Descriptive Measures for Populations; Use of Samples.

Recall: The sample mean is

$$\bar{x} = \frac{\sum x_i}{n}$$

for a sample of size n from a population.

The population mean or the mean of the variable X is the mean of all possible observations in the population. It is denoted μ_x or – if no confusion will arise – μ. For a finite population,

$$\mu = \frac{\sum x_i}{N}$$

where N is the population size.

Note:

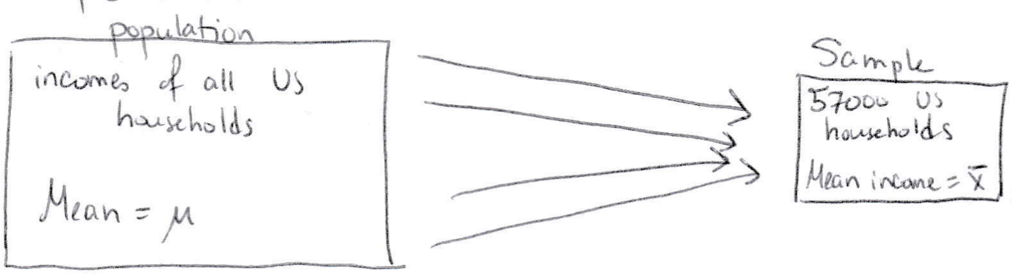
- There is only one possible value for μ with a given population.
- There are many possible value for $\bar{x}$ (depending on which sample is drawn)

Using a Sample Mean to Estimate a Population ~~Mean~~ Mean

Estimating mean household income:

It would be costly to do a census for this information.
It is much easier to collect a sample of households.

- Variable: income
- Population: US households
- Population data: incomes of all US households.
- Population mean: mean income μ of US households.
- Sample: 57000 US households
- Sample data: incomes of the 57000 sampled US households
- Sample mean: mean income $\bar{x}$ of the 57000 sampled US households.



The Population Standard Deviation

Recall: The sample standard deviation for a variable X and a sample of size n from a population is

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$$

The population standard deviation (for a finite population) or standard deviation of the variable X is denoted σ_x or, if no confusion will arise, σ. Then

$$\sigma = \sqrt{\frac{\Sigma (x_i - \mu)^2}{N}}$$

where N is the population size.

Where s^2 is the sample variance, σ^2 is the population variance.

Using a sample standard deviation to estimate a population standard deviation

A drug, tetrahydrolipstatin, is used to help manage weight loss in obese individuals with major risk factors such as diabetes, high blood pressure, and high cholesterol. It works by blocking dietary fat absorption in the intestines.

A standard dose comes in 120 mg capsules. Capsules may vary a bit from exactly 120mg and it is important to minimize this variation to keep doses consistent. We would like to know the population standard deviation.

Variable: capsule weight

Population: all "120 mg" tetrahydrolipstatin capsules

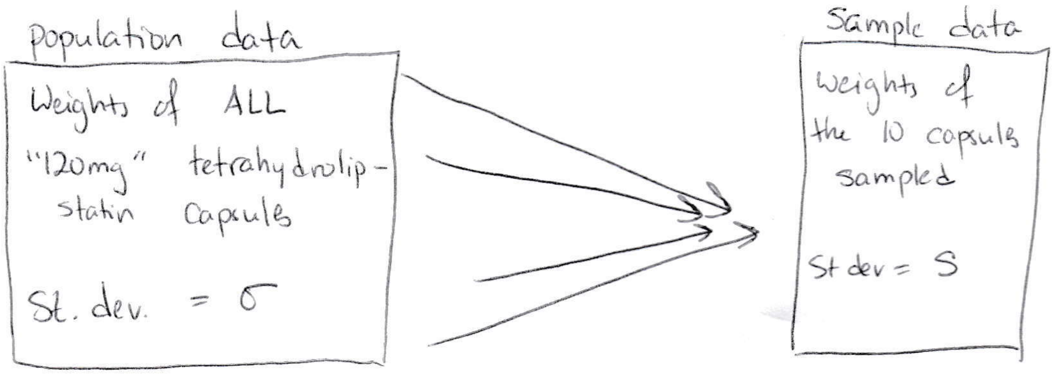
Population data: weights of all capsules

Population standard deviation: σ, the standard deviation of the weights of all capsules

Sample: 10 capsules chosen

Sample data: the weights of the 10 capsules in the sample

Sample standard deviation: standard deviation, s, of the weights of the 10 capsules.



We can use the sample standard deviation s as an estimate of σ.

<u>Def</u>: A <u>parameter</u> is a descriptive measure for a population.

<u>Def</u> A <u>statistic</u> is a descriptive measure for a sample.

So μ and σ are parameters

$\bar{x}$ and s are statistics

<u>Aside</u>: μ is the Greek letter "mu"

σ is the Greek letter "sigma"

Standardized Variables

For a variable X, the variable

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

is called the standardized version of X or the standardized variable corresponding to X.

Note: A standardized variable Z has mean 0 and standard deviation 1. This will make them very useful!

Ex

X	-1	3	3	3	5	5

Ex Standardize X.

X	-1	3	3	5	3	5
Z	-2	0	0	1	0	1

(X is a simple variable with all possible observations shown.)

$$\mu = \frac{-1+3+3+3+5+5}{6} = 3 \qquad \sigma = \sqrt{\frac{\Sigma (x_i-\mu)^2}{N}} = \sqrt{\frac{24}{6}} = 2$$

$x-\mu$	$(x-\mu)^2$
-4	16
0	0
0	0
2	4
0	0
2	4
	24

$$Z = \frac{X-\mu}{\sigma} = \frac{X-3}{2}$$

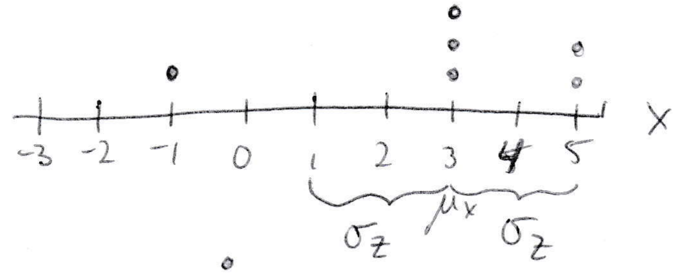
What are the mean and sd of Z?

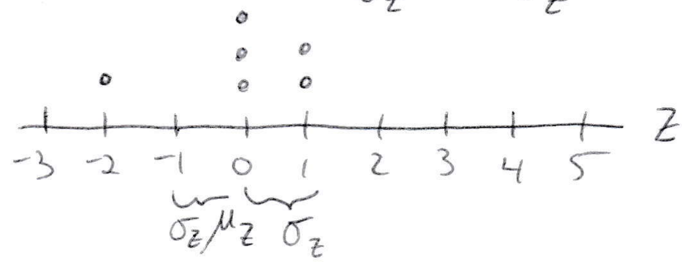
$$\mu_Z = \frac{-2+0+0+1+0+1}{6} = 0$$

$$\sigma_Z = \sqrt{\frac{4+1+1}{6}} = \sqrt{\frac{6}{6}} = 1$$

$Z-\mu$	$(Z-\mu)^2$
-2	4
0	0
0	0
1	1
0	0
1	1

X	-1	3	3	3	5	5
Z	-2	0	0	0	1	1





Notice: the shape of the distribution is the same, but the mean shifts to 0 and the scale changes to 1

<u>Z-Scores</u> For an observed value of a variable X, the corresponding value of the standardized variable Z is the <u>Z-score</u> of the observation. These may also be called <u>standard scores</u>.

- A negative Z-score means an observation is below the mean: $X < \mu$
- A positive Z-score means an observation is above the mean: $X > \mu$

<u>Ex</u> Carly got an ACT score of 21 while Bran got an SAT score of 1100. They want to know who performed better.
The mean SAT score is $\mu_S = 1060$ and the sd is $\sigma_S = 195$
The mean ACT score is $\mu_A = 18$ and the sd is $\sigma_A = 6$

<u>Soln</u> We can compare their scores by standardizing.

Carly's test Z-score is

$$Z_{carly} = \frac{X - \mu_A}{\sigma_A} = \frac{21 - 18}{6} = \frac{3}{6} = 0.5$$ std deviations <u>above</u> the mean ACT score.

Bran's test Z-score is

$$Z_{Bran} = \frac{X - \mu_S}{\sigma_S} = \frac{1100 - 1060}{195} = \frac{40}{195} = 0.2051$$ std devs. <u>above</u> the mean SAT score.

So Carly performed better than Bran on their respective tests.

Both Z-scores are positive, so both performed better than the average!

Since z-scores tell us how many standard deviations an observation is from the mean, we can use the three-standard-deviations rule, Chebyshev's Rule, and/or the empirical rule to think about how unusual an observation is.

how far from the mean, in standard deviations