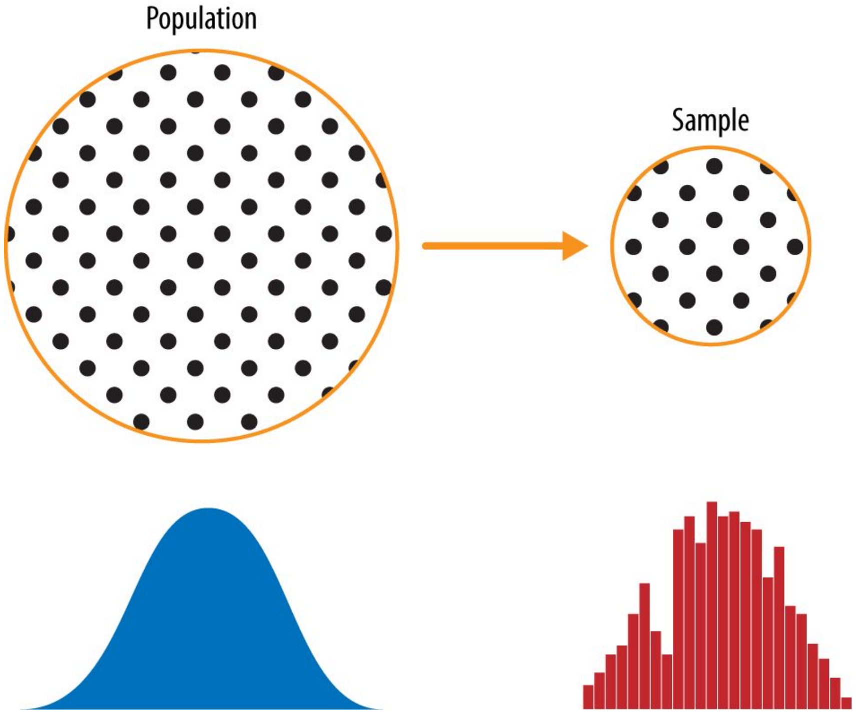
Interval Sampling and Sampling Distributions

Understanding a Population from a Sample

Sampling



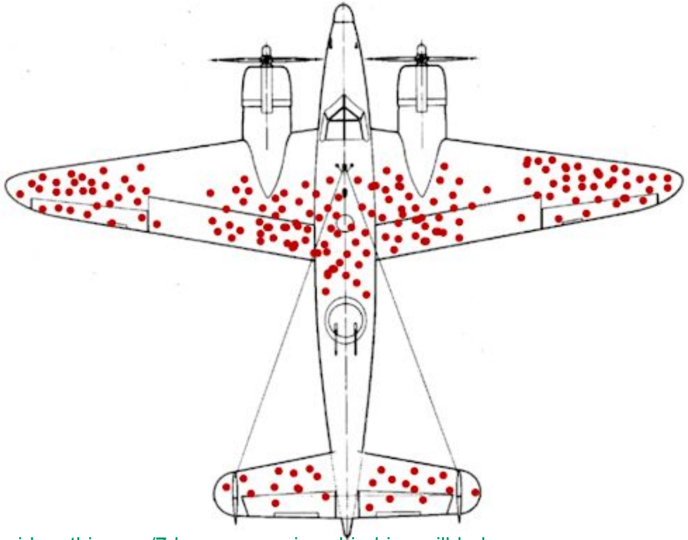
- Point Estimate
- Range of values

Why is it used?

- Point estimate of sample mean for population mean

http://www.accutrend.com/it-still-comes-down-to-garbage-in-garbage-out/

Common Sample Biases

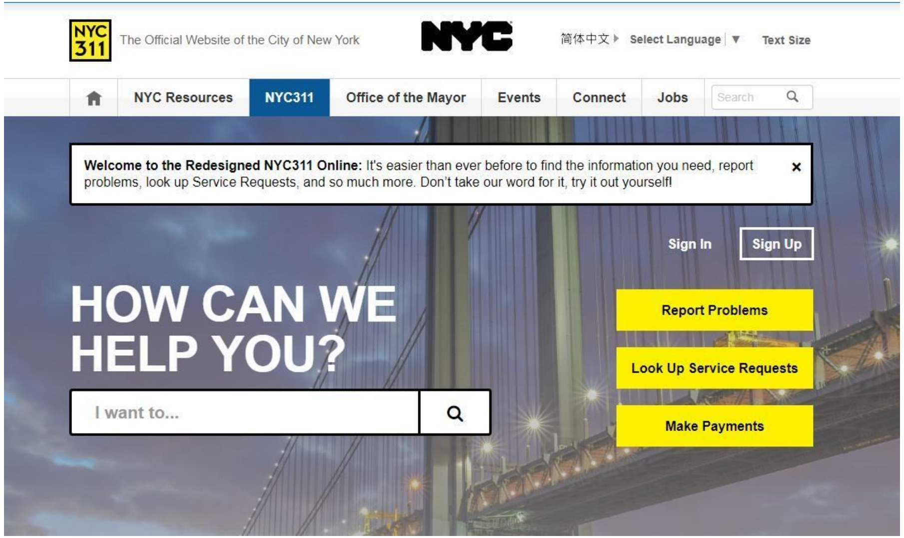


http://blog.idonethis.com/7-lessons-survivorship-bias-will-help-make-better-decisions/

Bias refers to the tendency of a sample statistic to systematically over/underestimate a population parameter

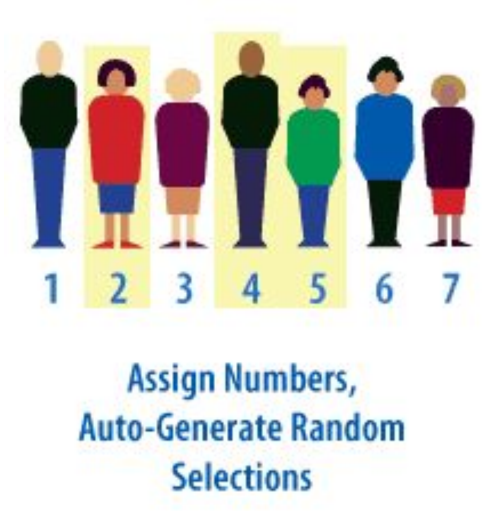
- Selection bias
- Non response bias
- Social-desirability bias

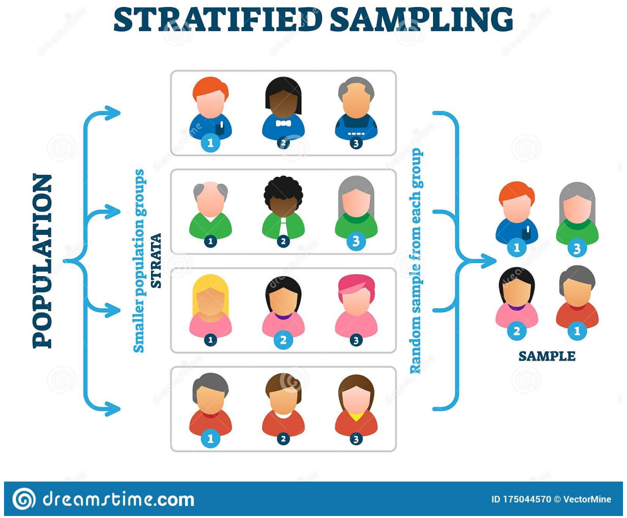
Bias in Data

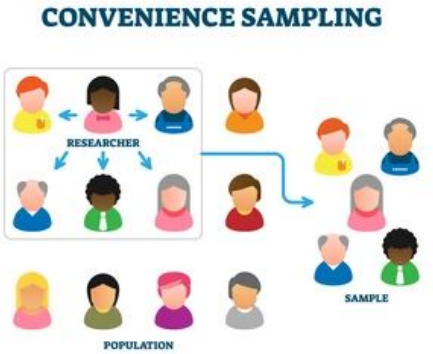


Consider 311 Service Request Dataset, what are some equity issues that might relate to the sample of service requests that are recorded on 311?

Sampling Methods







Simple Random Sample

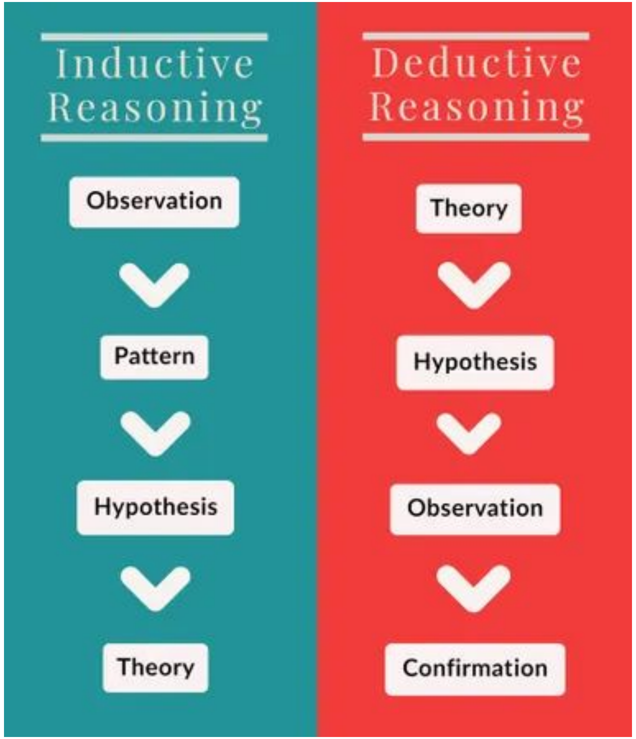
Stratified Random Sample

Convenience Sample

[Demo]

Why is Sampling Distribution Used?

- A sample mean is the point estimator of the population mean
- We use a statistic to estimate a parameter
  - “Statistic” is the random variable we measure
  - “Parameter” is the unknown constant we want to gain insight into
- Sampling distribution is when a sample point estimate is repeated many times and a frequency distribution can be calculated



THE EXPECTED VALUE AND THE STANDARD ERROR OF THE SAMPLE MEAN

The expected value of the sample mean $\overline{X}$ equals the population mean, or $E\left(\overline{X}\right) = \mu$. In other words, the sample mean is an unbiased estimator of the population mean.

The standard deviation of the sample mean $\overline{X}$ is referred to as the standard error of the sample mean. It equals the population standard deviation divided by the square root of the sample size; that is, $se\left(\overline{X}\right) = \frac{\sigma}{\sqrt{n}}$.

How to Calculate

Consider how ‘shrinkflation’ is changing the size/weight of their products, Doritos chips claims each bag to weigh 16 ounces. However, there is likely to be a inconsistency due to how the chips are packaged, resulting in a standard deviation of 0.8 ounces.

1. What is the expected value and the standard error of the sample mean if we took a sample of 2 bags?

How to Calculate

Consider how ‘shrinkflation’ is changing the size/weight of their products, Doritos chips claims each bag to weigh 16 ounces. However, there is likely to be a inconsistency due to how the chips are packaged, resulting in a standard deviation of 0.8 ounces.

1. What is the expected value and the standard error of the sample mean if we took a sample of 2 bags?

$\mu = 16$, $\sigma = 0.8$, we use $E(\bar{x}) = \mu$ and standard error $(\bar{x}) = \sigma/\sqrt{n}$

$$n = 2,\ E\left(\overline{X}\right) = 16 \text{ and } se\left(\overline{X}\right) = \frac{0.8}{\sqrt{2}} = 0.57.$$

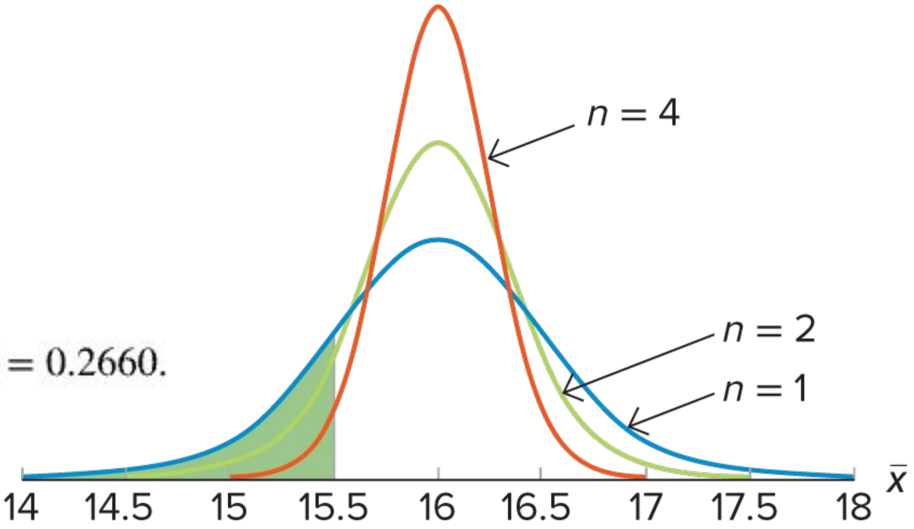
How to Calculate

μ = 16, σ = 0.8

1. What is the probability that a randomly selected bag is less than 15.5 ounces?

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}.$$

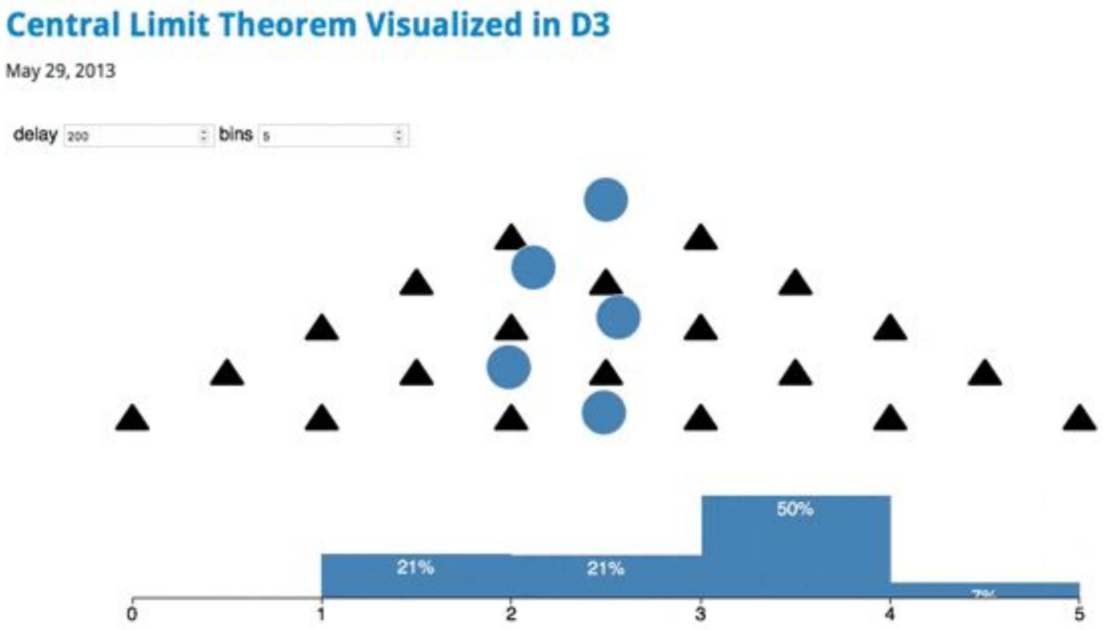
$$P(X < 15.5) = P\left(Z < \frac{15.5 - 16}{0.8}\right) = P(Z < -0.625) = 0.2660.$$



Demo

Central Limit Theorem

Sum of the average of a large number of independent samples observations from the same underlying distribution has an approximate normal distribution



source: @vicapow http://blog.vctr.me/posts/central-limit-theorem.html via @setosaio

Finite Population Correction Factor

Sample Size	FPC
1	1.000
10	1.000
25	0.999
50	0.998
100	0.995
500	0.975
1000	0.949
5000	0.707
8000	0.447

- How do we know our sample size is representative enough?
- Sample size is large relative to the population size
- We assume to be more precise the larger percentage of the population we sample from

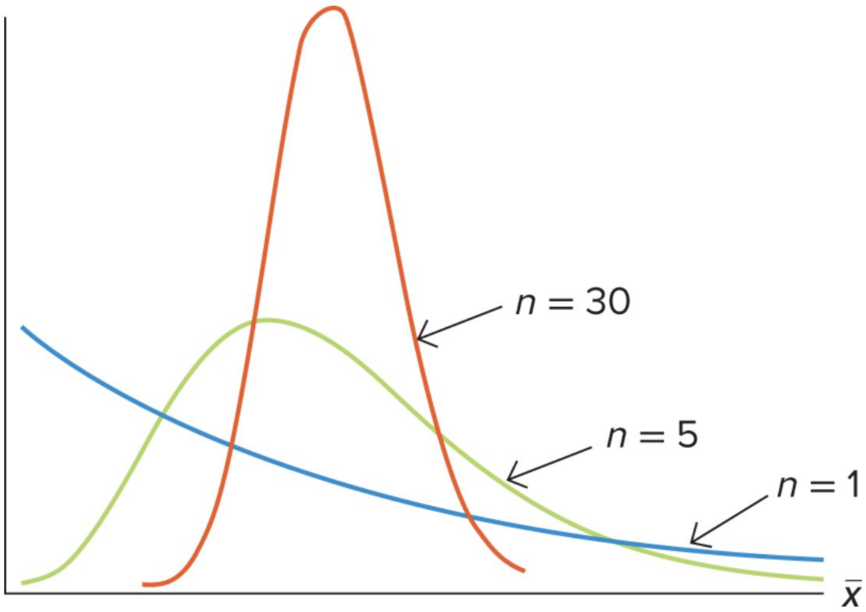
THE FINITE POPULATION CORRECTION FACTOR FOR THE SAMPLE MEAN

When the sample size is large relative to the population size ($n \geq 0.05N$), the finite population correction factor is used to reduce the sampling variation of the sample mean $\overline{X}$. The resulting standard error of $\overline{X}$ is $se\left(\overline{X}\right) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$. The transformation for any value $\overline{x}$ to its corresponding z value is made accordingly.

Why is it Used?

- Understand Population
- Evaluate Cost Effectiveness
- Quality Control

FIGURE 7.3 Sampling distribution of $\overline{X}$ when the population has an exponential distribution

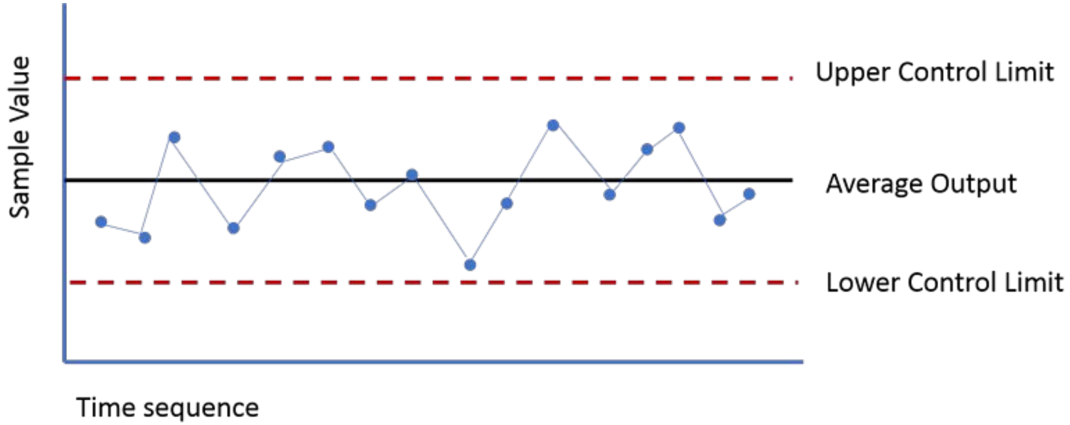


Statistical Quality Control

1. Acceptance sampling
   a. A portion of the product/asset/service is inspected to ensure it meets requirements
2. Detection Approach
   a. Production process is inspected
3. Variation is inevitable
   a. Chance variation
   b. Assignable variation

How to Measure Quality Control

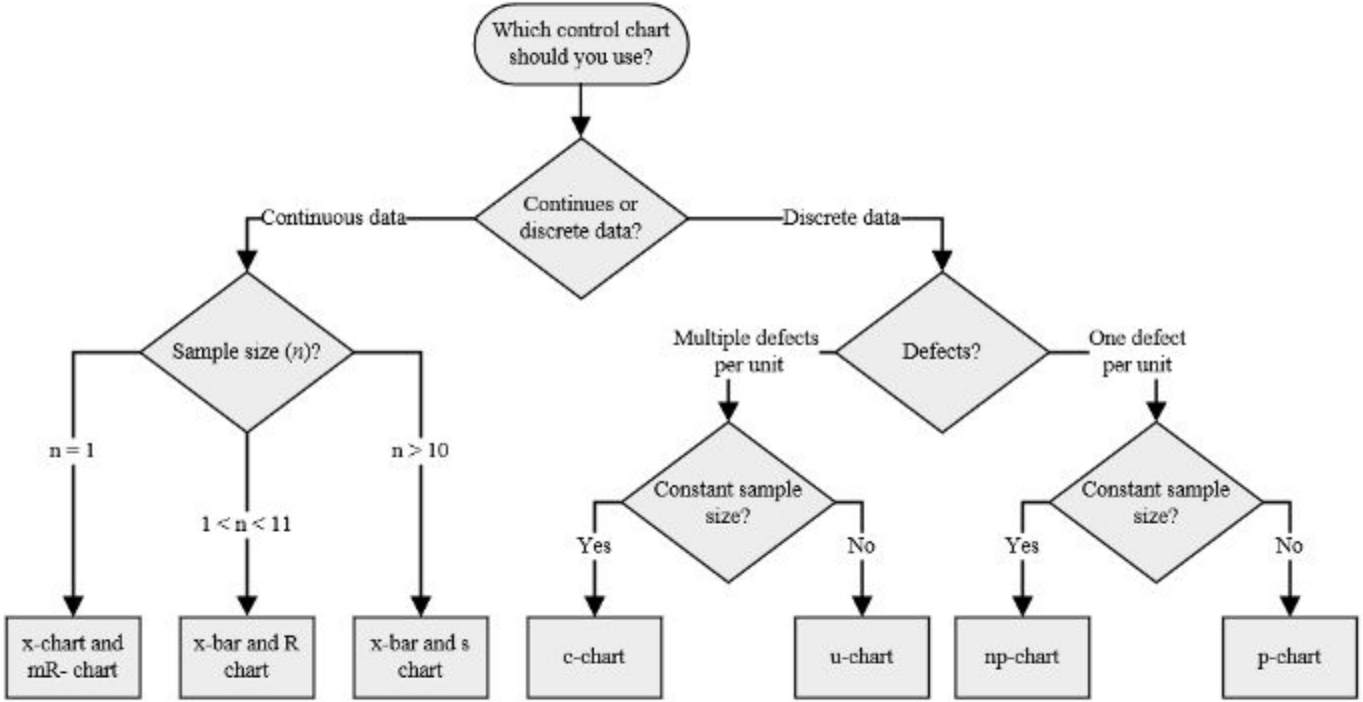
- Sample estimate taken over time
- Expected value is the baseline
- UCL and LCL are the limits indicating excessive deviation from the expected value



$$\text{UCL: Expected Value} + (3 \times \text{Standard Error})$$

$$\text{LCL: Expected Value} - (3 \times \text{Standard Error})$$

How to Calculate



Example

Manufacturer produces one-gallon jugs of milk, every two hours the company samples 25 jugs and calculates the following sample mean filling weights (in ounces). Assume that when the machine is properly operating, mu = 128, sigma = 2. And that filling weights follow the normal distribution.

Is the machine operating properly?

1. Calculate the upper and lower limits
2. See if any sample value falls outside of control limits
3. Plot the control chart

Demo