Hypothesis Testing

INFERENTIAL STATISTICS

Inferential statistics

- Inferential statistics allow us to directly test our hypothesis by evaluating our research question based on a sample, with the goal of generalizing the results to the larger population from which the sample was drawn.
- Statistical hypothesis testing is defined as assessing evidence provided by the raw data in favor of or against each hypothesis about the population.
- Probability is the underlying foundation of all statistical methods, so we must talk about it first.

Probability

- Ideally, the sample should be random so that it might better represent the entire population.
- Not all random samples are ideal, no random sample will be the exactly the same as any other.
- We cannot know how "ideal" a random sample from a population is. This uncertainty means that we must quantify how much we expect a random sample to vary.
- This allows us to draw conclusions about the population in the face of uncertainty when we use random samples.

Probability

- We are interested in estimating the percentage of US adults who favour the death penalty.
- We choose a random sample of 1,200 US adults and ask their opinion.
- 744 out of 1200, or 62% are in favour.
- Our goal is to infer/draw conclusions about the opinions of US adults regarding the death penalty.
- Based on 1200 US adults can we absolutely conclude that 62% of the population favours the death penalty?

Probability

- Another random sample could give a very different result, so we are uncertain.
- Since our sample is random we know the uncertainty is due to chance. Not due to sampling method.
- Therefore we can use probability to describe the likelihood that our sample is within the desired level of accuracy.
- How likely is it that our conclusion is within 3% of the ACTUAL percentage of ALL US adults who are in favour of the death penalty?

Probability

- The answer to that question has an important impact on the confidence we can attach to the inference step.
- Probability is the likelihood of something occurring; of an event occurring; the chance of something happening.
- It is a mathematical description of randomness and uncertainty.
- It is a way to measure or quantify uncertainty.
- It is the official name for chance.

Probability

- Gambling industry:
  - Cutting a deck of cards:
    - what are the chances that it going to be red?
    - what are the chances of it being a spade?
    - what are the chances of it being a King?
    - what are the chances of it being a King or a Club?
    - what are the chances of it being a King and being a Club?

Probability

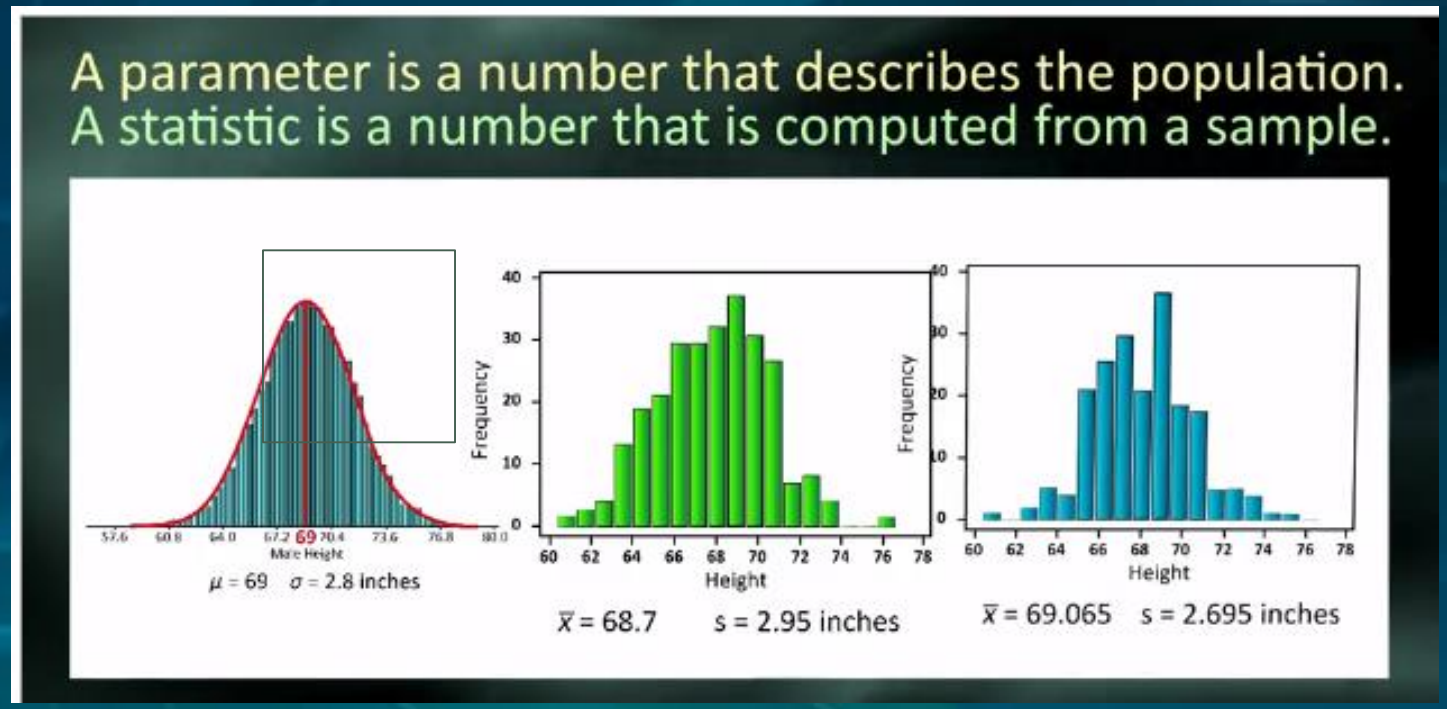
- The probability of an event ranges from 0 to 1.
- A probability of 0 means that the event has zero chances of happening. It will never occur.
- A probability of 1 means that the event will occur for certain.
- A probability of .5 indicates that the event has a 50% chance of happening. The event is as likely to occur as not to occur.

Probability

- Any probability that is greater than .5 indicates that the event is more likely to occur than it is not to occur.
- Any probability that is less than .5 indicates that the event is more likely not to occur than it is to occur.
- Expressing probability in percentages is very common.
- The chance that an event will occur is between 0% and 100%.

Central Limit Theorem

- The parameter is 69 inches.
- The statistic for sample 1 is 68.7 inches.
- The statistic for sample 2 is 69.25 inches

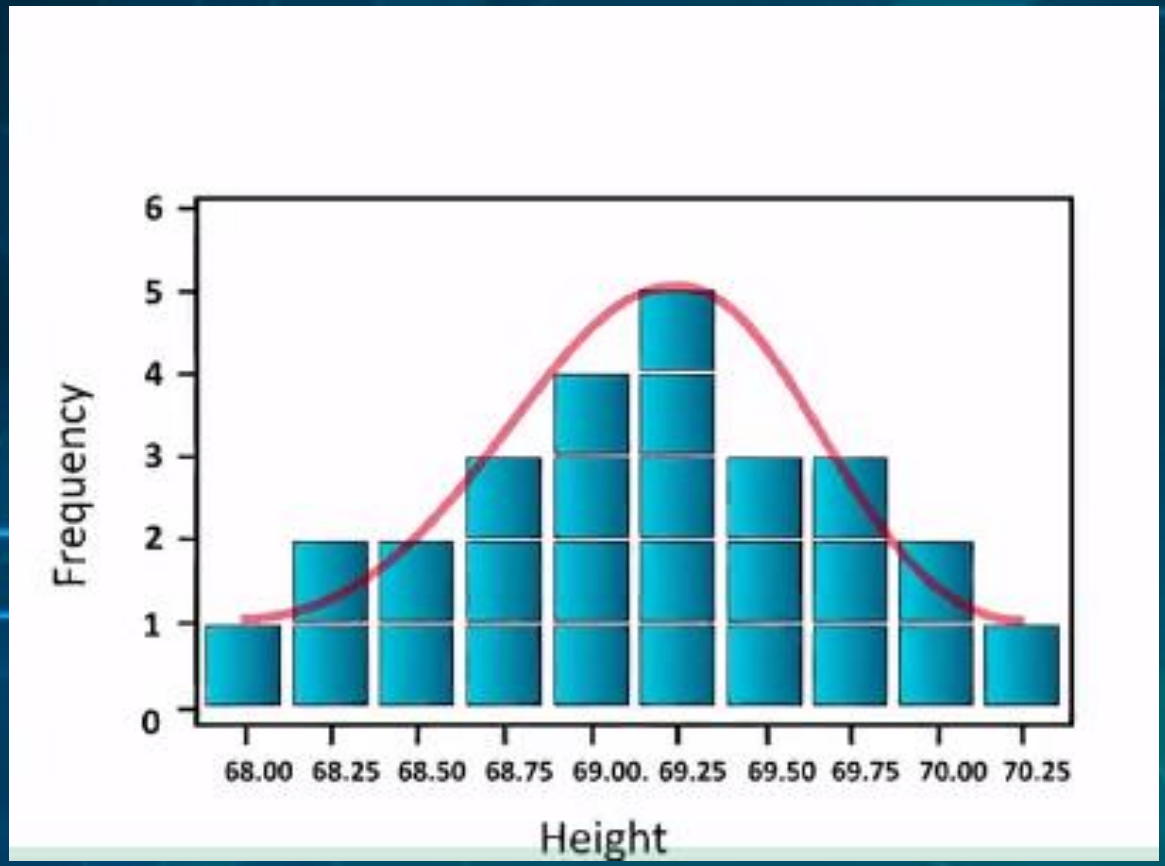


Parameters are typically unknown.

Central Limit Theorem

- Sampling variability
  - As we have already stated each random sample is different. The more random samples we use then the greater the picture of the population.
  - We calculate statistics from each sample, such as mean, std dev etc
  - Suppose we know the actual mean of the total male population height to be 69 inches. We select 30 random samples from the population, each sample size being 500.
  - Each sample would have slightly different mean, 68.5, 69.25 etc.
  - If we plotted these together we would have a normal distribution.

Central Limit Theorem



Central Limit Theorem

- As long as adequately large samples and an adequately large number of samples are drawn from a population, the distribution of the statistics of the samples, whether of mean, proportion, standard deviation, or any other statistic, will be normally distributed.

Hypothesis Testing

- This method is used when we want to make decisions concerning populations on the basis of only sample information.
- Analysis of variance test, ANOVA, Chi-Square test of independence.
- Same steps:
  - Specifying null hypothesis H_0
  - specifying the alternate hypothesis H_a
  - Choose sample, assess evidence, draw conclusion.

Hypothesis Testing

- NESARC data set
- Representative sample of 43,093 adults in U.S
- Evaluate if there is an association between a diagnosis of major depression and how much a person smokes.
- Step 1: Hypotheses
- H_0 There is no difference in smoking quantity between people with and without depression.
- H_a There is a difference in smoking quantity between people with and without depression.

Hypothesis Testing

- The null hypothesis says that nothing special is going on between the two variables, they are unrelated.
- The alternative says there is a relationship, it could be positive or negative.
- Step 2 Choose a sample:
- We are going to use the nesarc data set. We are going to evaluate the hypothesis among individuals who are smokers and who are younger (aged between 18 and 25).
- The sample size is 1320

Hypothesis testing

- Assess evidence – draw conclusion
- The p-value for a test is an estimate of how often we would get the obtained result by chance if in fact the null hypothesis is true.
- The mean number of cigarettes smoked is 13.9 std dev 9.2 for those with depression.
- the mean number of cigarettes smoked is 13.2 std dev 8.5 for those without depression.
- The probability of getting a difference of .7 cigarettes smoked per day in a random sample of 1320 is roughly 17%
- This means if we took 100 different random samples from the population and we rejected the null hypothesis each time, we would be wrong 17 times. (there would be 17 times the results would show a relationship but in fact there is none).
- Is 17% small enough to reject the null hypothesis?