This class is being conducted over Zoom. As the instructor, I will be recording this session. I have disabled the recording feature for others so that no one else will be able to record this session. I will be posting this session to the course's website.

If you have privacy concerns and do not wish to appear in the recording, you may turn video off (click "stop video") so that Zoom does not record you.

The chat box is always open for discussion and questions to the entire class. You may also send messages privately to the instructor or the TAs. Please note that Zoom saves all chat transcripts.

I create a live transcription of each session using Otter.ai. This means that Otter.ai will transcribe anything spoken over the Zoom audio. The transcript will be posted with the session video on the course website.

Continuous Probability Distributions

Stats 7

Mary Ryan

Aug. 25, 2020


Course website:

https://canvas.eee.uci.edu/courses/28451


Slides can be found at:

https://maryryan.github.io/stats7-SS2-2020-slides/stats7-SS2-2020-contDist/stats7-SS2-2020-contDist

Probability Distributions

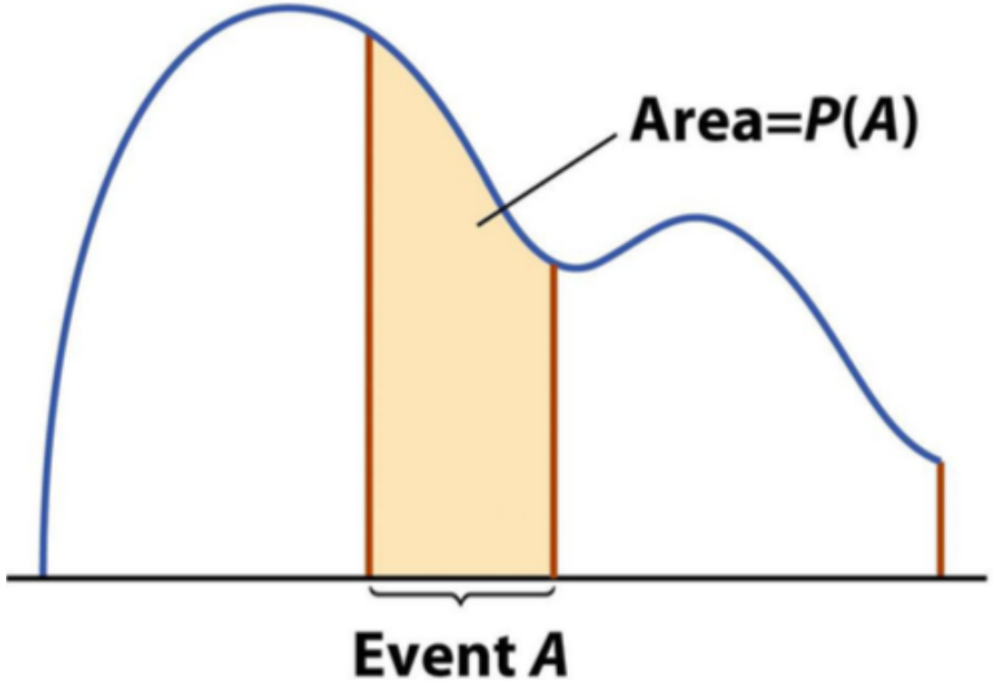
- Remember from last week: we might want to know the probability of an event, but have no data
  - Might be useful to apply a **probability model** or **probability distribution**
  - Existing framework with **known properties**, given that certain **conditions** apply
  - Each distribution has formulae for calculating probability
  - Different types of models for **discrete** and **continuous** variables
    - Today we'll focus on continuous random variables

Continuous Random Variables

- Like continuous data types, continuous random variables are variables that take on numerical values that can go out to **infinite decimal points**
- Continuous sample spaces have an **infinite number of outcomes** over an interval of values
  - But the probability over certain segments of those sample spaces are not necessarily constant
  - We use **density curves** to visualize where probability mass lies
  - The more probability mass an interval has, the more likely it is to happen!

Probability Density Curves

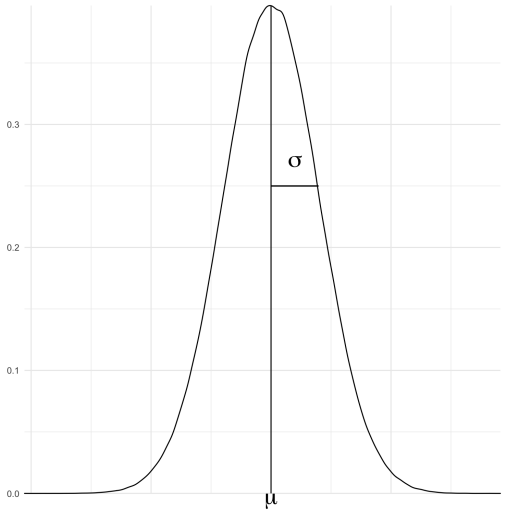
- Events are defined over an interval or range of values
  - $a \leq X < b, X \geq a, X < b$
- Probability is defined as the area under the curve of certain portions of the density curve
- The area under the entire curve will be equal to 1 (or 100% of some event in the entire sample space occurring)
- The area under one specific value is 0 -- therefore the probability of an event being exactly a certain value is 0.



Density Curves & Probability Density Functions

- The shape of a density curve is defined by something called a **probability density function**, or pdf
- The pdf is the continuous equivalent of the probability equations we saw for discrete distributions
- The pdf will also help us find the area under the curve for the specific area we want -- just integrate!
    - This might take quite a bit of math, though, so we often just have a computer calculate the probabilities for us

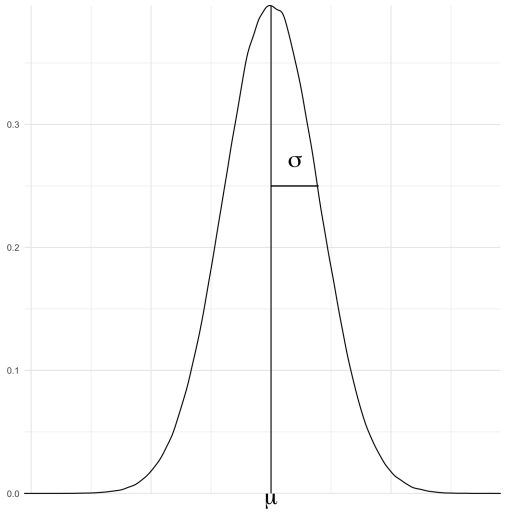
Normal Distribution



- Applies to continuous random variables
  - Variables can take on values that exist between ∞ and $-\infty$ (we fudge this conditic
  - Defined by the mean, μ, and variance, σ^2
  - Intervals near the mean more likely to happen than intervals near "tails"

- Some properties:
  - $P(X < u) = \int_{-\infty}^{u} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$
  - $E(X) = \mu$
  - $Var(X) = \sigma^2$

Normal Distribution

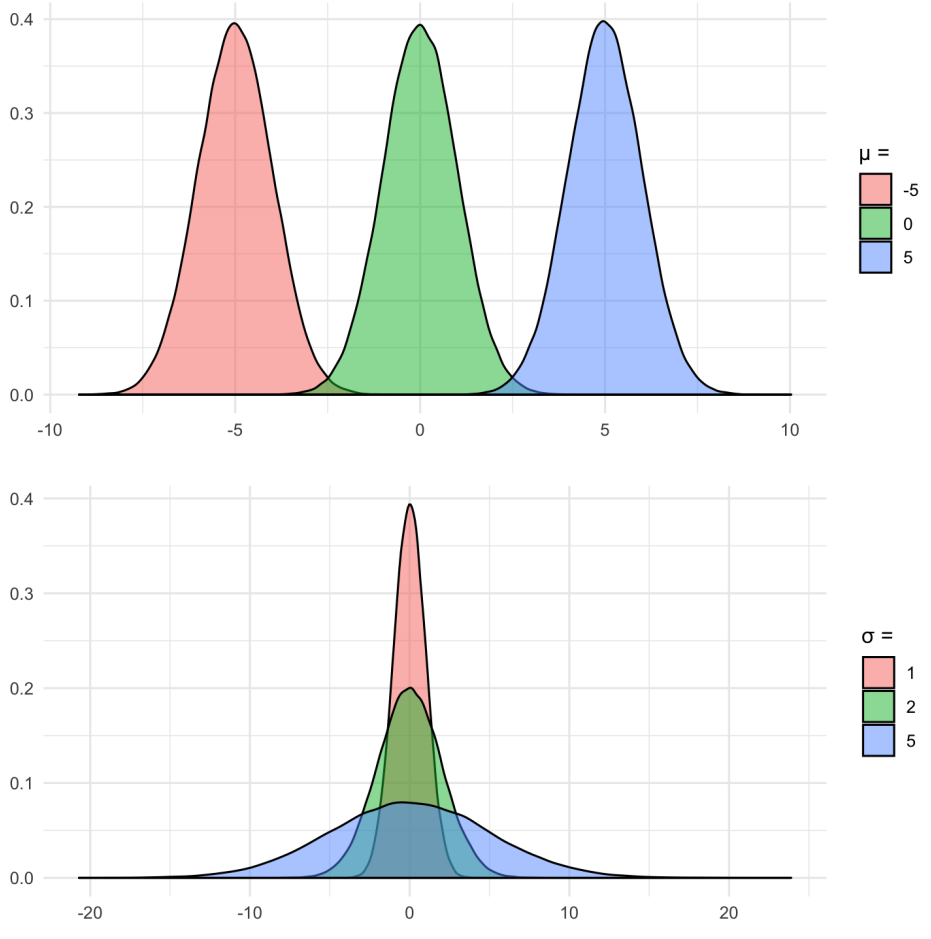


- Applies to continuous random variables
  - Variables can take on values that exist between ∞ and $-\infty$ (we fudge this conditic
  - Defined by the mean, μ, and variance, σ^2
  - Intervals near the mean more likely to happen than intervals near "tails"

- Some properties:
  - $P(X < u) = \int_{-\infty}^{u} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$
  - $E(X) = \mu$
  - $Var(X) = \sigma^2$

Normal Distribution

- μ and σ affect the shape of the distribution
  - Increasing or decreasing μ shifts the distribution right to left
  - Increasing or decreasing σ makes the distribution taller/narrower or shorter/fatter



Calculator: normalcdf()

To get to the calculator function on a TI-84:

- 2nd DISTR > 2: normalcdf(

To calculate P(X $\leq$ a):

- normalcdf(-1E99, a, μ , σ)

To calculate P(X $\geq$ a):

- normalcdf(a, 1E99, μ , σ)

To calculate P(a $\leq$ X $\leq$ b):

- normalcdf(a, b, μ , σ)

Example 1: SAT

The Scholastic Assessment Test (SAT) is a standardized test that's (was?🤞) widely used by colleges and universities to help make admission decisions. Scores can range 400 to 1600. SAT scores have a mean of 1,100 and a standard deviation of 200.

- What is the probability someones scores at least a 1,300?

Example 1: SAT

The Scholastic Assessment Test (SAT) is a standardized test that's (was?🤞) widely used by colleges and universities to help make admission decisions. Scores can range 400 to 1600. SAT scores have a mean of 1,100 and a standard deviation of 200.

- What is the probability someones scores lower than an 800?

Example 1: SAT

The Scholastic Assessment Test (SAT) is a standardized test that's (was?🤞) widely used by colleges and universities to help make admission decisions. Scores can range 400 to 1600. SAT scores have a mean of 1,100 and a standard deviation of 200.

- What is the probability someones scores between a 900 and a 1,200?

Calculator: invNorm()

If we have a percentile, and want to find the cut off value corresponding to that percentile, we use invNorm().

To get to the calculator function on a TI-84:

- 2nd DISTR > 3: invNorm(

To calculate the cut off value for the a-th percentile (a percent of the data is smaller than the cutoff):

- invNorm($\frac{a}{100}$, μ , σ)

Example 1: SAT

The Scholastic Assessment Test (SAT) is a standardized test that's (was?🤞) widely used by colleges and universities to help make admission decisions. Scores can range 400 to 1600. SAT scores have a mean of 1,100 and a standard deviation of 200.

- For what score are 90% of all other scores below?

Example: Hyena Bite Strength

Hyena bite strength is thought to follow a normal distribution, with a mean of 1,000 psi and a standard deviation of 300 psi. For comparison, humans have a bite strength of 171 psi when using the molars.

- What's the probability that a randomly selected hyena will have a bite strength greater than 852 psi?

Example: Hyena Bite Strength

Hyena bite strength is thought to follow a normal distribution, with a mean of 1,000 psi and a standard deviation of 300 psi. For comparison, humans have a bite strength of 171 psi when using the molars.

- What's the probability that a hyena will have a bite strength between 910 psi and 1,067 psi?

Example: Hyena Bite Strength

Hyena bite strength is thought to follow a normal distribution, with a mean of 1,000 psi and a standard deviation of 300 psi. For comparison, humans have a bite strength of 171 psi when using the molars.

- The top 15% of hyenas will have a bite strength greater than what?

Example: Hyena Bite Strength

Hyena bite strength is thought to follow a normal distribution, with a mean of 1,000 psi and a standard deviation of 300 psi. For comparison, humans have a bite strength of 171 psi when using the molars.

- The middle 40% of hyneas will have a bite strength between what psi's?

Standard Normal Distribution

- The **standard Normal distribution** is a Normal distribution with a mean of 0 and a variance of 1:

$$X \sim N(\mu = 0, \sigma^2 = 1)$$

- If a random variable *X* follows a different type of Normal distribution, we can always **transform** it so its transformation follows a standard Normal distribution

Standard Normal Distribution

- The **standard Normal distribution** is a Normal distribution with a mean of 0 and a variance of 1:

$$X \sim N(\mu = 0, \sigma^2 = 1)$$

- If a random variable *X* follows a different type of Normal distribution, we can always **transform** it so its transformation follows a standard Normal distribution

- We call this transformation a **Z score** or **Z statistic**:

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

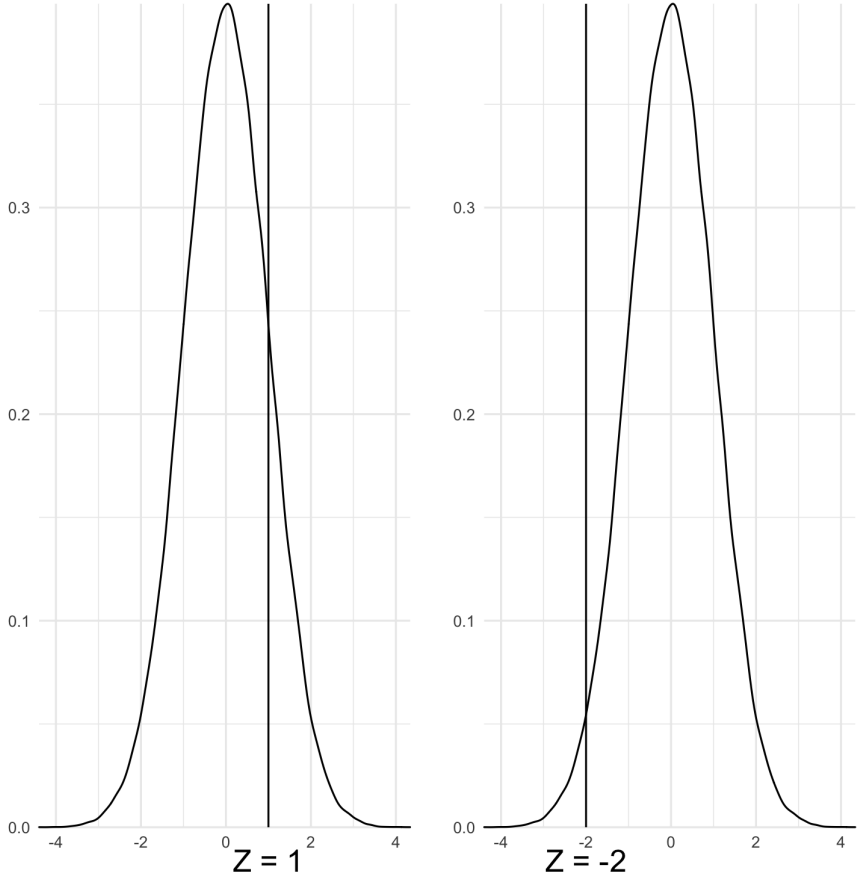
$$Z \sim N(0, 1)$$

Z Scores

- Z scores are nice because we can look at them and approximately know where it sits on its distribution
- If $\mu \neq 0$ or $\sigma^2 \neq 1$, it's harder to figure out where X is

Z Scores

- Z scores are nice because we can look at them and approximately know where it sits on its distribution
- If $\mu \neq 0$ or $\sigma^2 \neq 1$, it's harder to figure out where X is
- A Z score of 1 means that X is 1 standard deviation above its mean
- A Z score of -2 means means that X is 2 standard deviations below its mean



Why Z Scores?

- Before calculators, it was a real pain to find the probability of events with the Normal distribution
  - Integrating this by hand

$$\int_{-\infty}^{u} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

    was gnarly

- Statisticians came up with Z scores to get out of making those calculations by hand all the time
  - Simply have someone find and write down $P(Z < z)$ for a whole bunch of values of z where Z follows a N(0,1)
  - Now by transforming an observation from a different Normal distribution to a Z score, we can just look up the probabilty of the Z score!

Why Z Scores?

- Before calculators, it was a real pain to find the probability of events with the Normal distribution
  - Integrating this by hand

$$\int_{-\infty}^{u} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

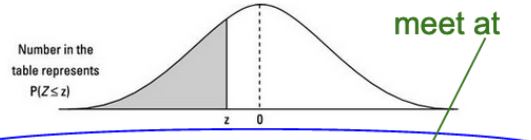
    was gnarly
- Statisticians came up with Z scores to get out of making those calculations by hand all the time
  - Simply have someone find and write down $P(Z < z)$ for a whole bunch of values of *z* where *Z* follows a N(0,1)
  - Now by transforming an observation from a different Normal distribution to a Z score, we can just look up the probabilty of the Z score!
- This $Z = \frac{x-\mu}{\sigma}$ was way easier to deal with than this $\int_{-\infty}^{u} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

Reading a Z Table

Then see what value this row and column meet at

Then find the second decimal value here

First, look for the main integer and first decimal value here



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Example 1

What is the Z-score for a random variable coming from a Normal distribution with a mean of 18 and standard deviation of 3, which takes on a value of 13?

Example 2

What is the Z-score for a random variable coming from a Normal distribution with a mean of -32 and standard deviation of 6.8, which takes on a value of -46?

Example 3

What is the Z-score for a random variable coming from a Normal distribution with a mean of 248 and standard deviation of 57, which takes on a value of 689?

Steps for a Normal Probability Problem

1. Draw the distribution
2. Write in the Z scale
3. Mark the probability interval you want to find
4. Calculate the Z-score
5. Use your calculator/Z-table to find the probability

Example: CD4+ Count

CD4+ count is an indicator of immune function. Decreasing CD4+ count indicates poor immune function. In the 1970s, men at high risk of contracting HIV were enrolled into a study to see if changes in their CD4+ count corresponded with seroconversion (time at which they officially contract HIV). CD4+ count is normally distributed in humans, with a mean of 1,300 and a standard deviation of 400.

- What is the probability that someone's CD4+ count is less than 700?

Example: CD4+ Count

CD4+ count is an indicator of immune function. Decreasing CD4+ count indicates poor immune function. In the 1970s, men at high risk of contracting HIV were enrolled into a study to see if changes in their CD4+ count corresponded with seroconversion (time at which they officially contract HIV). CD4+ count is normally distributed in humans, with a mean of 1,300 and a standard deviation of 400.

- What is the probability that someone's CD4+ count is greater than 1,400?

Example: CD4+ Count

CD4+ count is an indicator of immune function. Decreasing CD4+ count indicates poor immune function. In the 1970s, men at high risk of contracting HIV were enrolled into a study to see if changes in their CD4+ count corresponded with seroconversion (time at which they officially contract HIV). CD4+ count is normally distributed in humans, with a mean of 1,300 and a standard deviation of 400.

- What is the probability that someone's CD4+ count is between 1,200 and 1,600?

Variability in Sampling

- Two research groups are interested in the mean height of giraffes

Variability in Sampling

- Two research groups are interested in the mean height of giraffes
- Both groups go out independently and gather a simple random sample of 200 giraffes and measure them

Variability in Sampling

- Two research groups are interested in the mean height of giraffes
- Both groups go out independently and gather a simple random sample of 200 giraffes and measure them

Group 1 gets an sample mean of 17.32 feet.

Group 2 gets a sample mean of 18.75 feet

Variability in Sampling

- Two research groups are interested in the mean height of giraffes
- Both groups go out independently and gather a simple random sample of 200 giraffes and measure them

Group 1 gets an sample mean of 17.32 feet.

Group 2 gets a sample mean of 18.75 feet

Who is correct?

Sampling Distributions

- Because samples do not include every member of the population, there will be some **variation** in the value a sample's mean will take on
- As your sample size gets **bigger**, there should be **less variation** because you are included a larger portion of the population
    - When your sample includes the whole population, there should be **no variation** at all

Sampling Distributions

- Because samples do not include every member of the population, there will be some **variation** in the value a sample's mean will take on
- As your sample size gets **bigger**, there should be **less variation** because you are included a larger portion of the population
    - When your sample includes the whole population, there should be **no variation** at all
- In a way, your sample mean follows a distribution...
    - We call this a **sampling distribution**

Sampling Distributions

- Because samples do not include every member of the population, there will be some **variation** in the value a sample's mean will take on
- As your sample size gets **bigger**, there should be **less variation** because you are included a larger portion of the population
  - When your sample includes the whole population, there should be **no variation** at all
- In a way, your sample mean follows a distribution...
  - We call this a **sampling distribution**
- But what distribution does your sample mean follow?

Central Limit Theorem

When

1) observations are **independent**, and

2) the sample size is **sufficiently large**,

the sample mean $\bar{x}$ will follow a **Normal distribution** with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Central Limit Theorem

When

1) observations are **independent**, and

2) the sample size is **sufficiently large**,

the sample mean $\bar{x}$ will follow a **Normal distribution** with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Notice that the variance depends on your sample size, n
  - As **n increases**, the variance of the sampling distribution of $\bar{x}$ **decreases**

Example 1: SAT

Remember that SAT scores are Normally distributed with a population mean score of 1,100 and a population standard deviation of 200.

- If we gather a random sample of 40 scores, what distribution would the sample mean of those scores follow?

Example 1: SAT

Remember that SAT scores are Normally distributed with a population mean score of 1,100 and a population standard deviation of 200.

- If we gather a random sample of 100 scores, what distribution would the sample mean of those scores follow?

Example 1: SAT

Remember that SAT scores are Normally distributed with a population mean score of 1,100 and a population standard deviation of 200.

- If we gather a random sample of 500 scores, what distribution would the sample mean of those scores follow?

How Big Is Big Enough?

Well, it depends on your data...

- Nearly normal data, no clear outliers: $n \geq 10$
- Not sure what distributions the observations come from, no clear outliers: $n \geq 30$

CLT with Proportions

- One of the greatest things about CLT is that **we don't have to know what distribution our observations come from** in order to apply it
    - As long as the observations are independent and our sample size is large enough, we're good!
- Particularly, if we are dealing with count data to get proportions, this "large enough" means:

$$np \geq 10 \text{ and } n(1-p) \geq 10$$

- If we have binomial data and we want to estimate the sample proportion of successes, $\hat{p}$, we simply need to remember what the mean and variance of a binomial distribution is to find the sampling distribution.

CLT with Proportions

- One of the greatest things about CLT is that **we don't have to know what distribution our observations come from** in order to apply it
    - As long as the observations are independent and our sample size is large enough, we're good!
- Particularly, if we are dealing with count data to get proportions, this "large enough" means:

$$np \geq 10 \text{ and } n(1-p) \geq 10$$

- If we have binomial data and we want to estimate the sample proportion of successes, $\hat{p}$, we simply need to remember what the mean and variance of a binomial distribution is to find the sampling distribution.
    - Remember: E(X) = p, Var(X) = p(1-p)

CLT with Proportions

- One of the greatest things about CLT is that **we don't have to know what distribution our observations come from** in order to apply it
    - As long as the observations are independent and our sample size is large enough, we're good!
- Particularly, if we are dealing with count data to get proportions, this "large enough" means:

$$np \geq 10 \text{ and } n(1-p) \geq 10$$

- If we have binomial data and we want to estimate the sample proportion of successes, $\hat{p}$, we simply need to remember what the mean and variance of a binomial distribution is to find the sampling distribution.
    - Remember: E(X) = p, Var(X) = p(1-p)

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Example 1: Solar Energy

Suppose the proportion of American adults who support the expansion of solar energy is p = 0.88.

- If we get a random sample of 100 American adults, what distribution would the sample proportion follow?

Example 1: Solar Energy

Suppose the proportion of American adults who support the expansion of solar energy is p = 0.88.

- If we get a random sample of 1,000 American adults, what distribution would the sample proportion follow?

Z-Scores with CLT

- We can also calculate Z-scores to see where a particular sample mean (or sample proportion) lies on its sampling distribution:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

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

Estimating Variance

- All the distributions we've talked about so far have assumed we know what the population variance is

Estimating Variance

- All the distributions we've talked about so far have assumed we know what the population variance is
    - You're telling me we're estimating the population mean with a sample mean, but we always know the population variance?

Estimating Variance

- All the distributions we've talked about so far have assumed we know what the population variance is
    - You're telling me we're estimating the population mean with a sample mean, but we always know the population variance?

Estimating Variance

- All the distributions we've talked about so far have assumed we know what the population variance is
  - You're telling me we're estimating the population mean with a sample mean, but we always know the population variance?

- Then do we get a Z score if we replace σ with s?

$$\frac{x-\mu}{s} =? \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} =?$$

The T Distribution

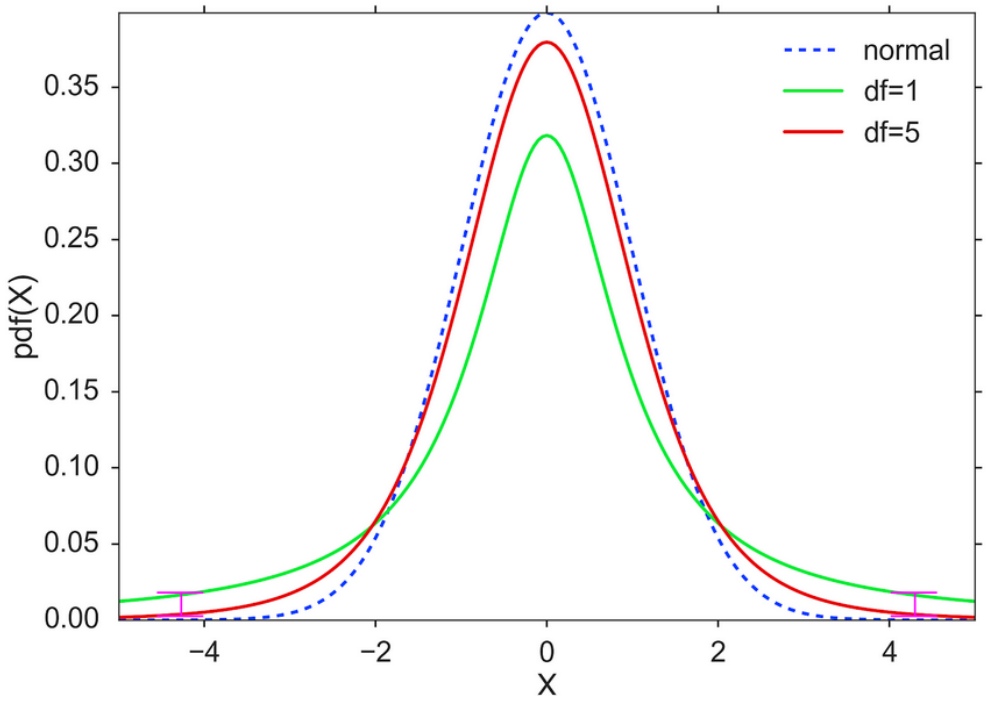
- When we replace σ with s, we get a **T score**:

$$T = \frac{x - \mu}{s}, T = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

- The T score follows the **T distribution**:

$$t \sim T(df = n - 1)$$

- Always centered with a mean of 0 (like a standard Normal)
- Shape defined by **degrees of freedom** (df) = n-1
  - The more degrees of freedom, the more the distribution looks like a standard Normal
- Designed to give you a little more wiggle room than a regular Normal distribution, to account for you also estimating the standard deviation



Calculator: tcdf()

To get to the calculator function on a TI-84:

- 2nd DISTR > 6: tcdf(

To calculate P(X $\leq$ a):

- tcdf(-1E99,a,df)

To calculate P(X $\geq$ a):

- tcdf(a,1E99,df)

To calculate P(a $\leq$ X $\leq$ b):

- tcdf(a,b,df)

Calculator: invT()

If we have a percentile, and want to find the cut off value corresponding to that percentile, we use invT().

To get to the calculator function on a TI-84:

- 2nd DISTR > 4: invT(

To calculate the cut off value for the a^{th} per entile (a% of the data is smaller than the cutoff):

- invT($\frac{a}{100}$, df)

Example 1: Mercury in Dolphins

Elevated mercury concentrations are an important problem for both dolphins and other animals. Say that we know the population mean mercury level in dolphin tissue is 5 micrograms of mercury per wet gram of muscle. We randomly sample 19 dolphins and test their muscle for mercury levels. We get a sample mean of 4.4, and a sample standard deviation of 2.3.

- Say one dolphin in our sample has a mercury level of 3. What is the T-score for this data point?

Example 1: Mercury in Dolphins

Elevated mercury concentrations are an important problem for both dolphins and other animals. Say that we know the population mean mercury level in dolphin tissue is 5 micrograms of mercury per wet gram of muscle. We randomly sample 19 dolphins and test their muscle for mercury levels. We get a sample mean of 4.4, and a sample standard deviation of 2.3.

- What is the probability that a dolphin will have a mercury level below 3?

Example 1: Mercury in Dolphins

Elevated mercury concentrations are an important problem for both dolphins and other animals. Say that we know the population mean mercury level in dolphin tissue is 5 micrograms of mercury per wet gram of muscle. We randomly sample 19 dolphins and test their muscle for mercury levels. We get a sample mean of 4.4, and a sample standard deviation of 2.3.

- What is the probability that a dolphin will have a mercury level between 3.5 and 5?

Example 1: Mercury in Dolphins

Elevated mercury concentrations are an important problem for both dolphins and other animals. Say that we know the population mean mercury level in dolphin tissue is 5 micrograms of mercury per wet gram of muscle. We randomly sample 19 dolphins and test their muscle for mercury levels. We get a sample mean of 4.4, and a sample standard deviation of 2.3.

- What is the T-score for our sample mean?

Example 1: Mercury in Dolphins

Elevated mercury concentrations are an important problem for both dolphins and other animals. Say that we know the population mean mercury level in dolphin tissue is 5 micrograms of mercury per wet gram of muscle. We randomly sample 19 dolphins and test their muscle for mercury levels. We get a sample mean of 4.4, and a sample standard deviation of 2.3.

- What is the probability that, if we took another sample like this one, we would get a sample mean above 4.4?