STA-1380 – Elementary Statistics
Week #5

Hello! Welcome to the additional online **Weekly Resources** for the course of **STA-1380**. Following a traditional calendar semester, these will be some of the topics your professors will go over. If you do not see material your section is going over for the week, please look at the other resources listed for this course. In addition to these resources, there might be **Group Tutoring** for this course, please see our website for more details. These sessions will go over these materials in more detail as well as any questions about the material.

Any additional help or services can be found through the **Baylor Tutoring Website**. Visit to schedule a free 30-minute private tutoring session, drop-in times for your course, the Baylor Tutoring YouTube channel, or any additional tutoring resources.

Contacts: Sid Rich M-Th 9am-8pm (Fall and Spring class days)   Office Phone: 254-710-4135

Topic of the Week:
"The Normal Distribution"

Key Points:

- Normal Distributions
- Z-Scores and Z_α
- Standardization

Last week's lesson introduced the continuous distribution family with the PDFs, now we will introduce the most important tool in your STA: 1380 course: the Normal distributions. We will use this graph to estimate several variables from height to weight to many other common and uncommon inputs. The Normal Distribution can be specific to the sample size or 'Standardized' in which the spread of data follows a uniform shape.

This week will be the groundwork for the rest of this course and is vital for understanding how the graphs, tables, and named functions of this course are used. We will go over how to Identify a Normal Distribution, how to Standardize, and the various notations used to ask for probability or a Z-score.

Highlight #1
"Normal Distribution"

Definition: A continuous distribution with a unimodal symmetrically bell-shaped curve centered at the population's mean and shaped by the population's standard deviation standardized by the Empirical rule.

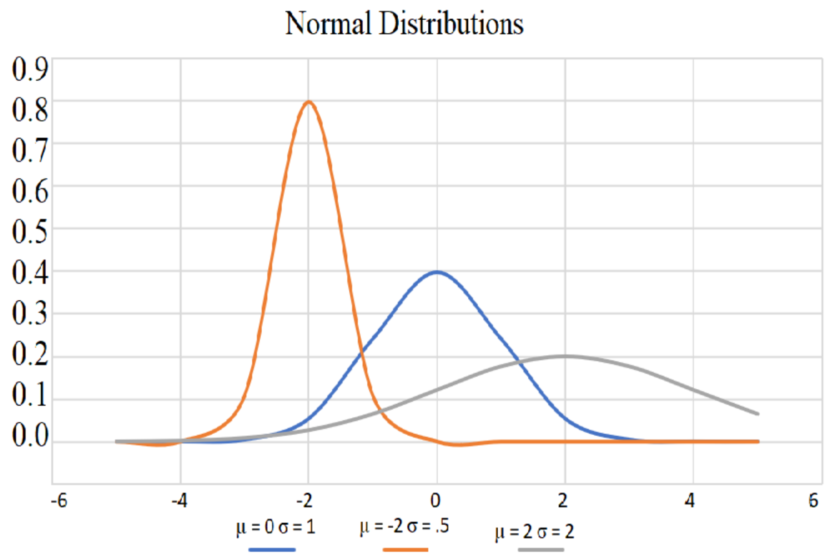
Notation: $X \sim N(\mu,\sigma)$

Example: PDF Graphs and CDF Tables
* A Standardized Normal CDF Table

	Z-Score Normal Distribution CDF									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Without a doubt the most important distribution in STA-1380, the Normal Distribution stands above the rest with its versatility and practical inference applications. This distribution uses an entire population's mean and standard deviation to accurately display the spread of the individuals. From frog weights to the proportion of females in a workplace, the bell-shaped curve is a perfect distribution form for most types of continuous single variable data sets.

Remember that the distribution uses the population mean and standard deviation or the parameters of a population. These are set values and are often unknown in the real world but will often be given in this course. These parameters are gained through censuses and other means of polling larger masses. If you see the symbols Mu or Sigma, (μ or σ), then intuitively, that will be your way of recognizing that the data that has been given will be a Normal Distribution.



As shown to the right with the PDF graphs, every Normal Distribution will be centered at the mean of the data, this is why you may also hear the mean be called the 'location' parameter. The varying 'strechiness' of a graph is determined by its 'shape' parameter or the standard deviation. A smaller std. dev. will result in a higher peak with steep slopes and tiny tails. A larger std. dev. will have wider tails and a lower peak. Remember that the term 'tails' is in reference to one or both ends of a distribution that tapper off away from the mean.

These PDFs provide a visual understanding of the population spread, but are lacking in a quantifiable way to understand them together. Saying, "I have a data value at -1" means little in this cirumstance since we cannot quantify what "-1" means in terms of the distribution. To find a quantifiable amount of probability associated with this value, we will use the process of Standardization.

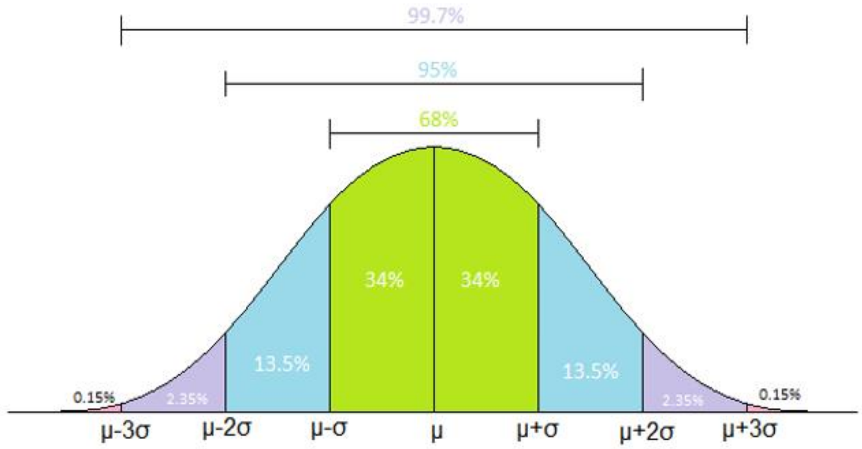
Highlight #2
"The Empirical Rule and Z-Scores"

Definition: The rule that separates a normal Distribution into 8 sections and provides a uniform notation of expressing a value's distance away from the mean.

Notation: **"68, 95, 99.7"** $Z = \frac{x - \mu}{\sigma}$

In order to properly gage the probabilities of a normal distribution and form a table, a process called the Emperical Rule is used on all Normal Distributions to 'Standardize' them. After a sucessful standardization, the Normal Distribution should have 50% of the probability to the left and right of the mean with four equidistant parts. The first section is + or – 1 Standard Deviation from the mean, the second section is + or – 2 Standard Deviations from the mean, etc.

Due to the fact that most values fall around the mean, most of the area under the curve falls within the first section. As the distribution tapers off to the tails, the sections drastically decrease in size until only a combined .003 of the area under the curve is beyond + or – 3 standard deviations.



The process creates a ‘Z-score,’ which is the equivalent to a certain amount of standard deviations below or above the mean. This process uses 3 data points: The shape parameter, location parameter, and test statistic.

Recall that μ is the location parameter and σ is the shape parameter. The test statistic, ‘*x*’ in this example, is what you are ‘testing’ for. Since the statistic can be changed, this is your variable. Do not confuse $\bar{x}$ with *x* in later chapters. You can use a sample mean’s results as x against the population mean, but $\bar{x}$ will also function as a part of a parameter in later distributions. Use caution and properly analyze what the question is asking before answering.

After a sucessful standardization, the Normal Distribution should have 50% of the probability to the left and right of the mean with four equidistant parts. The first section is + or – 1 Standard Deviation from the mean. Because

Using the formula listed in the notation section of the guide, the standardization process is quick and simple. For this example, look at the first of the PDF graphs. It has a μ value of -2 and a standard deviation of .5. In this example, we wish to find what the probability is of having a value of -1 or fewer on this distribution. -1 would be the test statistic ‘*x*’ for this example. Using the formula and the data points given, we would calculate the Z-score as such:

$$Z = \frac{{}^{-}1 - {}^{-}2}{.5} = \frac{1}{.5} = 2\ \sigma$$

Highlight #3
"Z-Scores and Zα"

Definition: The amount of 'Standard Deviations' away from the mean a particular point is, and the specific Z-score associated with having 'α' amount of probability to the right of the value.

Notation: $Z_{\alpha} = P(X \geq \alpha)$

The Emperical Rule standardizes every Normal distribution into the distribution *X* ~ *N*(0,1). Every value that is produced from the standardization is a Z-score, which can be read as "_ standard deviations away from the mean." That means that a Z-score of 2.00 will be 2.00 standard deviations away from the mean of 0. Using the CDF table, the probability of having less than 2 standard deviations away from the mean is equal to .9772 or 97.7%. If you are given the probability, you can check a CDF table to find the nearest Z-score. If you are given the Z-score and one of the data points, you can solve for what the original

Typically, a Normal distribution CDF will show incremental additions to the probability starting from -∞ and going up until the decided z-score. However, there is an alternative form of the Z-score called Z-alpha, this is the first symbol in the 3rd formula, that shows probabilities from the decided z-score up to +∞ (Or reading from right to left). Graphically, a normal z-score will start at the left tail and go towards the point, while Z-alpha will start at the point and go towards the right tail. Z-alpha is meant to be a Z-score that has an 'alpha' amount of area to the right of the point. Z.0228 would have 2.28% of the probability to the right of it.

There are two ways to find the Z-scores associated with Z-alpha with a CDF table: Find the Z-score associated with the Probability of (1-alpha) or, find the Z-score associated with P(X < alpha). For example. If I wanted to find the Z-score associated with $Z_{.05}$, I can look to the CDF table and find the Z-score associated with .9500. The Z-score closest to this would be 1.64. The other method takes advantage of the fact that the Normal Distribution is a bell curve and that the probabilites on one side of the curve is equal to the other side. Using a CDF calculator, the Z-score associated with a probabiltiy of .05 is -1.64. When using the second method, remember to switch the negative sign around before reporting your answer.

There are many ways to find the values of a CDF. Your professor may use JMP and have the computer calculate it. Other professors may ask you to use websites or a graphing calculator with statistic functions. A calculator or website will give you a more accurate Z-score or Z-alpha. However, if you are using the CDF table and the value you are looking for is inbetween two points, pick the point with less probability. It is always better to round and loose some of the probability than risk claiming that there is a higher probability for an event than there actually is. If your professor allows you to chose between methods, use whichever statistical tool is the easiest for you to use.

Check Your Learning

1. After the 20X0 census, the state of Texas found that the average number of people living in a household was 4.2 with a standard deviation of .7.
   a. Write the Function Notation of this data set.
   b. Use Standardization to find the Z-score of a household with 5 people.

2. Avery is 5 foot 10 inches, or 70 inches, and she is in the 85^{th} percentile of her STA-1380 class.
   a. If the mean height of her course was 5'5", what is standard deviation assuming height is normally distributed?
   b. Using the Standard deviation, you found in Part A and find the probability associated with the heights between 5'2" and 5'9".

Things Students Struggle With

1. Converting words into Z-scores
   a. Whenever you are working through a statistics problem, A large portion of what the problem is asking you to solve for will be in words. Every question dealing with probability will have 5 parts.

      Probability, 'X', Sign, Z-score, and Equals. One way I learned how to set up these parts of the problem was the phrase, "P Exercised into Q" Phonetically, the word 'Exercise' is the combination of the sounds of "XSZ" This clue helps set up the nomenclature of $P(X < Z) =$. Always read the question and find out what the P and XSZ trio are.

      First is the Parameters or 'P' of the problem. Since this clue revolves around the fact that the question is asking for probability, there will have to be parameters set. These will be the Population standards and will be the means and standard deviations of the data set.

      Next is to find what the range of 'X' is for the problem. Does it include a tail? Between two points in the curve? Look for the range of 'X' first.

This range will help you determine the Sign or 'S' of the problem. Is the question asking for greater than or less than the z-score? Is it between two Z-values?

Finally, use the range of 'X' as the statistic and the parameters found from 'P' to find the Z-score for the probability.

After all the parts of the problem have been identified, use a CDF to find the probability that equals, 'Q' the Z-score you have found.

2. Statistic vs Parameter:

  a. As mentioned in the Normal Distribution Highlight, the parameters are the population standards. These are expected and a single value. The statistic is the term to describe the '*x*' value that you are testing for. This value can be any value that '*X'* can take on. A sample's mean value can be chosen as the statistic and can be compared to the population mean, but a sample mean itself is the parameter of the sample. Take precautions and learn to identify whether a piece of information is a statistic or parameter.

Concluding Comments

That's it for this week! Please reach out if you have any questions and don't forget to visit the Tutoring Center website for further information at https://www.baylor.edu/tutoring.

Answers to CYL

1. **a.** *X ~ N(4.2,.7)*
b. Z-score = 1.14

2. **a.** 5.88 standard deviations.
b. *P(X = -.51 < x < .68)* CDF: .7517 - .305 = .4467 or 44.67%