

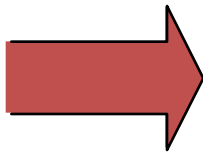


Hypothesis Test

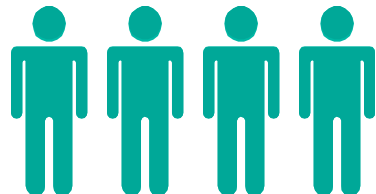
Hypothesis Testing Process in a common man's language



Assume the population mean age is 40.
(Null Hypothesis)

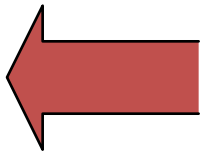
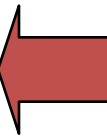


Population



Sample

The Sample Mean is 18



No, not likely!
REJECT
Null Hypothesis

Let us take an example with little statistics..



We make chocolates containing 0.04 coco units on an average with a standard deviation of 0.004

A chocolate manufacturing Company (CMC)





**CMC makes 10K
chocolates and
passes to a bakery**

Wont accept chocolates if it
contains too much or too less
coco, than claimed by CMC.

Bakery

Is it possible for the bakery to examine all the 10 K chocolates..?



A sample of 100 chocolates being checked



Bakery

Average = 0.0408 coco units

Can we accept the batch of 10000 chocolates



Your Answer ..?

Interpreting the scenario



- To determine whether the population mean is actually 0.04
 - i.e. If the true mean is 0.04 and standard deviation is 0.004, what are the chances of getting a sample mean that differs from 0.04 by 0.008 ($0.008 = 0.04 - 0.408$) or more
 - i.e. check whether the probability (of selecting the random sample with a mean of 0.0408 from the population with a mean of 0.04 and SD of 0.004) is far low or not.



- What do you mean if probability is
 - too low?
 - too high?
- For Answers refer Slide No.11

Solution



- Standard Error of Mean $\sigma_x = \sigma/\sqrt{N}$
 $-0.004/\sqrt{100}=0.0004$
- Determine Z value $z = (x_{\text{bar}} - \mu)/\sigma_x$
 $-(0.0408 - 0.04)/0.0004 = 2$

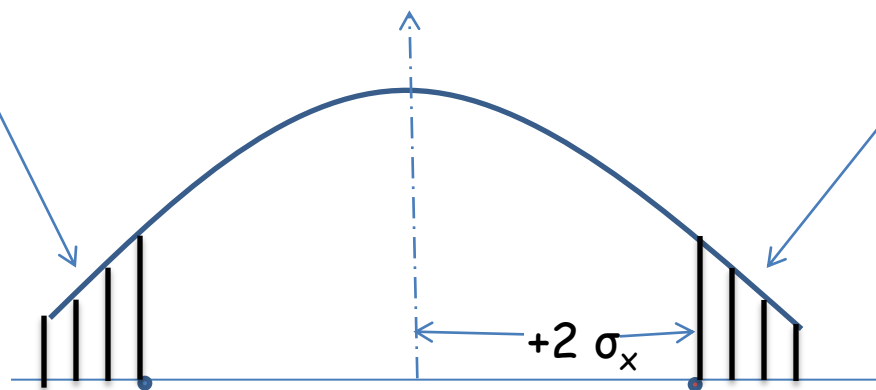


- Determine the chance that the sample mean would be 0.0408 or larger or 0.0392 or smaller
 - Refer Z look up table or use Minitab.
 - From look up table
 - $P(Z \geq 2 \text{ or } Z \leq -2) = 2 * (.5 - 0.4777) = 4.5 \%$
 - Probability is very low

Probability is very low..??
What does it mean..??



2.25 % area
under curve



2.25 % area
under curve

0.0392

0.04

0.0408

Hypothesized population mean

Sample mean



Probability is very low..??
What does it mean..??

- Probability to get a such a sample mean from the given population mean is too low.
 - It means the anticipated population mean is wrong !!!!!



We don't want the
chocolates from CMC

Bakery



In this case the difference between the sample mean and the hypothesized population mean is too large and the chance that the population would produce such a random sample is far too low. Hence the bakery reject the batch of chocolates from CMC

Why we treat 4.5 % as low..?



- It is a judgment for decision makers to make
 - Sometimes even 1 % may be too high
- In each decision determine the **costs** resulting from an incorrect decision and the **precise** level of risk, we are willing to accept.



- Consider the previous example of CMC
 - The bakery rejected the CMC contention that population mean is 0.04
 - Suppose the population mean was actually 0.04
 - If then stick to the rejection rule of 2 standard errors or more (4.5 %), the bakery had rejected a good batch of chocolates.
 - Hence the minimum standard for an acceptable probability (4.5%) is also the **risk** we take of rejecting a judgment that is true.
- In this or in any case there can be no risk free trade-off

Some Terms



- Hypothesis
 - Null Hypothesis
 - Alternate Hypothesis
- Significance level
- Type 1 error
- Type 11 error

What is a Hypothesis?



- A hypothesis is an assumption about the population parameter.
 - A parameter is a characteristic of the population, like its mean or variance.
 - The parameter must be identified before analysis.

I assume the mean GPA of this class is 3.5!



• Hypothesis testing is about making inferences about population from a sample

© 1984-1994 T/Maker Co.

Null Hypothesis



- The assumption we wish to test
 - H_0
 - *CMC* example → population mean is actually 0.04 and it is nearer to 0.0408

Null Hypothesis



- The term Null
 - Arises from earlier agricultural and medical applications of statistic.
 - In order to test the effectiveness of a new drug/fertilizer, the tested hypothesis (null hypothesis) was that it had no effect
 - i.e. there was no difference between treated and untreated samples.
- The choices are: =
 - No change/relationship/difference (as a result of the study).

Alternate Hypothesis



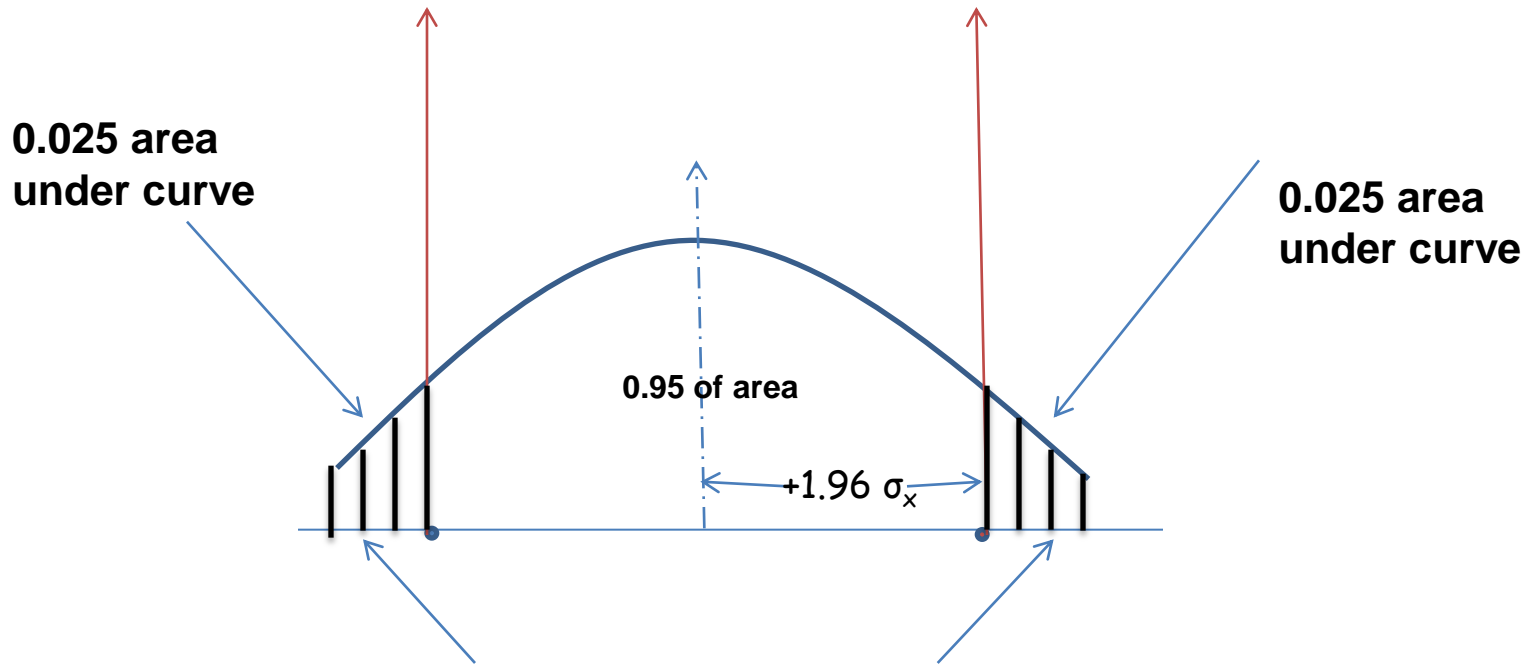
- If our sample test results fail to support null hypothesis, we must conclude that something else is true
 - Whenever we reject the null hypothesis, the conclusion we accept is known as alternate hypothesis. (H_1)
- The choices are: $<$, $>$, or \neq .
- Example
 - $H_0: \mu = 200$ (Population mean = 200)
 - $H_1: \mu \neq 200$
 - $H_1: \mu > 200$
 - $H_1: \mu < 200$

Significance level



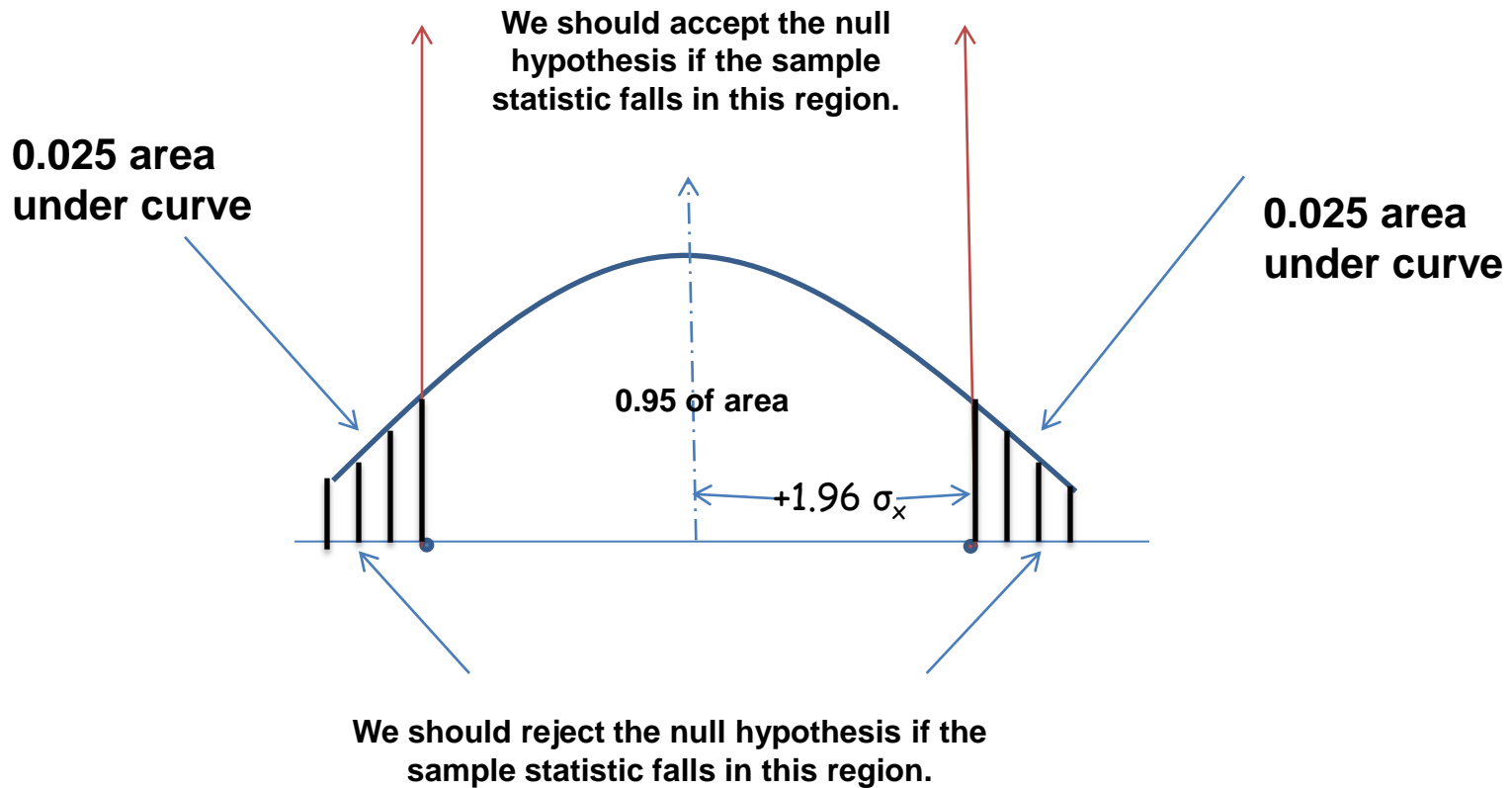
- Refer Previous CMC Example
- Rejected the chocolate batch after observing 100 samples when the probability was 4.5 %
 - i.e. the criterion to decide whether to accept or reject null hypothesis.
 - Known as significance level.

How to interpret a 5 % level of Significance



In these regions there is a significant difference between the sample statistic and hypothesized population parameter

Even if the sample statistic falls in this region, this does not prove that our null hypothesis, it does not provide statistical evidence to reject it





- Assume the hypothesis is correct, then what does the significance level indicate?
 - The % sample means that is outside certain limits.
 - The higher the significance level, the higher the probability of rejecting a H_0 , when it is true.

Errors in Making Decisions



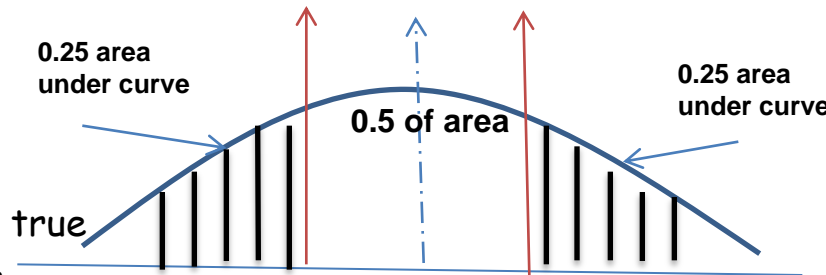
- Type 1 \rightarrow reject a null hypo. when true
- Type 11 \rightarrow accept a null hypo. When false
- There is a tradeoff between these two errors

Type 1 and 11 Errors

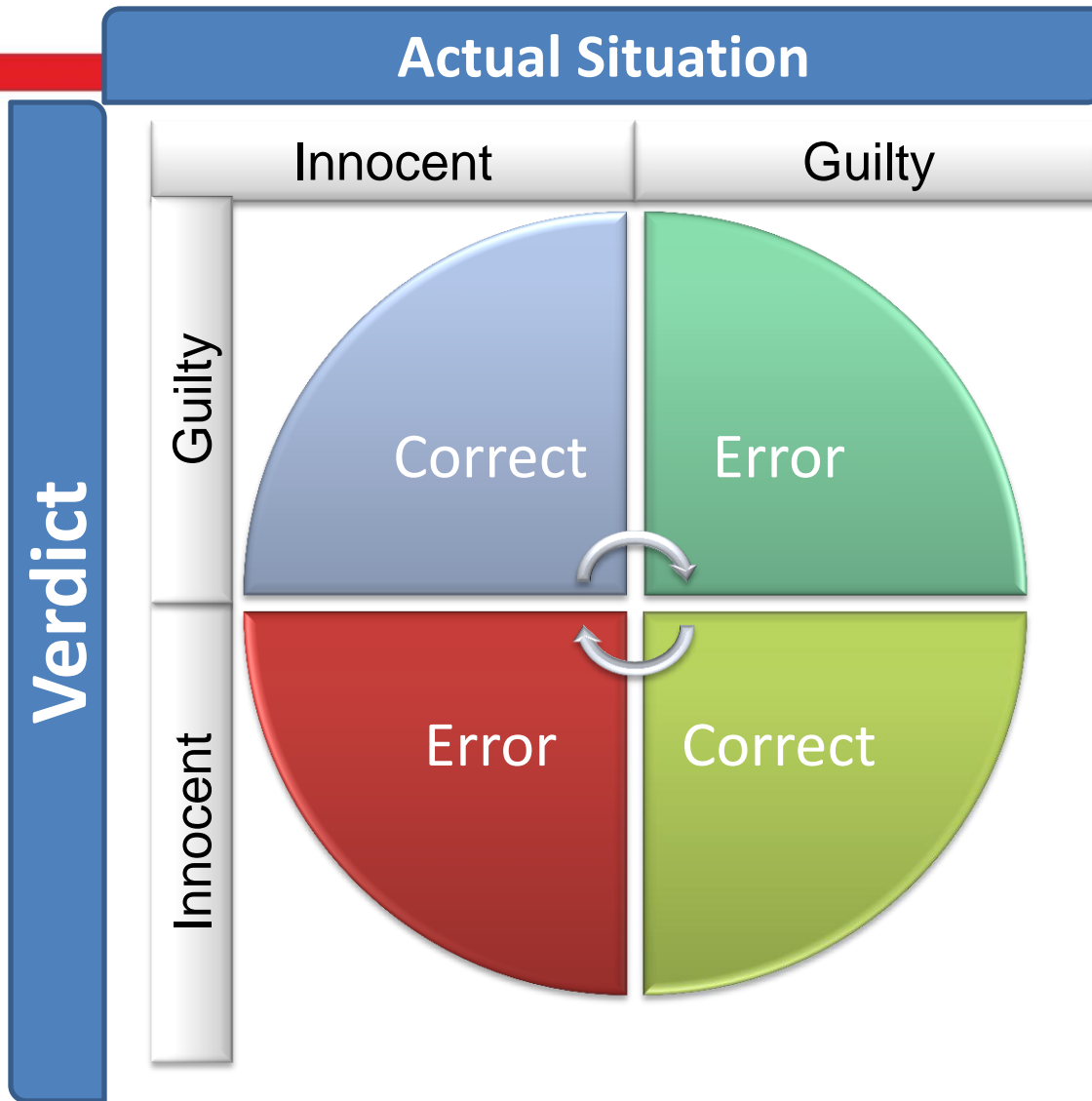


Example

- Acceptance region is 50 % of area
 - We will rarely accept a null hypo when it is not true
 - We will often reject a null hypo when it is true
- In order to get a low type 1 error we have to put large type 2 error
- Decision taken based on cost involved.



Type 1 and 11 Errors: example using a judiciary system.



H_0 : Innocent



Preference for a Type 1 Error

- Validating sample drugs from a batch
- Examining aircraft parts from a batch

Preference for a Type 11 Error

- Examining Chappals from a batch

Summary



Descriptive

used to organize and describe a sample

Statistics

Inferential

used to extrapolate from a sample to a
larger population

Summary

Whether the population is likely to produce a sample like this..?



- **Hypothesis testing:** Make inferences about a population from only a sample.
 - i. Starts with an assumption called hypothesis that we makes about a population parameter
 - ii. Collect sample data, make sample statistics
 - iii. Use sample statistics to decide how likely it is that our hypothesized population parameter is correct.
 - iv. P-values provide a sense of the strength of the evidence against the null hypothesis.



Thank you