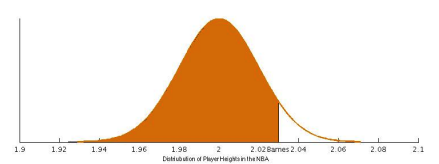
***Udacity Data Analyst Track***

**I. Into to Descriptive Stats**

5. Standardizing

* **Z score/Standard Score** = Given an observed value x, Z score finds the # of SD’s that x is away from the mean 🡪 Z = (x – mu) / SD
* **Z** = # of SD’s away from mean
* **Standard Normal Curve =** curve that is the resulting distribution when we standardize scores.
* Used along w/ a **Z table** to compute percentages above, below, or in between observations
* Area under this curve = 1 = sum of all frequencies of all the proportion bins in a histogram
* Mean = median = mode 🡺 symmetrcail curve (rare in real life)
* If concerned w/ finding a proportion less/greater than some value in a distribution, we should use **relative frequencies**
* Plots of distributions look almost alike (counts vs. proportions)
* Smaller bin size = more detail
* The average height of a professional basketball player was 2.00 meters w/ SD = 0.02 meters.
* Harrison Barnes is a basketball player who is 2.03 meters tall. How many SD’s from the mean is he?
* 1st sketch the normal curve that represents the distribution of basketball player heights.
* 
* Notice the mean height = 2.00 is right in the middle w/ tick marks 1 SD/0.02 meters in length in both directions.
* Next compute the standard score/z score 🡪 (2.03 - 2.00) / 0.02 = 3/2 = **1.5**
* *Barnes’ height is 1.5 SDs from the mean* 🡪 is 1.5\*SD + µ = Barnes’ Height
* The average height of a professional hockey player is 1.86 meters w/ SD = 0.06 meters. Tyler Myers is the same height as Harrison Barnes. Which of the 2 is taller in their respective league?
* Z-score = 2.03 - 1.86 / 0.06 = 0.17 / 0/06 = **2.833**
* Comparing the 2 z-scores we see Tyler Myers score of 2.833 SD’s above the mean is larger than Barnes’ score of 1.5 SD’s above the mean.
* This tells us that there are more hockey players shorter than Myers than there are basketball players shorter than Barnes
* Find the Z-score given the following info
* µ = 54, s = 12, x = 68 🡪 **14/12 = 1.1667**
* µ = 25, s = 3.5, x = 20 🡪 **-5/3.5 = -1.429**
* µ = 0.01, s = 0.002, x = 0.01 🡪 = **0**
* The average GPA of students in a local high school is 3.2 w/ SD = 0.3. Jenny has a GPA = 2.8. How many standard deviations away from the mean is Jenny’s GPA?
* 2.8 – 3.2 / 0.3 = **-1.3333**
* Jenny’s trying to prove to her parents that she is doing better in school than her cousin. Her cousin goes to a different high school where the average GPA = 3.4 w/ SD = 0.2. Jenny’s cousin has a GPA of 3.0. Is Jenny performing better than her cousin based on standard scores?
* 3 – 3.4 / 0.2 = **-2 = Yes**
* Kyle’s score on a recent math test was 2.3 SD above the mean score of 78%. If the SD of the test scores were 8%, what score did Kyle get on his test?
* 2.3 = x – 78 / 8 0🡪 18.4 = x – 78 🡪 **x = 96.4**
* Unpopularity
* Katie’s FB friends = 63, mean friends = 190, SD = 36 🡪 Z-score = 63 – 190 / 36 = **-3.52778 = 3.5 SDs below the mean = more unpopular**
* Andy’s Twitter followers = 54, mean = 208, SD = 60 🡪 Z-score = 54 –208 / 60 =- **2.5667** = **2.5 SDs below the mean = less unpopular**
* Different distributions (see mean + SD) 🡪 can put them on the same scale in terms of their unique SD’s = **standardize** the distributions, using 0 as the reference point
* Standardize Katie and Andy’s scores, Katie is further from mean 🡪 Standardizes scores show the proportion of values that are lower or higher in that distribution 🡪 greater proportion of people w/ more friends than Katie than there is of people w/ more followers than Andy
* *When we convert any value in a distribution to a Z-score, we standardize the distribution* 🡪 can start w/ any distribution and then standardize it
* If we standardize a distribution by converting *every* value in it to a Z-score, the *new mean of the distribution would be = 0* and then the *new SD of the distribution would be = 1*
* Z-score of mean = (mean – mean) / SD = **0 = new standardized mean**
* After standardizing the distribution, mean = 0 🡪 so Z-score of SD = (SD – 0) / SD = **1 = new standardized SD**
* **Standard Normal Distribution** = mean – 0, SD = 1
* Z-score = 1 🡪 1 SD above mean 🡪 50% + (68/2)% = 84% of data is below this value, 16% above

