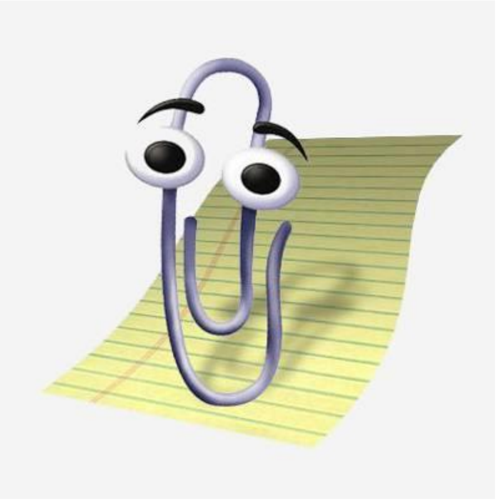


Quiz 2 Review

Lectures 6-8

(Normal Curve, Z Scores, Confidence Intervals, Hypothesis Testing)

April 1, 2016

https://www.directive.com/images/easyblog_images/16037/b2ap3_thumbnail_clippy400.jpg

Quiz 2
(Lectures 6-8)

DATE: Tuesday, April 5*

WHEN: 24-hour window opens 7:40 pm (ET)
and closes on April 6 at 7:40 pm (ET)

Location: Online, MATH E-3 Canvas course site
(no proctor needed)

*No class meeting or on campus section

You will have 75 minutes within the 24-hour window to complete the quiz.

Agenda

Jessica

Normal Curve

Z scores

Q & A

Sue

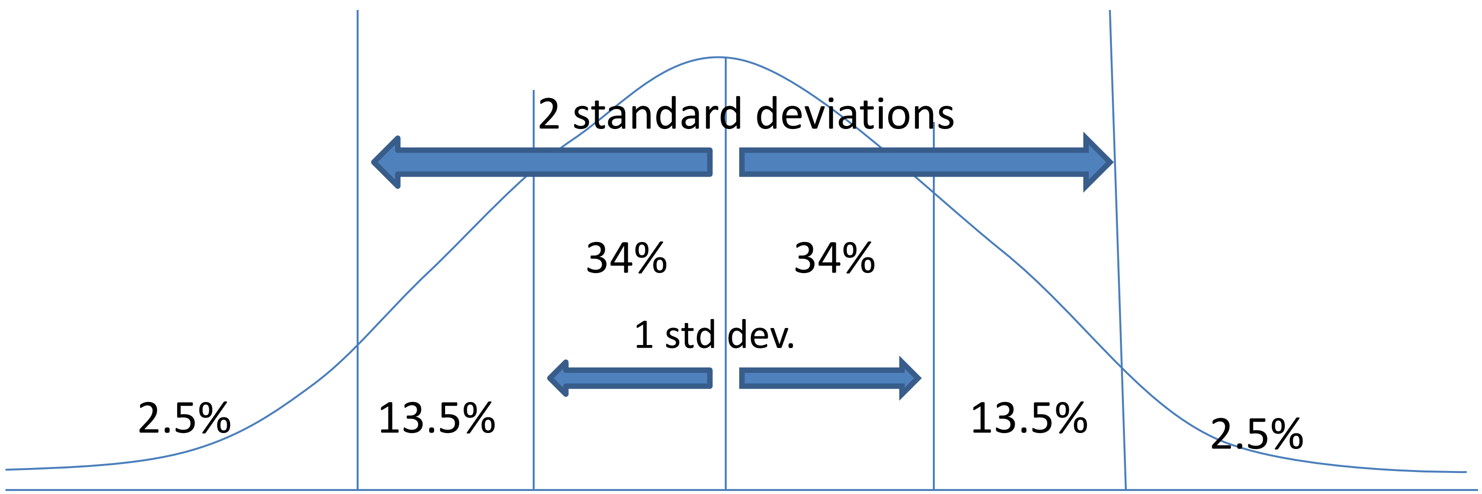
Confidence Intervals

Hypothesis Testing

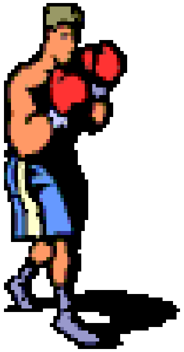
Q & A

Normal Distribution Curve

- Also called the 'Bell Curve'
- Data distributed normally



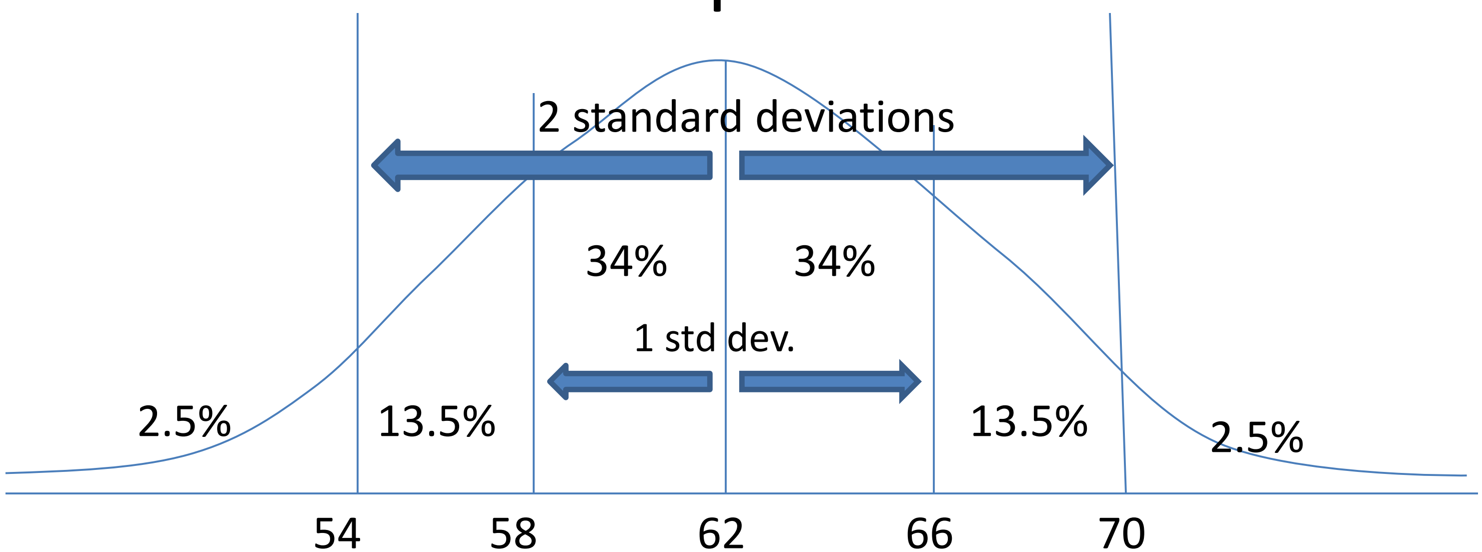
Rough and Ready Rules



- 68% of data lie within 1 standard deviation
- 95% of data lie within 2 standard deviations
- 99.7% of data lie within 3 standard deviations
- 100% of the data under the entire curve

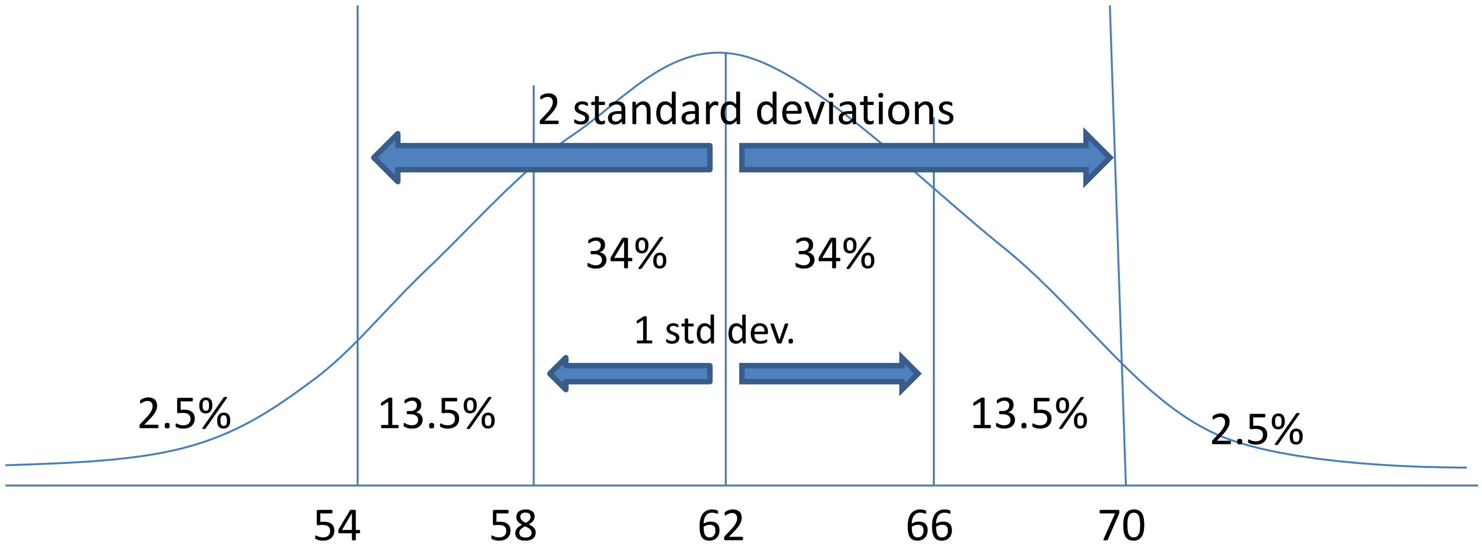
Using Rough and Ready

- Weights of Australian Shepherd dogs
- Mean = 62 pounds
- Standard deviation 4 pounds



Using Rough and Ready

- What % of Aussies are greater than 66 pounds?
- What % of Aussies are between 58 and 66 pounds?
- What percent of Aussies are less than 54 pounds?



Aussies continued

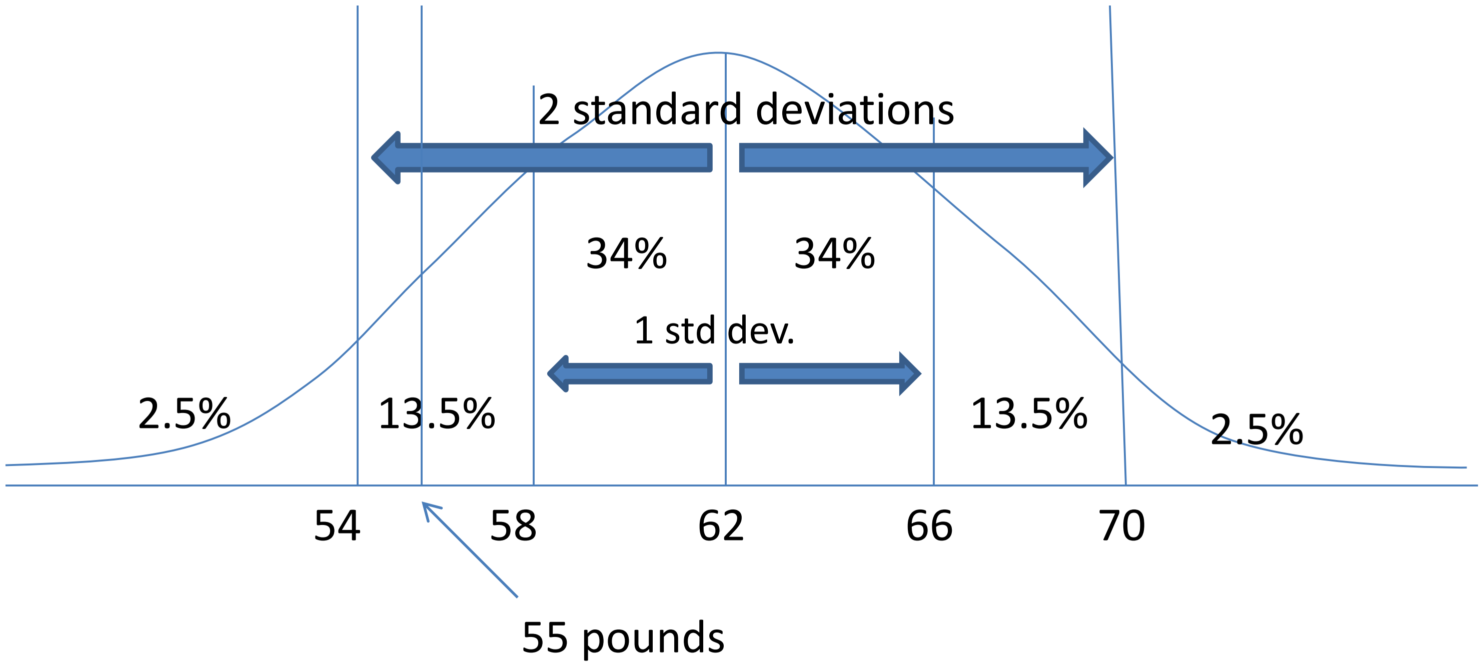
- What % of Aussies are greater than 66 pounds?
  - ✓ Answer: 13.5%+2.5%=16%
- What % of Aussies are between 58 and 66 pounds?
  - ✓ Answer: 34% +34%=68%
- What percent of Aussies are less than 54 pounds?
  - ✓ Answer: 2.5%

So, what if we need to calculate % but not at exactly 1 or 2 or 3 s.ds?

- We use z scores
- Z scores will give the amount under the normal curve when we are not dealing with neat increments of standard deviations
- The z score tables are based on the number of standard deviations we are away from the mean

Aussie example using z scores

For our Aussie example, what if we wanted to know what percent of Aussies are between 55 and 62 pounds?



First calculate how many standard deviations you are from the mean

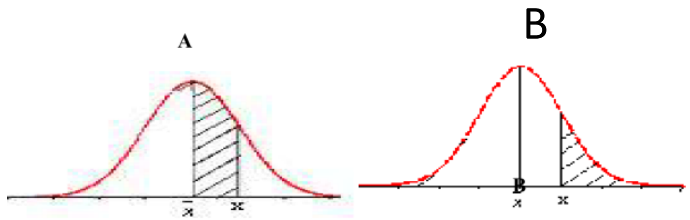
- $Z = \left| \frac{x - \bar{x}}{\sigma} \right|$
- The X is our data value = 55 pounds
- The σ or standard deviation =4 pounds
- The mean, X bar, is 62 pounds
- So let's calculate z

Z scores

$$\frac{|55-62|}{4} = \frac{7}{4} = 1.75 = 1.8 \text{ (1 dp)}$$

Now look at table A of z scores at 1.8

The z score tables



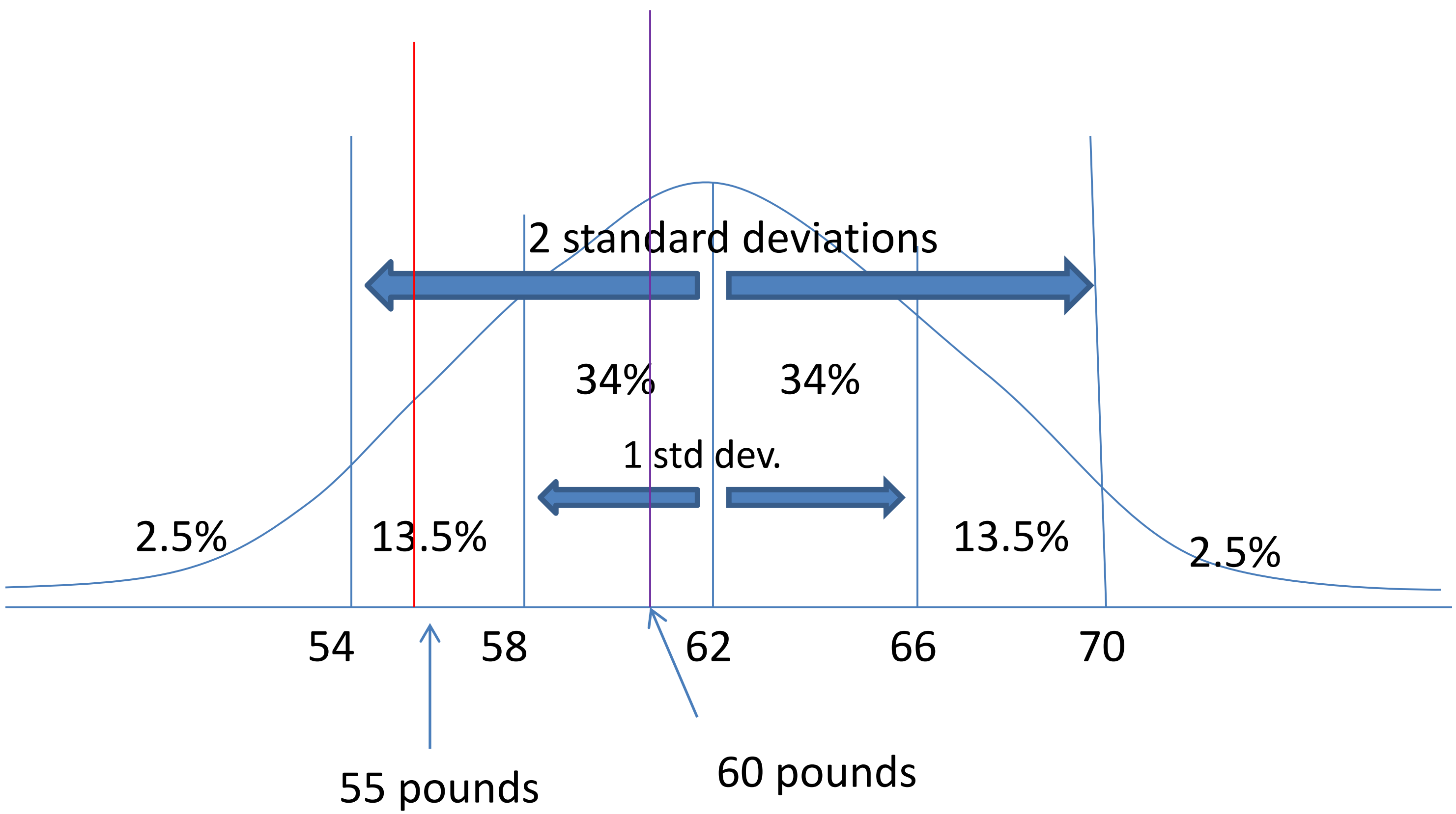
Z	Area between the Mean and x (curve A)	Area beyond x (curve B)
0	0.0000	0.0000
0.1	0.0398	0.4602
0.2	0.0793	0.4207
0.3	0.1179	0.3821
0.4	0.1554	0.3446
0.5	0.1915	0.3085
0.6	0.2257	0.2743
0.7	0.2580	0.2420
0.8	0.2881	0.2119
0.9	0.3159	0.1841
1	0.3413	0.1587
1.1	0.3643	0.1357
1.2	0.3849	0.1151
1.3	0.4032	0.0968
1.4	0.4192	0.0808
1.5	0.4332	0.0668
1.6	0.4452	0.0548
1.7	0.4554	0.0446
1.8	0.4641	0.0359
1.9	0.4713	0.0287
2	0.4772	0.0228

Z scores

- At 1.8, the z score table says .4641, or 46.41%
- So the answer here is 46.4% of the Aussies are between 55 pounds and 62
- Okay, great but what if you weren't measuring from the mean to 55?
- What if you wanted to know what percent were between 55 and 60 pounds?

Z scores

What if we want to know what % of Aussies are between 55 and 60 pounds …



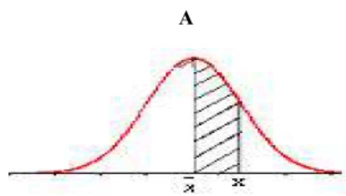
Z scores

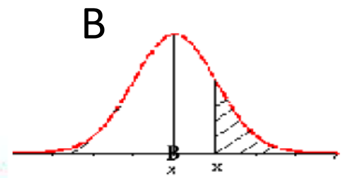
- We already calculated how many standard deviations 55 was from the mean: that was 1.8
- Now we need to calculate how many 60 is from the mean:
- $\frac{60-62}{4} = \frac{2}{4} = .5$

Z scores

- Between the mean of 62 and 1.8 standard deviations there is 46.41% of the data
- Now look at .5 standard deviations
- There is .1915 or 19.15% of the data
- Subtract .1915 from .4641 to get the amount between 55 and 60 pounds
- .4641-.1915 =.2726 =27.26%

The z score tables





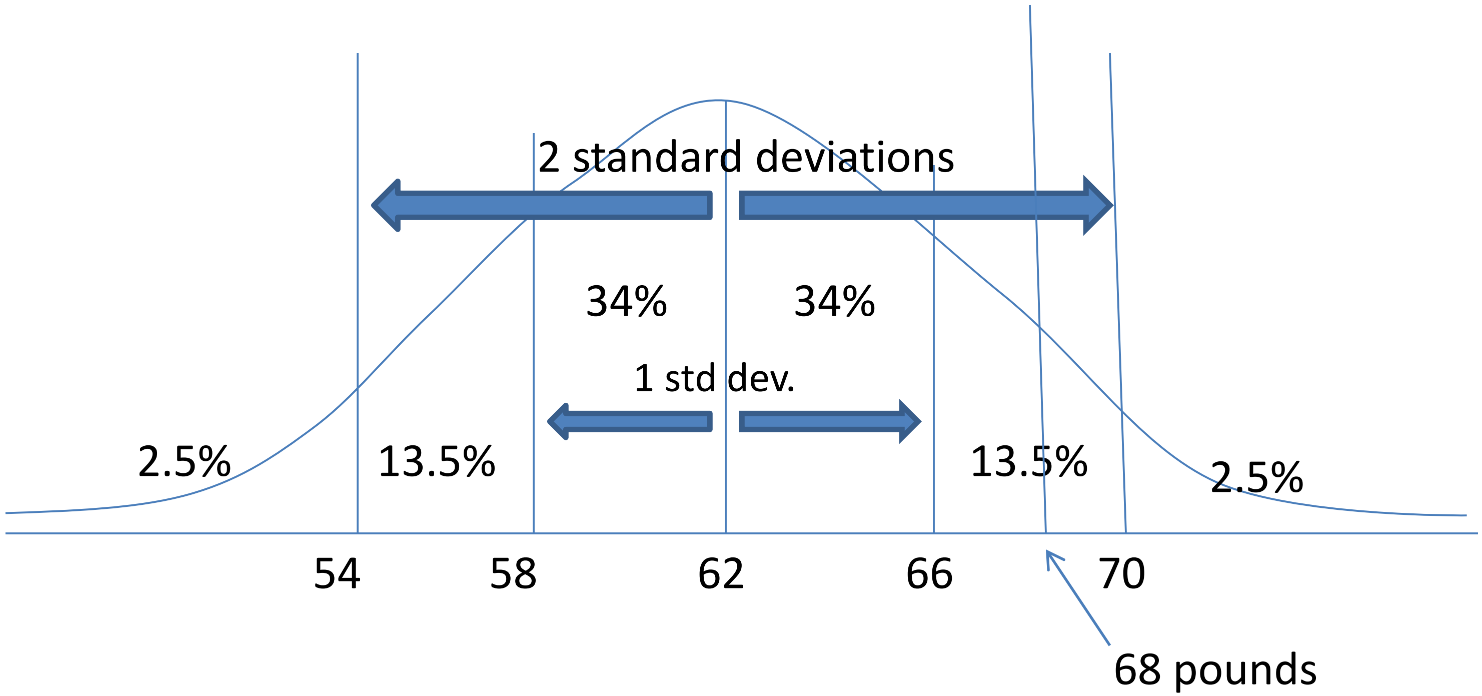
Z	Area between the Mean and x (curve A)	Area beyond x (curve B)
0	0.0000	0.0000
0.1	0.0398	0.4602
0.2	0.0793	0.4207
0.3	0.1179	0.3821
0.4	0.1554	0.3446
0.5	0.1915	0.3085
0.6	0.2257	0.2743
0.7	0.2580	0.2420
0.8	0.2881	0.2119
0.9	0.3159	0.1841
1	0.3413	0.1587
1.1	0.3643	0.1357
1.2	0.3849	0.1151
1.3	0.4032	0.0968
1.4	0.4192	0.0808
1.5	0.4332	0.0668
1.6	0.4452	0.0548
1.7	0.4554	0.0446
1.8	0.4641	0.0359
1.9	0.4713	0.0287
2	0.4772	0.0228

Z scores

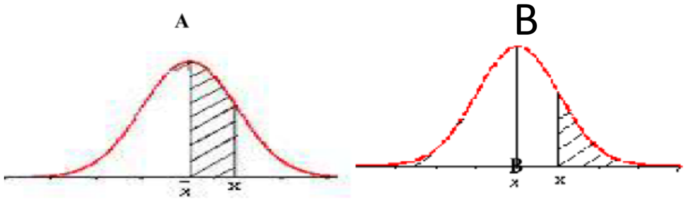
- What do we use Table B for ?
- That is the % of data under the curve BEYOND whatever value you are looking for
- So if we want to know what % of Aussies are OVER 68 pounds
- Calculate z score: $\frac{68-62}{4} = \frac{6}{4} = 1.5$

Aussie example using z scores

For our Aussie example, what if we wanted to know what percent of Aussies are over 68 pounds?



The z score tables



Z	Area between the Mean and x (curve A)	Area beyond x (curve B)
0	0.0000	0.0000
0.1	0.0398	0.4602
0.2	0.0793	0.4207
0.3	0.1179	0.3821
0.4	0.1554	0.3446
0.5	0.1915	0.3085
0.6	0.2257	0.2743
0.7	0.2580	0.2420
0.8	0.2881	0.2119
0.9	0.3159	0.1841
1	0.3413	0.1587
1.1	0.3643	0.1357
1.2	0.3849	0.1151
1.3	0.4032	0.0968
1.4	0.4192	0.0808
1.5	0.4332	0.0668
1.6	0.4452	0.0548
1.7	0.4554	0.0446
1.8	0.4641	0.0359
1.9	0.4713	0.0287
2	0.4772	0.0228

Z scores

- At 1.5 standard deviations for Table B, we get .0668 or 6.68%
- So the percentage of Aussies over 68 pounds =6.68%

Sample quiz questions

- Normal Distribution
- Z Scores

Normal Distribution

The New England Patriots' heights are normally distributed with a mean height of 6'2" and a standard deviation of 3"

Normal Distribution Questions

What percent of the Patriots have a height:

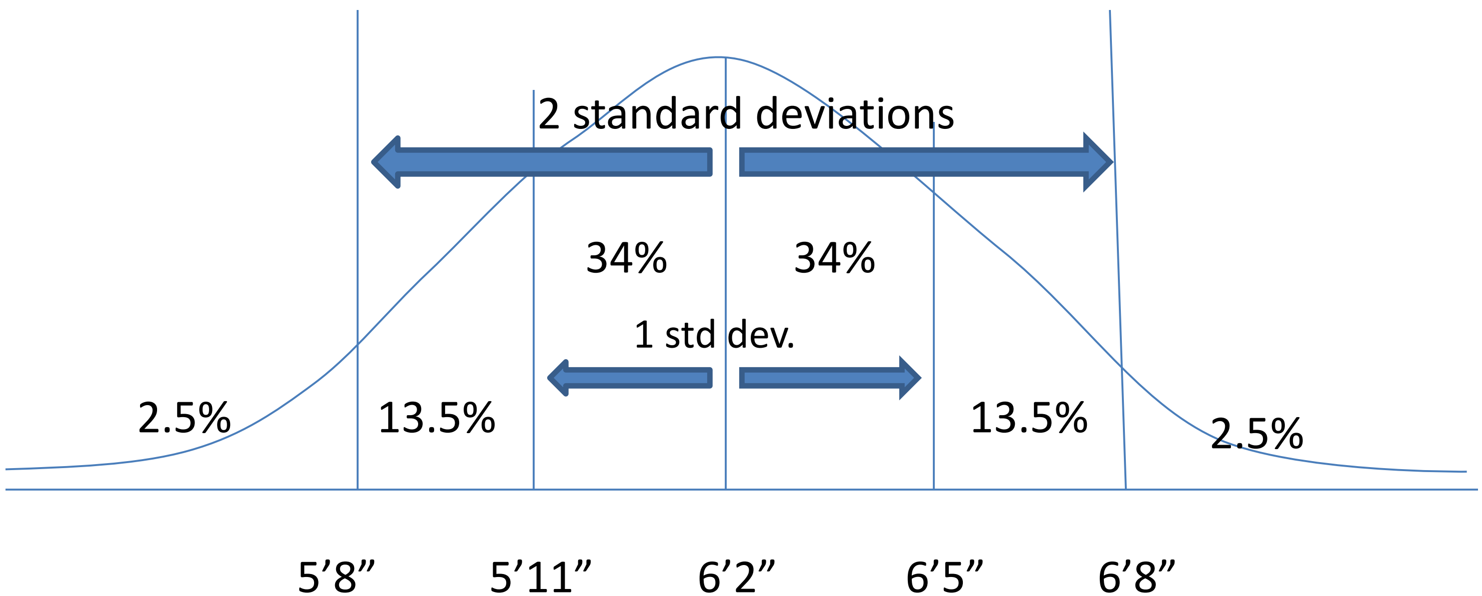
- between 5’11” and 6’5”?
- over 6’2”?
- under 5’11”?
- between 6’5” and 6’8”?

What to do first?

Draw the curve:

Patriots heights

- between 5’11” and 6’5”?
- over 6’2”?
- under 5’11”?
- between 6’5” and 6’8”?



Normal Distribution Questions

What percent of the Patriots have a height:

- between 5’11” and 6’5”? 68%
- over 6’2”? 50%
- under 5’11”? 16%
- between 6’5” and 6’8”? 13.5%

Z score problem

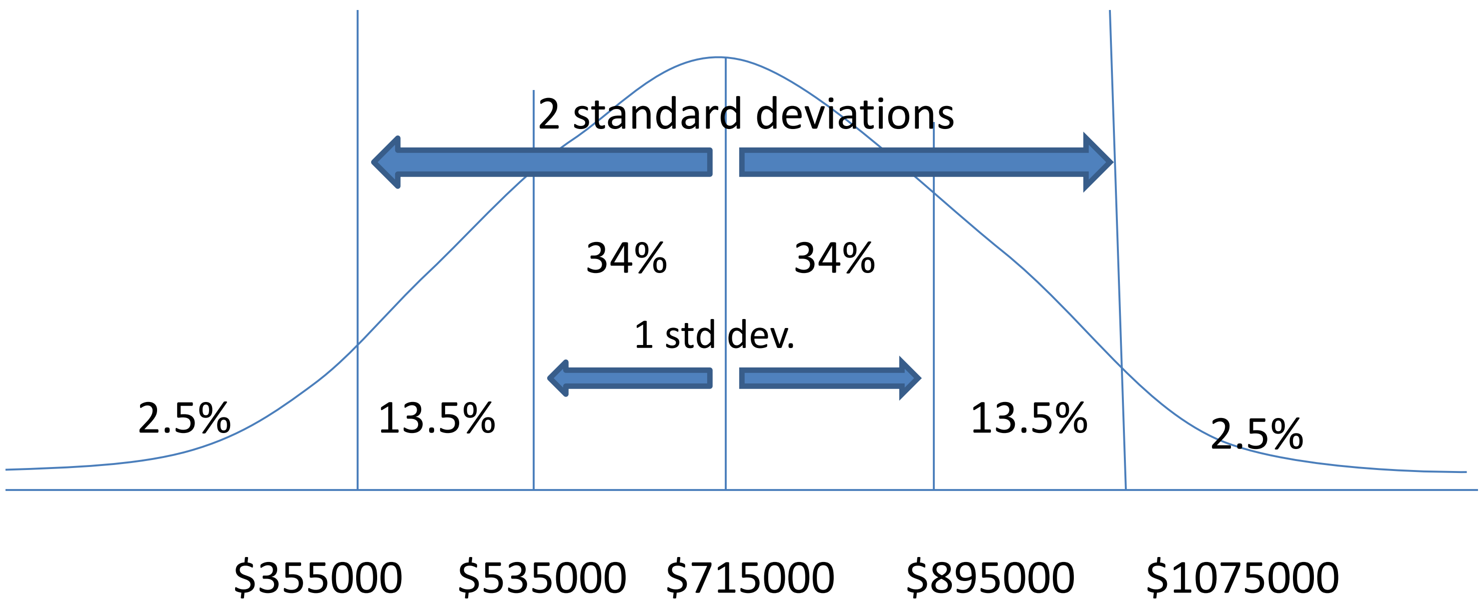
- The price of homes in a pricey suburb of Boston is normally distributed with a mean price of \$ 715,000 and a standard deviation of \$180,000.
- What percentage of the homes cost between 750,000 and 850,000? Give as percent rounded to nearest whole percent?

A)42% B)34% C)21% D)50%

What to do first?

Draw the curve:

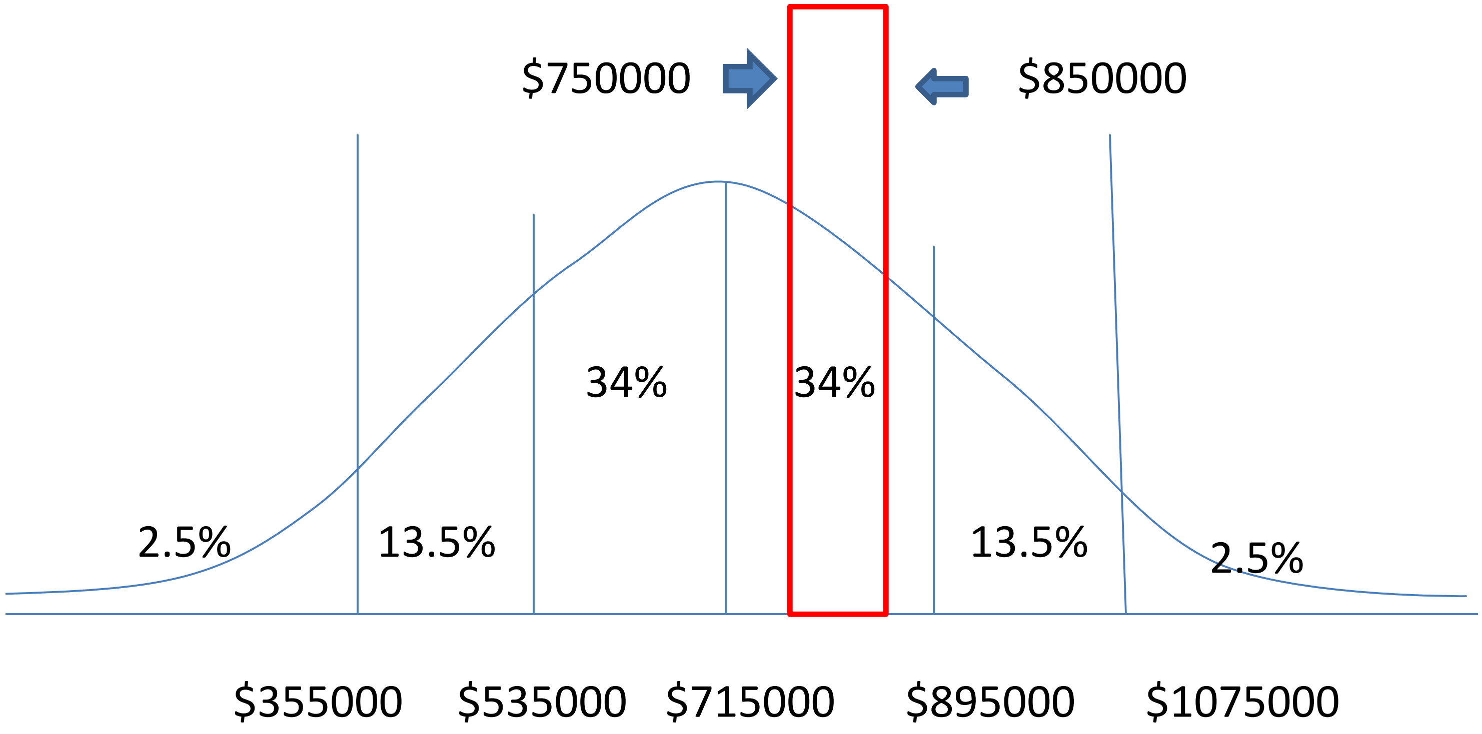
Prices of homes in Way- to- Well-Offsly



What are we looking for ?

Try to answer this without doing the calculation.

A)42% B)34% C)21% D)50%



How do we find the percentage using Z scores?

- The Z score tables are calculated on the distance from the mean
- We are inside the first standard deviation so we know percentage must be less than 34%
- Calculate percentage from mean to $850000
- Calculate percentage from mean to $750000
- Subtract second percentage from first

Calculate z scores

For 750,000:

$$\frac{750000-715000}{180000} = \frac{35000}{180000} = .1944..$$

Round to .2

For 850000:

$$\frac{850000-715000}{180000} = \frac{135000}{180000} = .75$$

Round to .8

Z score table

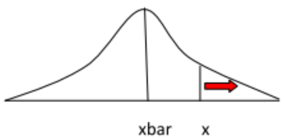
Z -Scores*

$$Z = \frac{x - x\ bar}{\sigma}$$

x bar=mean

A

B



Z	Area between the Mean and X (curve A)	Area beyond X (curve B)
.1	.0398	.4602
.2	.0793	.4207
.3	.1179	.3821
.4	.1554	.3446
.5	.1915	.3085
.6	.2257	.2743
.7	.2580	.2420
.8	.2881	.2119
.9	.3159	.1841
1	.3413	.1587
1.1	.3643	.1357
1.2	.3849	.1151

Calculate the percentage

- .2881 -.0793 =.2088 =20.88% or 21%
- 21% of the homes cost between $750000 and $850000
- A)42% B)34% C)21% D)50%

What to do first?

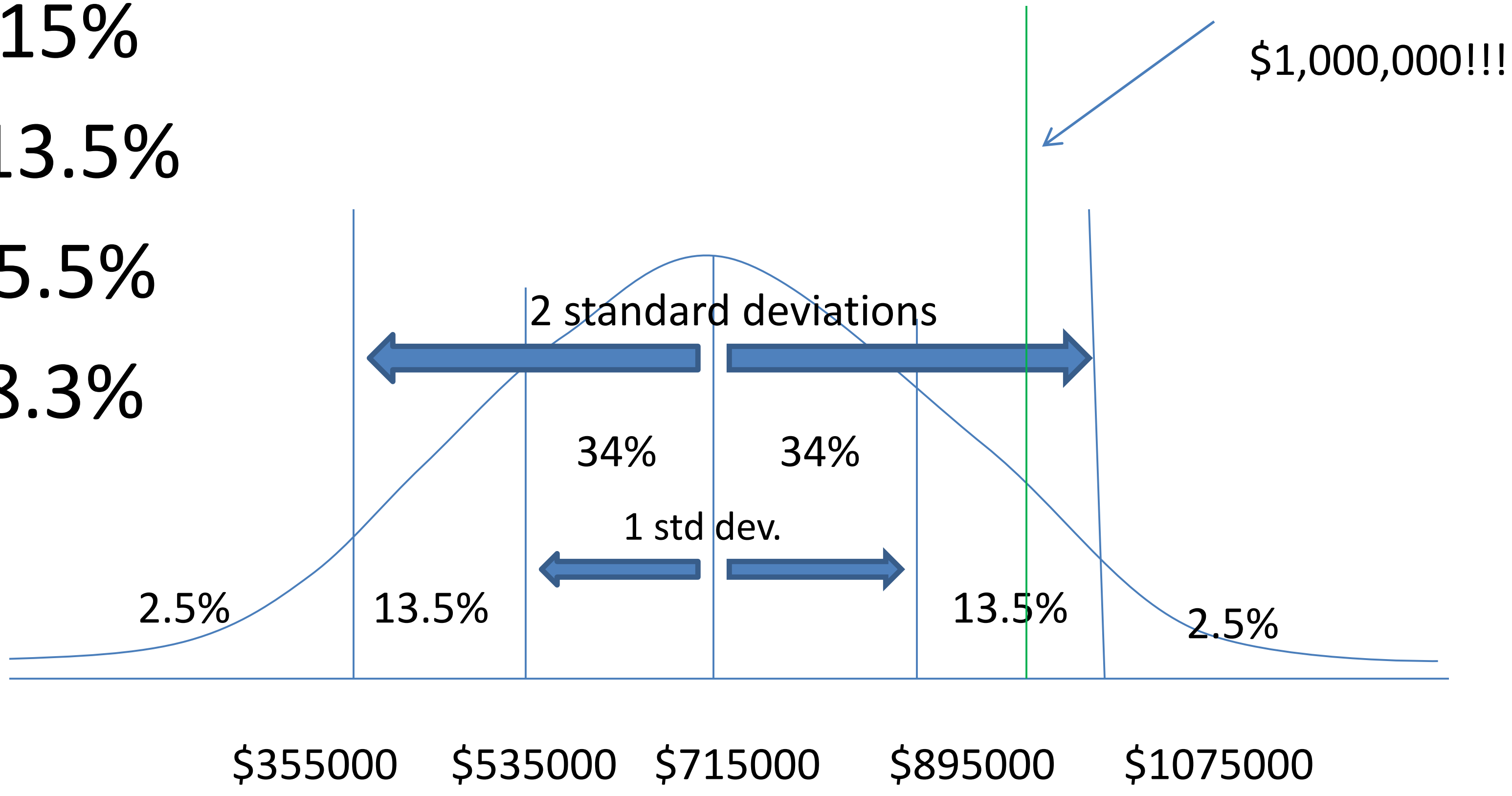
How many cost **more** than $1,000,000?

A) 15%

B)13.5%

C) 5.5%

D)8.3%



Z scores

How many cost more $1,000,000?

A) 15%

B)13.5%

C) 5.5%

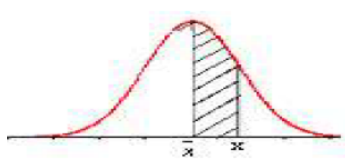
D)8.3%

How did we get that?

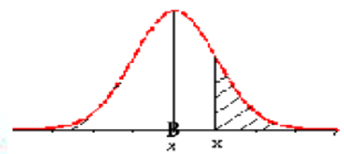
- Calculate the z score:
- $$\frac{1{,}000{,}000 - 715{,}000}{180{,}000} = \frac{285{,}000}{180{,}000} = 1.58333$$
- Round to 1.6
- Look at Table B

The z score tables

A



B



Z	Area between the Mean and x (curve A)	Area beyond x (curve B)
0	0.0000	0.0000
0.1	0.0398	0.4602
0.2	0.0793	0.4207
0.3	0.1179	0.3821
0.4	0.1554	0.3446
0.5	0.1915	0.3085
0.6	0.2257	0.2743
0.7	0.2580	0.2420
0.8	0.2881	0.2119
0.9	0.3159	0.1841
1	0.3413	0.1587
1.1	0.3643	0.1357
1.2	0.3849	0.1151
1.3	0.4032	0.0968
1.4	0.4192	0.0808
1.5	0.4332	0.0668
1.6	0.4452	0.0548
1.7	0.4554	0.0446
1.8	0.4641	0.0359
1.9	0.4713	0.0287
2	0.4772	0.0228

The Answer is…

- .0548 =
- 5.48%
- Rounds to 5.5%

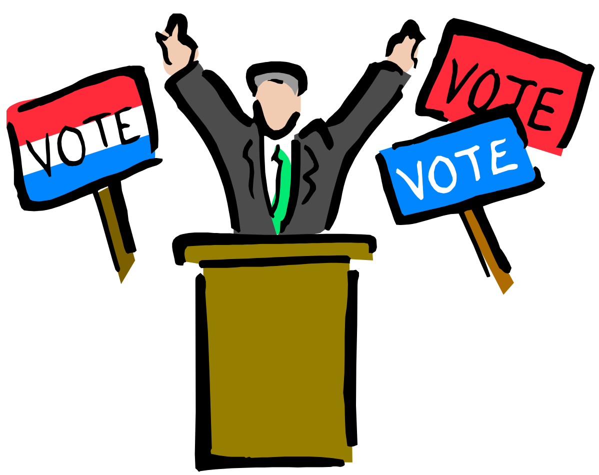
Q&A

Confidence Intervals

What are Confidence Intervals?

Confidence intervals provide a **range of possible values** for the true value of the population with which we are interested.

We frequently use confidence intervals to determine the popularity of a political candidate through polling.

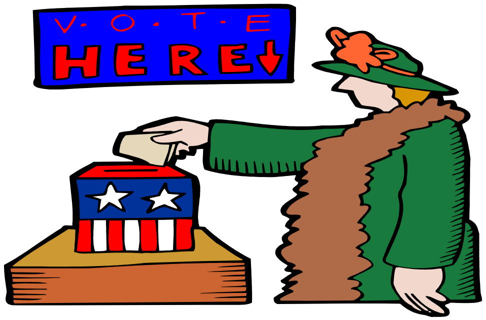


Random Sample

- In order to be able to construct a confidence interval we must start with a **random sample** from a population of interest.
- A **sample** is random if each member of the **population** has an equal chance of being selected.
- In other words, the sample must be **representative** of the entire population and not **be biased**.

Why we use confidence intervals simply stated…

If a campaign manager wants to know the popularity his candidate for governor, Mr. Here We Go Again, three months before the election.   He would have to draw a random sample of voters from the entire state – not just from the cities or the rural areas. His question would be, "Would you vote for my candidate if the election were held today?"



Why we use confidence intervals simply stated…

- Let’s say that out of a random sample of 500 voters from his home state of Euphoria, 47% of them said that they, indeed, would vote **Mr. Here-We-Go-Again**.

- This could be disappointing news, but all is not lost. Keep in mind that his percentage is based upon one sample of 500 voters.

- Another sample certainly would yield a different result. A third sample would yield yet another percentage. This is called sample variability.

Why we use confidence intervals simply stated…

- In this case, the campaign manager would use the 47% from his sample to create a confidence by “hedging his bets” with a **margin of error** on either side of the 47%.

- Collectively, the margin of error above and below his sample percentage of 47% is the confidence intervals.

- Any percentage within this interval is a possible percentage of the population of voters from his state.

New Standard Deviation formula

Remember that for confidence intervals (and hypothesis testing), we will be using the new standard deviation formula.

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

Steps to Constructing a Confidence Interval

1. **Obtain** **a value for p based upon your sample**
2. **Calculate** **the standard deviation based on this observed percentage, p. Use the new standard deviation formula.**
3. **Calculate** **p plus or minus 2 standard deviations. Remember that "two times the standard deviation" is the Margin of Error.**
4. **Construct** **the confidence interval and explain what the confidence interval tells us.**

Let’s Construct a Confidence Interval

- We have taken an unbiased poll of 500 people from **Mr. Here-We-Go-Again’s** home state of Euphoria to find out the popularity at this point in the campaign.

- Of these 500 people that we have polled, 235 of them said they would likely vote for him in the next election.

- Let’s construct a confidence interval for the true proportion of Massachusetts voters who approve of Mr. Here-We-Go-Again and would vote for him.

Step 1: Obtain a value for P based upon your sample

Obtain a value for p based upon our sample.   We polled 500 randomly selected voters and 235 of them said that they would likely vote for Mr. Here-We-Go-Again. Thus, our sample percent is 47%.

$$p = \frac{235}{500} = 47\%$$

N = 500

2. Calculate the standard deviation based on the observed percentage.

σ = √ $\frac{p(1-p)}{500}$ = √ $\frac{.47(.53)}{500}$ = 0.0223 = **2.2% (1 dp)**

Step 3: Calculate p plus or minus 2 standard deviations

47% plus or minus 2(2.2%) or

47% plus or minus 4.4% or

42.6% to 51.4%

Step 4: Construct the 95% Confidence Interval and explain what the interval tells us

42.6% $\longleftrightarrow$ 51.4%

The 95% confidence interval for the true proportion of voters who would likely vote for Mr. Here-We-Go-Again is between 42.6% and 51.4%.

PROBLEM:

The Dean of the Math Department of a major university wants to hire more male math teachers. The women professors are opposed to this. They tell the Dean that too many MALE math teacher succumb to the pressures of teaching math and die before reaching the age of 55. The Dean consults with several of his colleagues across the country and collects his own data. He finds that out of a random sample of 350 math teachers, only 21 die before the age of 55.

Question 1: What is the margin of error for this information statement?

A. 2.6%
B. 1.3%
C. 3.6%
D. 26%
E. 10.8%

Info from problem to do the confidence interval:
The Dean consults with several of his colleagues across the country and collects his own data. He finds that out of a random sample of 350 math teachers, only 21 die before the age of 55.

Question 1: What is the margin of error for this information statement?

What information do you need to know?

the sample percentage : 21/350 = .06 or 6%

the standard deviation $\sigma = \sqrt{\frac{.06(1-.06)}{350}}$ = .0126 = 1.3% (1 dp)

The Dean consults with several of his colleagues across the country and collects his own data. He finds that out of a random sample of 350 math teachers, only 21 die before the age of 55.

Question 1: What is the margin of error for this information statement?

What information do you need to know?

the sample percentage : 21/350 = .06 or 6%

the standard deviation $\sigma = \sqrt{\frac{.06(1-.06)}{350}}$ = .0126 = 1.3% (1 dp)

margin of error (MOE) = 2 * standard deviation

The Dean of the Math Department of a major university wants to hire more male math teachers. The women professors are opposed to this. They tell the Dean that too many MALE math teacher succumb to the pressures of teaching math and die before reaching the age of 55. The Dean consults with several of his colleagues across the country and collects his own data. He finds that out of a random sample of 350 math teachers, only 21 die before the age of 55.

Question 1: What is the margin of error for this information statement?

A. 2.6%

B. 1.3%

C. 3.6%

D. 26%

E. 10.8%

The Dean of the Math Department of a major university wants to hire more male math teachers. The women professors are opposed to this. They tell the Dean that too many MALE math teacher succumb to the pressures of teaching math and die before reaching the age of 55. The Dean consults with several of his colleagues across the country and collects his own data. He finds that out of a random sample of 350 math teachers, only 21 die before the age of 55.

Question 1: What is the margin of error for this information statement?

A. 2.6%
B. 1.3%
C. 3.6%
D. 26%
E. 10.8%

We found the sample percentage was 6%, the standard deviation was 1.3%, and the margin of error was 2.6%.

Question 2: Shown below are five 95% Confidence Intervals supposedly representing the true proportion of the male math teachers that succumb before the age of 55. Circle the 95% CI that the Dean reported to the women professors. The 95% confidence interval is:

A. 34% ⟷ 86%
B. 6% plus or minus 1.3%
C. 49.6% ⟷ 60.4%
D. 55% plus or minus 1.8%
E. 3.4% ⟷ 8.6%

The Dean of the Math Department of a major university wants to hire more male math teachers. The women professors are opposed to this. They tell the Dean that too many MALE math teacher succumb to the pressures of teaching math and die before reaching the age of 55. The Dean consults with several of his colleagues across the country and collects his own data. He finds that out of a random sample of 350 math teachers, only 21 die before the age of 55.

Question 2: Shown below are five 95% Confidence Intervals supposedly representing the true proportion of the male math teachers that succumb before the age of 55. Circle the 95% CI that the Dean reported to the women professors. The 95% confidence interval is:

A. 34% ⟷ 86%

B. 6% plus or minus 1.3%

C. 49.6% ⟷ 60.4%

D. 55% plus or minus 1.8%

E. 3.4% ⟷ 8.6%

(sample % = 6% -/+ 2.6)

Hypothesis Testing

What is Hypothesis Testing?

Hypothesis testing is a statistical procedure in which we test the validity of a **claim** that has been made that we suspect may not be true.




Hypothesis Testing

- Step 1: State your Null Hypothesis
  - Important to have a claim percentage
- Step 2: Calculate your standard deviation using new formula
  - Important to use the claim percentage in your standard deviation calculation
- Step 3: Draw your normal curve with your claim percentage as the mean

Hypothesis Testing

- Step 4: See if your observed percentage is inside the likely region
- Step 5: State your conclusion in proper statistical language
  - If your observed result is **outside $p \pm 2\sigma$**, we can **reject** the NH at a 5% los (level of significance)
  - If your observed result is inside $p \pm 2\sigma$, we cannot reject the NH at a 5% los

Hypothesis Testing

- Step 6: conclusion in your own words

Hypothesis Testing

- A car company claims only 2% of their cars needs repairs in the first 2 years of operation. You test it and find that the 35 out 1600 need service.
- What is the claim percentage?
- What is the observed percentage?

Hypothesis Testing

- A car company claims only 2% of their cars needs repairs in the first 2 years of operation. You test it and find that the 35 out 1600 need service.
- What is the claim percentage? 2%
- What is the observed percentage?

$$\frac{35}{1600} = 2.2\%$$

Hypothesis Testing

- Sometimes you need to do a little bit of thinking to get to the claim percentage
- Let’s do an example

Hypothesis Testing Example

- A college admissions officer is worried about the image of her school in relation to other comparable ones in her state, and whether she will be able to attract the most promising students.

- She feels that her school may be slipping.

College Admissions Example

- She decides to ask a grad student in the Statistics department to test whether her school, Excellence U, is equally attractive to the most sought-after students as The University of Exclusivity, the College of the Cerebral, and the University of OB (Only the Best).
- The grad student decides to conduct a hypothesis test.

Grad Student Survey

He does a survey of 460 likely high school seniors, and gets the following results:

School	# of students who prefer	% of students who prefer
Excellence U	105	105/460=22.8%
U of Exclusivity	120	120/460=26.1%
College of the Cerebral	125	125/460=27.2%
University of Only the Best	110	110/460=23.9%

College Admissions Hypothesis Test

What is his null hypothesis?

a)Students prefer Excellence University 22.8% of the time.

b) Students are equally impressed by these prestigious institutions and would choose them equally, meaning 33.3% per university.

c)Students would choose Excellence U 50% or more of the time.

d) Students would choose Excellence U 25% of the time.

College Admissions Hypothesis Test

What is his null hypothesis?

a) Students prefer Excellence University 22.8% of the time.

b) Students are equally impressed by these prestigious institutions and would choose them equally, meaning 33.3% per university.

c) Students would choose Excellence U 50% or more of the time.

d) Students would choose Excellence U 25% of the time.

Why is it d?

Well, it’s not:

a) Students prefer Excellence University 22.8% of the time.

- 22.8% is an observed percentage
- The Null Hypothesis should contain your claim percentage
- The observed percentage is used to compare against your likely region
- The observed percentage would never be your claim percentage

Why d???

Well it’s not:

b)Students are equally impressed by these prestigious institutions and would choose them equally, meaning 33.3% per university

There 4 schools including your school, Excellence U

If you are trying to see if the schools are equally attractive to students, you would have an equal percentage of those surveyed choosing them

100%/4 choices =25% apiece

Why d?

And, it’s not c

c)Students would choose Excellence U 50% of the time

SEE LAST SLIDE, AND THE WINNER IS:

d) Students would choose Excellence U 25% of the time.

Hypothesis Testing

Important stuff:

Never use the word accept –it is “cannot reject”

Use your claim percentage when calculating your standard deviation

Our standard deviation for College Admissions would use what?

Hypothesis Testing

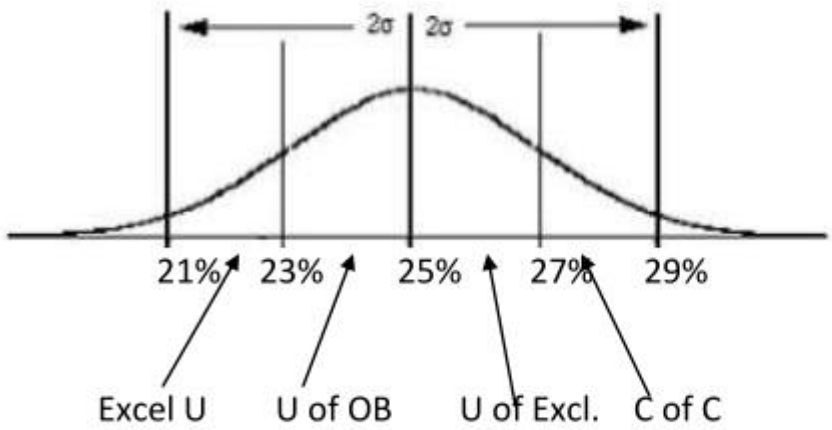
- 25%
- So, calculate the standard deviation using our formula:
- σ= $\sqrt{\frac{p(1-p)}{n}}$

Hypothesis Test

- Here it is:
- σ= $\sqrt{\dfrac{.25(1-.25)}{460}}$
- .020189321 =2.0189321% =2.0% = S.D.
- Change this to a percentage always
- Round to 1 dp unless otherwise instructed
- Always round last – in other words, after you convert to a percentage

Draw the curve

Draw curve:



Calculate observed %s:

Excellence U	105/460 =22.8%
U of Exclusivity	120/460=26.1%
College of the Cerebral	125/460 =27.2%
University of OB	110/460=23.9%

So answer this…

Which conclusion is correct?

a. We can reject the NH at a 5% los because our observed % is inside p±2σ
b. We cannot reject the NH at a 5 % los because our observed % is outside the p±2σ
c. We cannot reject the NH at a 5% los because our observed % is inside the p±2σ
d. We accept the NH at a 5% los because our observed % is inside the p±2σ

So answer this…

Which conclusion is correct?

a. We can reject the NH at a 5% los because our observed % is inside p±2σ
b. We cannot reject the NH at a 5 % los because our observed % is outside the p±2σ
c. We cannot reject the NH at a 5% los because our observed % is inside the p±2σ
d. We accept the NH at a 5% los because our observed % is inside the p±2σ

Quiz 2 TIPS

- **Review** readings, homework assignments, lecture slides
- **Watch** the review section video and review section slides
- **Complete** the Math E-3 Practice Test

GOOD LUCK!