MKTG 352

Principles of Marketing Research

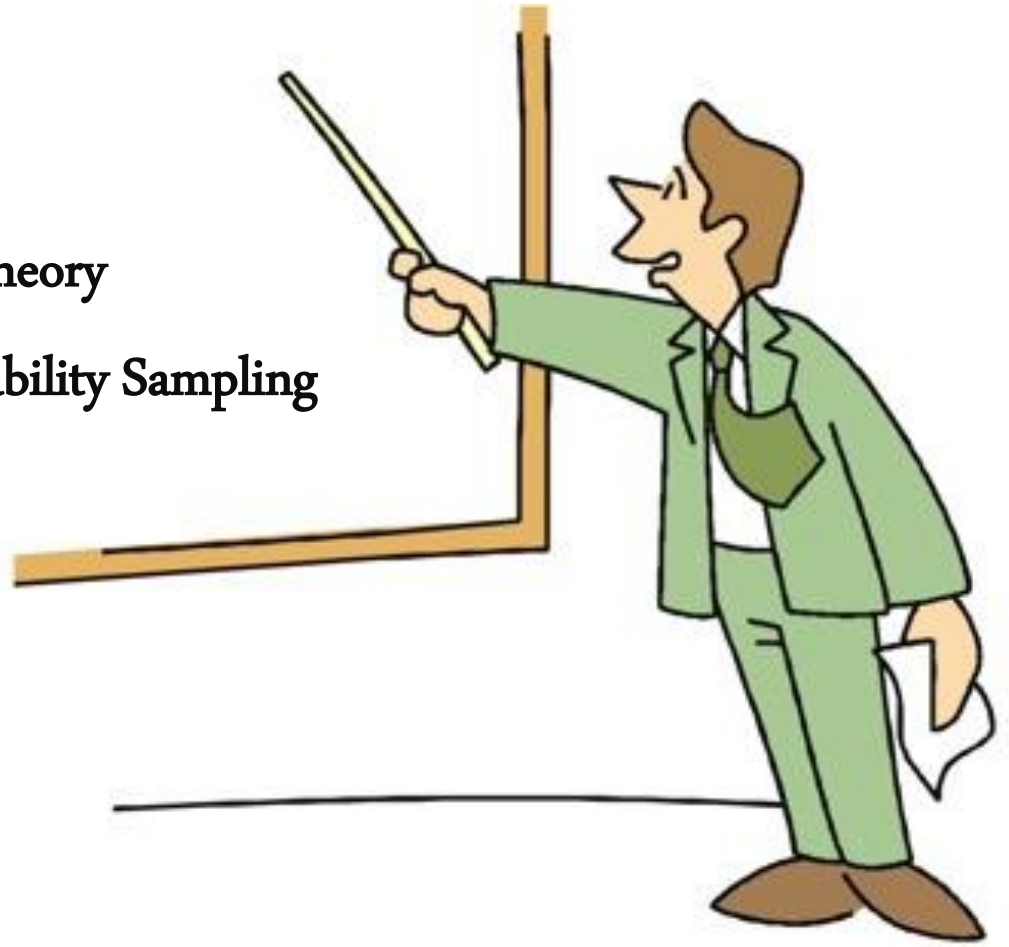
Instructor: Sina Aghaie
Email: sina.aghaie@grad.moore.sc.edu
Office: DMSB, 4th Floor Marketing Department
Office hours: Mon & Fri 11-1pm, or by appointment

Session #15

Topics for today…

The Basics of Sampling Theory

Probability and Nonprobability Sampling



Announcement

Mid term Exam:

Wed., Feb. 28th

Cluster Sampling

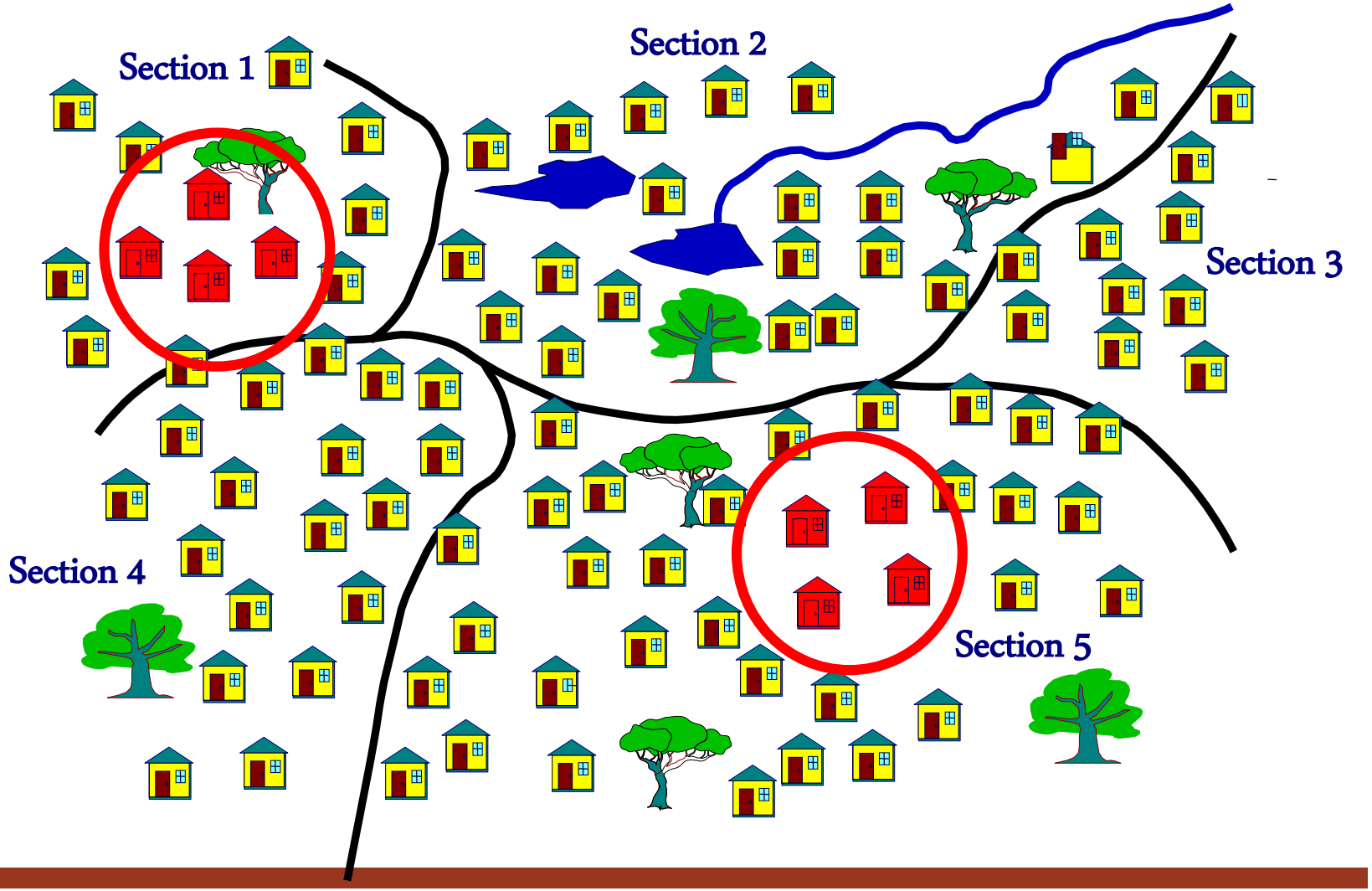
- The target population is first divided into **mutually exclusive** and **collectively exhaustive** subpopulations, or clusters.
- Elements **within** a cluster should be as **heterogeneous** as possible, but clusters themselves should be as **homogeneous** as possible.
- Then a **random sample of clusters** is selected, based on a probability sampling technique such as SRS.
- For each selected cluster, either all the elements are included in the sample (one-stage) or a sample of elements is drawn probabilistically (two-stage).

Randomly select 3 clusters, B, D and E. Within each cluster, randomly select one or two elements. The resulting sample consists of population elements 7, 18, 20, 21, and 23.

A	B	C	D	E
1	6	11	16	**21**
2	**7**	12	17	22
3	8	13	**18**	**23**
4	9	14	19	24
5	10	15	**20**	25

Note, no elements are selected from clusters A and C.

Probability Sampling



- In the stratified sampling all groups are included
- In the cluster sampling we use a random selection of groups

Probability Sampling

Ideally, each cluster should be a small-scale representation of the population.

In **probability proportionate to size sampling**, the clusters are sampled with probability proportional to size.

Probability Sampling

Advantages:

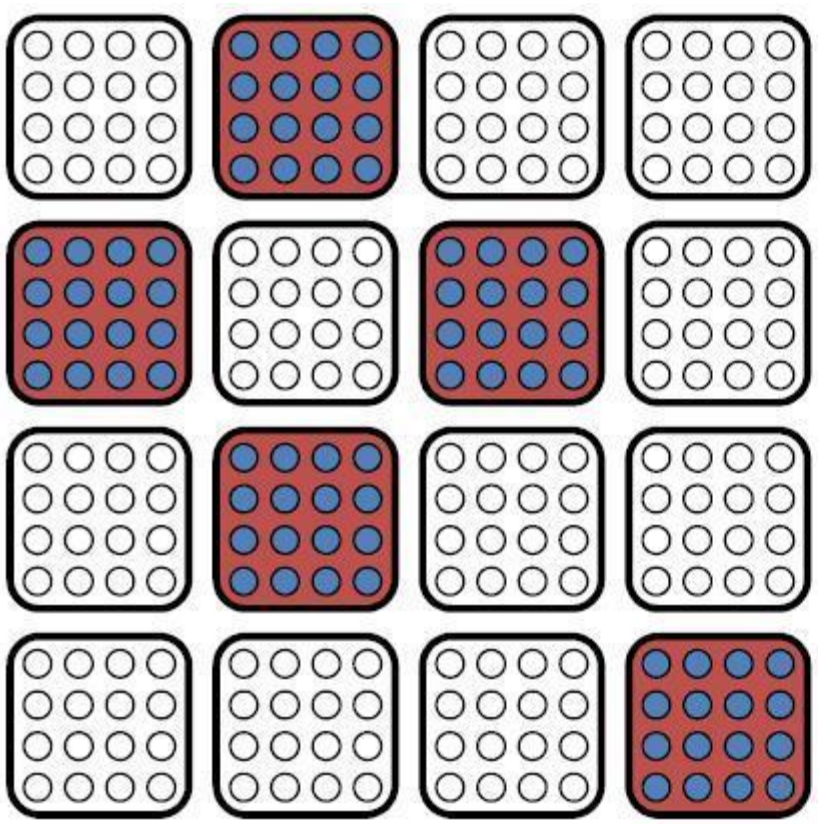
Is cost effective and easy to implement (easy to divide the defined target population into clusters).

Disadvantages:

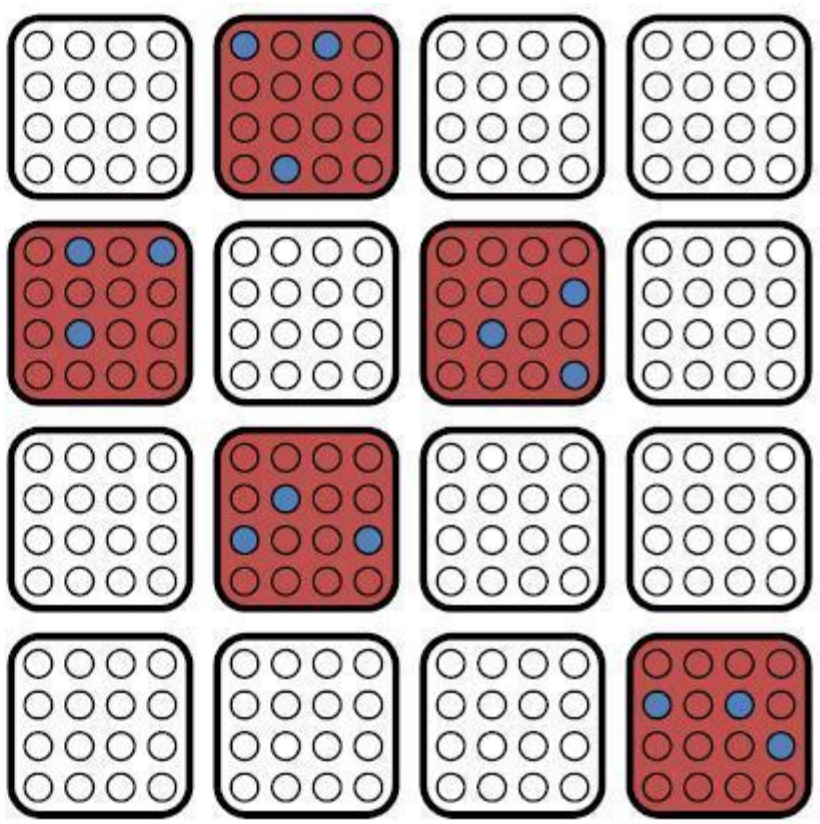
The clusters are often homogenous, i.e. people in a particular city might hold similar preferences and have similar lifestyles.

Hence, generalizing the results to the entire population may be inaccurate.

Probability Sampling



Randomly select Clusters and select all subjects



Randomly select Clusters and select subjects randomly

- In the stratified sampling all groups are included
- In the cluster sampling we use a random selection of groups

Probability and Nonprobability Sampling

When you design your sampling:Two basic type

- **Probability Sampling: Each sampling unit has a known chance (probability) of being included in the sample.**

  I know the probability of each passenger from Southwest's list of 10,000 customers being included in my sample of 100.

- **Non-Probability Sampling: The probability of each sampling unit being included in the sample is not known**

  I am interviewing people at the mall about iPod usage. I do not know what the probability is of each person from the target population being selected.

Non-probability Sampling

Nonprobability Sampling Designs

Four types:

Convenience Sampling

Judgment Sampling

Quota Sampling

Snowball Sampling

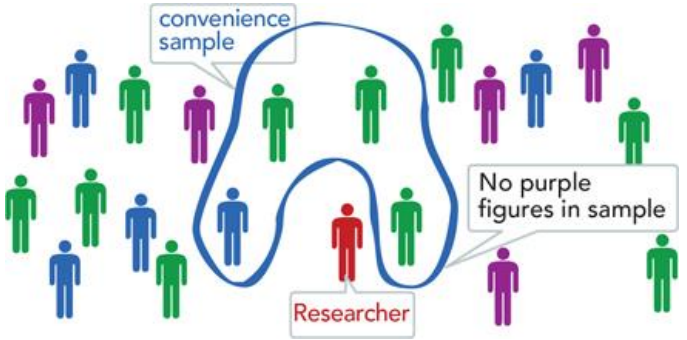
Non-probability Sampling

Convenience Sampling: Samples are drawn based on convenience (ease of administering a study).

Often, respondents are selected because they happen to be in the right place at the right time.

Interviewing individuals at malls. It is easy and quick.

The assumption is that people who shop at the mall would be representative of the defined target population as a whole.



Non-probability Sampling

Advantages:

A large number of respondents can interviewed in a short time.

Can be used in the early stages of research (e.g. to refine the questionnaire and pre-test certain scales used).

Disadvantages:

The convenience sample might not be representative of the true population and it can be hard to determine representativeness.

Data **is not** generalizable to the defined target population.

Non-probability Sampling

Judgment Sampling: A sampling procedure in which an experienced researcher selects the sample based on some appropriate characteristics of the sample members to serve a purpose.

- Instead of interviewing customers, a researcher may interview its sales representatives since they might know more about how the product is being received in the field.
- Hence, the researcher believes the opinions of a certain group of individuals will be most beneficial for the study
- The assumption is that the opinions of a group of experts will be representative of the target population.

Non-probability Sampling

Advantages:

If the researcher's judgment is accurate, the results will be better than those obtained by convenience sampling, as the individuals were selected specifically because they could provide better information.

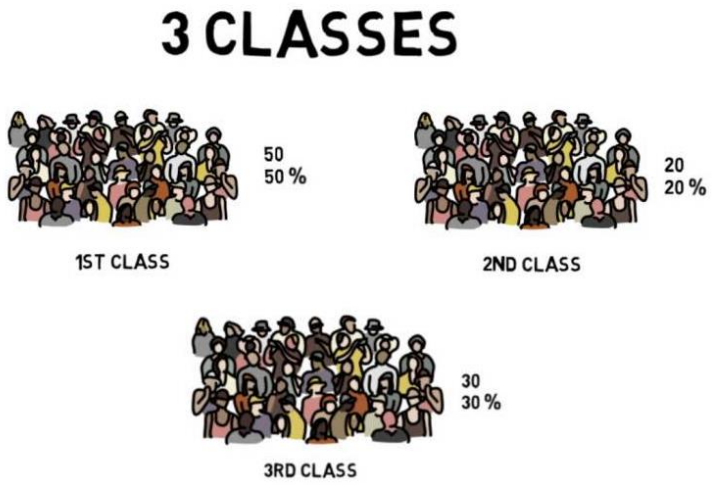
Disadvantages:

Again, difficult to know **how representative** this sample is of the target population. Results should be interpreted cautiously.

Non-probability Sampling

Quota Sampling: Participants are selected according to a pre-specified quota regarding demographics, attitudes or behavior.

- May be viewed as two-stage restricted judgmental sampling.
- Researchers may have a pre-specified number of males to include in the research, or a pre-specified number of people who do not own the product to be included in the research.



Non-probability Sampling

Population Characteristics

	<35 Years	35 Years and Above	Total	Percentage
Working women	300	200	500	50
Nonworking women	200	300	500	50
Total	500	500	1,000	100
Percentage	50	50	100	

Sample Characteristics

	<35 Years	35 Years and Above	Total	Percentage
Working women	50	0	50	50
Nonworking women	0	50	50	50
Total	50	50	100	100
Percentage	50	50		

Non-probability Sampling

Advantages:

Appropriate subgroups are identified and included in the survey, according to certain pre-set proportions.

Selection bias is reduced as sub-groups are ensured participation.

Disadvantages:

The accuracy of the results are dependent upon whether or not the **pre-set quotas** were appropriately set by the researcher.

Generalizing the results beyond the sample would be difficult.

Non-probability Sampling

Snowball Sampling: Identifying an initial group of respondents who can help the researcher identify additional respondents to include in the study.

- After being interviewed, these respondents are asked to identify others who belong to the target population of interest.
- Subsequent respondents are selected based on the referrals.

Used when the defined target population is small and unique.

Non-probability Sampling

Advantages:

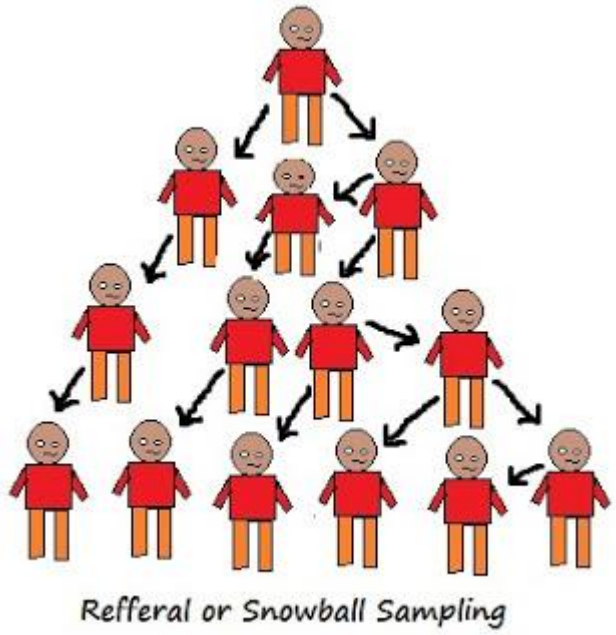
A good method to survey **hard to reach** populations.

Mostly useful in qualitative research where small samples are used.

Disadvantages:

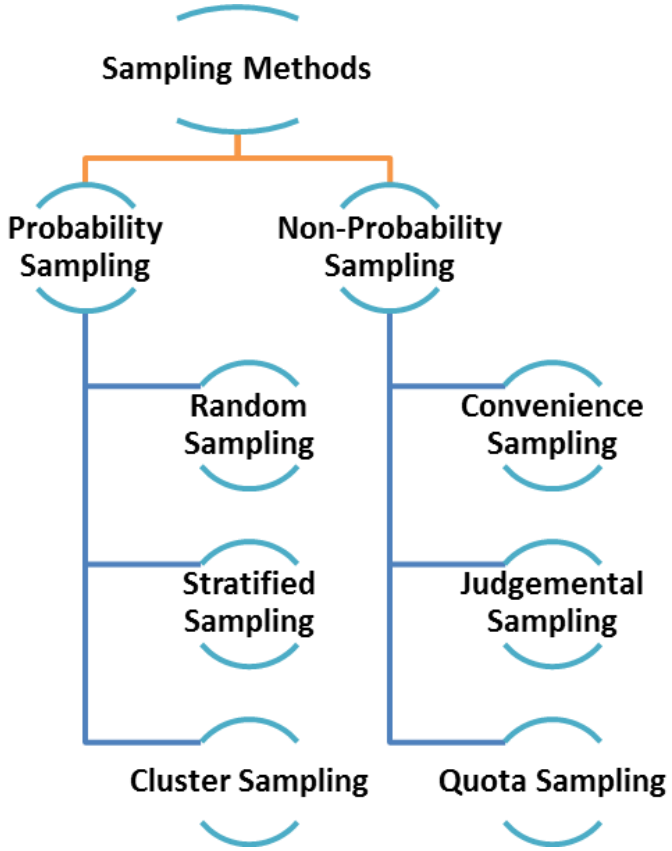
Possible bias can enter the study.

Generalizability to the target population would again be difficult (this problem occurs in **all non probability** sampling methods).



Sampling Techniques

Technique	Strengths	Weaknesses
Nonprobability Sampling		
Convenience sampling	Least expensive, least time-consuming, most convenient	Selection bias, sample not representative, not recommended for descriptive or causal research
Judgmental sampling	Low cost, convenient, not time-consuming	Does not allow generalization, subjective
Quota sampling	Sample can be controlled for certain characteristics	Selection bias, no assurance of representativeness
Snowball sampling	Can estimate rare characteristics	Time-consuming
Probability sampling		
Simple random sampling (SRS)	Easily understood, results projectable	Difficult to construct sampling frame, expensive, lower precision, no assurance of representativeness.
Systematic sampling	Can increase representativeness, easier to implement than SRS, sampling frame not necessary	Can decrease representativeness
Stratified sampling	Include all important subpopulations, precision	Difficult to select relevant stratification variables, not feasible to stratify on many variables, expensive
Cluster sampling	Easy to implement, cost effective	Imprecise, difficult to compute and interpret results



Determining Sample Size

How large should the sample be?

Sample Size

Nonprobability samples

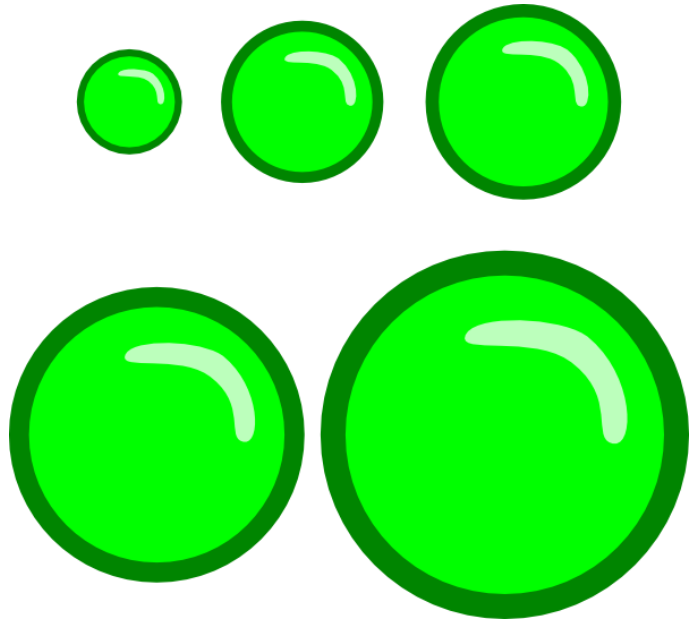
Size less important

Researchers know that they cannot generalize to population

Probability samples

Size more important

Researchers want a **big enough** sample to be confident that the results are an accurate reflection of the population



Ad Hoc Methods (non-statistical)

Budget constraints

Calculate the cost of interview and data analysis per respondent. Divide total budget by this amount to get maximum sample size.

Rule of Thumb

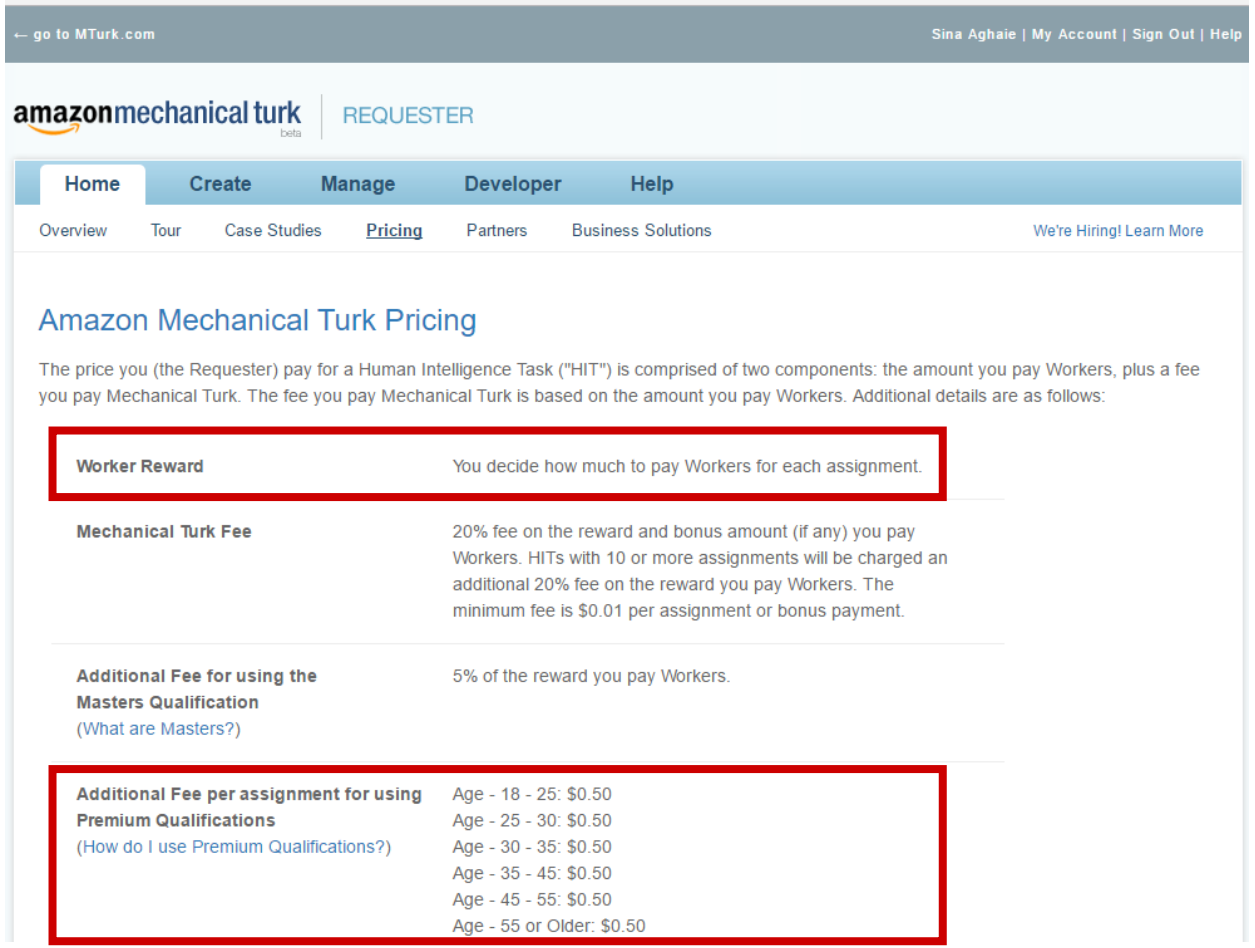
Sample should be large enough, so that when divided into groups, each group will have a minimum sample size of 100 or more.

If analysis involves comparison between subgroups, sample size in each subgroup should be 20 to 50

Comparative Studies

Find similar studies which are successful and getting sufficiently reliable results then use their sample sizes as a guide

Budget constraints



Amazon calls tasks that are difficult for computers but easy for humans HITS (human intelligence tasks).

A person completing a HIT might be asked to **write a product description**, respond to a mobile voice search query or **choose the best photograph of a single subject**

Determining Sample Size (Probability Sampling)

Sticky notes

Three factors to consider when determining the sample size in a probability design:

- **Variability** of the population characteristic under investigation (σ)
- Level of **confidence** (certainty) desired in the estimate (Z)
- Degree of **precision** (acceptable amount of error) desired in estimating the population characteristic (sampling error, e)

Qualitative Approaches

Arbitrary Approach: a heuristic: "5% of the population."

Conventional Approach: Industry Standards

Cost Approach: "All you can afford"

Determining Sample Size

Population Variance

This is a measure of the dispersion of the population.

The **greater** the variability in the data, the **larger** the sample size required.

Determining Sample Size

Degree of Precision

Precision is the acceptable amount of error in the sample estimate.

- The more precise you want your sample results to be, the smaller will be your desired error, or precision range.
- The **smaller** the precision range you set, the **larger** your sample size will be.

Let's say we want to estimate the likelihood that someone will purchase our product on 1 to 7 scale

(1=Not at all likely, 7=Extremely likely)

Is it acceptable to get a value which is within ±2 scale points of the true value?

Hence, we set the amount of error we would be willing to accept.

Determining Sample Size

Confidence Level

Confidence is the certainty that the true value of what we are estimating falls within the **precision range** we have selected.

- The **higher** the level of confidence desired, the **larger** the sample size needed.
- So if we set a range of our result being within ±1 scale points of the true value, how confident are we that our result falls within this range?
- Typically, researchers use a 90 or 95 percent confidence level.

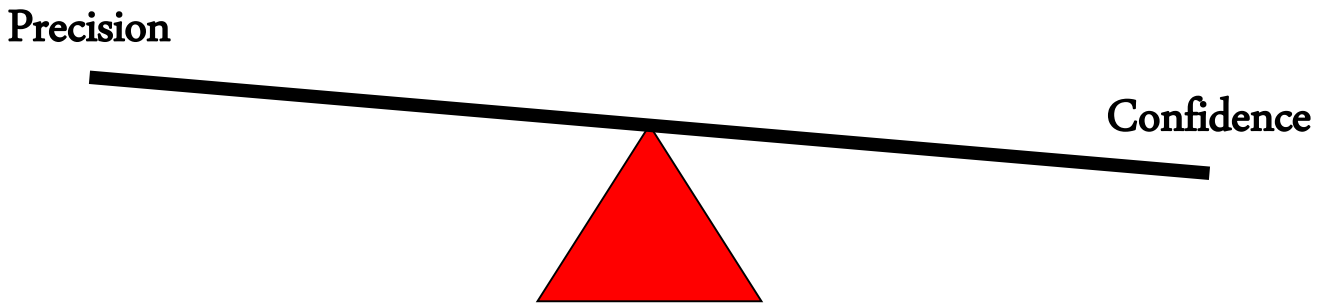
Determining Sample Size

Trade Off

You cannot always set a high degree of confidence and high degree of precision.

There is a trade off between the two. The higher your required degree of confidence, the lower your degree of precision will need to be, and vice versa.

But if both need to be high, then the sample will need to be quite large too.



Determining Sample Size

Based on these three factors, researchers can statistically compute the required sample size.

Alternatively, researchers may also use rules of thumb and their own experience to judge how large a sample size should be.

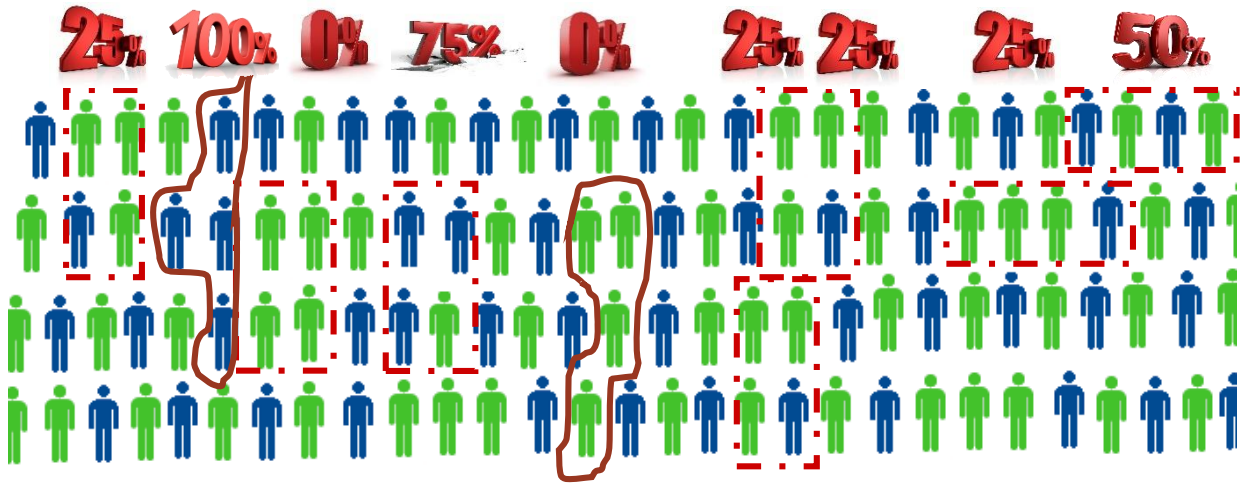
The fundamental goal is to ensure that the sample size is sufficient to yield accurate and **reliable** results.

Question …

Suppose that you are working at a cookie company as a marketing expert and a CMO asks you the following question:

- **What proportion of the consumers like/dislike the new cookie launched last month?**

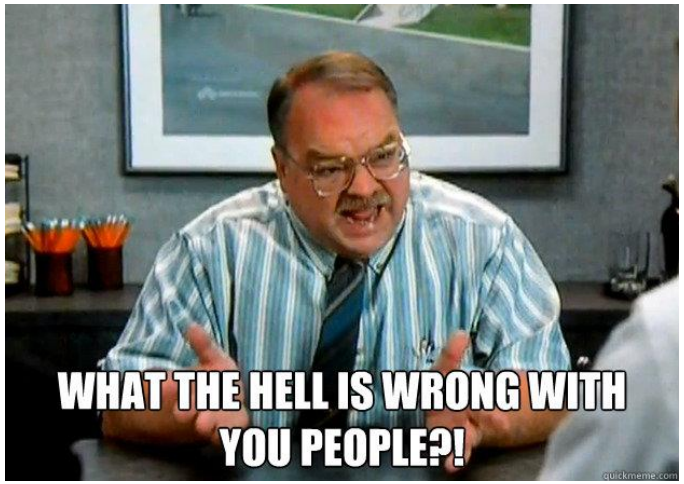
Question …



BLUE : Likes the cookie

GREEN: Dislikes the cookie

: represents 10 individuals



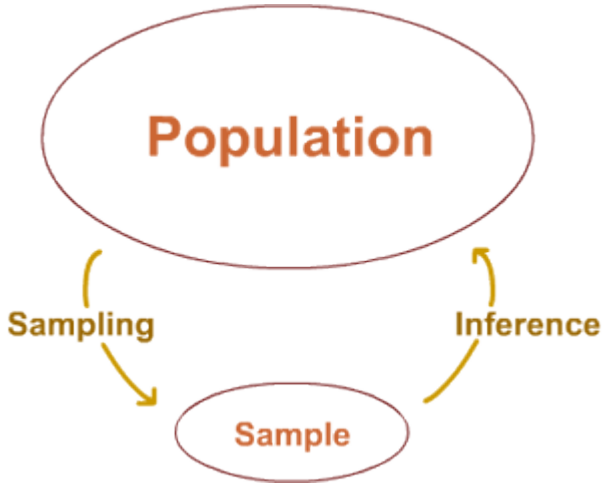
Sample vs. Population

There is always a sampling error

In this example sample means are :

25%, 100%, 0%, 75%, 0%, 25%, 25%, 25%, 50%

The average of sample means is 36%



- The sample mean ($\bar{X}$) is used to estimate the unknown population mean "μ".
- Because $\bar{X}$ varies from sample to sample, it is not equal to the population mean

Sampling error

Since there is always a **sampling error**, it is useful to provide an interval estimate around ($\bar{X}$) that reflects our judgement of the extent of this sampling error:

$$\bar{X} \pm z\sigma_{\bar{X}} = \text{interval estimate of } \mu$$

sampling error

- Z is the value associated with the confidence level.
- $\sigma_{\bar{X}}$ is standard deviation of the sample means (33%).

Interval estimate of population mean :
(assuming 95% confidence)

100 ± 1.96(33) = (34,166) = (34,100)

0 ± 1.96(33) = (-66,66) = (0,66)

50 ± 1.96(33) = (-16,116) = (0,100)

Finding the "Z"- 3 Common Scenarios:

90% Confidence Level = 1.65

95% Confidence Level = 1.96

99% Confidence Level = 2.58

Determining Sample Size

More on e

Formula used to calculate a sample size for a mean

$z\sigma_{\bar{X}} = \text{Sampling error (e)}$

$\sigma_{\bar{X}} = {}^{\sigma}/_{\sqrt{n}}$ (σ *is the population SD*)

$z \times {}^{\sigma}/_{\sqrt{n}} = e$

$$n = (Z^2) \times {}^{\sigma^2}/_{e^2}$$

Z : Standardized z-value associated with the level of confidence

σ : Estimate of the population standard deviation (σ) based on some type of prior information

e : Acceptable tolerance level of error

(is expressed in terms of the units we are estimating)

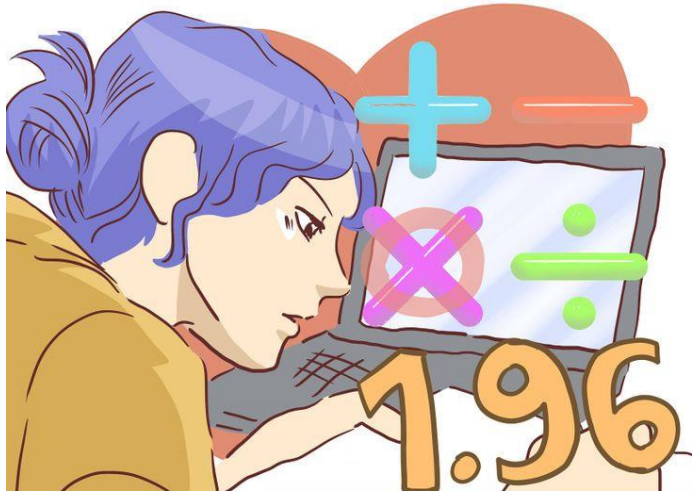
Examples:

- If we are measuring attitudes on a 1-7 scale, we may want our error to be no more than + or - .5 scale units.
- If we are estimating dollars consumers are willing to pay for a product, we may want our error to be no more than + or - $3.00.

Determining Sample Size

How to decide on the level of confidence desired

- Researchers should work with managers to make this decision. The **more confidence**, the **larger the sample size**
- The more important the decision, the more likely the manager will want more confidence.



Finding the "Z"- 3 Common Scenarios:

90% Confidence Level = 1.65

95% Confidence Level = 1.96

99% Confidence Level = 2.58

Sample Size Example

The management of a major brewery wanted to determine the average ounces of beer consumed per resident in Florida. Past trends indicated that the variation in beer consumption (SD) was 4 ounces.

- If A 95% confidence level is required, and the error is not to exceed ±0.5 ounce then what sample size would you recommend?

Management now wants a calculation twice as precise as the initial precision level and an increase in the level of confidence to 99%.

- Now what sample size do you recommend?

Answer

$$- n = \frac{(1.96)^2(4)^2}{(.5)^2} = 246$$

$$- n = \frac{(2.58)^2(4)^2}{(.25)^2} = 1704$$

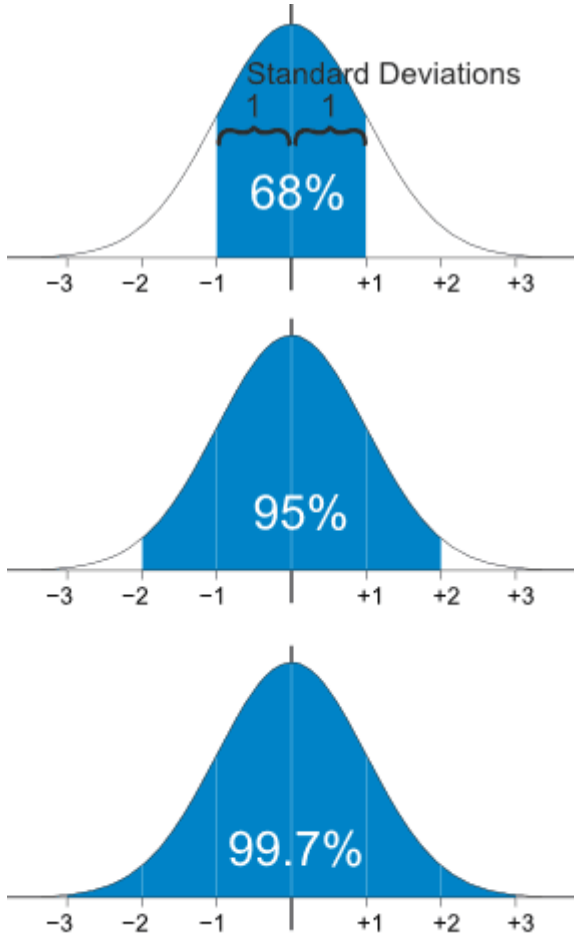
We have another trade-off: Precision and confidence for sample size.

Estimating Sample Size without "SD"

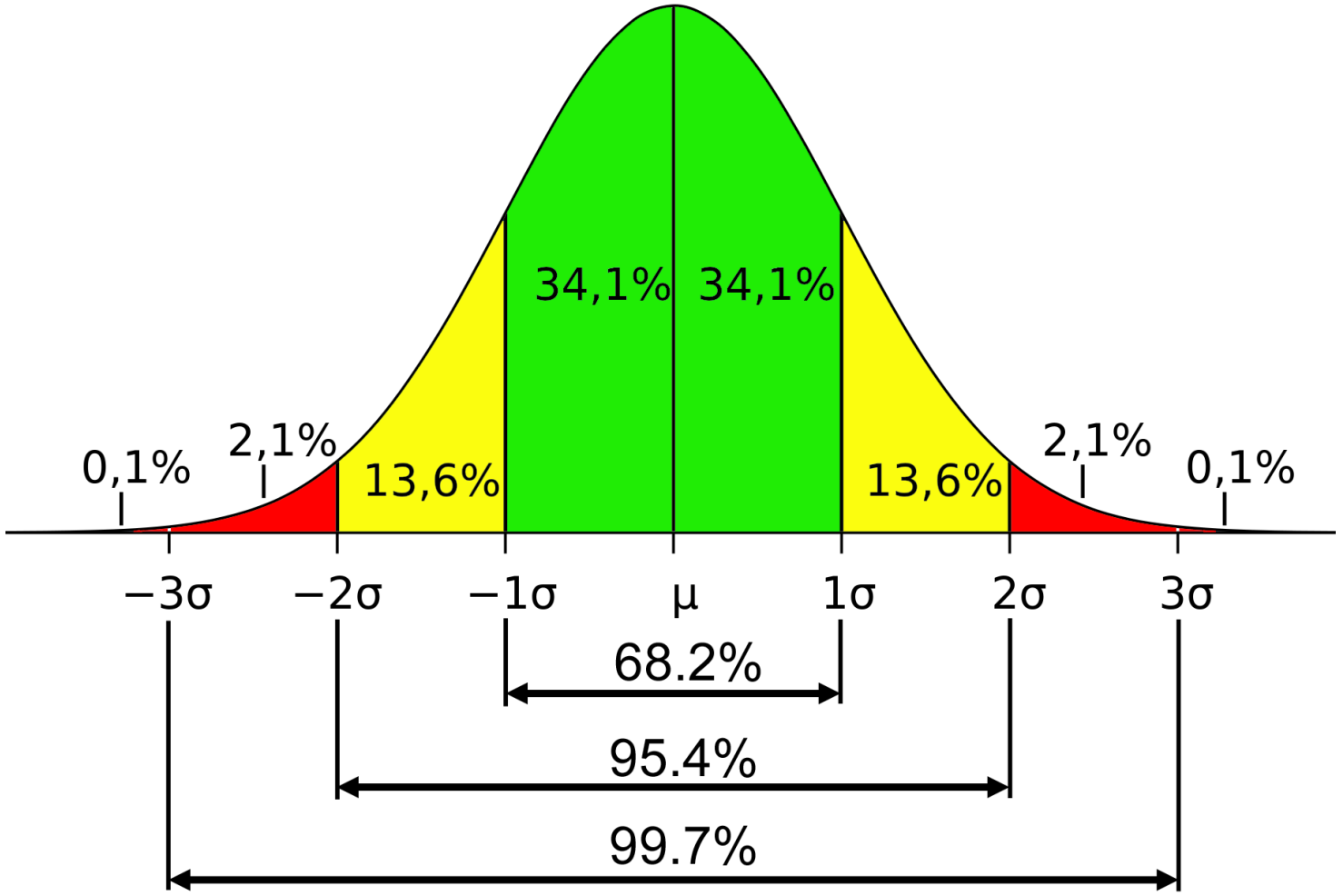
- **Range : normal variable = ±3 standard deviations**

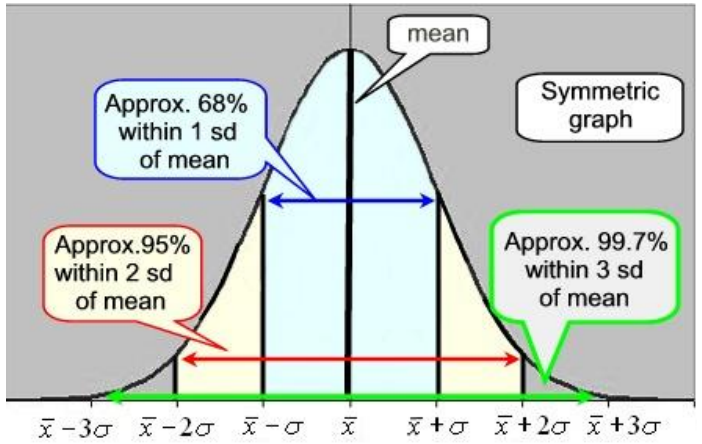
  So we can divide the range by 6

- **Estimate it with a Pilot study**
- **Use Previous Research when available**
- **Use the Number of Scale Points**



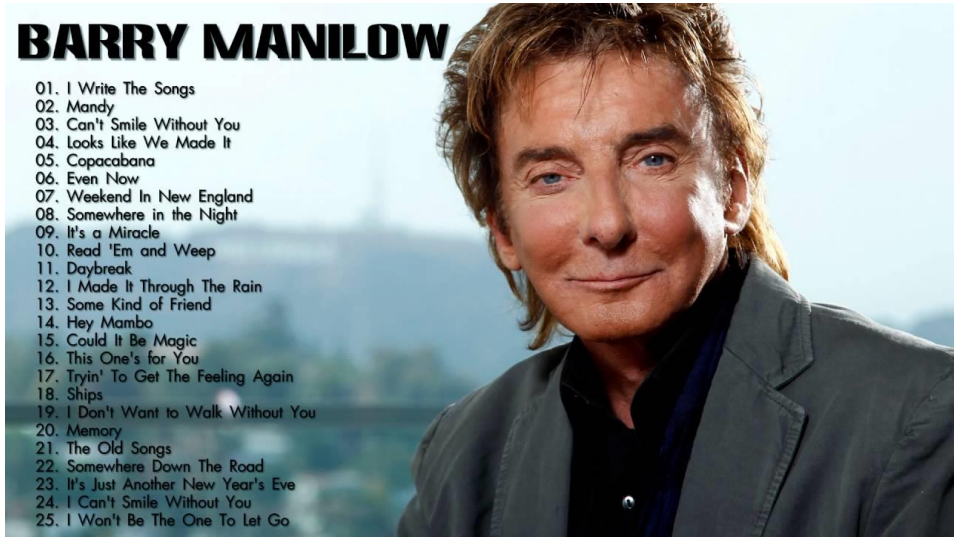
Normal Distribution





Sample Size Example

Estimate the age of Barry Manilow listeners. Want to be 95% confident in the results and do not want the error to exceed 1.5 years. It was determined that the oldest Manilow fan is 80 and the youngest is 20.



Answer

95% confidence interval

e = 1.5 years

range = 60 years

$$n = \frac{(1.96)^2(10)^2}{(1.5)^2} = 171$$