities, take better advantage of university resources, etc.?

- Answer: All of the above!!

sta130-4

## And Now the Details

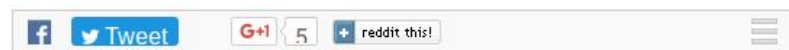
- Course info (evaluation, etc.): [www.probability.ca/sta130](http://www.probability.ca/sta130)
- 

- So, how should we learn about statistics?
- Well, statistics are everywhere.
- What's in the news?
- For example . . .

sta130-5

### Justin Trudeau enjoying a heartfelt political honeymoon, poll suggests

Prime Minister Justin Trudeau is getting a lot of love from voters during his political honeymoon, a new poll suggests.



CHRIS WATTIE / REUTERS FILE PHOTO

Prime Minister Justin Trudeau is enjoying a 60 per cent approval rating, according to a new Forum Research poll, and almost three-quarters of those surveyed say they are satisfied with the outcome of the election.

By: **Robert Benzie** Queen's Park Bureau Chief, Published on Wed Nov 11

sta130-6

- November 11, 2015: “Justin Trudeau enjoying a heartfelt political honeymoon, poll suggests”.
  - “Prime Minister Justin Trudeau’s approval has soared to 6-in-10 (60%)”.
  - “If an election were held today, the Liberals would take well more than half the vote (55%), to one quarter for the Conservatives (25%) and just more than one tenth for the NDP (12%).”
- Poll details (<http://poll.forumresearch.com/post/2424/gender-balanced-cabinet-very-popular>):
  - “a random sampling of public opinion taken by the Forum Poll among 1256 Canadian voters”.
  - “Results based on the total sample are considered accurate  $\pm 3\%$ , 19 times out of 20.”
  - What does this mean?? (Coming soon!)

sta130-7

### Experiment: Do You have ESP?

- I have two different cards.
- Suppose you try to “guess” which card it is.
- If you guess correctly once, does that show you have ESP?
  - What about twice? three times? four times?
- Let’s do an experiment!
  - I will hold up a 2 or a 4, and “concentrate” on it.
  - Then you guess, by raising 2 or 4 fingers. Okay?
  - How many times did you get it right?
  - What does that show?
- What about the TAs? Do they have ESP? Let’s check!
- What conclusions can we draw? (Coming soon!)

sta130-8

## Flu vaccine helps reduce hospitalizations due to influenza pneumonia: study

by Paul Govern | Tuesday, Oct. 6, 2015, 8:50 AM



The study team includes (front row, from left) Kathryn Edwards, M.D., Carlos Grijalva, M.D., MPH, Marie Griffin, M.D., MPH, (back row, from left) Yuwei Zhu, M.D., M.S., Derek Williams, M.D., MPH, and Wesley Self, M.D., MPH. (photo by Susan Urmey)

More than half of hospitalizations due to influenza pneumonia could be prevented by influenza vaccination, according to a study led by investigators at Vanderbilt University Medical Center, published this week in the *Journal of the American Medical Association*.

"We estimated that about 57 percent of influenza-related pneumonia hospitalization could be prevented through influenza vaccination," said Carlos Grijalva, M.D., MPH, associate professor of Health Policy.

sta130-9

- "Flu vaccine helps reduce hospitalizations due to influenza pneumonia: study".
  - "We estimated that about 57 percent of influenza-related pneumonia hospitalization could be prevented through influenza vaccination."
  - Details: *Journal of the American Medical Association* 314(14): 1488-1497, October 2015.
  - Studied 2,767 pneumonia patients at 4 U.S. hospitals.
  - Of the 2,605 non-influenza-related pneumonia patients, 766 (29%) had gotten the flu vaccine.
  - Of the 162 influenza-related pneumonia patients, 28 (17%) had gotten the flu vaccine.
  - Seems (?) to suggest that getting the flu vaccine reduces your chance of getting hospitalised for pneumonia.

sta130-10

- Recall: 766 of 2,605 non-influenza-related patients (29%), and 28 of 162 influenza-related patients (17%), had gotten the flu vaccine.
- Does this show that getting the flu vaccine reduces your chance of getting hospitalised for pneumonia?
  - Or was it just random luck??
- In their table, they say “P Value < .001”.
  - That’s supposed to make the result “statistically significant”.
  - What does this mean??
  - What is a “P-value”??
  - Does it show anything??
- Coming soon!

sta130–11

### Experiment: Two-Headed Coin?

- I have one regular coin, and one two-headed coin. [show]
- I will choose one of the two coins, and flip it.
  - I will tell you (honestly) if it’s heads or tails. [TA?]
- Question: Did I choose the two-headed coin?
  - If the flip is tails: No!
  - If the flip is heads: Maybe!
- Flip it again? How many times?
- Suppose I get Heads every time.
  - Does that show I’m using the two-headed coin?
  - After one head? two heads? three? five? ten?
  - For sure?? How to decide?

sta130–12

## Statistical Approach: P-values

- In any experiment, the P-value is the probability that we would have gotten such an extreme result by pure chance alone, i.e. if there is no actual effect or difference, i.e. under the “null hypothesis” that there is no underlying cause.
- In the coin experiment, the P-value is the probability that we would have gotten so many heads in a row, under the “null hypothesis” that it’s just a regular coin (not two-headed).
  - After one head:  $P\text{-value} = 1/2 = 50\%$ .
  - After two heads:  $P\text{-value} = 1/2 \times 1/2 = 1/4 = 25\%$ .
  - (We multiply the individual probabilities because the coins are independent, i.e. the first head doesn’t affect the probabilities of the second head. So, it’s “one half of one half”, i.e.  $1/2 \times 1/2$ .)
  - After three heads:  $P\text{-value} = 1/2 \times 1/2 \times 1/2 = 1/8 = 12.5\%$ .

sta130-13

## P-values for Coin Experiment (cont’d)

- Recall: P-value after one head = 50%. After two heads = 25%. After three heads = 12.5%.
- After five heads:  $P\text{-value} = 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = 1/32 \doteq 3.1\%$ .
- After ten heads:  $P\text{-value} = 1/2 \times 1/2 \times \dots \times 1/2 = (1/2)^{10} = 1/1,024 \doteq 0.001 = 0.1\%$ . Small!
- The stronger the evidence (that it’s a two-headed coin, i.e. that the null hypothesis is false), the smaller the P-value.
- When is it “statistically significant”, i.e. when does it “show” that the coin is two-headed? When the P-value is ... small enough?
- How small? Usual cut-off: 5%, i.e. 0.05, i.e.  $1/20$ .
  - (So, above, 5 heads suffices to “show” it’s two-headed.)
  - Good choice of cut-off? Should it be smaller?

sta130-14

## Back to ESP Example

- Suppose you get it right five times in a row, on your first try.
- Then,  $P\text{-value} = 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = 1/32 \doteq 3.1\%$ .  
(Just like for the coin.)
- Does that “show” you have ESP?
- What if you try many times, before finally getting five in a row?  
Does it still “count”?
- What if many people try, and one of them gets five in a row.  
Does that show they have ESP?
- What could we do to test this further?
  - (More about this in tutorial on Jan 20 ... ?)

sta130–15

## ESP Example (cont'd)

- Suppose you don't guess the right card every time, just most of the time. Does that still show you have ESP?
- e.g. suppose you guess 9 right out of 10.
- What is the P-value now?
- Well, the P-value is the probability of “such an extreme result”.
- Here “such an extreme result” means: 9 or 10 guesses correct, out of 10 total.
- Then the P-value is the probability of that happening, under the “null hypothesis” of just random guesses (i.e., of no ESP).
- Need to compute the probability!

sta130–16



## ESP Example: Computing Probabilities

- Write “C” if guess is correct, or “I” if incorrect.
- Then the probability of getting all 10 guesses correct is  $P(\text{CCCCCCCCCC}) = (1/2)^{10} = 1/1,024 \doteq 0.001 = 0.1\%$ .
- What about the probability of getting exactly 9 correct?
- Well,  $P(\text{CCCCCCCCCI}) = (1/2)^{10} = 1/1,024$ .
- But also  $P(\text{ICCCCCCCCC}) = (1/2)^{10} = 1/1,024$ .
- And  $P(\text{CICCCCCCCC}) = (1/2)^{10} = 1/1,024$ .
- Ten different “positions” for the “I”, each have probability  $1/1,024$ . So,  $P(\text{get 9 correct}) = \text{Total} = 10/1,024$ .
- So,  $P\text{-Value} = P(9 \text{ or } 10 \text{ correct}) = P(9 \text{ correct}) + P(10 \text{ correct}) = 10/1,024 + 1/1,024 = 11/1,024 \doteq 0.0107$ .
- Less than 0.05, so still statistically significant!

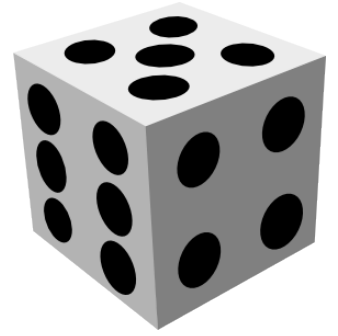
sta130-17

## ESP Example: More Probabilities

- Great, but what if you guess 8 out of 10? Significant?
  - Or, what about just 6 out of 10? Significant?
  - Or, 17 out of 20? Or 69 out of 100?
- Can we compute all these P-value probabilities?
  - Well, there is a formula for each individual probability: the binomial distribution.
  - But adding them all up (to get a P-value) is hard.
- Alternative: find a pattern and approximation.
  - (Coming soon.)

sta130-18

## P-Value Practice: Too Many Twos



- An ordinary six-sided die should be equally likely to show any of the numbers 1, 2, 3, 4, 5, or 6.
- Suppose a casino operator complains that his six-sided die isn't fair: it comes up "2" too often!
- To test this, you roll the die 60 times, and get "2" on 17 of the rolls.
  - What can we conclude from this?
- Here the "null hypothesis" is: the die is fair, i.e. it is equally likely to show 1, 2, 3, 4, 5, or 6.
- And, the P-Value is: the probability, under the null hypothesis (i.e. if the die is fair), of getting such an extreme result, i.e. of getting "2" on 17 or more rolls if you roll the die 60 times.

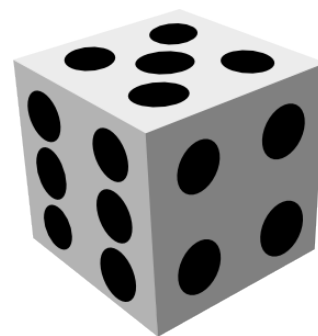
sta130-19

## P-Value Practice: Too Many Twos (cont'd)

- So, the P-Value is the probability of getting "2" on 17 or 18 or 19 or 20 or ... or 60 of the 60 rolls.
  - This can be calculated on a computer.
  - e.g. in R: `pbinom(16, 60, 1/6, lower.tail=FALSE)` (later)
  - Or it can be approximated (coming soon).
- This P-Value turns out to equal: 0.01644573
- This P-Value is less than 0.05. So, by the "usual standard", this "shows" (or "demonstrates" or "establishes" or "confirms") that the casino operator was correct: the die comes up "2" too often!
  - Do you agree?

sta130-20

## P-Value Practice: Not Enough Fives



- Suppose the casino operator comes back later, and complains that his new six-sided die isn't fair: it doesn't come up "5" often enough!
- To test this, you roll the new die 90 times, and get "5" on 12 of the rolls.
- Here the "null hypothesis" is: the die is fair, i.e. it is equally likely to show 1, 2, 3, 4, 5, or 6. (Same as before.)
- And, the P-Value is: the probability, under the null hypothesis (i.e. if the die is fair), of getting such an extreme result, i.e. of getting "5" on 12 or fewer rolls (i.e., on 12 or 11 or 10 or 9 or 8 or ... or 1 or 0 rolls) if you roll the die 90 times.
  - So what does it equal?

sta130-21

## P-Value Practice: Not Enough Fives (Cont'd)

- This P-Value turns out to equal: 0.2448148
- This P-Value is more than 0.05.
- So, by the "usual standard", this does not show that the casino operator was correct, i.e. the die does not come up "5" insufficiently often, i.e. there is no evidence that the die is not fair, i.e. as far as we can tell, the die is fair.
- So the casino operator was wrong!
  - (How do we tell him?)
- Let's do one more ...

sta130-22

## P-Value Practice: Roulette Wheel

- A standard (“American-style”) roulette wheel has 38 different numbers, which should all be equally likely.
- Suppose a casino operator complains that it is showing “22” too often.
- To test this, you spin the roulette wheel 200 times, and get “22” on 11 of the spins.
- In this example:
  - What is the null hypothesis?
  - How should the P-Value be computed?
  - What does the P-Value equal?
  - Does this show that the casino operator was correct?
- Try it yourself! (In fact, the P-Value is: 0.007164461)



sta130–23

## P-Value Application: New Medical Drug

- Suppose a serious disease kills 60% of all the patients who get it.
- Then a pharmaceutical company invents a new drug, to try to help patients with the disease.
- They test the drug on 12 patients with the disease. Only 1 of those patients dies; the other 11 live.
- Does this show that their drug is effective?
- Here the null hypothesis is: The drug has no effect, i.e. even with the drug, 60% of all patients will die (and 40% will live).
- Then the P-Value is the probability, under the null hypothesis, i.e. if the drug has no effect, that 11 or more out of 12 patients will live, i.e. that 11 or 12 out of 12 patients will live.
- What is this probability?

sta130–24

## P-Value Application: New Medical Drug (cont'd)

- Well, the probability that all 12 patients live is:  $(0.40)^{12}$ .
- And (similar to the two headed coin example), the probability that the first 11 patients live but the 12th patient dies is  $(0.40)^{11} \times (0.60)$ .
- And (again similar to the two headed coin example), there are 12 different patients each of whom could have been the one who died. So, the total probability that exactly 11 out of 12 patients live is:  $12 \times (0.40)^{11} \times (0.60)$ .
- So, here the P-Value is the probability that all 12 live, plus the probability that exactly 11 live, i.e. it is equal to:  $(0.40)^{12} + 12 \times (0.40)^{11} \times (0.60) = 0.0003187671$ .
- This P-Value is much less than the 0.05 cutoff.
- So, yes, give this drug to all patients right away!

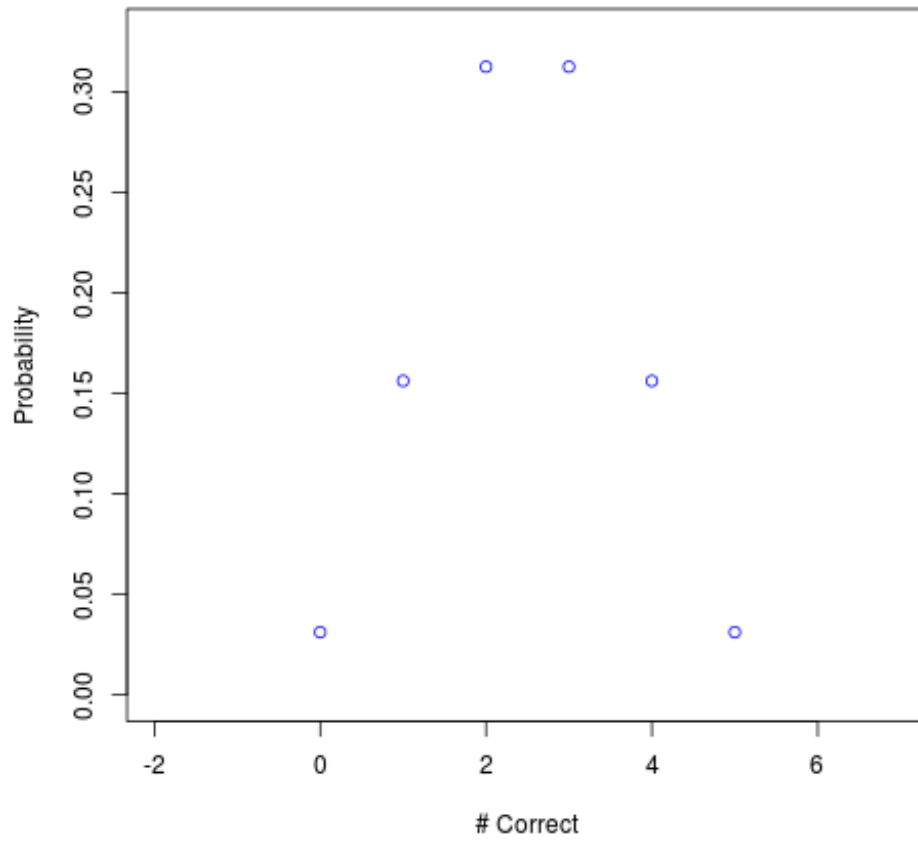
sta130–25

## P-Values and Probabilities

- In all of these examples, we need to compute a P-Value.
- This P-Value is the probability of “such an extreme result”, under the null hypothesis of “no effect”.
- It might involve adding up lots of individual probabilities.
- R can help with this. But sometimes it gets messy.
- Alternative: find a pattern and approximation.
- Consider the ESP experiment (guessing between two cards).
- If you get 69 correct out of 100, does this show anything?
- Let's begin by graphing some probabilities.
  - (I used R; see [www.probability.ca/sta130/Rbinom](http://www.probability.ca/sta130/Rbinom) )

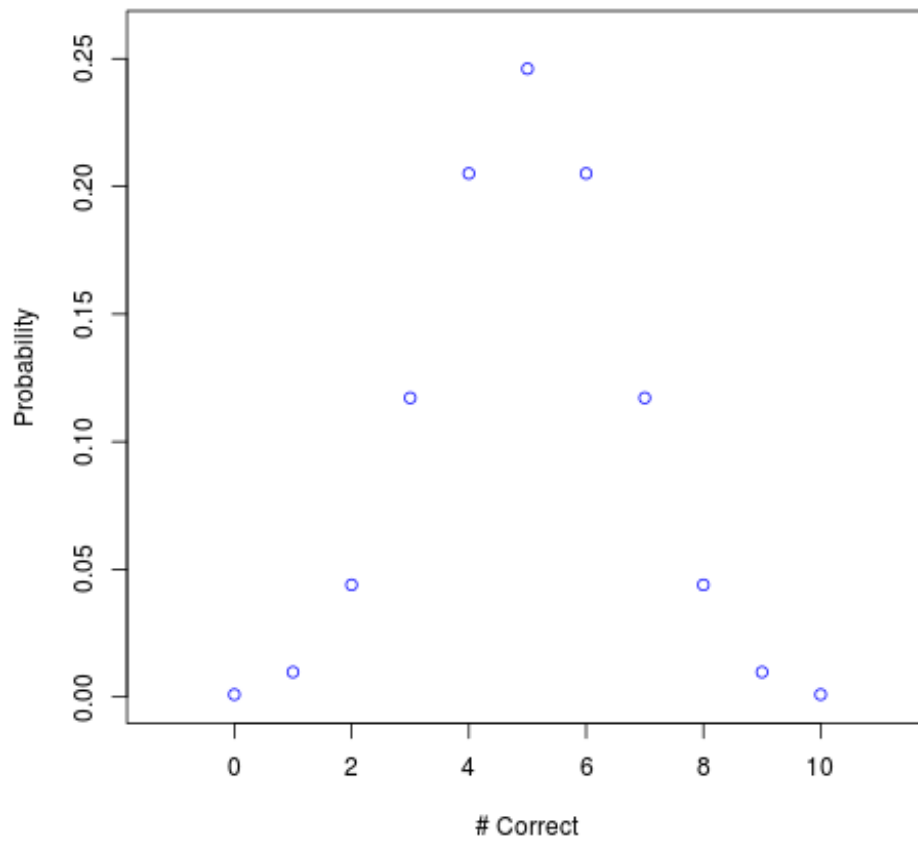
sta130–26

**ESP: Probabilities For 5 Guesses**



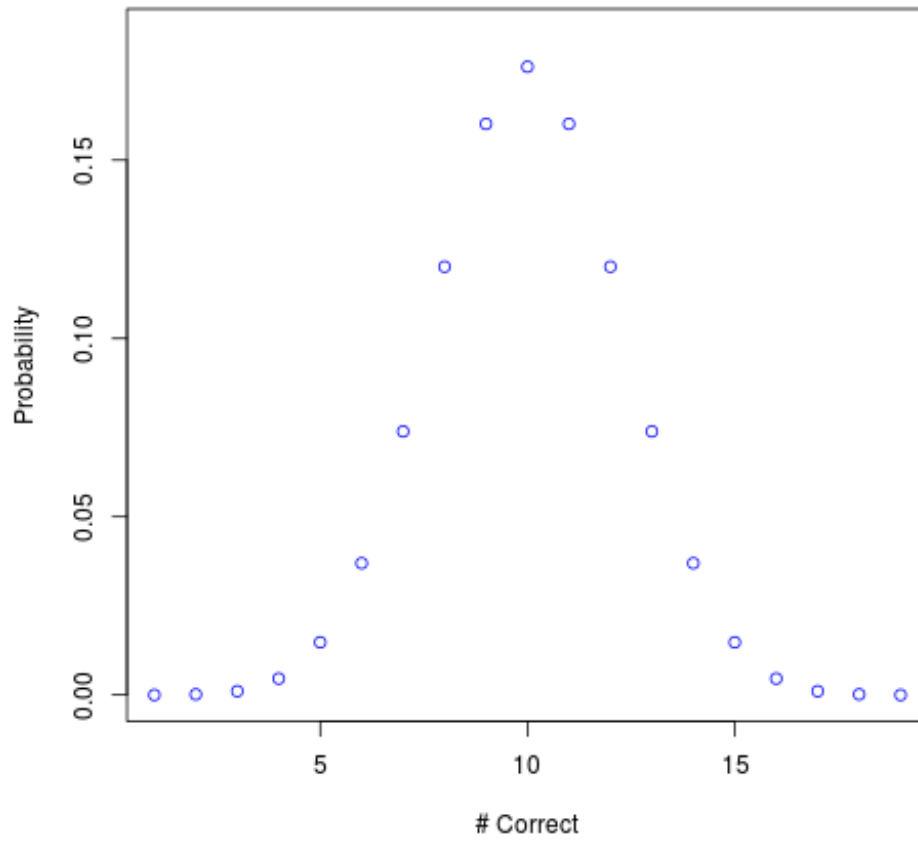
sta130-27

**ESP: Probabilities For 10 Guesses**



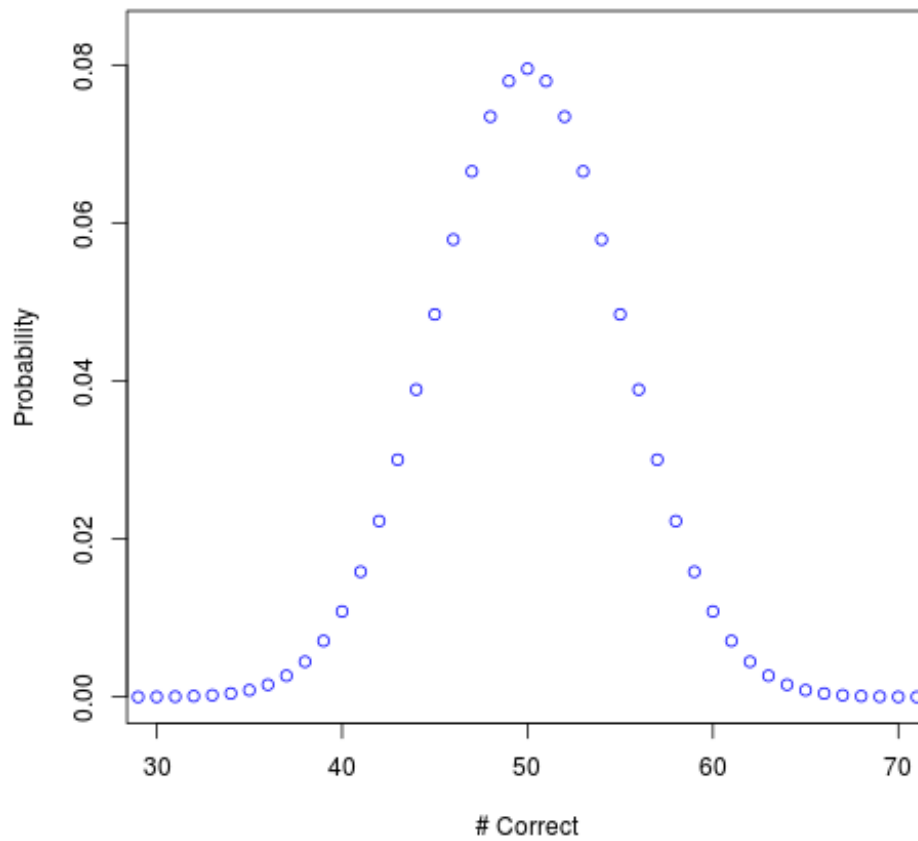
sta130-28

ESP: Probabilities For 20 Guesses



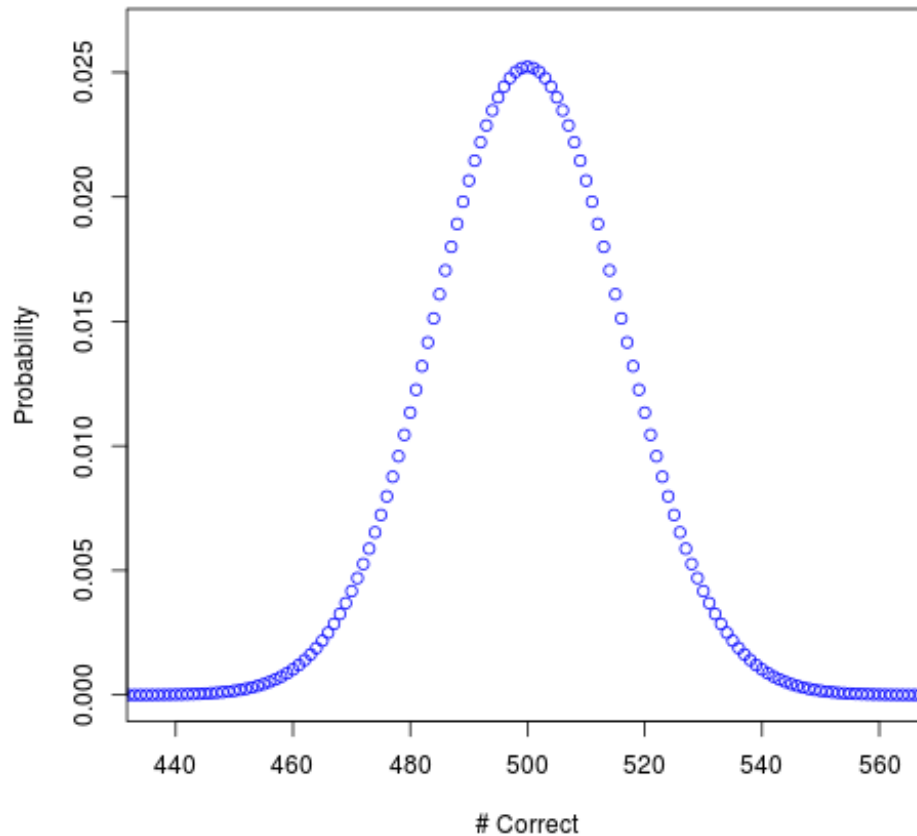
sta130-29

ESP: Probabilities For 100 Guesses



sta130-30

## ESP: Probabilities For 1000 Guesses



sta130-31

### ESP Probabilities: Conclusion?

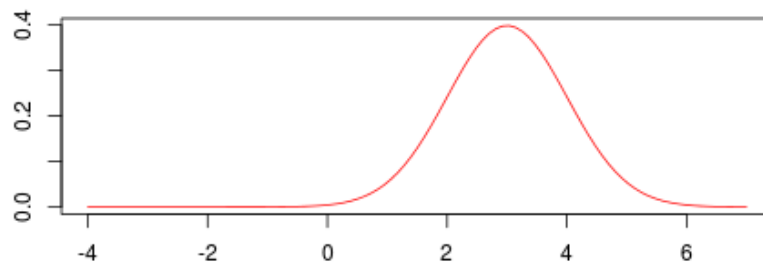
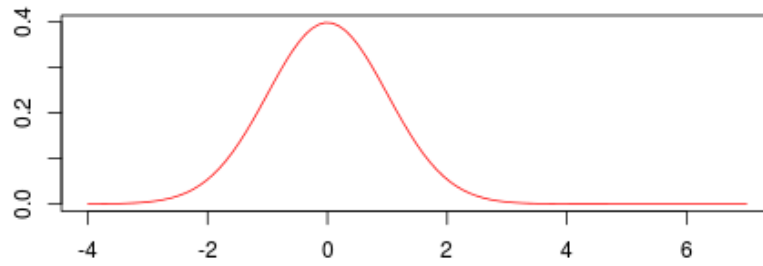
- The various probabilities seem to be getting close to a certain specific “shape”. Namely, the ... “normal distribution”.
  - Also called the “Gaussian distribution”, or “bell curve”.
- This turns out to be true very generally!
  - Whenever lots of identical experiments are repeated (like ESP tests), the probabilities for the number correct, or the average, or the sum, get closer and closer to a normal distribution!
    - This fact is the “central limit theorem”.
    - It guides many statistical practices. Important!
- So, we’d better learn more about normal distributions.

sta130-32



## Normal Distributions

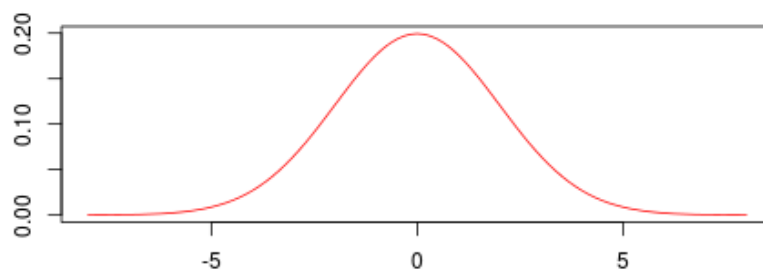
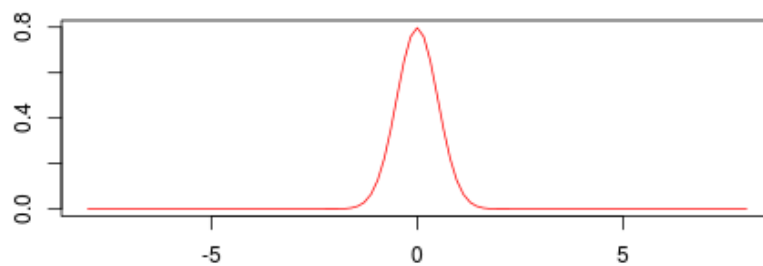
Normal distributions can have different “locations”:



sta130-33

## Normal Distributions (cont'd)

Normal distributions can also have different “widths”:



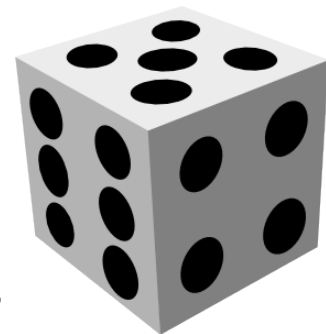
sta130-34

- So which normal distribution do we want?
  - Which location? Which width?
- Well, the location should match up with the “mean” or “expected value” or “average value” of our quantity.
- And, the width should match up with the “standard deviation” of our quantity. (Later.)
- Let’s start with location.
- Suppose you repeat the ESP experiment 100 times, each time guessing one of two cards.
- And suppose you have no ESP powers, so the probability of being correct each time is 1/2.
- What is the “mean” number of correct guesses?
  - Well, it must be 50. But why?

sta130–35

### Working With Expected Values (Means)

• Example: Suppose you roll a six-sided die (which is equally likely to show 1, 2, 3, 4, 5, or 6). What is the expected (average) value you will see?



• FACT: The expected value of any “discrete” random quantity can be found by adding up all the possible values, each multiplied by their probability:  $E(X) = \sum_x x P(X = x)$ .

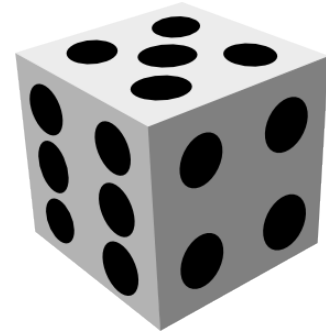
• For the die, mean =  $1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 + 6 \times 1/6 = 21/6 = 3.5$ . (Not 3!)

• Suppose I will pay you \$30 if it lands on 1 or 2, and I will pay you \$60 if it lands on 3 or 4 or 5 or 6. Then what is the expected (average) value that you will receive?

– The expected value is:  
 $\$30 \times (2/6) + \$60 \times (4/6) = \$10 + \$40 = \$50$ .

sta130–36

- Suppose instead I will pay you \$30 if it lands on 1 or 2, but you have to pay me \$60 if it lands on 3 or 4 or 5 or 6. Then what is the expected value of your net gain?



- The expected value is:  

$$\$30 \times (2/6) + (-\$60) \times (4/6) = \$10 - \$40 = -\$30.$$

- Example: Suppose that when you take a free throw in basketball, you score 40% of the time. Suppose you bet \$20 that you will score your next free throw (so you will win \$20 if you score, or lose \$20 if you don't score).



- Then your expected net gain is:  

$$\$20 \times (40\%) - \$20 \times (60\%) = \$8 - \$12 = -\$4,$$
 i.e. you lose \$4 on average.

sta130-37

### Back to ESP Example

- Suppose you guess 100 times, and have no ESP power.
- What is your expected number of correct guesses?
- Let  $X_1$  be a random quantity, which equals 1 if you guess the first card correctly, otherwise it equals 0.
- Then since you have no ESP (null hypothesis),  $P(X_1 = 1) = 1/2$ , and  $P(X_1 = 0) = 1/2$ .
- So what is  $E(X_1)$ , the expected value of  $X_1$ ?
  - It must be 1/2, right?
- Well, in this case,  $E(X_1) = 1 \times P(X_1 = 1) + 0 \times P(X_1 = 0) = 1 \times 1/2 + 0 \times 1/2 = 1/2$ . (Of course.)
- How does this help?

sta130-38

- Recall:  $X_1 = 1$  if you guess the first card correctly, otherwise  $X_1 = 0$ . And, if no ESP (null hypothesis),  $E(X_1) = 1/2$ .

- Similarly, let  $X_2 = 1$  if you guess the second card correctly, otherwise  $X_2 = 0$ .

- Then, under the null hypothesis,  $P(X_2 = 1) = 1/2$ , and  $P(X_2 = 0) = 1/2$ , same as for  $X_1$ .

- And, again,  $E(X_2) = 1/2$  too (same calculation).

- Similarly can define  $X_3, X_4, \dots, X_{100}$ .

- In terms of these  $X_i$  random variables, your total number  $T$  of correct guesses is equal to the sum:  $T = X_1 + X_2 + \dots + X_{100}$ .

- What is the expected value of this complicated sum?

- Well ...

sta130–39

- FACT: The expected value of any sum of random quantities, is equal to the sum of the individual expected values. (“linear”)

- So,  $E(X_1 + X_2 + \dots + X_{100}) = E(X_1) + E(X_2) + \dots + E(X_{100}) = (1/2) + (1/2) + \dots + (1/2) = 100 \times (1/2) = 50$ .

- That is,  $E(T) = 50$ .

- So, the normal distribution that approximates the probabilities when guessing at ESP 100 times, has location or mean equal to 50. (Of course.)

- (Or, if we guessed  $n$  times instead of 100 times, then the mean would be:  $n \times (1/2) = n/2$ .)

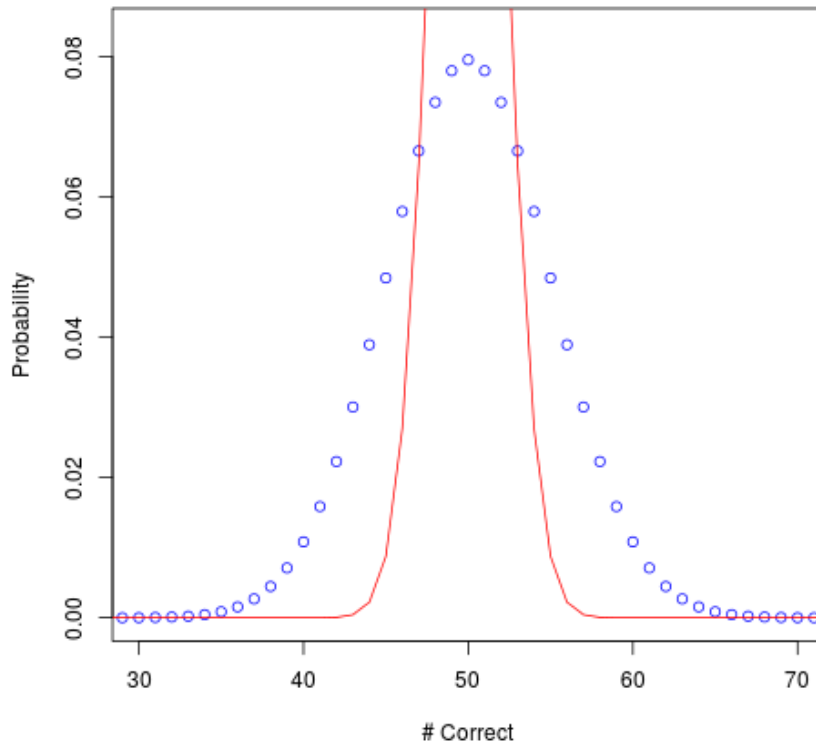
- But what width? Too thin? Too fat? Just right?

- Look at some pictures, or try it ourselves!

[www.probability.ca/sta130/Rnormtry](http://www.probability.ca/sta130/Rnormtry)

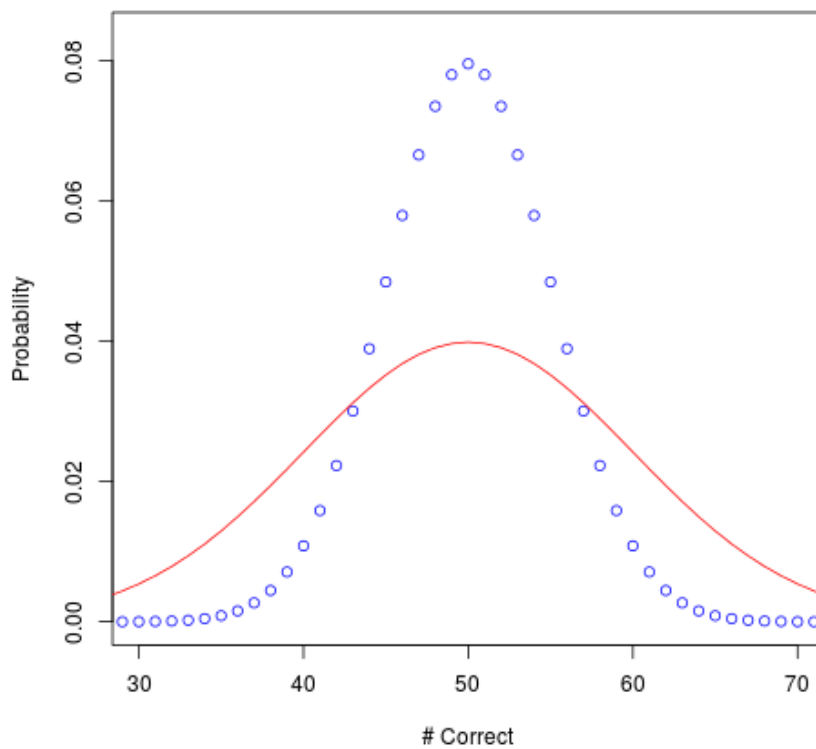
sta130–40

## Normal Fit: Too Thin



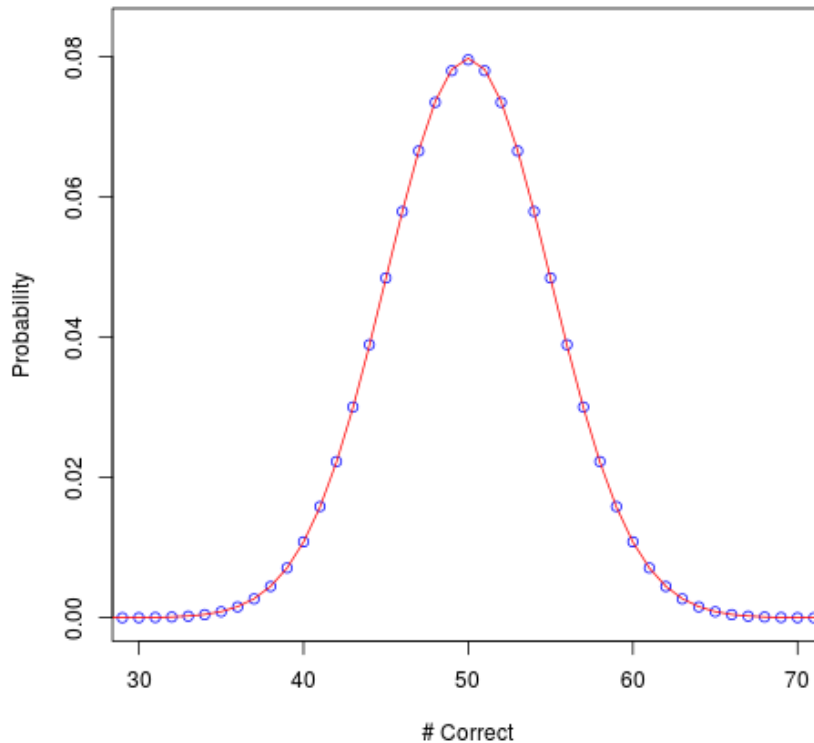
sta130-41

## Normal Fit: Too Fat



sta130-42

## Normal Fit: Just Right



sta130-43

## Summary So Far

- Statistics combines math, data analysis, computing, oral & written communication, the news, . . . So does this course!
- The “P-value” is the probability of getting such an extreme result (i.e., as extreme as observed or more extreme) just by luck, i.e. under the “null hypothesis” that there is no actual effect (e.g., the coin is fair, there is no ESP, the drug doesn’t work, . . .).
- An observed result is “statistically significant” if the P-value is sufficiently small (usual: less than 0.05), indicating that the result was probably not just by luck.
- So, we need to compute these probabilities! Can approximate them by a normal distribution (bell curve), if we know the correct “mean” (“location”, “expected value”, “average value”) [DONE!] and “width” (related to “variance”) [NEXT!].

sta130-44

## Learning About Width: Variance

- Width is closely related to variance, a form of uncertainty.
- For any random quantity  $X$ , the variance is defined as the expected squared difference from the mean.
- That is, if the mean is  $m = E(X)$ , then the variance is  $v = \text{Var}(X) = E([X - m]^2) = \sum_x (x - m)^2 P(X = x)$ .
- Example: Flip one fair coin, and  $X = 1$  if it's heads, or  $X = 0$  if it's tails.
  - Then mean is  $m = E(X) = 1 \times 1/2 + 0 \times 1/2 = 1/2$ .
  - Then variance is  $v = \text{Var}(X) = E([X - (1/2)]^2) = [1 - (1/2)]^2 \times 1/2 + [0 - (1/2)]^2 \times 1/2 = [1/2]^2 \times 1/2 + [-1/2]^2 \times 1/2 = 1/4 \times 1/2 + 1/4 \times 1/2 = 1/8 + 1/8 = 2/8 = 1/4$ .
  - Got it?

sta130-45

## Variance Example

- Example: Roll a fair six-sided die.
- We know  $m = 3.5$ .
- So,  $v = E([X - 3.5]^2) = [1 - 3.5]^2 \times 1/6 + [2 - 3.5]^2 \times 1/6 + [3 - 3.5]^2 \times 1/6 + [4 - 3.5]^2 \times 1/6 + [5 - 3.5]^2 \times 1/6 + [6 - 3.5]^2 \times 1/6 = [-2.5]^2 \times 1/6 + [-1.5]^2 \times 1/6 + [-0.5]^2 \times 1/6 + [0.5]^2 \times 1/6 + [1.5]^2 \times 1/6 + [2.5]^2 \times 1/6 = 17.5/6 \doteq 2.917$ .
- (Phew!)
- Gets pretty messy, pretty quickly.
- But very useful!
- And, has nice properties!



sta130-46

## Variance for ESP Example

- Suppose you guess 100 times, and have no ESP power.
- Then the number correct is  $T = X_1 + X_2 + \dots + X_{100}$ , where  $X_i = 1$  if your  $i^{\text{th}}$  guess was correct, otherwise  $X_i = 0$ .
- And we know that  $E(X_i) = 1/2$ , so  $E(T) = E(X_1) + E(X_2) + \dots + E(X_{100}) = 1/2 + 1/2 + \dots + 1/2 = 100/2 = 50$ .
- What about the variance  $\text{Var}(T)$ ? Difficult to compute?
- FACT: If random quantities are independent, i.e. they don't affect each other's probabilities (like separate card guesses), then the variance of a sum is equal to the sum of the variances.
- So,  $\text{Var}(T) = \text{Var}(X_1 + X_2 + \dots + X_{100}) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_{100})$ .
- Does this help? Yes!

sta130-47

## Variance for ESP Example (cont'd)

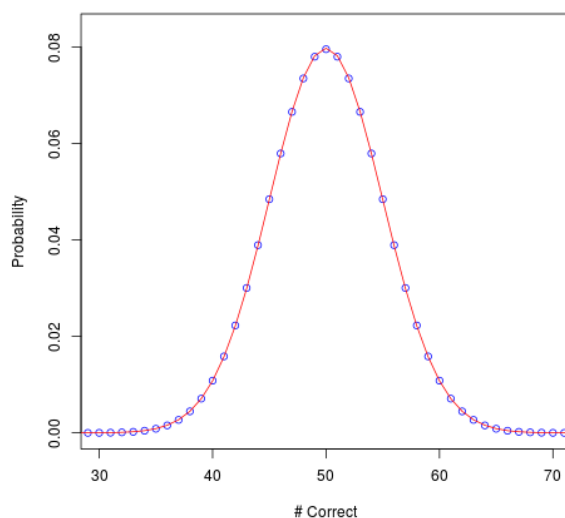
- Know  $\text{Var}(T) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_{100})$ .
- Now,  $X_1$  has  $m = 1/2$ . So,  $\text{Var}(X_1) = E([X_1 - (1/2)]^2) = [1 - (1/2)]^2 \times 1/2 + [0 - (1/2)]^2 \times 1/2 = [-1/2]^2 \times 1/2 + [1/2]^2 \times 1/2 = 1/4 \times 1/2 + 1/4 \times 1/2 = 1/8 + 1/8 = 2/8 = 1/4$ , just like for flipping one coin (before). Similarly  $\text{Var}(X_2) = 1/4$ , etc.
- So,  $\text{Var}(T) = 1/4 + 1/4 + \dots + 1/4 = 100 \times 1/4 = 25$ .
- (Or, if we made  $n$  guesses instead of 100, then we would have  $\text{Var}(T) = n \times 1/4 = n/4$ .)
- Summary: If you make 100 guesses, each having probability  $1/2$  of success, then the total number of correct guesses has mean=50, and variance=25.
- Or, with  $n$  guesses, mean= $n/2$ , and variance= $n/4$ .
- So now we know variance! Can it help us? Yes! (soon)

sta130-48



## Taking the Square Root: Standard Deviation

- Since variance is defined using “squares”, for width we take the square root of it instead, called the standard deviation (sd or s).
- So, with the ESP example, if you guess 100 cards, then mean = 50, and variance = 25, so standard deviation =  $\sqrt{25} = 5$ .
- Normal distribution, with mean=50 and sd=5, gives a great fit!



sta130-49

## Variance Practice: Number of Fives

- Let  $T$  be the number of times a “5” appears when we roll a fair six-sided die 12 times. What are  $E(T)$  and  $\text{Var}(T)$  and  $sd(T)$ ?
- Let  $X_i = 1$  if the  $i^{\text{th}}$  roll is a 5, otherwise  $X_i = 0$ .
- Then  $E(X_i) = 1 \times 1/6 + 0 \times 5/6 = 1/6$ .
- And,  $\text{Var}(X_i) = [1 - 1/6]^2 \times 1/6 + [0 - 1/6]^2 \times 5/6 = [25/36] \times 1/6 + [1/36] \times 5/6 = 30/216 = 5/36$ .
- Then,  $T = X_1 + X_2 + \dots + X_{12}$ .
  - So, mean =  $E(T) = E(X_1) + E(X_2) + \dots + E(X_{12}) = 1/6 + 1/6 + \dots + 1/6 = 12 \times 1/6 = 2$ .
  - And,  $\text{Var}(T) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_{12}) = 5/36 + 5/36 + \dots + 5/36 = 12 \times 5/36 = 5/3$ .
  - And, width =  $sd(T) = \sqrt{\text{Var}(T)} = \sqrt{5/3} \doteq 1.29$ . Phew!



sta130-50

## Back to ESP Example, Again

- In the ESP example, suppose you correctly guess 69 out of 100.
  - Does this show you have ESP?
- We need to compute the P-Value, i.e. the probability that you would get 69 or more correct (under the null hypothesis, i.e. no ESP, i.e. just guessing randomly, with probability 1/2 each time).
- (This is the same as the probability, if you flip 100 fair coins, that you will get 69 or more heads.)
- What is the probability? Guesses? Try a simulation? [Rcoins]
- But how to actually compute the probability?
- Use the normal approximation!
  - As follows ...

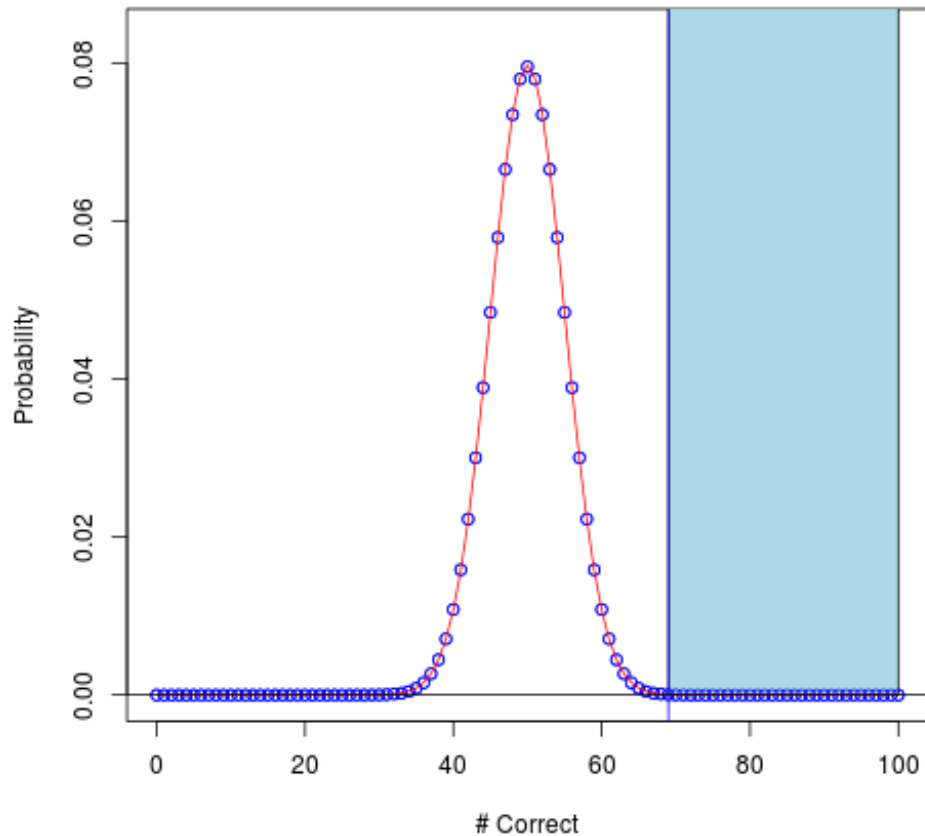
sta130-51

## Normal Approximation for ESP Example

- We know that the number of correct guesses has mean = 50, and variance = 25, so width = sd = 5.
- So, our P-value is approximately the same as the probability that a normal distribution, with mean = 50 and sd = 5, is more than 69.
- Can compute this in R (see `?pnorm`):  
`pnorm(69, 50, 5, lower.tail=FALSE)`
- Answer is about 0.0000723.
  - Much less than 0.05.
  - So yes, 69 correct would indicate ESP abilities (or cheating).
  - Or, to put it differently, I don't think you'll get 69 correct!
  - Go ahead and try! (Wanna bet?)

sta130-52

### ESP Example: Diagram



sta130-53

### PM Trudeau: Still Popular? (Friday Headline)

More than half of Canadians approve of Trudeau: poll

**CHRIS HANNAY**

Ottawa — The Globe and Mail

Published Friday, Jan. 15, 2016 8:34AM EST

> The Justin Trudeau honeymoon continues: an **Abacus Data poll** this morning shows 53 per cent of respondents give a thumbs-up to the Liberal government, and even more rate the Prime Minister positively.

sta130-54

## “53%”? Really “More Than Half”? Just Luck??

- Poll details at: <http://abacusdata.ca/how-do-we-feel-about-the-trudeau-government/>
  - (January 15, 2016) “Today, 53% say they approve of the performance of the federal government”.
  - “Our survey was conducted online with 1,500 Canadians aged 18 and over from January 8 to 12, 2016”.
  - (Also: “The margin of error for a comparable probability-based random sample of the same size is  $\pm 2.6\%$ , 19 times out of 20.” Huh? Later!)
- So does this demonstrate that “more than half” approve?
- Or, could it be that half or less approve, but just by luck they polled more people who approved than who didn't?
- To consider this properly, need ... a P-value! How?

sta130–55

## Statistical Analysis of Polls

- How to model a poll in terms of probability? What null hypothesis? Etc. Well ...
- Suppose that some (unknown) fraction  $p$  of Canadians approve of the government (or, at least, would say that they do).
- The poll selects Canadians uniformly at random (hopefully! online??). It then asks each one of them if they approve (Yes/No).
- Can consider this as repeating a random “experiment” each time: select a Canadian at random, and ask them if they approve (Yes/No). Then repeat many times (e.g. 1,500).
- Kind of like repeatedly flipping a coin, or repeatedly guessing cards in the ESP experiment, or repeatedly rolling a fair six-sided die. Main difference: this time we don't know  $p$ . (Before:  $p = 1/2$ , or  $p = 1/6$ , or ...)

sta130–56

## Statistical Analysis of the Trudeau Poll

- For the Trudeau poll, one possible null hypothesis: does  $p = 50\%$ ? Versus alternative hypothesis: is  $p > 50\%$ ?
- So, here we define the P-value as the probability, assuming  $p = 50\%$ , that if we sample 1,500 people and ask each one of they approve, that 53% or more will answer Yes.
- Now, 53% of 1,500 is: 795.
- So, the question is, if you repeatedly conduct an experiment, which has probability  $p = 50\%$  of getting Yes each time, a total of 1,500 times, then what is the probability that you will get Yes 795 or 796 or 797 or 798 or 799 or 800 or 801 or ... or 1,500 times.
- (Phew!) How to compute?
- Use normal approximation! What mean? What variance or sd?

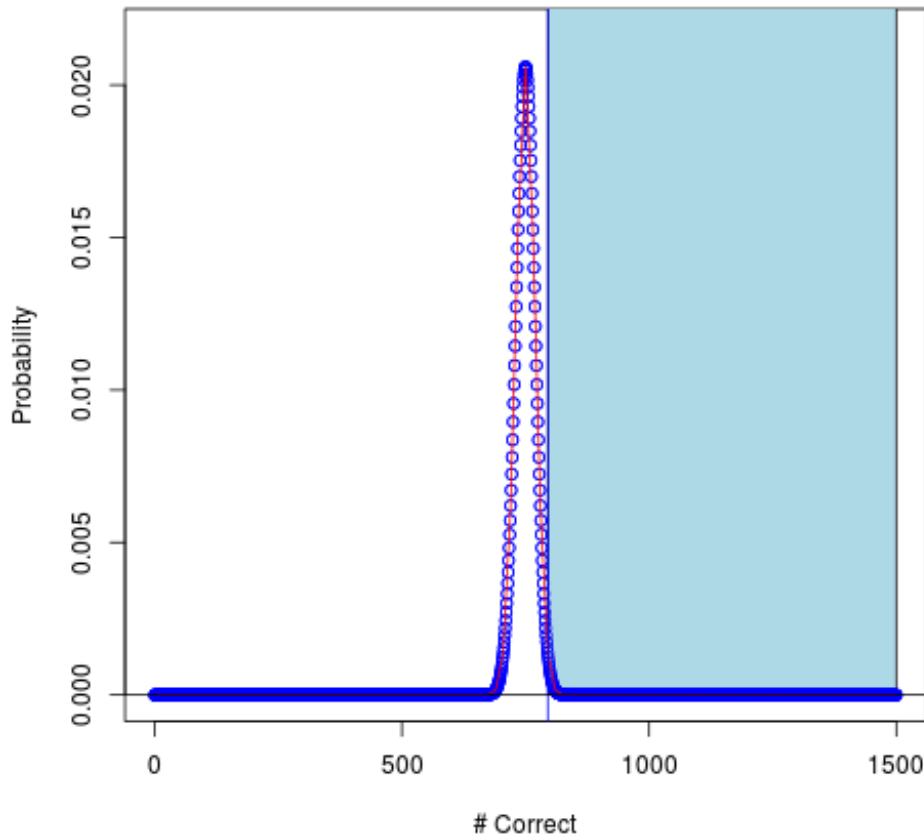
sta130-57

## P-value for the Trudeau Poll

- Recall: out of 1,500 Canadians, 795 approved.
- Well, we know that if you do an experiment  $n = 1,500$  times, with probability  $1/2$  of success each time (like flipping a fair coin, or guessing between two cards), then mean =  $n/2 = 1,500/2 = 750$ , and variance =  $n/4 = 1,500/4 = 375$ . So,  $sd = \sqrt{375} \doteq 19.36$ .
- So, the P-value is approximately the probability that a normal distribution with mean=750 and variance=375 (i.e., with width =  $sd = \sqrt{375}$ ), will equal 795 or more.
- There is an R command for this:  
`pnorm(795, 750, sqrt(375), lower.tail=FALSE)`
- Answer is: 0.01006838. Much less than 0.05. So, this seems to show that  $p > 50\%$ , i.e. that more than half of Canadians would say they approve. (But does it show that  $p > 51\%$ ? Later.)

sta130-58

### Poll Example: Diagram



sta130-59

### New Babies: Boy or Girl?

- When a new baby is born, is it equally likely to be a boy or girl? Or is one more likely than the other?
- Suppose the probability of a new baby being a girl is “ $p$ ”.
- Null hypothesis:  $p = 1/2$ . Alternative hypothesis:  $p \neq 1/2$ . (Not sure if  $p > 1/2$  or  $p < 1/2$ , so include both possibilities: this is called a “two-sided” test.)
- To proceed, we need some ... data!  
<http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/health103a-eng.htm>



— In 2011 (the last year available), of 377,636 total births in Canada, 184,049 were females (girls). (48.74%) Just luck? Or does this “demonstrate” that girls are less likely?

sta130-60

## Boy or Girl? (cont'd)

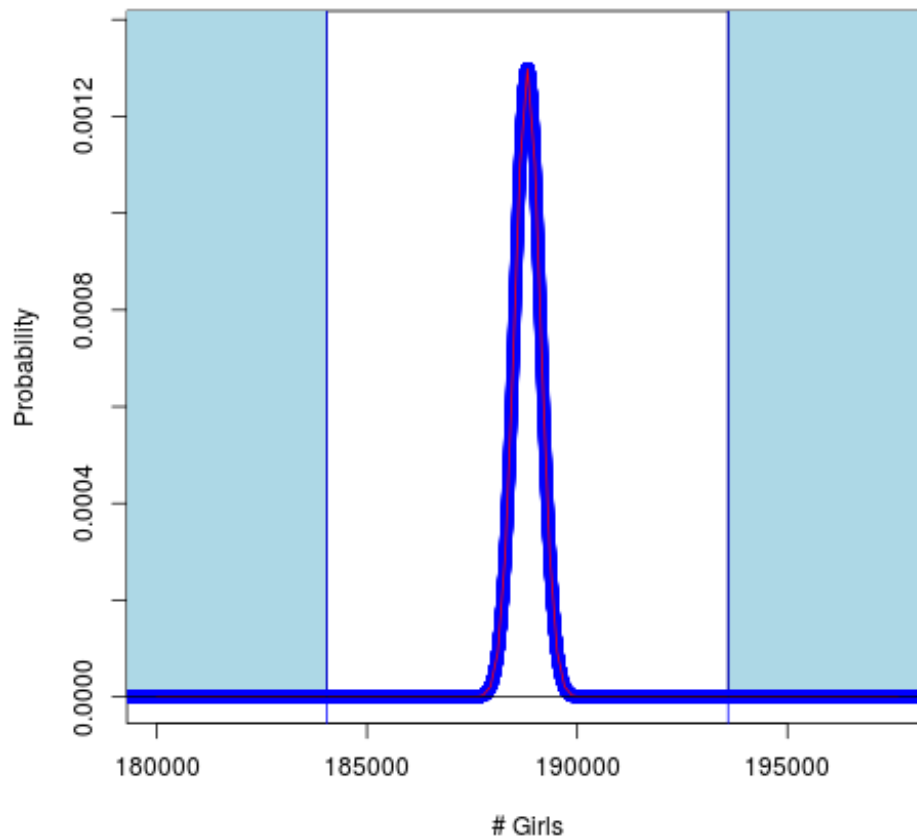
- Recall: of 377,636 total births, 184,049 (48.74%) were females.
- If  $p = 50\%$ , would expect  $50\% \times 377,636 = 188,818$  females. So is 184,049 too few? Or just random luck?
- Here the P-value is the probability, assuming the null hypothesis that  $p = 50\%$ , that out of 377,636 total births, there would be 184,049 or fewer females. Or, to make it two-sided, we should also include the probability of correspondingly too many females.
  - Here  $|\text{observed} - \text{expected}| = |184,049 - 188,818| = 4,769$ .
  - So, correspondingly too many females would be the expected number plus 4,769, i.e.  $188,818 + 4,769 = 193,587$  or more.
  - So, perhaps we should really ask, what is the probability that out of 377,636 total births, the number of females would be either 184,049 or fewer, or 193,587 or more. (This is approximately twice the probability of just 184,049 or fewer females.)

sta130-61

- Recall: of 377,636 total births, 184,049 (48.74%) were females.
- P-value? Well, under the null hypothesis that  $p = 50\%$ , the number of females has mean  $= n/2 = 377,636/2 = 188,818$ , and variance  $= n/4 = 377,636/4 = 94,409$ , so sd  $= \sqrt{94,409} \doteq 307.26$ .
- So, the P-value is approximately the probability that a normal distribution, with mean 188,818, and sd  $\sqrt{94,409}$ , is either 184,049 or fewer, or 193,587 or more.
- R: `pnorm(184049, 188818, sqrt(94409), lower.tail=TRUE) + pnorm(193587, 188818, sqrt(94409), lower.tail=FALSE)`
  - Answer is:  $2.5 \times 10^{-54}$ . Astronomically tiny!!
- So, we REJECT the null hypothesis that  $p = 50\%$ .
- Instead, we conclude that  $p \neq 50\%$ , i.e. boys and girls are not equally likely.

sta130-62

### Babies Example: Diagram



sta130-63

### Boy or Girl: Conclusion

- We've demonstrated: boys and girls are not equally likely.
- The data certainly indicates that: girls are less likely than boys.
  - (However, there are actually more women than men in Canada: 17,147,459 female – 16,869,134 male. Why? Women live longer on average! [Statscan: 83 versus 79 years, if born today](#). So, more baby boys, but more old women!)
- But formally, all we can actually conclude from our test is that boys and girls are not equally likely, not which one is more likely.
  - (Could try to clarify this with a “conditional hypothesis test” .)
  - (Simpler is to use confidence intervals – coming soon!)
  - (But intuitively, we have shown that girls are less likely.)

sta130-64



## One More Example: Maple Leafs Hockey!

- So far this year (as of Jan 24, 2016), out of 46 hockey games, the Toronto Maple Leafs have won ... 17. (36.9%)

(<http://www.sportsnet.ca/hockey/nhl/standings>)



- Maybe they're "really" a 50–50 team, i.e. maybe their probability of winning each game is "really" 50%, they've just had bad luck so far! Right??

- Suppose their probability of winning each game is  $p$ .
- Here the null hypothesis is:  $p = 0.5$  (or  $p = 1/2$  or  $p = 50\%$ ).  
Alternative:  $p \neq 0.5$ . (two-sided)
  - And, out of 46 games, if  $p = 0.5$ , then expect 23 wins.
  - So, what is the P-value?

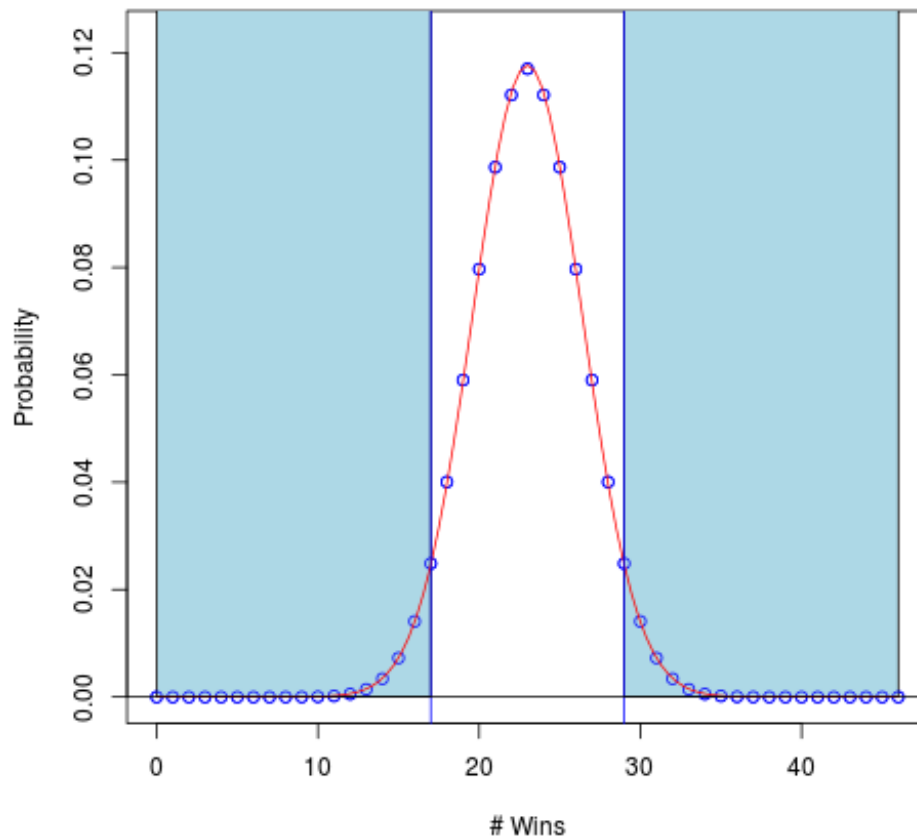
sta130–65

## Maple Leafs Hockey (cont'd)

- Recall: won 17 out of 46 games. If  $p = 0.5$ , expect 23.
- So, here the P-value (two-sided) is the probability, assuming  $p = 0.5$ , of them winning 17 or fewer, or 29 or more. (Since  $17 = 23 - 6$ , while  $29 = 23 + 6$ .)
- Normal approximation! Mean =  $n/2 = 46/2 = 23$ . Variance =  $n/4 = 46/4 = 11.5$ . So  $sd = \sqrt{11.5} \doteq 3.39$ .
- R: `pnorm(17, 23, sqrt(11.5), lower.tail=TRUE) + pnorm(29, 23, sqrt(11.5), lower.tail=FALSE)`
- Answer is: 0.07684324. This is still more than 0.05.
- So YES, it is POSSIBLE that  $p = 0.5$ . Phew!
- But I wouldn't count on it! (The last time the Maple Leafs won the Stanley Cup championship was: May 2, 1967.)

sta130–66

## Maple Leafs Example: Diagram



sta130-67

## Maple Leafs Hockey: One-Sided Test?

- Suppose instead that we started out by believing that the Maple Leafs are worse than being a 50–50 team, and we wanted to test that theory. In this case, we might use a one-sided hypothesis test. That is, our null hypothesis would still be  $p = 0.5$ , but our alternative hypothesis would be:  $p < 0.5$ .
- Then our P-value would just be the probability, assuming  $p = 0.5$ , of a team winning 17 or fewer games out of 46.
- R: `pnorm(17, 23, sqrt(11.5), lower.tail=TRUE)`
- This equals: 0.03842162. (Exactly half of the two-sided P-value ... makes sense since normal distribution is symmetric.)
- This is less than 0.05, so we could reject the null hypothesis that  $p = 0.5$ , and we could conclude that  $p < 0.5$ . But a two-sided test is more conservative, and more appropriate if we don't know which way in advance.

sta130-68

## P-Value Comparison: Why “Or More”

- Recall: The “P-value” is the probability (under the null hypothesis) of getting a result as extreme as observed or more extreme. Why is this important?
- Example: Suppose you do the ESP experiment (guessing with two cards), and get 502 correct out of 1000. Impressive? No!
- Here the P-value is the probability, assuming  $p = 1/2$ , of getting 502 or more correct out of 1000.
- Here mean =  $n/2 = 1000/2 = 500$ , and variance =  $n/4 = 1000/4 = 250$ , so sd =  $\sqrt{250} \doteq 15.8$ , so using the normal approximation, the P-value can be computed in R as:  
`pnorm(502, 500, sqrt(250), lower.tail=FALSE)`  
which equals  $0.4496716 \doteq 0.45$ , i.e. about 45%.

– Much more than 5%. So, it demonstrates nothing. Of course!

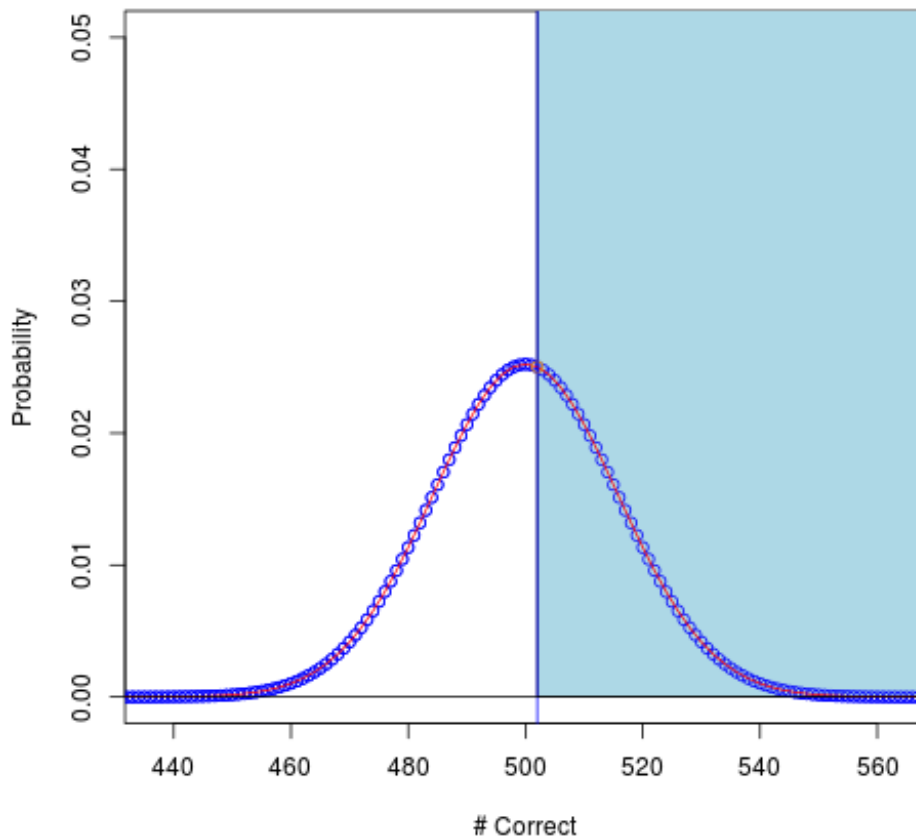
sta130–69

## Why “Or More” (cont’d)

- Recall: in ESP experiment, got 502 out of 1000 correct.
  - P-value  $\doteq 0.45$ , much more than 0.05.
  - Not statistically significant.
- However, the probability of getting exactly 502 out of 1000 correct (under the null hypothesis that  $p = 0.5$ ) can be computed in R by:  
`dbinom(502, 1000, 1/2)`  
(see `?dbinom`) which equals 0.02502422 i.e. is less than 0.05.
- But this doesn’t prove anything!
- We don’t care about getting exactly 502 correct, what we care about is getting 502 or more correct.
- (Or, for a two-sided test: 502 or more, or 498 or less.)

sta130–70

## Why 'Or More': Diagram



sta130-71

## P-Value Comparison: Normal Approximation

- Recall: in ESP experiment, got 502 out of 1000 correct.
- The P-value is the probability, assuming  $p = 1/2$ , of getting 502 or more correct.
- Normal approximation: `pnorm(502, 500, sqrt(250), lower.tail=FALSE)`, which equals 0.4496716.
- Can also compute this P-value exactly, using the “binomial” distribution. (See `?pbinom`.)
  - In R: `pbinom(501, 1000, 0.5, lower.tail=FALSE)`,
  - (Why 501? Because “502 or more” means “not 501 or less”.)
  - Answer is: 0.4622128.
- Quite close, but not exact: normal approximation.

sta130-72

## Normal Approximation: Continuity Correction

- Recall: Exact P-value = 0.4622128, Normal Approximation = 0.4496716. Can we improve this normal approximation a little?
- Well, for the exact (binomial) probabilities, saying “not 501 or less”, or “not 501.5 or less”, or “not 501.9 or less”, is exactly the same, since the number is always an integer (whole number).
- But for the normal distribution, it does matter which one. So, which one should we use? 501? 501.5? 501.9? other?
- Answer: The one right in the middle! 501.5!
- That is, a better normal approximation might use 501.5. (“continuity correction”) In R: `pnorm(501.5, 500, sqrt(250), lower.tail=FALSE)`. Answer = 0.4622097. Better approximation!
- (We won’t always bother with continuity corrections in this class. But you should still understand them.)

sta130–73

## P-Value Comparison: Sample Size

- Consider again that same ESP experiment.
- Suppose you get 11 out of 20 correct (55%). Impressive? No! P-value: `pnorm(11, 20/2, sqrt(20/4), lower.tail=FALSE)`. Answer is: 0.3273604. Not significant at all.
- Suppose you get 22 out of 40 correct (55%). Impressive? No! P-value: `pnorm(22, 40/2, sqrt(40/4), lower.tail=FALSE)`. Answer is: 0.2635446. Not significant at all. (But a bit smaller ...)
- Suppose you get 55 out of 100 correct (55%). Impressive? Here P-value is: `pnorm(55, 100/2, sqrt(100/4), lower.tail=FALSE)`. Answer is: 0.1586553. Not significant. (But smaller ...)
- Suppose you get 550 out of 1000 correct (55%). Impressive? Here P-value is: `pnorm(550, 1000/2, sqrt(1000/4), lower.tail=FALSE)`. Answer is: 0.0007827011. Very significant!

sta130–74

## Sample Size Comparison (cont'd)

- So, even with the same percentage correct (55%), if the sample size is larger, then the result has a smaller P-value, i.e. is more significant, i.e. is more surprising. Why?
- Intuition: getting 550/1000 correct, is kind of like 11/20 correct, repeatedly, 50 times in a row. (Since  $11 \times 50 = 550$ , and  $20 \times 50 = 1000$ .)
- More precisely: If you repeat an experiment (e.g. taking a sample) many times, then the average result will get closer and closer to the true value (e.g.  $p = 1/2$ ). “Law of Large Numbers”
- So, with more samples, even minor percentage differences become more significant.
- (The issue of “how big a sample size do we need” is very important in statistical consulting! More later.)

sta130–75

## Computing Other P-Values: Roulette Wheel

- Recall: roulette wheel has 38 different numbers, which should all be equally likely.
- Suppose a casino operator complains that it is showing “22” too often. How to test this?
- To investigate, you spin the roulette wheel 200 times, and get “22” on 11 of the spins. Too many? (Check:  $200/38 \doteq 5.26$ .)
- Want to test null hypothesis that  $p = 1/38$ , against the alternative hypothesis that  $p > 1/38$ .
- Want to use a normal approximation.
  - Mean =  $200/38 \doteq 5.26$ , of course.
  - Variance = ?? sd = ?? How to compute them?



sta130–76

## P-Values for Roulette Wheel (cont'd)

- We know how to solve this!
- Let  $X_i = 1$  if the  $i^{\text{th}}$  spin is a “22”, otherwise  $X_i = 0$ .
- Then  $E(X_i) = 1 \times 1/38 + 0 \times 37/38 = 1/38$ . (Of course.)
- And,  $Var(X_i) = E[(X_i - 1/38)^2] = [1 - 1/38]^2 \times 1/38 + [0 - 1/38]^2 \times 37/38 = 37/1444 \doteq 0.0256233$ .
- Then,  $T = X_1 + X_2 + \dots + X_{200}$ .
  - So, mean =  $E(T) = E(X_1) + E(X_2) + \dots + E(X_{200}) = 200 \times 1/38 = 200/38 \doteq 5.26$ .
  - And,  $Var(T) = Var(X_1) + Var(X_2) + \dots + Var(X_{200}) = 200 \times 37/1444 = 1850/361 \doteq 5.12$ .
  - And,  $sd(T) = \sqrt{Var(T)} = \sqrt{1850/361} \doteq 2.26$ .
- P-value? `pnorm(11, 200/38, sqrt(1850/361), lower.tail=FALSE)`. Answer: 0.005635243. Very significant!!

sta130-77

## Expectation and Variance: More General Formula

- More generally, suppose we repeat an experiment “ $n$ ” times, with probability “ $p$ ” of success each time, and let “ $T$ ” be the total number of successes. What are  $E(T)$ ,  $Var(T)$ , and  $sd(T)$ ?
  - (We already solved this when  $n = 100$  and  $p = 0.5$  [ESP], or  $n = 12$  and  $p = 1/6$  [dice], or  $n = 200$  and  $p = 1/38$  [roulette], ...)
- Well, let  $X_i = 1$  if the  $i^{\text{th}}$  experiment is a success (e.g. a Head, or a Correct Guess, or a Five, or a “22”, or ...), otherwise  $X_i = 0$ .
- Then  $E(X_i) = 1 \times p + 0 \times (1 - p) = p$ .
- And,  $Var(X_i) = E[(X_i - p)^2] = [1 - p]^2 \times p + [0 - p]^2 \times (1 - p) = (1 - 2p + p^2)p + p^2(1 - p) = (p - 2p^2 + p^3) + (p^2 - p^3) = p - p^2 = p(1 - p)$ .
- Then,  $T = X_1 + X_2 + \dots + X_n$ .
  - So, mean =  $E(T) = E(X_1) + E(X_2) + \dots + E(X_n) = n \times p = np$ .

sta130-78

## More General Formula (cont'd)

- Know:  $T = X_1 + X_2 + \dots + X_n$ , and  $E(T) = np$ .
  - Then,  $Var(T) = Var(X_1) + Var(X_2) + \dots + Var(X_n) = n \times p(1 - p) = np(1 - p)$ .
  - So,  $sd(T) = \sqrt{Var(T)} = \sqrt{np(1 - p)}$ .
- ESP Example: Guess  $n = 100$  times, had probability  $p = 0.5$  of success each time, got mean = 50, and variance = 25. Check:  $np = 100 \times 0.5 = 50$  (same), and  $np(1 - p) = 100 \times 0.5 \times (1 - 0.5) = 25$  (same). Good!
- Dice Example: Rolled a die  $n = 12$  times, with probability  $p = 1/6$  of getting a “5” each time. For the number of 5’s, got mean = 2, and variance = 5/3. Check:  $np = 12 \times 1/6 = 2$  (same), and  $np(1 - p) = 12 \times 1/6 \times (1 - 1/6) = 5/3$  (same).

sta130–79

## More General Formula (cont'd)

- Roulette Example: Spun  $n = 200$  times, with probability  $p = 1/38$  of getting a “22” each time. For number of 22’s, got mean =  $200/38$ , and variance =  $1850/361$ . Check:  $np = 200 \times 1/38 = 200/38$  (same), and  $np(1 - p) = 200 \times 1/38 \times (1 - 1/38) = 1850/361$  (same).
- So, for any repeated Success/Failure experiment (e.g. Yes/No, or Heads/Tails, or Correct/Incorrect, or Five/Not-Five, or 22/not-22, or ...), repeated  $n$  times, with probability of success  $p$  each time, the number of successes has mean  $np$ , and variance  $np(1 - p)$ , hence sd  $\sqrt{np(1 - p)}$ .
- Can use this for normal approximations, P-values, etc.
- We can apply this to polls! For example ...

sta130–80



## Application: that Trudeau Government Poll again

- Recall: out of 1,500 Canadians surveyed in a poll, 795 of them (53%) approved of the government.
- We verified earlier that this “demonstrates” that more than half of Canadians approve. (P-value = 0.01006838.)
  - (Well, at least assuming that it was a truly random sample, everyone replied honestly, etc. Let’s assume that for now.)
- But does it demonstrate that over 51% approve? Let’s test that!
- Null hypothesis:  $p = 0.51$ , i.e. 51% approve.
  - Alternative hypothesis:  $p > 0.51$ , i.e. over 51% approve.
- Under the null hypothesis, the number (out of 1,500) who approve has mean =  $np = 1,500 \times 0.51 = 765$ , and variance =  $np(1 - p) = 1,500 \times 0.51 \times 0.49 = 374.85$ , so sd =  $\sqrt{374.85}$ .

sta130–81

## Trudeau Government Poll (cont’d)

- Then the P-value should be the probability, assuming the null hypothesis that  $p = 0.51$ , that out of 1,500 people, 795 or more would approve.
- Use normal approximation, with mean = 765, and sd =  $\sqrt{374.85}$ :  
`pnorm(795, 765, sqrt(374.85), lower.tail=FALSE)`
- Answer is: 0.06063039.
- This is more than 0.05, so cannot reject the null, i.e. cannot claim that more than 51% approve.
- So what precisely can we claim? More than 50.5%??
- Should we just keep doing different hypothesis tests?
- Better: use confidence intervals. (The poll said: “The margin of error for a comparable probability-based random sample of the same size is +/- 2.6%, 19 times out of 20.” Huh?)

sta130–82

## Confidence Interval for the Trudeau Poll

- Recall: of  $n = 1,500$  people surveyed, 795 (53%) approve.
- Suppose the true fraction of all Canadians who approve is equal to some value  $p$  (unknown).
- What are the “plausible” values for  $p$ ?
  - 40%? (no!) 50%? (no!) 51%? (yes!) 50.5%? 57%?
- Let  $T$  be the number of respondents (out of 1,500) who approve.
  - Then we know that here  $T = 795$ .
  - But what were the probabilities for  $T$ ?
- If the true fraction is  $p$ , then the quantity  $T$  should be approximately normal, with mean  $= np = 1500 p$ , and variance  $= np(1 - p) = 1500 p(1 - p)$ , so  $sd = \sqrt{np(1 - p)}$ .

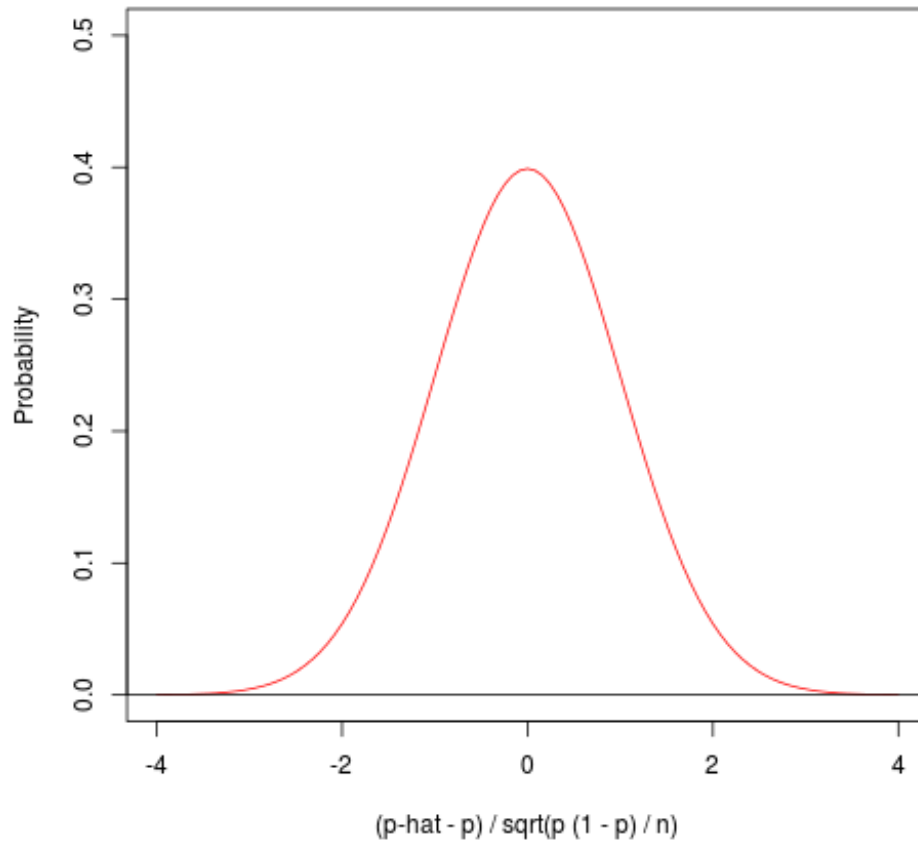
sta130–83

## Confidence Interval for Trudeau Poll (cont'd)

- Recall:  $T$  normal, mean  $= np = 1500 p$ ,  $sd = \sqrt{np(1 - p)}$ .
- Let  $\hat{p} = T/n = T/1,500$  be the fraction who approve.
  - Observed value:  $\hat{p} = 0.53$ . But what about the probabilities?
  - The probabilities for  $\hat{p}$  should be approximately normal, with mean  $= np / n = p$ , and  $sd = \sqrt{np(1 - p)} / n = \sqrt{p(1 - p)/n}$ .
  - (We divide sd by  $n$ , i.e. divide var by  $n^2$ . Check this!)
  - Then  $\hat{p} - p$  is normal, with mean  $= 0$ , and  $sd = \sqrt{p(1 - p)/n}$ . (Subtracting  $p$  shifts the mean, not the sd.)
  - So  $(\hat{p} - p) / \sqrt{p(1 - p)/n}$  is (approximately) normal, with mean  $= 0$ , and  $sd = 1$ . (“standard normal”)
- For the standard normal, we can find the required probabilities by measuring. What is a 95% range? Let's see ...

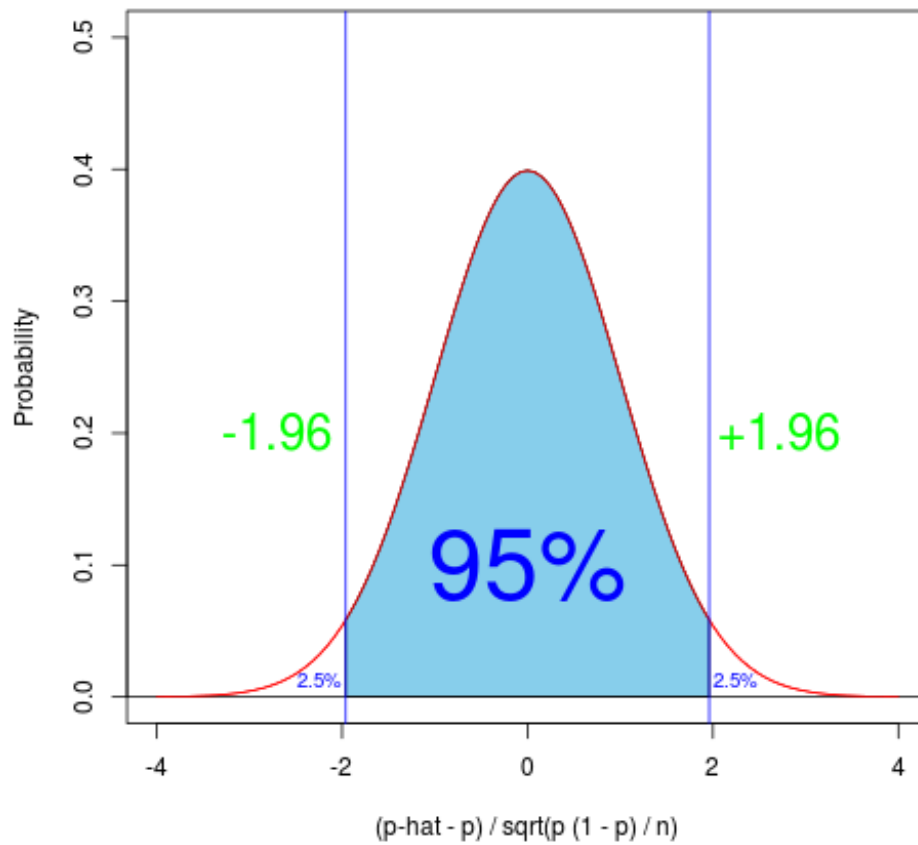
sta130–84

### Standard Normal: How Wide is 95%?



sta130-85

### Probabilities for the Standard Normal



sta130-86

## Confidence Interval for Trudeau Poll (cont'd)

- That is: the standard normal has probability about 95% of being between  $-1.96$  and  $+1.96$ .

- Check: `pnorm(+1.96,0,1,lower.tail=TRUE) - pnorm(-1.96,0,1,lower.tail=TRUE)` equals 0.9500042. good!

- (What if we wanted probability 99%? Replace “1.96” by about: 2.58. Then get: 0.99012.)

- So, if  $Z$  has the standard normal distribution (i.e., probabilities), then  $P[-1.96 < Z < +1.96] \doteq 0.95$ .

So, apply this to:  $Z = (\hat{p} - p) / \sqrt{p(1-p)/n}$

- Conclusion:  $P[-1.96 < (\hat{p} - p) / \sqrt{p(1-p)/n} < +1.96] \doteq 0.95$ .

So,  $P[-1.96\sqrt{p(1-p)/n} < \hat{p} - p < +1.96\sqrt{p(1-p)/n}] \doteq 0.95$ .

So,  $P[\hat{p} - 1.96\sqrt{p(1-p)/n} < p < \hat{p} + 1.96\sqrt{p(1-p)/n}] \doteq 0.95$ .

- That is, with probability 95%,  $p$  is within  $\pm 1.96\sqrt{p(1-p)/n}$  of the observed fraction  $\hat{p}$ . This is what we want!

sta130-87

## Confidence Interval for Trudeau Poll: Conclusion?

- With probability 95%,  $p$  is within  $\pm 1.96\sqrt{p(1-p)/n}$  of  $\hat{p}$ .

- Here  $1.96\sqrt{p(1-p)/n}$  is the “(95%) margin of error”.

- And, the interval from  $\hat{p} - 1.96\sqrt{p(1-p)/n}$  to  $\hat{p} + 1.96\sqrt{p(1-p)/n}$  is the “(95%) confidence interval”.

- Good! Just one problem: “ $p$ ” is unknown! So we can’t calculate this margin of error! Bad! What to do?

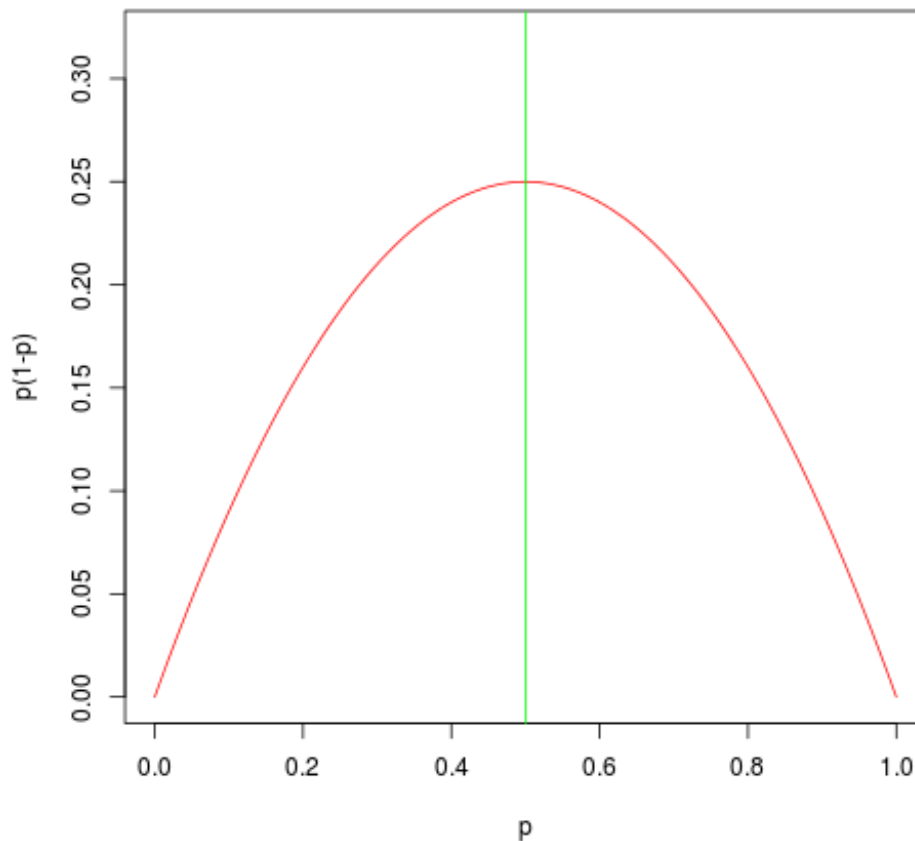
- Option #1: replace  $p$  (unknown) by  $\hat{p}$  (known; close(?) to  $p$ ). Get  $1.96\sqrt{\hat{p}(1-\hat{p})/n} = 196\%\sqrt{\hat{p}(1-\hat{p})/n}$ . (“bold option”)

- Option #2: figure out the largest value that  $p(1-p)$  can ever be. (“conservative option”)

- What value is that? Let’s check ...

sta130-88

"p(1-p)" Diagram



sta130-89

### Confidence Interval for Trudeau Poll: Conclusion!

- So, always have  $p(1 - p) \leq (1/2)(1 - 1/2) = 1/4 = 0.25$ .
- So,  $1.96\sqrt{p(1 - p)/n} \leq 1.96\sqrt{(1/4)/n} = 1.96/\sqrt{4n} = 1.96/2\sqrt{n} = 0.98/\sqrt{n} = (98\%)/\sqrt{n}$ .
- This margin of error,  $(98\%)/\sqrt{n}$ , is what polls usually use!
- Trudeau poll: “The margin of error for a comparable probability-based random sample of the same size is +/- 2.6%, 19 times out of 20.” Check:  $(98\%)/\sqrt{1,500} \doteq 2.530349\% \doteq 2.6\%$ .
  - Yep! (Though “rounded up”, to be safe ...)
  - Conclude that true support  $p$  is between  $53\% - 2.6\% = 50.4\%$ , and  $53\% + 2.6\% = 55.6\%$ . Confidence interval is:  $[50.4\%, 55.6\%]$ .
- Or, could use the bold option:  
 $196\%\sqrt{\hat{p}(1 - \hat{p})/n} = 196\%\sqrt{0.53(1 - 0.53)/1,500} \doteq 2.52579\%$ .  
Very similar! (since  $\hat{p} \approx 0.5$  in this case ...) Other polls?

sta130-90

## Majority of Canadians support decriminalizing marijuana: poll

Ipsos found 65 per cent of Canadians (29 per cent strongly/36 per cent somewhat) support the decriminalization of marijuana so that possession of small amounts would no longer carry a penalty or fine.

*was conducted between August 13 and August 17, with a sample of 1,000 Canadians, from Ipsos' online panel. The poll reported above is accurate to within 3.5 percentage points 19 times out of 20.*

- Check:  $98\%/\sqrt{n} = 98\%/\sqrt{1000} \doteq 3.099\% \doteq 3.1\%$ .  
(Which they rounded up to 3.5%.) Or, use bold option:  
 $196\%\sqrt{\hat{p}(1-\hat{p})/n} = 196\%\sqrt{0.65(1-0.65)/1,000} \doteq 2.9563\%$ .  
(A bit smaller. They didn't use that.) Let's use 3.1%.
- Then, 95% confidence interval is: [61.9%, 68.1%].

sta130-91

### But Do Poll Companies Really Use This?

- To check this, I went to the main web site for Forum Research Inc., a leading Canadian pollster.
- For information about margins of error, they refer you to this web page:  
<http://www.forumresearch.com/tools-margin-of-error.asp>
- That page gives various margins of error, based on the Sample Size ( $n$ ) and the Observed Proportion ( $\hat{p}$ ).
- Does it follow our formula? Let's check. Oh yeah!
- (In practice, they usually use the "Observed Proportion = 50%" row – corresponding to the ... conservative option.)
- Can we use margins of error and confidence intervals for other experiments besides polls? Yes!

sta130-92

## Confidence Intervals for ESP Example

- Suppose you guess between two cards, 100 times, and get 69 correct. Then  $\hat{p} = 0.69$ , and margin of error =  $98\%/\sqrt{n} = 98\%/10 = 9.8\% = 0.098$  (conservative option), or  $1.96\sqrt{\hat{p}(1-\hat{p})/n} = 1.96\sqrt{0.69 \times 0.31/100} \doteq 0.090$  (bold).

- So, 95% confidence interval (conservative option) for  $p$  is: [0.592, 0.788]. Or, bold option: [0.600, 0.780].

- Or, if you guess 1000 times, and get 550 correct:  $\hat{p} = 0.55$ , and 95% M.O.E. (cons.) =  $0.98/\sqrt{1000} \doteq 0.031$ , and 95% conf. int. (cons.) = [0.519, 0.581]  $\doteq$  [0.52, 0.58] = [52%, 58%]. (Not 50%!)

- Or, if you guess 20 times, and get 14 correct:  $\hat{p} = 0.70$ , and 95% M.O.E. (cons.) =  $0.98/\sqrt{20} \doteq 0.22$ , and 95% conf. int. (cons.) = [0.48, 0.92]. Or, bold option: M.O.E. =  $1.96\sqrt{0.7 \times 0.3/20} \doteq 0.20$ , conf. int. = [0.50, 0.90].

sta130-93

## Confidence Intervals for our Other Examples

- Boy or Girl? Of  $n = 377,636$  births, 184,049 were female, so  $\hat{p} = 184,049/377,636 \doteq 0.4874$ . M.O.E. (cons.) =  $0.98/\sqrt{n} \doteq 0.0016$ . So, conf. int. = [0.4858, 0.4890]. So, the probability of a girl is significantly less than 0.5.

- Maple Leafs: Out of  $n = 46$  games, won 17, so  $\hat{p} = 17/46 \doteq 0.369$ . M.O.E. (cons.) =  $0.98/\sqrt{46} \doteq 0.144$ . So, conf. int. (cons.) for  $p$  is [0.225, 0.513]. So, we still might have  $p > 0.5$ . (Right??) [Hm, they lost two more games since then ... ]

- Dice: Suppose roll a die 12 times, get “5” on 7 rolls. Then  $\hat{p} = 7/12 \doteq 0.58$ . M.O.E. (bold) =  $1.96 \times \sqrt{\hat{p}(1-\hat{p})/n} = 1.96 \times \sqrt{0.58(0.42)/12} \doteq 0.28$ . So, conf. int. (bold) is [0.30, 0.86]. Does this interval include  $1/6 \doteq 0.17$ ? No! Conclusion: we got too many 5’s!

sta130-94

- Roulette: Spin wheel 200 times, get “22” on 11 spins. Here  $\hat{p} = 11/200 = 0.055$ . M.O.E. (cons.) =  $0.98/\sqrt{200} \doteq 0.069$ . So, conf. int. (cons.) for  $p$  is  $[-0.014, 0.124]$ . Does it include  $1/38 \doteq 0.024$ ? Yes! In fact it’s clearly too large an interval –  $p$  can’t be negative.

- Instead, better use bold option here (since  $\hat{p}$  is far away from 0.5). M.O.E. (bold) =  $1.96 \times \sqrt{\hat{p}(1 - \hat{p})/n} = 1.96 \times \sqrt{0.055(0.945)/200} \doteq 0.032$ . So, conf. int. (bold) for  $p$  is  $[0.023, 0.087]$ . Does this one include  $1/38 \doteq 0.024$ ? Yes! But just barely. (A “one-sided” version would exclude  $1/38$ .)

- If instead had gotten “22” on 13 spins out of 200, then  $\hat{p} = 13/200 = 0.065$ , and M.O.E. (bold) =  $1.96 \times \sqrt{\hat{p}(1 - \hat{p})/n} = 1.96 \times \sqrt{0.065(0.935)/200} \doteq 0.034$ . So, conf. int. (bold) for  $p$  is  $[0.031, 0.099]$ . Does this one include  $1/38 \doteq 0.024$ ? No! Too many 22’s! [M.O.E. test: **Rcoins**]

sta130–95

## Clarification: Properties of Mean and Variance

- Example: Roll a fair six-sided die. Let  $X$  be the value showing.



- Then mean =  $E(X) = 1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 + 6 \times 1/6 = 21/6 = 3.5$ .

- And variance =  $Var(X) = E([X - 3.5]^2) = [1 - 3.5]^2 \times 1/6 + [2 - 3.5]^2 \times 1/6 + [3 - 3.5]^2 \times 1/6 + [4 - 3.5]^2 \times 1/6 + [5 - 3.5]^2 \times 1/6 + [6 - 3.5]^2 \times 1/6 = 17.5/6 \doteq 2.917$ .

- And,  $sd(X) = \sqrt{Var(X)} = \sqrt{17.5/6} \doteq 1.708$ .

- Now, let  $Y = X + 10$ , i.e.  $Y$  equals ten more than the value showing. So,  $Y$  could be 11, or 12, or 13, or 14 or 15 or 16.

- What are  $E(Y)$  and  $Var(Y)$  and  $sd(Y)$ ?

sta130–96



- Here  $E(Y) = E(X+10) = 11 \times 1/6 + 12 \times 1/6 + 13 \times 1/6 + 14 \times 1/6 + 15 \times 1/6 + 16 \times 1/6 = 13.5 = 3.5 + 10 = E(X) + 10$ .

- And  $Var(Y) = E([Y - 13.5]^2) = [11 - 13.5]^2 \times 1/6 + [12 - 13.5]^2 \times 1/6 + [13 - 13.5]^2 \times 1/6 + [14 - 13.5]^2 \times 1/6 + [15 - 13.5]^2 \times 1/6 + [16 - 13.5]^2 \times 1/6 = 17.5/6 = Var(X)$ .

- And,  $sd(Y) = \sqrt{Var(Y)} = \sqrt{Var(X)} = sd(X)$ .

- CONCLUSION: If you add a constant number (like 10) to a random quantity (like  $X$ ), the mean increases by the same amount, and the variance and sd are unchanged.

- But what if we multiply or divide instead?

- Well, let  $W = 2 \times X$ , i.e.  $W$  equals twice the value showing. So,  $W$  could equal 2, or 4, or 6, or 8 or 10 or 12.

- What are  $E(W)$  and  $Var(W)$  and  $sd(W)$ ?

sta130-97

- Here  $E(W) = E(2 \times X) = E(2X) = 2 \times 1/6 + 4 \times 1/6 + 6 \times 1/6 + 8 \times 1/6 + 10 \times 1/6 + 12 \times 1/6 = 7 = 2 \times 3.5 = 2 \times E(X)$ .

- And  $Var(W) = E([W - 7]^2) = [2 - 7]^2 \times 1/6 + [4 - 7]^2 \times 1/6 + [6 - 7]^2 \times 1/6 + [8 - 7]^2 \times 1/6 + [10 - 7]^2 \times 1/6 + [12 - 7]^2 \times 1/6 = 35/3 = 4 \times 17.5/6 = 4 \times Var(X) = 2^2 \times Var(X)$ . (Makes sense.)

- And,  $sd(W) = \sqrt{Var(W)} = \sqrt{4 \times Var(X)} = 2 \times sd(X)$ .

- CONCLUSION: If you multiply a random quantity (like  $X$ ) by a constant positive number (like 2), then the mean is multiplied by the same amount, and the sd is multiplied by the same amount, and the variance is multiplied by the square of the amount.

- Summary: Adding (or subtracting) affects the mean by the same amount, but not the variance or sd. Multiplying (or dividing) affects the mean and sd by the same factor (if pos.), and affects the variance by the square of the factor. (Useful! Already used this!)

sta130-98

## Standardised Variables (Z-Scores)

- Above, to derive the confidence interval for the poll, we used a trick: We “adjusted”  $\hat{p}$  into a “standard normal” quantity, i.e. a normal with mean 0 and variance 1, also sometimes written  $\mathcal{N}(0, 1)$ . How did we do this? By subtracting the mean ( $p$ ), and dividing by the sd ( $\sqrt{p(1-p)/n}$ ). This is a good trick!
- For any random quantity, with any mean and sd, if you subtract off its mean, and then divide by its sd, you get a quantity with mean=0, and sd=1.
  - (So if it was approximately normal before, then now it is approximately standard normal.)
  - The standardised value is sometimes called a “Z-score”.
- Intuition: the Z-score measures “how many standard deviations is it above (or below) the mean”.

sta130–99

## Standardised Variables (cont’d)

- Example: If we roll a fair die, then the result has mean 3.5, and variance  $17.5/6 \doteq 2.917$ , hence sd  $\sqrt{17.5/6} \doteq 1.708$ .
- Suppose on our next roll, we get a 6. What is the Z-score?
  - Answer:  $(6 - 3.5)/1.708 \doteq 1.464$ .
  - Interpretation: “It was about 1.464 standard deviations above the mean.”
- Suppose instead we get a 2. What is the Z-score?
  - Answer:  $(2 - 3.5)/1.708 \doteq -0.878$ .
  - Interpretation: “It was about 0.878 standard deviations below the mean.”
- This reduces all probabilities, to “standard” ones.
  - And, it reduces normal probabilities to standard normals!

sta130–100

## Using Z-Scores to Compare: Elephants and Mice

- Suppose you meet a “large” African elephant which weighs 5000 kg, and a “large” white mouse which weighs 38 g.
- Relatively speaking, which is larger?
  - How to compare them??
- For a relative comparison, should use Z-scores!
  - Apparently(?) African elephants’ weights have mean 4550 kg, and standard deviation 150 kg.
  - And, white mice’s weights have mean 30 g, and standard deviation about 2 g.
- So, the elephant has Z-score =  $(5000 - 4550)/150 = 3.0$ .
- And, the mouse has Z-score =  $(38 - 30)/2 = 4.0$ .
- So, on a relative scale, the mouse is larger! Z-scores can provide a comparison between two very different quantities. sta130–101

## Converting to Standard Normal: Example

- Suppose a random quantity  $T$  is approximately normal, with mean 50 and sd 10. Then in terms of a standard normal, what is the probability that  $T$  is more than 70, i.e. what is  $P[T > 70]$ ?
  - Well,  $P[T > 70] = P[(T - 50)/10 > (70 - 50)/10] = P[(T - 50)/10 > 2] \doteq P[Z > 2]$  (where  $Z$  is a standard normal random quantity).
- So what? Well, for one thing, we can look up this probability in a standard normal table (e.g. on midterm/exam!), see [www.probability.ca/sta130/normaltable.png](http://www.probability.ca/sta130/normaltable.png).
  - Need to use that:  $P[Z > 2] = 1 - P[Z < 2] = 1 - P[Z \leq 2]$ .
- From the table:  $P[Z \leq 2] \doteq 0.9772$ . So,  $P[T > 70] \doteq P[Z > 2] \doteq 1 - 0.9772 = 0.0228$ . Check in R: `pnorm(70, 50, 10, lower.tail=FALSE)`. Ans: 0.02275013. Good!

## Standard Normal: ESP Example

- ESP experiment (again!). Suppose you guess 59 right out of 100. Does this show that you have ESP?
  - We know that the number correct (assuming  $p = 1/2$ ) has mean = 50, variance = 25, and sd = 5.
  - So, Z-score is:  $(59-50)/5 = 9/5 = 1.8$ . (Meaning?)
  - P-value =  $P[\text{would have gotten 59 or more right}] \doteq P[\text{normal with mean 50 and sd 5 would be 59 or more}] = P[\text{standard normal would be } (59-50)/5 \text{ or more}] = P[Z \geq (59 - 50)/5] = P[Z \geq 1.8] = 1 - P[Z \leq 1.8]$ .
  - From table: This is about  $1 - 0.9641 = 0.0359$ .
  - Less than 0.05. So, statistical significant!
  - (But a two-sided test would not be significant ...)

sta130-103

## Standard Normal: Dice Example

- Suppose (again) that you roll a fair die 60 times, and get “2” on 17 of the rolls. Too many 2’s?
- Know that number of 2’s has mean =  $np = 60(1/6) = 10$ , and variance =  $np(1 - p) = 60(1/6)(5/6) = 25/3$ , so sd =  $\sqrt{25/3}$ .
  - So, Z-score is  $(17 - 10)/\sqrt{25/3} = 7/\sqrt{25/3} \doteq 2.425$ .
  - And,  $P[\text{number of “2”} \geq 17] \doteq P[\text{normal with mean 10 and sd } \sqrt{25/3} \text{ is } \geq 17] = P[\text{standard normal } \geq (17 - 10)/\sqrt{25/3}] \doteq P[\text{standard normal } \geq 2.425] = 1 - P[\text{standard normal } \leq 2.425] \doteq 1 - 0.992$  (from Table) = 0.008. Less than 0.05! So, yes: too many 2’s!
- Tricks for standard normal table: if  $Z$  is standard normal, then  $P[Z \leq -0.6] = 1 - P[Z \geq -0.6] = 1 - P[Z \leq +0.6] = 1 - 0.7257 = 0.2743$ . (And so on.)

sta130-104

## More Standard Normal Practice

- Suppose  $Z$  has standard normal probabilities. Using only the standard normal table and a calculator, what is  $P[Z < -0.43]$ ?
  - Well, by symmetry,  $P[Z < -0.43] = P[Z > 0.43] = 1 - P[Z \leq 0.43] = 1 - 0.6664 = 0.3336$ . Check in R?
    - `pnorm(-.43, 0, 1, lower.tail=TRUE)` = 0.3335978. Good!
- Suppose  $X$  has normal probabilities with mean 80 and sd 6. With just table and calculator, what is  $P[X > 90]$ ?
  - Here  $Z$  has standard normal probabilities if  $Z = (X - 80)/6$ .
  - Then,  $P[X > 90] = P[(X - 80)/6 > (90 - 80)/6] = P[Z > (90 - 80)/6] \doteq P[Z > 1.67] = 1 - P[Z \leq 1.67] \doteq 1 - 0.9525 = 0.0475$ .
  - Check in R? `pnorm(90, 80, 6, lower.tail=FALSE)` = 0.04779035. Pretty good.

sta130-105

## “Drop”? “Increase”? Comparing Polls

### Pollster finds rapid drop in public trust in police

By: Jackie Hong Staff Reporter, Published on Sun Jan 31 2016

In poll done on Jan. 20, 78 per cent of respondents said they trusted Toronto police officers.

But, by Jan. 29, that number had fallen to 63 per cent, according to a second, more

The initial poll was conducted Jan. 20 by Forum Research and surveyed 546 randomly selected voters in Toronto. The second poll was conducted in the Jan. 29 and surveyed 586 randomly selected voters in Toronto via an interactive voice

Results based on the total sample are considered accurate +/- 4 per cent, 19 times out

- Check M.O.E.:  $98\% / \sqrt{586} \doteq 4.048\%$ . Yep!
- But can we compare polls? Is it a statistically significant “drop”? Or just luck? How to test? (Tutorial!)

sta130-106

## Comparing Polls: Marijuana Support in U.S.

CBS NEWS / April 20, 2015, 7:00 AM

# **Poll: Support for legal marijuana use reaches all-time high**

## Should Marijuana Use be Legal?

	Now	2/2014	4/2013	11/2012	10/2011	7/2009	7/1979
Yes	53%	51%	45%	47%	40%	41%	27%
No	43%	46%	45%	47%	51%	52%	69%

SOURCE: CBS News

- True increases? Just luck? How do we test?

sta130-107

## Comparing Polls: Marijuana Support in U.S. (cont'd)

- 2012 poll (<http://www.cbsnews.com/news/poll-nearly-half-support-legalization-of-marijuana/>): out of 1,100 adults surveyed, 47% supported legalisation.
- 2014 poll (<http://www.cbsnews.com/news/majority-of-americans-now-support-legal-pot-poll-says/>): out of 1,018 adults surveyed, 51% supported legalisation.
- 2015 poll (<http://www.cbsnews.com/news/poll-support-for-legal-marijuana-use-reaches-all-time-high/>): out of 1,012 adults surveyed, 53% supported legalisation.
- Which differences are “real”, and which are “just luck”?
- Start with 2012 versus 2014 ...

sta130-108

## Comparing Polls (cont'd)

- Write  $p_1$  for the true fraction of support in 2012, and  $n_1$  for the number of people surveyed in 2012, and  $T_1$  for the total number of supporters among the surveyed people in 2012, and  $\hat{p}_1 = T_1/n_1$  for the observed fraction of support in 2012. Similarly  $p_2$  and  $n_2$  and  $T_2$  and  $\hat{p}_2$  for 2014.
- Null hypothesis?  $p_1 = p_2$ . Alternative?  $p_1 \neq p_2$ .
- Now,  $T_1$  is approximately normal, with mean  $n_1 p_1$ , and variance  $n_1 p_1 (1 - p_1)$ , so sd  $\sqrt{n_1 p_1 (1 - p_1)}$ .
- So,  $\hat{p}_1$  is approximately normal, with mean  $n_1 p_1 / n_1 = p_1$ , and sd  $\sqrt{n_1 p_1 (1 - p_1)} / n_1 = \sqrt{p_1 (1 - p_1) / n_1}$ , and variance  $n_1 p_1 (1 - p_1) / (n_1)^2 = p_1 (1 - p_1) / n_1$ .
- Similarly,  $\hat{p}_2$  is approximately normal, with mean  $p_2$ , and sd  $\sqrt{p_2 (1 - p_2) / n_2}$ , and variance  $p_2 (1 - p_2) / n_2$ .

sta130-109

## Comparing Polls (cont'd)

- What about the negative of  $\hat{p}_1$ , i.e.  $-\hat{p}_1$ ? It has mean  $-p_1$ , and variance  $p_1 (1 - p_1) / n_1$  [not  $-p_1 (1 - p_1) / n_1$ , because it involves squares, and  $(-1)^2 = 1$ ]. So, sd =  $\sqrt{p_1 (1 - p_1) / n_1}$ .
- What about the difference,  $\hat{p}_2 - \hat{p}_1$ ?
  - Its mean is:  $p_2 - p_1$ . Unknown! But under the null hypothesis, it equals: 0.
  - And, its variance is:  
$$\text{Var}(\hat{p}_2) + \text{Var}(-\hat{p}_1) = p_1 (1 - p_1) / n_1 + p_2 (1 - p_2) / n_2.$$
  - So, its sd is:  $\sqrt{p_1 (1 - p_1) / n_1 + p_2 (1 - p_2) / n_2}$ .
  - Note: Not  $\sqrt{p_1 (1 - p_1) / n_1} + \sqrt{p_2 (1 - p_2) / n_2}$ . You get to add the variances, but not the standard deviations!
- So how does this help?

sta130-110

## Comparing Polls (cont'd)

- We see that under the null hypothesis that  $p_1 = p_2$ , the quantity  $\hat{p}_2 - \hat{p}_1$  should be approximately normal, with mean 0, and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ .
- But even under the null hypothesis,  $p_1$  and  $p_2$  are unknown!
- Bold option: estimate sd by  $\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$ .
- Conservative option: estimate sd as  $\leq \sqrt{1/2(1-1/2)/n_1 + 1/2(1-1/2)/n_2} = \sqrt{1/n_1 + 1/n_2} / 2$ .
- Then can compute a P-value under the null hypothesis.
  - The P-value should be the probability, assuming that  $p_1 = p_2$ , that  $\hat{p}_2 - \hat{p}_1$  would be as “extreme” or more “extreme” than the observed value.
  - That is, the probability that  $|\hat{p}_2 - \hat{p}_1|$  would be as large or larger than the observed value.

sta130–111

## Back to the U.S. Marijuana Surveys

- In 2012, had  $n_1 = 1,100$ , and  $\hat{p}_1 = 0.47$ . In 2014, had  $n_2 = 1,018$ , and  $\hat{p}_2 = 0.51$ .
- So, the observed value of  $|\hat{p}_2 - \hat{p}_1|$  was 0.04.
- Under the null hypothesis, that quantity should be approximately normal, with mean 0, and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ .
- Conservative option: estimate sd as  $\leq \sqrt{1/n_1 + 1/n_2} / 2 = \sqrt{1/1,100 + 1/1,018} / 2 \doteq 0.02174517$ .
- So the P-value should be the probability that a normal quantity with mean 0 and sd  $\doteq 0.02174517$  is as large or larger than 0.04, or as small or smaller than  $-0.04$ . In R: `pnorm(0.04, 0, 0.02174517, lower.tail=FALSE) + pnorm(-0.04, 0, 0.02174517, lower.tail=TRUE)`. Answer = 0.0658433. [Or use table??] More than 0.05! So, we cannot conclude that more people in the U.S. supported legalisation of marijuana in 2014 than in 2012.

sta130–112



## U.S. Marijuana Surveys (cont'd)

- Could perhaps try bold option instead. Would it help much? No, because  $\hat{p}_1, \hat{p}_2 \approx 1/2$  anyway.

- Let's instead compare 2012 to 2015! In 2012, had  $n_1 = 1,100$ , and  $\hat{p}_1 = 0.47$ . In 2015, had  $n_2 = 1,012$ , and  $\hat{p}_2 = 0.53$ . So, the observed value of  $|\hat{p}_2 - \hat{p}_1|$  was 0.06. Under the null hypothesis, that quantity should be approximately normal, with mean 0, and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ . Conservative option: estimate sd as  $\leq \sqrt{1/n_1 + 1/n_2} / 2 = \sqrt{1/1,100 + 1/1,012} / 2 \doteq 0.02177862$ .

- So the P-value should be the probability that a normal quantity with mean 0 and sd  $\doteq 0.02177862$  is as large or larger than 0.06, or as small or smaller than  $-0.06$ . In R: `pnorm(0.06, 0, 0.02177862, lower.tail=FALSE) + pnorm(-0.06, 0, 0.02177862, lower.tail=TRUE)`. Answer = 0.005869294. Much less than 0.05 (or 0.01)! Conclusion: more people in the U.S. supported legalisation of marijuana in 2015 than in 2012.

sta130-113

## Comparing Polls: Confidence Intervals

- Can we get a confidence interval for  $p_2 - p_1$ ?
- Well,  $\hat{p}_1$  has mean  $p_1$ , and variance  $p_1(1-p_1)/n_1$ . And  $\hat{p}_2$  has mean  $p_2$ , and variance  $p_2(1-p_2)/n_2$ . So observed  $\hat{p}_2 - \hat{p}_1 = 0.06$ .
- So, without assuming anything about  $p_1 = p_2$  or anything, we can still say that  $\hat{p}_2 - \hat{p}_1$  should have mean  $p_2 - p_1$ , and variance  $p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$ .
- Can again estimate the sd of  $p_2 - p_1$  as  $\sqrt{1/n_1 + 1/n_2} / 2$  (conservative), or  $\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$  (bold).
- Then the difference  $(\hat{p}_2 - \hat{p}_1)/sd - (p_2 - p_1)/sd$  is approximately normal, with mean 0, and sd 1.
- So, with probability 95%, this difference should be between  $-1.96$  and  $+1.96$ . So, with probability 95%, the quantity  $p_2 - p_1$  should be between  $\hat{p}_2 - \hat{p}_1 - 1.96 \times sd$  and  $\hat{p}_2 - \hat{p}_1 + 1.96 \times sd$ .

sta130-114

## Confidence Interval for U.S. Marijuana Surveys

- In 2012, had  $n_1 = 1,100$ , and  $\hat{p}_1 = 0.47$ . In 2015, had  $n_2 = 1,012$ , and  $\hat{p}_2 = 0.53$ .
- So, without assuming any null hypothesis, the quantity  $\hat{p}_2 - \hat{p}_1$  should have mean  $p_2 - p_1$ , and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ .
- Conservative option: estimate sd as  $\leq \sqrt{1/n_1 + 1/n_2} / 2 = \sqrt{1/1,100 + 1/1,012} / 2 \doteq 0.02177862$ .
- So,  $p_2 - p_1$  should be between  $\hat{p}_2 - \hat{p}_1 - 1.96 \times 0.02177862$  and  $\hat{p}_2 - \hat{p}_1 + 1.96 \times 0.02177862$ . That is, between about  $0.06 - 1.96 \times 0.02177862$  and  $0.06 + 1.96 \times 0.02177862$ . That is, between about  $0.0173139 \doteq 1.7\%$  and  $0.1026861 \doteq 10.3\%$ .
- This gives a 95% confidence interval for  $p_2 - p_1$ .
- Conclusion: From 2012 to 2015, U.S. support for legalising marijuana probably increased by between about 1.7% and 10.3%.

sta130–115

## Summary: Comparing Polls

- Suppose Poll #1 has sample size  $n_1$ , and true probability  $p_1$ , and observed fraction  $\hat{p}_1$ . Similarly  $n_2$  and  $p_2$  and  $\hat{p}_2$  for Poll #2.
- $\hat{p}_1$  is approximately normal, with mean  $p_1$ , and sd  $\sqrt{p_1(1-p_1)/n_1}$ . Similarly  $\hat{p}_2$  has mean  $p_2$  and sd  $\sqrt{p_2(1-p_2)/n_2}$ .
- So, the difference,  $\hat{p}_2 - \hat{p}_1$ , has mean  $p_2 - p_1$ , and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ . (You add variances, not sd. And,  $-\hat{p}_1$  has the same variance and sd as  $\hat{p}_1$  since  $(-1)^2 = 1$ .)
- Can estimate this sd by  $\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$  (bold option) or as  $\sqrt{1/n_1 + 1/n_2} / 2$  (conservative option).
- Then how do we test if  $p_1 = p_2$ ?
- And, how do we get confidence intervals for  $p_2 - p_1$ ?

sta130–116

## Summary: Comparing Polls (cont'd)

- To test the null hypothesis  $p_1 = p_2$ , versus the alternative that  $p_1 \neq p_2$ , we compute the P-value as the probability that a normal with mean 0 (since if  $p_1 = p_2$ , then mean =  $p_2 - p_1 = 0$ ), and sd from the above estimate, would be as large or larger than the observed difference, or as small or smaller than the negative of the observed difference.
- Or, for a 95% confidence interval for  $p_2 - p_1$ , we can take  $\hat{p}_2 - \hat{p}_1 \pm 1.96 sd$ , i.e. from  $\hat{p}_2 - \hat{p}_1 - 1.96 sd$  to  $\hat{p}_2 - \hat{p}_1 + 1.96 sd$ , with sd from the above estimate.
- Very similar to testing and confidence intervals for a single poll.
  - Here we're interested in  $p_2 - p_1$ , instead of just a single  $p$ .
  - The main difference is that the sd is now  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$  instead of just  $\sqrt{p(1-p)/n}$ .

sta130-117

## U.S. Marijuana Surveys Again: 2012 v. 2014

- In 2012, had  $n_1 = 1,100$ , and  $\hat{p}_1 = 0.47$ . In 2014, had  $n_2 = 1,018$ , and  $\hat{p}_2 = 0.51$ .
  - So, the observed value of  $|\hat{p}_2 - \hat{p}_1|$  was 0.04.
- So,  $\hat{p}_2 - \hat{p}_1$  should be approximately normal, with mean  $p_2 - p_1$ , and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ .
- Conservative option: estimate sd as  $\leq \sqrt{1/n_1 + 1/n_2} / 2 = \sqrt{1/1,100 + 1/1,018} / 2 \doteq 0.02174517$ .
- To test the hypothesis that  $p_2 = p_1$ , the P-value is the probability that a normal with mean  $p_2 - p_1 = 0$ , and sd  $\doteq 0.02174517$  is as large or larger than 0.04, or as small or smaller than  $-0.04$ . In R: `pnorm(0.04, 0, 0.02174517, lower.tail=FALSE) + pnorm(-0.04, 0, 0.02174517, lower.tail=TRUE)`. Answer = 0.0658433. More than 0.05! Cannot reject the null hypothesis!

sta130-118

## Confidence Interval for 2012 v. 2014

- In 2012, had  $n_1 = 1,100$ , and  $\hat{p}_1 = 0.47$ . In 2014, had  $n_2 = 1,018$ , and  $\hat{p}_2 = 0.51$ .
- So, the quantity  $\hat{p}_2 - \hat{p}_1$  should have mean  $p_2 - p_1$ , and sd  $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ .
- Conservative option: estimate sd as  $\leq \sqrt{1/n_1 + 1/n_2} / 2 = \sqrt{1/1,100 + 1/1,018} / 2 \doteq 0.02174517$ .
- So,  $p_2 - p_1$  should be between  $\hat{p}_2 - \hat{p}_1 - 1.96 \times 0.02174517$  and  $\hat{p}_2 - \hat{p}_1 + 1.96 \times 0.02174517$ . That is, between about  $0.04 - 1.96 \times 0.02174517$  and  $0.04 + 1.96 \times 0.02174517$ . That is, between about  $-0.002620533$  and  $0.08262053$ .
- Conclusion: From 2012 to 2014, U.S. support for legalising marijuana probably decreased by at most 0.26%, or increased by at most 8.26%; we can't say which.

sta130-119

## General Quantities, Besides Yes/No

- So far, we have mostly done statistics on Yes/No quantities. (Do you support the government? Is the coin heads? Does the die show 5? Did the roulette spin come up 22? etc.)
- Then we could study proportions or fractions or probabilities, and compute P-values and confidence intervals for them, and (now) compare them to each other, etc. Good!
- But what about quantities that don't involve just Yes/No? (Medicine: blood pressure, life span, weight gain, etc. Economics: GDP, stock price, company profits, etc. Social policy: number of accidents, amount of congestion, etc. Weather: amount of rain, wind speed, temperature, etc. Environment: global warming, ocean levels, contamination levels, atmospheric concentrations, etc. Sports: number of goals, time of possession, etc. Science: number of particles, speed of chemical reaction, etc.) Next!

sta130-120