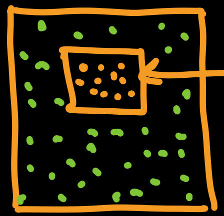
statistics

descriptive
- → central tendancy
- → variations
- → asymmetry
- → correlation

inferential
- → hypothesis testing



population

sample → study → Experiment

descriptive stats

↓

draw conclusion

generalize the conclusion

inferential stats

Hypotheses Testing

assumptions

- → accept
- → Reject

1] make the assumption (hypotheses) – Null + alternate

2] Decide the target population

3] Get a sample from population → Random / → Non-random

4] Study the sample (using descriptive stats)

5] Draw the conclusion after performing experiment

6] Decide hypothesis testing strategy → Parametric / → Non parametric

7] Calculate critical value and t-value and compare them

8] Using one-tailed or two-tailed test, accept / reject the hypothesis

Introduction

[0 ... 1]

- Statistical inference is that branch of statistics which is concerned with using probability concept to deal with uncertainty in decision-making → conclusion
- It refers to the process of selecting and using a sample statistic to draw inference about a population parameter based on the set of attributes
- It treats two classes of problems
  - Hypotheses testing i.e. to test some hypothesis about parent population from which sample is drawn
  - Estimation i.e. to use the statistics obtained from the sample as estimate of the unknown parameter of the population from which ~~population~~ sample is drawn

* population parameter → attribute of a population
* sample statistic → attribute of sample

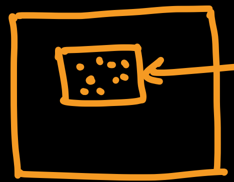
Hypothesis Testing

- It begins with assumption called as hypothesis, that we make about the population parameter
- A hypothesis is supposition made as a basis for reasoning
- A hypothesis in statistics is simply a quantitative statement about a population
- There can be several types of hypotheses
- E.g.
  - A coin may be tossed 200 times and we may get heads 80 times and tails 120 times. We may now be interested in testing the hypothesis that the coin is unbiased. → Qualitative → textual
  - We may study the average weight of 100 students of a college and get result as 110 lb. We may now be interested in testing the hypothesis that the sample has been drawn from a population with average weight of 115 lb.

Quantitative statement

$p(\text{Head}) = 0.5$ or

$p(\text{tail}) = 0.5$



sample of 100 students ⇒ study → $\bar{x} = 110$ lb

population → μ = mean of population = 115 lb

$\mu == \bar{x}$

Procedure of Testing Hypothesis

Procedure of Testing Hypothesis

- Set up a hypothesis → null hypothesis and alternative hypothesis
- Set up a suitable significance level → (α) → 0.01 or 0.05 or 0.10
- Setting a test criterion → set the test strategy → student T, chi-square
- Doing computations → Experiment using one/two tailed test
- Making Decisions → accept the hypothesis or reject the hypothesi

Set up a hypothesis → frame the hypotheses

- The first thing in hypothesis testing is to setup a hypothesis about the population parameter → assumption
- Then we collect sample data, produce sample statistics
- Use this information to decide how likely it is that out hypothesized population parameter is correct
- E.g.
  - We assume certain value of a population mean ($\mu = 115$ lb)
  - To test the validity of the assumption, we gather the sample data (100 students, $\bar{x} = 110$)
  - Then we determine the difference between hypothesize mean (population mean) and actual value of sample mean
  - Then we judge whether the difference is significant → test criteria & α
  - The smaller the difference, greater the likelihood that our hypothesized mean is correct or greater the difference, smaller the likelihood

Hypotheses Types

- The conventional approach to hypothesis testing is not to construct a single hypothesis about the population parameter, but rather to set up two hypotheses
- These hypotheses must be so constructed that if one hypothesis is accepted, the other is rejected by default and vice-a-versa
- The two hypotheses are normally referred as
  - Null hypothesis (H_0)
  - Alternative hypothesis (H_1)

Null hypothesis & alternative hypothesis are opposite of each other

Null Hypothesis = true statement = H_0

- It is a very useful tool in testing the significance of difference
- In its simplest form, the hypothesis asserts that there is no real difference in the sample and the population in under consideration
- Hence the word null, which means invalid, void or amounting to nothing that the difference found is accidental and unimportant arising out of fluctuations of sampling
- It is skin to the legal principle that a man is innocent until he is proved guilty
- It constitutes a challenge, and the function of the experiment is to give the facts to a chance to refute or failed to refute this challenge
- The rejection of null hypothesis indicates that the differences have statistical significance
- Acceptance of null hypothesis indicates that the differences are due to chance

$H_0 = p(\text{Head}) = 0.5$, $H_0 = \mu = 115$

Alternate Hypothesis

H_1 or H_a

- It specifies these values that the researcher believes to hold true and of course, he hopes that the sample data lead to acceptance of this hypothesis true
- It may embrace the whole range of values rather than a single value
- E.g.
  - A certain psychologist who wishes to test whether or not a certain class of people have mean IQ higher than 100 might establish following hypotheses
  - Null hypothesis
    - $H_0: \mu = 100$

$p(\text{Head}) = 0.5$, $\mu = 115$

  - Alternate hypothesis
    - $H_a: \mu \neq 100$

$p(\text{Head}) \neq 0.5$, $\mu \neq 115$

Set up significance level (α) → set by researcher

- Having setup the hypotheses, the next step is to test the validity of null hypothesis against that of alternative one at a certain level of significance
- The confidence with which an experimenter rejects or remains a null hypothesis depends upon the significance level accepted
- It is expressed as percentage which is probability of rejecting the null hypothesis if it is true
- E.g. when a null hypothesis is accepted at the 5% level, the statistician is running the risk that in long run, he will be making the wrong decision about 5% of the time

level of confidence = 90% = 0.9

level of significance = 10% = 0.1 = 1 - level of confidence (α)

α:
- 0.1 = 10%
- 0.05 = 5%
- 0.01 = 1%

margin of error

Setting up test criterion → set by researcher

test

- The third step in hypothesis testing is to construct the test criterion → parametric testing
- It involves selecting an appropriate probability distribution for the particular test ⇒ Normal dist.
- That is the probability distribution which can properly be applied for testing
- Some probability distributions which are commonly used
  - T distribution → student -T - test
  - F distribution
  - Chi Square distribution
- E.g. if only small sample information is available, the use of normal distribution would be inappropriate

Doing computations

- This step involves performing the calculations from the random sample n which are necessary for the selected test
- These calculations include testing statistic and standard error of testing statistic

Making Decisions

- Finally with respect to the test calculations, we may draw the statistical conclusions and take decisions
- A statistical conclusion or decision is a decision either to reject or to accept the null hypothesis
- The decision will be dependent on whether the computed value of the test criterion falls in the accepted region or rejection region

Testing Error = Type I + Type II

- When a statistical hypothesis is tested, there are four possibilities
  - The hypothesis is true, but our test rejects it (Type I error)
  - The hypothesis is false, but our test accepts it (Type II error)
  - The hypothesis is true, and our test accepts it (correct decision)
  - The hypothesis is false, and our test rejects it (correct decision)
- While testing the hypothesis, the aim is to reduce both the types of error

$$\alpha + \beta = 1$$

$$\beta = 1 - \alpha$$

$$\alpha = 1 - \beta$$

	accept	reject
true Null hypothesis	Correct	Type I Error (α)
false Null hypothesis	Type II Error (β)	Correct

	accept	reject
lady is pregnant	correct	Error α
lady is Not pregnant	Error β	correct

Confusion Matrix

Type I Error

- In hypothesis testing, the Type I error is committed by rejecting the null hypothesis when it is true
- The probability of committing a Type I error is denoted by α where
  - α = probability (Type I error)
  - α = probability (rejecting H_0 | H_0 is true)

Type II Error

- The Type II error is committed by not rejecting (i.e. accepting) the null hypothesis when it is false
- The probability of committing a Type II error is denoted by β where
  - β= probability (Type I error)
  - β= probability (not rejecting H_0 | H_0 is false)

Trade-off between errors

- The probability of making one type of error can only be reduced if we are willing to increase the probability of making the other type of error
- In other words, to get low β, we will have to put up with a high α
- It is more dangerous to accept a false hypothesis (Type II error) than to reject a correct one
- Hence, we keep the probability of Type I error at a certain level called as level of significance
- The level of significance also known as size of the rejection region or size of critical region or simply the size of the test is traditionally denoted by α
- In most cases, the level of significance is generally fixed as 5% which means that the probability of accepting true hypothesis is 95%

Hypothesis Tests

* set the test criterion → student-T test → uses T-distribution table

Set the test criterion:

- ↳ formula to calculate (p-value)
- ↳ get critical value using distribution

Student-T test:

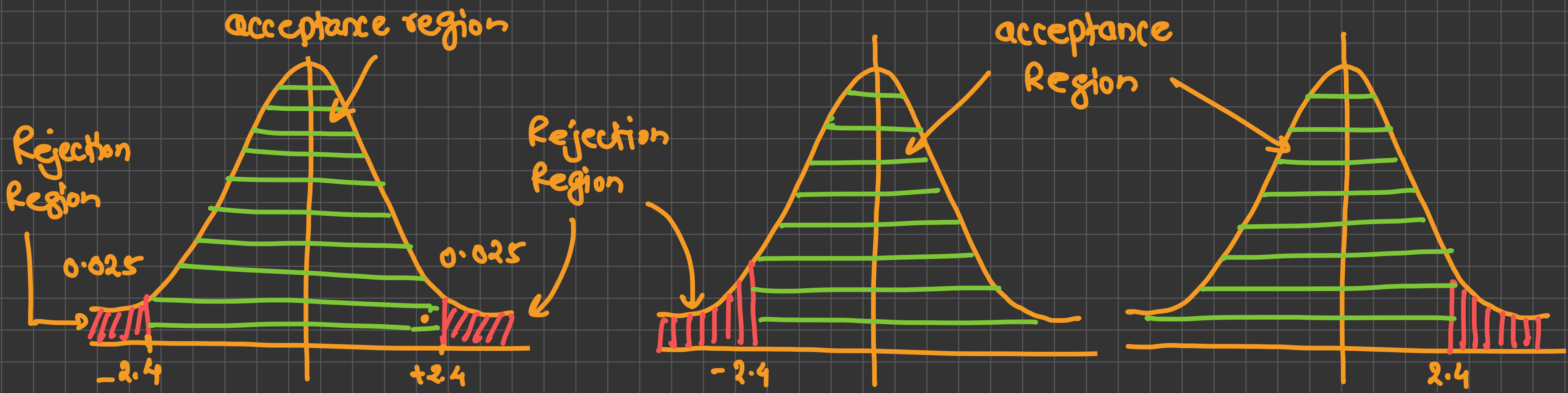
- ↳ has its own formula to calculate t-value

  $t = 2.3$

- ↳ calculate critical value using T-distribution

  2.4

$\alpha = 0.05 = 5\%$



Two-tailed test | one-tailed test | one-tailed test

Two-tailed tests of hypothesis

- A two tailed hypothesis test will reject the null hypothesis, if the sample statistic is significantly higher or lower than the hypothesized population parameter
- Thus in the two tailed test, the rejection region is located on both the sides
- If we want to reduce the risk of committing type I error, we have to reduce the size of rejection region
- For this the hypothesis may be tested at lower level of significance → 0.05 / 0.01 / 0.10
- As we decrease the size of rejection region, we increase the probability of accepting hypothesis
- Two-tailed test is appropriate when the alternative hypothesis has equal-to or not-equal-to sign
  - $H_a: \mu \neq 100$

One-tailed test

- It is so called because the rejection region will be located only on one side which may be either right or left depending on the test and hypothesis formed
- One-tailed test is used when alternate hypothesis has >, >=, < or <= sign
  - $H_a: \mu > 100$
  - $H_a: \mu < 100$

Student T Test

Introduction

- A t-test compares the average values of two data sets and determines if they came from the same population
- Mathematically, the t-test takes a sample from each of the two sets and establishes the problem statement
- It assumes a null hypothesis that the two means are equal
- Using the formulas, values are calculated and compared against the standard values
- The assumed null hypothesis is accepted or rejected accordingly
- If the null hypothesis qualifies to be rejected, it indicates that data readings are strong and are probably not due to chance

Assumptions

- The first assumption is concerned with the scale of measurement. Here assumption for a t-test is that the scale of measurement applied to the data collected follows a continuous or ordinal scale.
- The second assumption is regarding simple random sample. The Assumption is that the data is collected from a representative, randomly selected portion of the total population.
- The third assumption is the data, when plotted, results in a normal distribution, bell-shaped distribution curve.
- The fourth assumption is a that reasonably large sample size is used for the test. Larger sample size means the distribution of results should approach a normal bell-shaped curve.
- The final assumption is the homogeneity of variance. Homogeneous, or equal, variance exists when the standard deviations of samples are approximately equal.

T-Test Formula

- Calculating a t-test requires three fundamental data values
  - Difference between the mean values from each data set, or the mean difference
  - Standard deviation of each group
  - Number of data values of each group
- This comparison helps to determine the effect of chance on the difference, and whether the difference is outside that chance range
- The t-test questions whether the difference between the groups represents a true difference in the study or merely a random difference
- The t-test produces two values as its output:
  - T-value or T-Score
  - Degrees of freedom

T-Value or T-Score

- The t-value, or t-score, is a ratio of the difference between the mean of the two sample sets and the variation that exists within the sample sets
- The numerator value is the difference between the mean of the two sample sets
- The denominator is the variation that exists within the sample sets and is a measurement of the dispersion or variability
- This calculated t-value is then compared against a value obtained from a critical value table called the T-distribution table
- Higher values of the t-score indicate that a large difference exists between the two sample sets
- The smaller the t-value, the more similarity exists between the two sample sets

Degrees of Freedom

- Degrees of freedom refer to the values in a study that has the freedom to vary and are essential for assessing the importance and the validity of the null hypothesis
- Computation of these values usually depends upon the number of data records available in the sample set

Paired Sample T-Test

- The correlated t-test, or paired t-test, is a dependent type of test and is performed when the samples consist of matched pairs of similar units, or when there are cases of repeated measures
- This method also applies to cases where the samples are related or have matching characteristics, like a comparative analysis involving children, parents, or siblings

$$T = \frac{mean1 - mean2}{\frac{s(diff)}{\sqrt{n}}}$$

- Where
  - mean1 and mean2 = The average values of each of the sample sets
  - s(diff) = The standard deviation of the differences of the paired data values
  - n = The sample size (the number of paired differences)
  - Degrees of freedom = n -1

Equal Variance or Pooled T-Test

- The equal variance t-test is an independent t-test and is used when the number of samples in each group is the same, or the variance of the two data sets is similar

$$T = \frac{mean1 - mean2}{\frac{(n1-1)*var1^2 + (n2-1)var2^2}{n1+n2-2} * \sqrt{\frac{1}{n1} + \frac{1}{n2}}}$$

- Where
  - mean1 and mean2 = Average values of each of the sample sets
  - var1 and var2 = Variance of each of the sample sets
  - n1 and n2 = Number of records in each sample set
  - Degrees of Freedom: n1 + n2 - 2

Unequal Variance T-Test

- The unequal variance t-test is an independent t-test and is used when the number of samples in each group is different, and the variance of the two data sets is also different
- This test is also called Welch's t-test

$$T = \frac{mean1 - mean2}{\sqrt{\frac{var1}{n1} + \frac{var2}{n2}}}$$

- Where
  - mean1 and mean2 = Average values of each of the sample sets
  - var1 and var2 = Variance of each of the sample sets
  - n1 and n2 = Number of records in each sample set
- Degrees of Freedom

$$DoF = \frac{\left(\frac{var1^2}{n1} + \frac{var2^2}{n2}\right)^2}{\frac{\left(\frac{var1^2}{n1}\right)^2}{n1 - 1} + \frac{\left(\frac{var2^2}{n2}\right)^2}{n2 - 1}}$$

Which T-Test to use ?

- If two sample sets are same or related => Paired T-Test
- If two sample sets are of same size => Equal Variance T-Test
- If two sample sets have same variance => Equal Variance T-Test
- If two sample sets do not have same variance => Unequal Variance T-Test

Example

- S1 = 19.7, 20.4, 19.6, 17.8, 18.5, 18.9, 18.3, 18.9, 19.5, 21.95
- S2 = 28.3, 26.7, 20.1, 23.3, 25.2, 22.1, 17.7, 27.6, 20.6, 13.7, 23.2, 17.5, 20.6, 18, 23.9, 21.6, 24.3, 20.4, 23.9, 13.3