Week Six: The Normal Distribution and The Central Limit Theorem

CS 217

Discrete Distributions

- Last week we talked about the **probability mass function** and **cumulative distribution function** for discrete distributions
- The **probability mass function** gives us the probability that a discrete distribution is **equal** to a given variable
- The **cumulative distribution function** gives us the probability that a discrete distribution is **less than or equal** to a given variable

Continuous Distributions

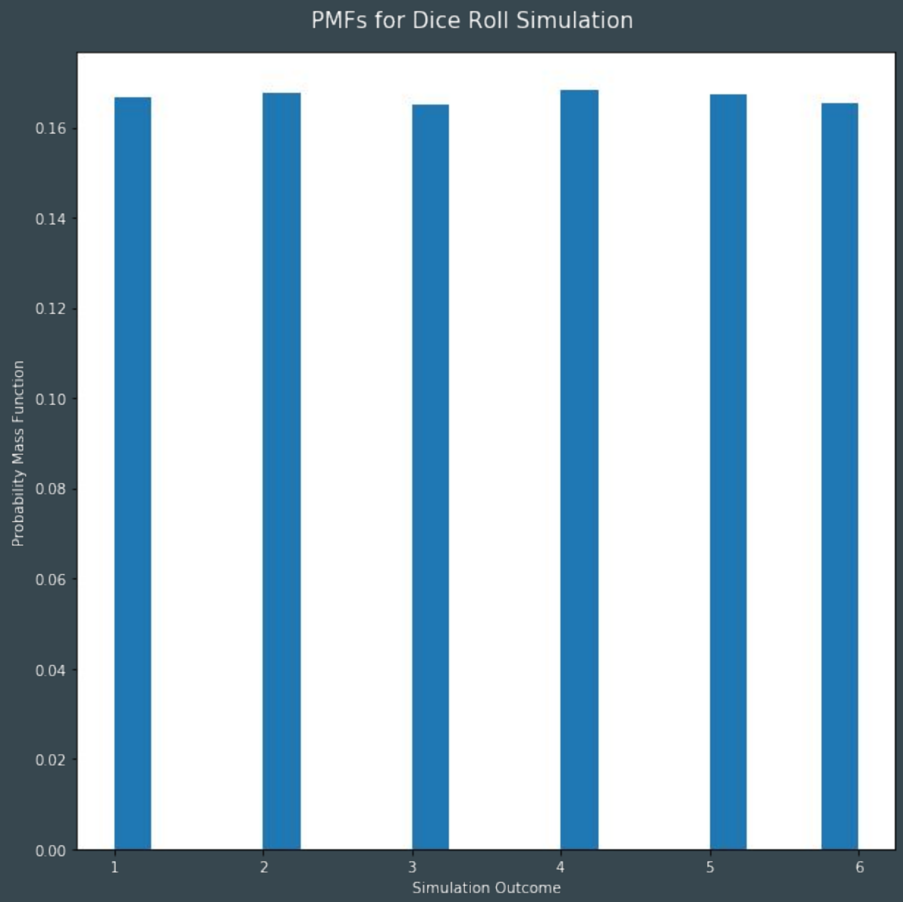
- The **probability mass function** is intuitively easy to understand - it's the probability that if we flip a coin three times, we will get two heads. The outcome is discrete and unequivocal.
- Say I were to measure the height of every student in this class. What is the probability that I will get three students who are 5'8"?
- If I were to round every students' height to the nearest inch, I could treat height as a discrete measurement
- So technically, want to find the probability that I will get three students who are between 5'7.5" and 5.8.5" tall
- I can do so with the **probability density function**

Continuous Distributions

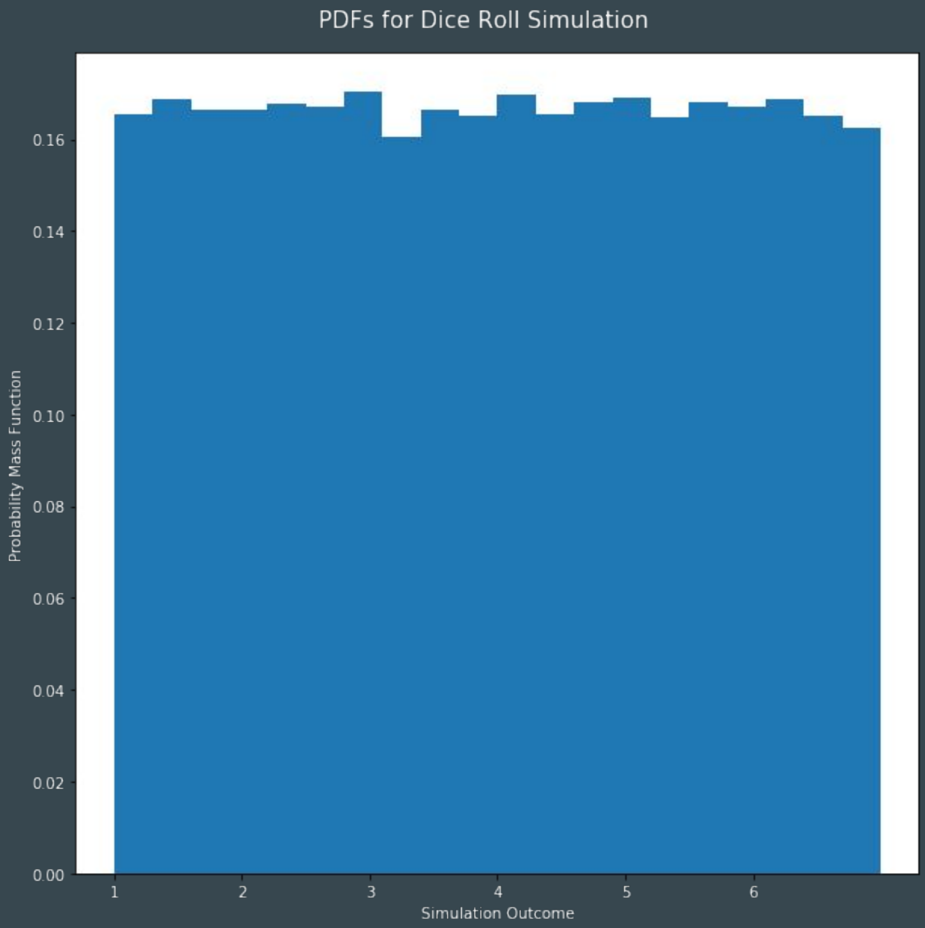
- The **probability density function** is the continuous analogue of the probability mass function
- The probability of something fallen in a given range (i.e. the probability of a student being between 5’7.5” and 5’8.5”) is the integral of the probability density function
- Continuous distributions, like discrete distributions, have cumulative distribution functions that represent the probability of something being **less than or equal to** a given value

The Uniform Distribution

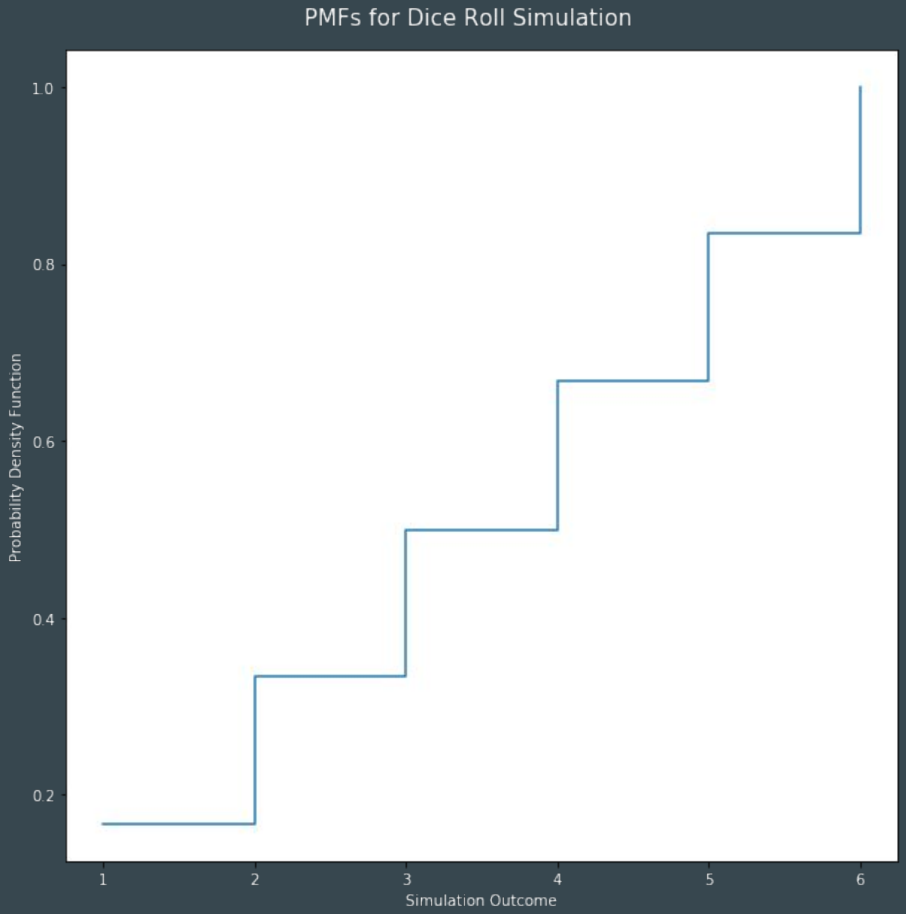
- Let’s take a **uniform** distribution as an example of the difference between discrete and continuous distributions
- Last class we had the example of a single die as a uniform distribution between 1 and 6, since those are the only possible outcomes
- We could also set up a uniform distribution of all decimals between 1.0000000 and 6.99999
- The probability that a single die will land on 1 is equivalent to the probability that a computer will randomly choose a number between 1.0000 and 2.000000

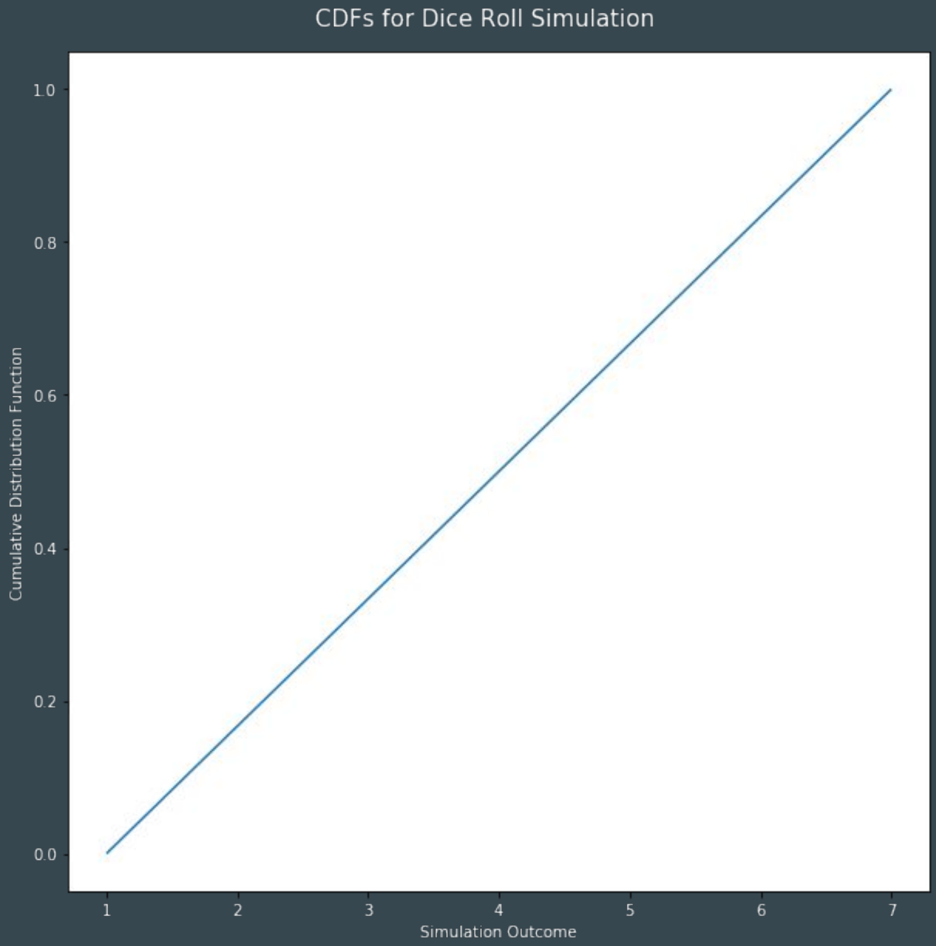


PMFs for Dice Roll Simulation
Probability Mass Function
Simulation Outcome
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
1
2
3
4
5
6



PDFs for Dice Roll Simulation
Probability Mass Function
Simulation Outcome
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
1
2
3
4
5
6


PMFs for Dice Roll Simulation
Probability Density Function
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
Simulation Outcome


CDFs for Dice Roll Simulation
Cumulative Distribution Function
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
Simulation Outcome

The Exponential Distribution

- Last week we looked at the **Poisson Distribution**, which measures the probability of a given number of times happening in a fixed interval of time, with the example of the number of trains that will arrive at a platform in a given hour
- Say an average of six trains will arrive every hour - this is a Poisson distribution with a **lambda** value of six
- The exponential distribution takes the inverse of this - that a train will arrive every ten minutes - as an input. Our **theta** value is ⅙ (it will take ⅙ of an hour for the next train to arrive)
- It may model a question, such as, given that a train arrives every ten minutes, what is the probability that a train arrives in the next **five** minutes?

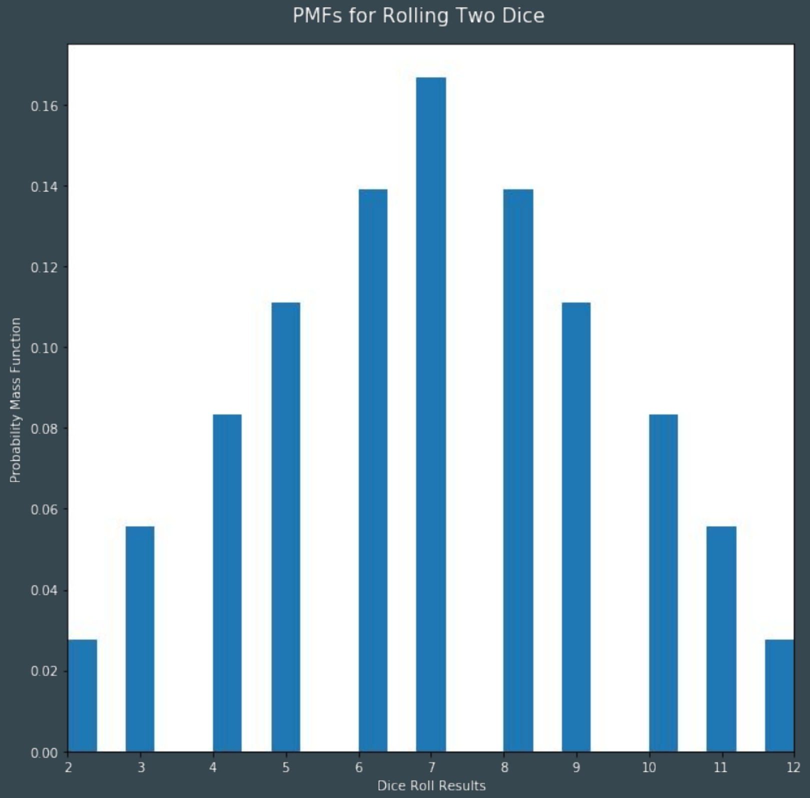
The Exponential Distribution

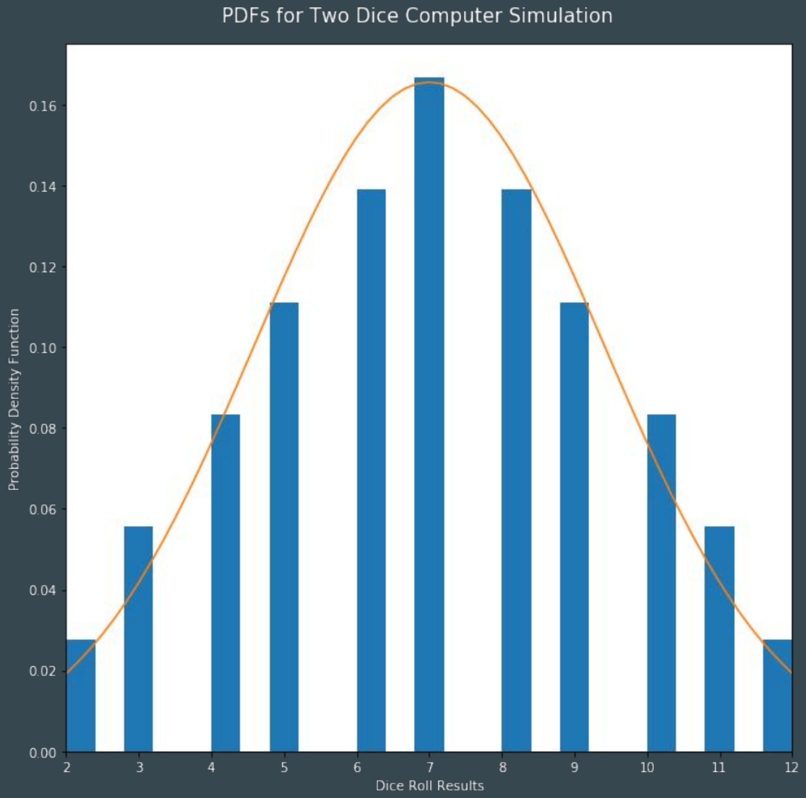
- Both the Poisson and exponential distribution are **memoryless** in that what has happened in the past does not affect what happens in the future
- If a train hasn’t arrived in the past fifteen minutes, the probability that a train arrives in the next five minutes will be unaffected
- If ten trains have arrived in the past fifteen minutes, the expected number of trains that will arrive in the next hour will remain the same

The Normal Distribution

- The **normal distribution** is by far the most important continuous distribution, for reasons we will explore throughout the course
- A normal distribution is a continuous distribution where the the data tends to cluster around a central value with no skew or bias
- Common examples of the **normal distribution** include height, SAT scores, or the sum of the rolls of two die (like we covered next week)
- It has two parameters, the mean and standard deviation

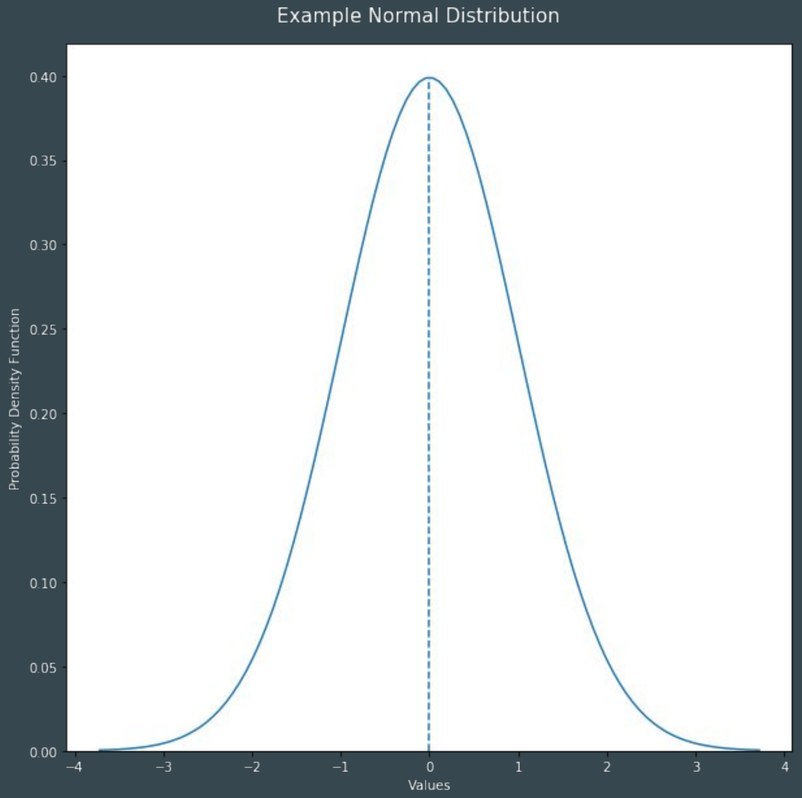
The Normal Distribution





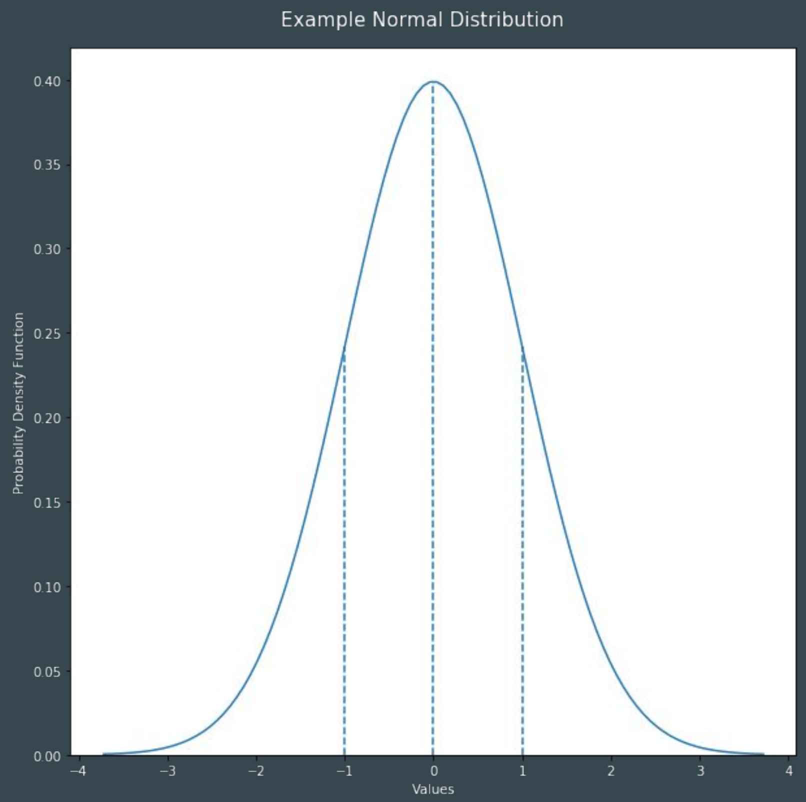
The Normal Distribution

- The **standard normal distribution** has a mean of 0 and a standard deviation of 1



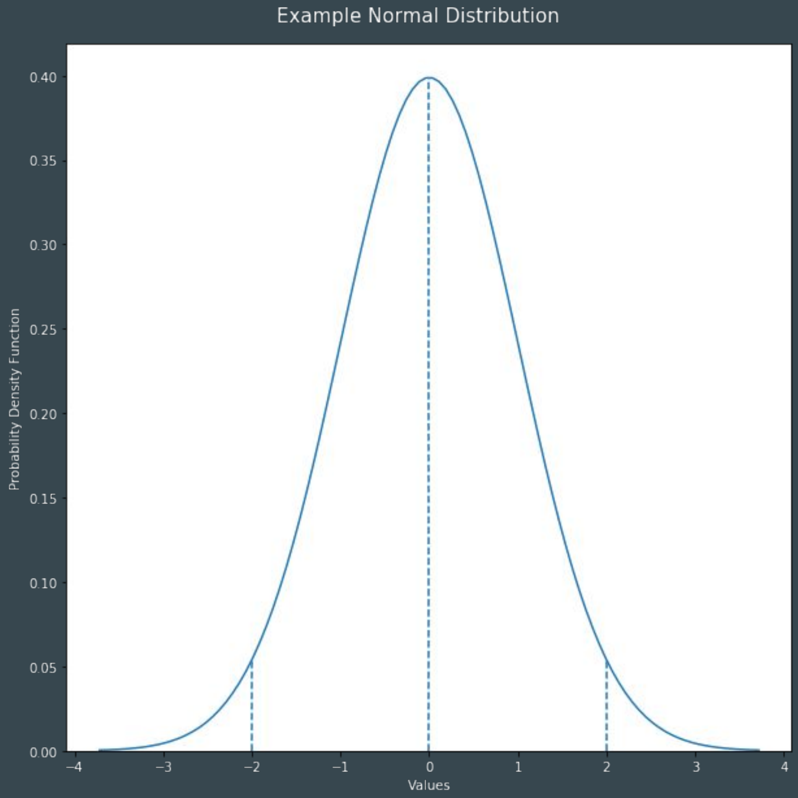
The Normal Distribution

- 68% of the data in a normal distribution will fall within **one standard deviation** of the mean
- With a mean of 0 and a standard deviation of 1, this means that 68% of the data will fall between **-1** (0 -1) and **1** (0 + 1)



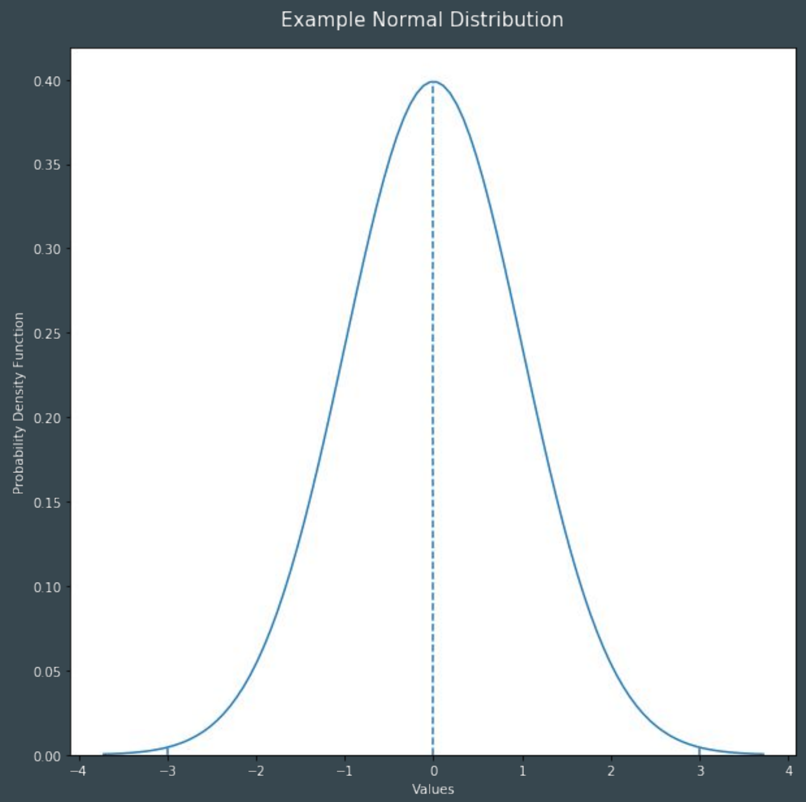
The Normal Distribution

- 95% of the data in a normal distribution will fall within **two standard deviations** of the mean
- With a mean of 0 and a standard deviation of 1, this means that 68% of the data will fall between **-2** (0 -2) and **1** (0 + 2)



The Normal Distribution

- 99% of the data in a normal distribution will fall within **three standard deviations** of the mean
- With a mean of 0 and a standard deviation of 1, this means that 68% of the data will fall between **-3** (0 -3) and **3** (0 + 3)



The Z-Score

- Given these principles, we can find the CDF of any given point on a normal distribution if we have the **mean** and **standard deviation** of that normal distribution
- We can do this by seeing how many standard distributions away from the mean a given point is
- This value is called the **Z-Score**

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

The Z-Score

- A Z-score of less than 0 means that that value is less than the mean of the distribution, and a Z-score of more than 1 means that the value is more than the mean of the distribution
- The value of the Z-score is the **number of standard distributions away** a value is from the mean of a distribution

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

The Z-Score

- For example, say that SAT scores are distributed with a mean of 1060 and a a standard deviation of 195
- If I get a 1120, what is my Z score?

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

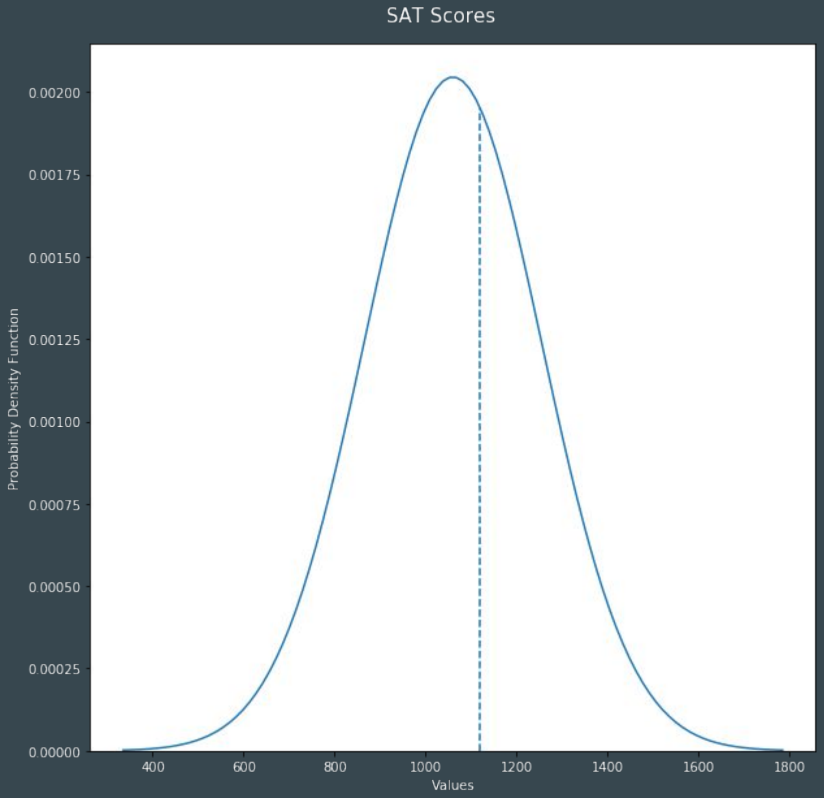
The Z-Score

- For example, say that SAT scores are distributed with a mean of 1060 and a a standard deviation of 195
- If I get a 1120, what is my Z score?
- My Z-score is (1120-1060)/195, or 0.30
- This means that I scored 0.3 standard distributions above the mean score

$$Z = \frac{1120 - 1060}{195}$$

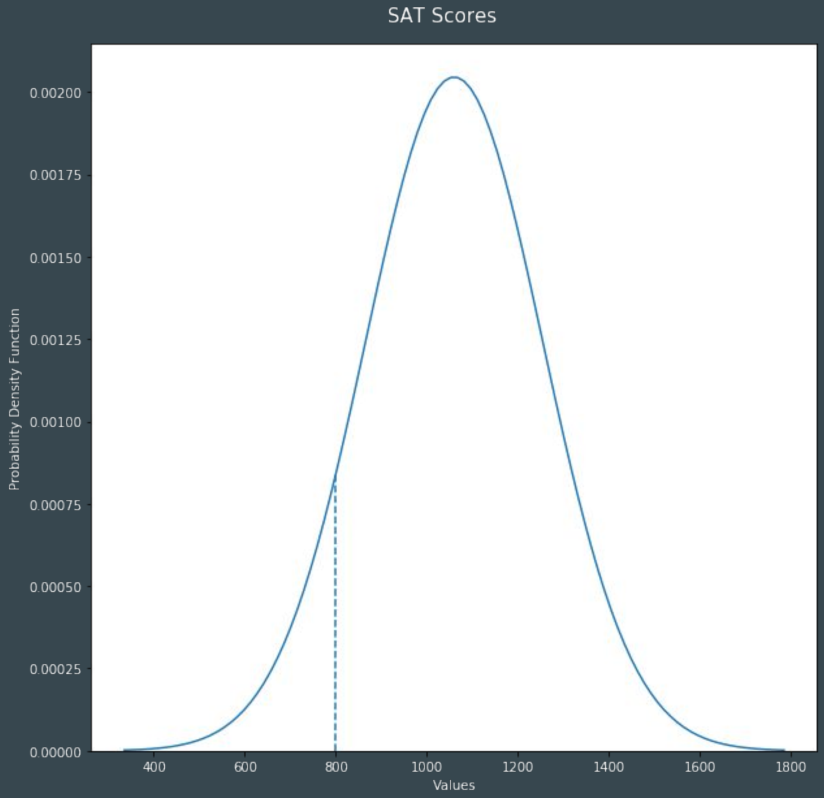
The Z-Score

- For example, say that SAT scores are distributed with a mean of 1060 and a a standard deviation of 195
- If I get a 1120, what is my Z score?
- My Z-score is (1120-1060)/195, or 0.30
- This means that I scored 0.3 standard distributions above the mean score



The Z-Score

- For example, say that SAT scores are distributed with a mean of 1060 and a a standard deviation of 195
- If I get an 800, my Z score is (800-1060)/195, or -1.33
- This means that I scored 1.33 standard distributions below the mean score



The Z-Score

- With a Z-score of -1.33, what percentile of scores does that put me in?
- This would make my score below average, specifically I would be in the 9th percentile of scores.
- Traditionally we would find this by looking up the CDF value in a Z-table
- Now we have computers :)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–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

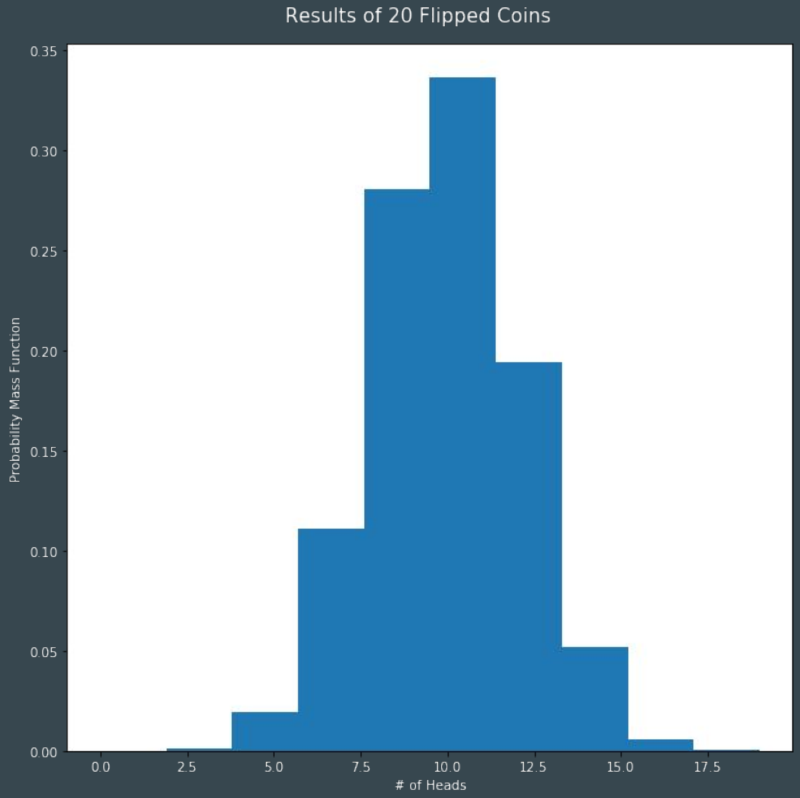
The Z-Score

- Intuitively, it's important to know the relationship between these Z-scores and their relative percentile scores

Z-Score	Cumulative Percentile
-2	2%
-1	15%
0	50%
1	84%
2	98%

The Normal Approximation

- The Normal Distribution can also be used as an approximation to the **binomial distribution**.
- Let's flip 20 coins. What is the PMF for getting 0 - 20 heads?
- The results look quite similar to a normal distribution!



The Normal Approximation

- Say we wanted to find the odds of getting 7 or less heads in 20 flips.
- One way to approach this would be to add up the PMFs of the flips for 0, 1, 2, etc.... This is possible but annoying to do by hand.
- *However*, we could approximate a normal distribution and find this probability using a Z-score

The Normal Approximation

- In this example X is 7, the mean is 10, and the standard deviation is 2.23
- However, because we are approximating a continuous distribution with a discrete distribution, we add a **continuity correction** of 0.5, so that the X is 7.5

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

The Normal Approximation

- In this example X is 7, the mean is 10, and the standard deviation is 2.23
- However, because we are approximating a continuous distribution with a discrete distribution, we add a **continuity correction** of 0.5, so that the X is 7.5
- In this case the Z-score will be -1.12, which translates to a probability of around 13.11%

$$Z = \frac{7.5 - 10}{2.23}$$

The Normal Approximation

- It turns out that the CDF of the underlying binomial distribution is 13.15%.
- The normal approximation is extremely accurate when the n * p >= 5, and n*(1-p) >= 5
- Here, both p and 1-p are 0.5, so n * either is 10. Thus the normal approximation is applicable, and useful!

$$Z = \frac{7.5 - 10}{2.23}$$

The Law of Large Numbers

- If we flip ten coins, what is the expected number of heads we will see?
- Do we expect to see this result every time we flip ten coins?
- Do we expect to see this result as an average if we flip ten coins one-hundred times?

The Law of Large Numbers

- If we flip ten coins, what is the expected number of heads we will see?
    - *5*
- Do we expect to see this result every time we flip ten coins?
    - *No, specifically the PMF of five successes in ten trials gives us a 25% chance of this happening*
- Do we expect to see this result as an average if we flip ten coins one-hundred times?
    - *Yes*

The Law of Large Numbers

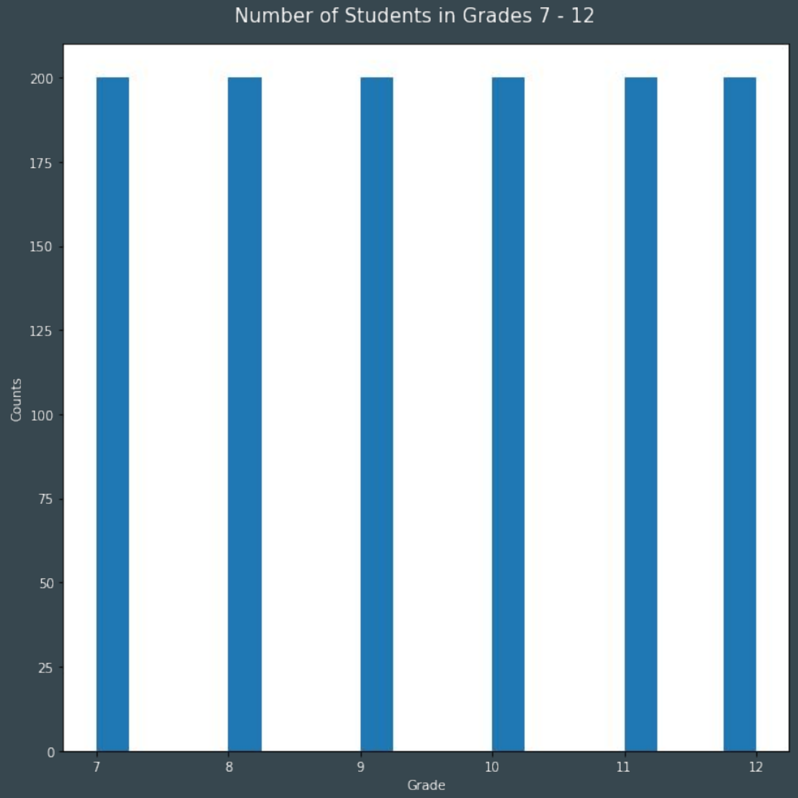
- The **law of large numbers** states that if we perform the same experiment a large number of times, the average of our results will approach the expected value of the experiment's underlying distribution

The Central Limit Theorem

- The **central limit theorem** tells us if we take large enough samples from **any** distribution, the distribution of the resulting sample means will resemble a normal distribution
- This normal distribution will have the same mean as the underlying distribution, and have a standard deviation approximately close to the underlying distribution's standard deviation divided by the square root of each sample size
- This typically works well when we have sample sizes greater than 30, though it could be more for especially skewed distributions, or less for normal distributions
- This means that we can use the normal distribution to quantify uncertainty about a population's mean given a sample mean

The Central Limit Theorem

- Say we have a school where there are exactly 200 students in grades 7-12.
- The average grade for a student is 9.5 and the standard deviation is approximately 1.7



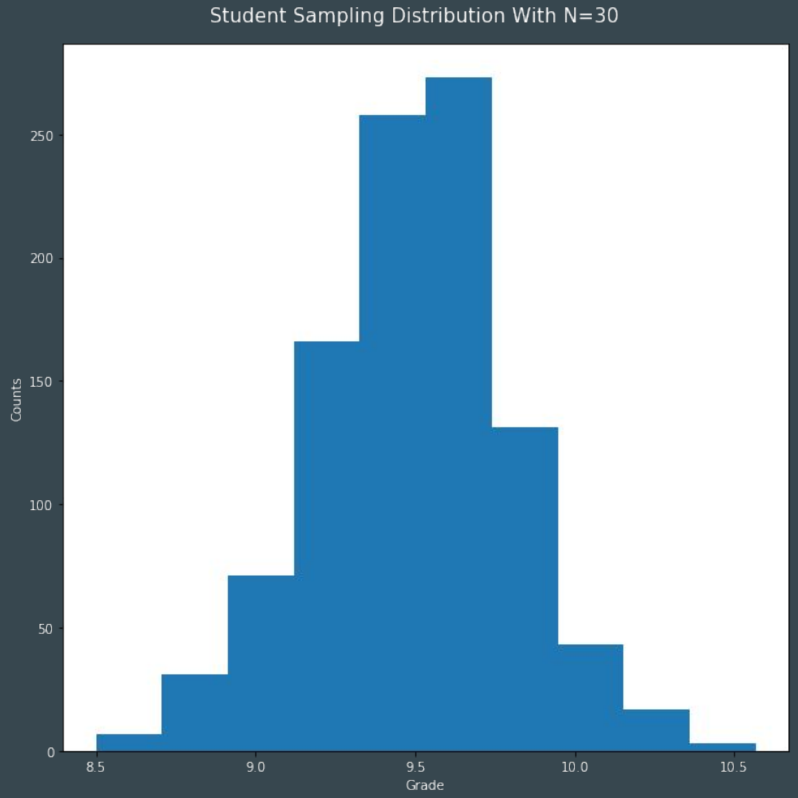
The Central Limit Theorem

- Say we take samples of 30 students each from this student population. We find that generally the average grade of each sample of students is fairly close to the actual average of 9.5

Sample	Average Grade
Sample #1	9.13
Sample #2	9.33
Sample #3	9.9
Sample #4	9.6
Sample #5	9.53

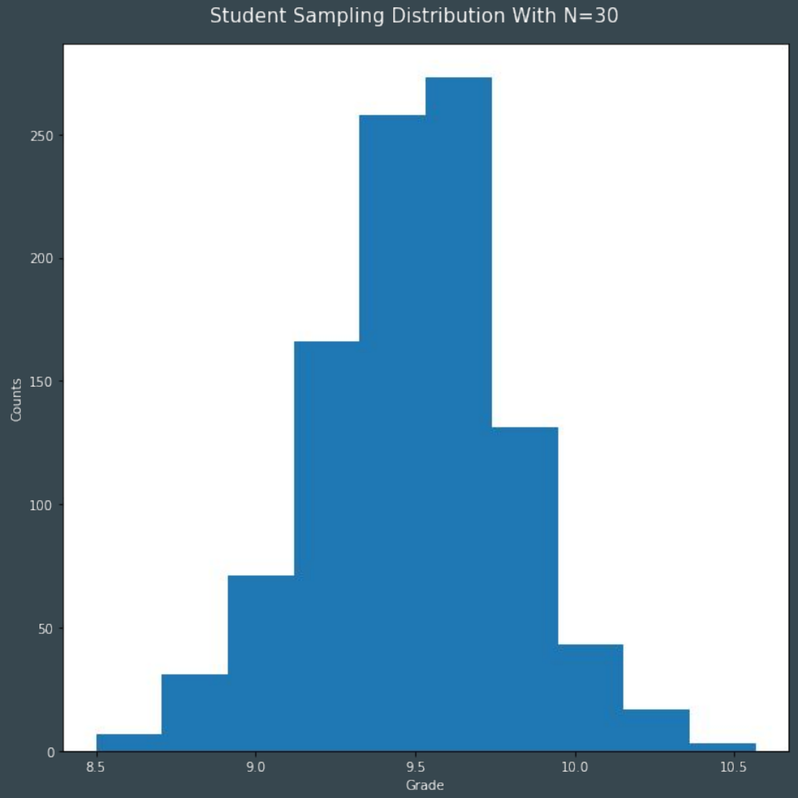
The Central Limit Theorem

- If we take 1,000 samples of 30 from the population, we'll find that the resulting distribution of grade averages is actually a normal distribution.
- The normal distribution has an average of 9.5, the same average as the population



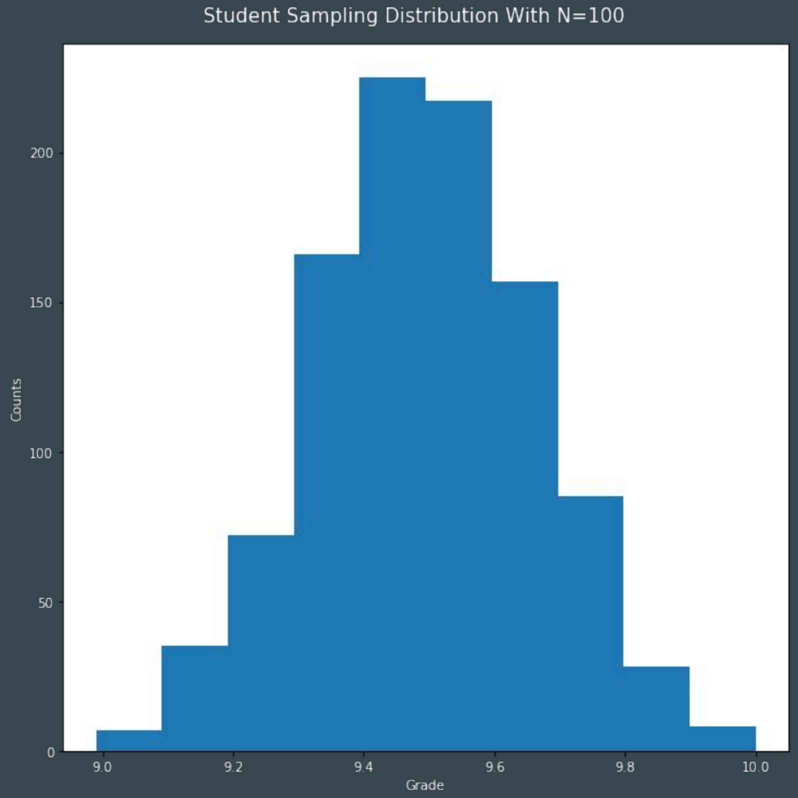
The Central Limit Theorem

- If we take 1,000 samples of 30 from the population, we'll find that the resulting distribution of grade averages is actually a normal distribution.
- And it has a standard deviation of 0.30, which is approximately 1.7 divided by 30, the size of each random sample
- The standard deviation of the **sampling distribution** is also known as the **standard error**



The Central Limit Theorem

- If we take 1,000 samples of 100 instead from the population, our standard error will decrease to 0.16
- The larger our sample size, the smaller our standard error is.
- Intuitively this makes sense: if we were polling random students about which grade they were in, we would want to poll as many as possible to get as little error as possible



The Central Limit Theorem

- The Central Limit Theorem is important because of these two beliefs:
  - The mean of a random sample will be the same as the mean from the overall population
  - The standard error of a random sample will be smaller the greater you increase your sample size
- As we do **hypothesis testing** in the second half of the course, the central limit theorem will be a core tool in understanding whether a random sample is part of the distribution or not
- If we poll 30 random students and find that they are in grade 11 on average, do we believe that they are telling the truth, given what we know about the distribution of students in the school? Is it statistically feasible?