4/12/2022

Normal Distribution (Chapter 6)

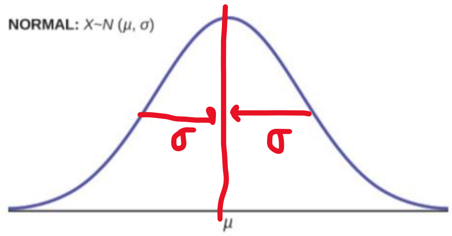


Figure 6.2

Defined by the mean, and the standard deviation.
Distribution is symmetric, so mean = median = mode

Standard normal distribution == allows us to compare variables with different normal distributions
Z-score: the number of standard deviation from the mean.

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

μ mu is the mean, σ sigma is the standard deviation (based on the theoretical distribution)
The equivalent version for data:

$$z = \frac{x - \bar{x}}{s}$$

Suppose the SAT (English and Math) has a mean of 1000 (500 on each test) and a standard deviation of around 199 (100 on each test). And the ACT has a mean of around 21 (for each test), the standard deviation is around 5.2.

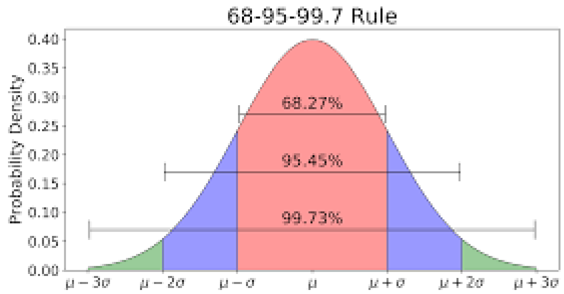
Suppose you take both tests. On the SAT, you got a score 1280. On the ACT you got a score of 29. Which is the better score to send to colleges?

$$z_{SAT} = \frac{1280 - 1000}{199} = 1.4 \ldots$$

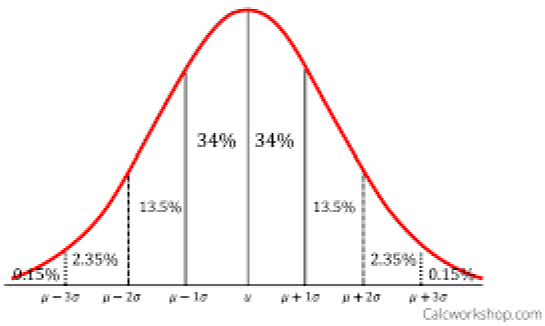
$$z_{ACT} = \frac{29 - 21}{5.2} = 1.5 \ldots$$

Both scores are fairly similar, but the ACT score is a little higher.

Empirical Rule
68-95-99.7 Rule
In a normal distribution, there are approximately 68% of the population between 1 standard deviation of the mean. There is roughly 95% of the population between 2 standard deviations of the mean. And there is roughly 99.7% of the population between 3 standard deviations of the mean.



Outside 2 standard deviations is considered "unusual". 5% of the population or less.



The SAT Math test has a mean of roughly 500 and a standard deviation of 100. What percent of test takers would be expected to receive scores between 400 and 600?
400 is 1 standard deviation below the mean $z = \frac{400-500}{100} = -1$
600 is one standard deviation above the mean. $z = \frac{600-500}{100} = 1$
68% is expected to be between these two values.

What percent of the test takes would we expect to get results between 400 and 700?
400 is 1 standard deviation below the mean
700 is 2 standard deviations above the mean $z = \frac{700-500}{100} = 2$
34+34+13.5=81.5%

What percent of the test takers would be expected to receive a score above 700?
700 = 500 + 2(100) so 2 standard deviations above the mean
2.35%+0.15% = 2.5%

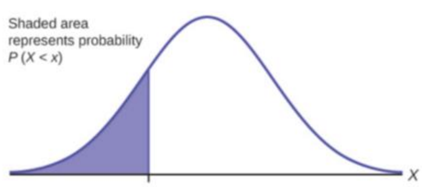


Figure 6.4

Using a general normal distribution to find probabilities.
Typically when we are using normal distributions, the standard normal distribution (mean of 0 and a standard deviation of 1) uses the letter z. While a general normal distribution (with some other mean and some other standard deviation which have to be specified) is given the x label.

The probability is considered to be the area under the curve (by default) to the left of the value given (is less than the value). If we want the area (probability) to the right (above the value) we will have to subtract the area below from 1. (using the complement rule).

$P(X < x)$ this is the probability of being less than x, and the area to the left.
$1 - P(X < x) = P(X > x)$ this is the probability of being greater than x, or the area to the right.

Go to Excel for examples.
To calculate the in-between case: calculate the probability of being less than the larger value, and the probability of being less than the smaller value, and then subtract.

Going backwards: How do I find an observation from a probability?
Inverse operation. "Inverse Norm".

The probability that goes into the inverse norm function must be the probability below the desired value.
So if you want the probability (or the percentage of the population) that is in the bottom 10%, you put 10% or 0.1 into the formula. But if you want the top 10%, you need to put in 1-0.1 = 0.9, or 90% that is below.

Excel for examples.

If you have the observation and want a probability, use the dist function (in calc this is cdf)
If you have a probability and want an observation, use the inv function

Next time: we will look at the central limit theorem, and t-distribution