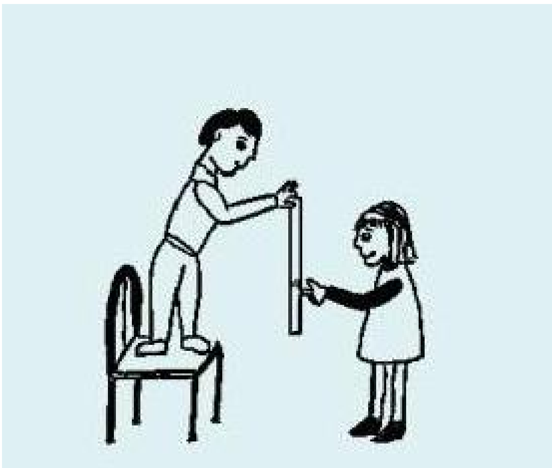

INSAID

Commonly used T-tests

- **Student's t-test** is a **method** of **testing hypotheses** about the **mean of a small sample** drawn
  - from a **normally distributed** population
  - when the population **standard deviation** is unknown.

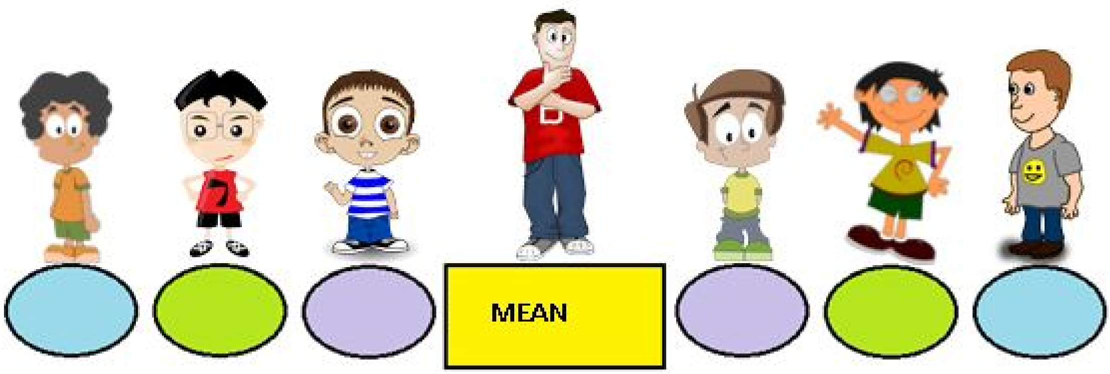
- In any **statistical hypothesis testing** situation, if the test statistic follows a **Student's t-test distribution** under **null hypothesis**, it is a **t-test.**



- **Most frequently used t-tests are**:
  - For comparison of mean in single sample.
  - Two samples related.
  - Two samples unrelated tests.

- In **one-sample location test**, it is tested whether or not the mean of the population has a value as specified in a **null hypothesis**.

- In **two independent sample location test**, equality of **means** of **two populations** is tested.



- To compare the **mean delta**(difference between two related samples) against **hypothesized value of zero** in a null hypothesis.
- It is also known as **paired t-test** or **repeated-measures t-test**.

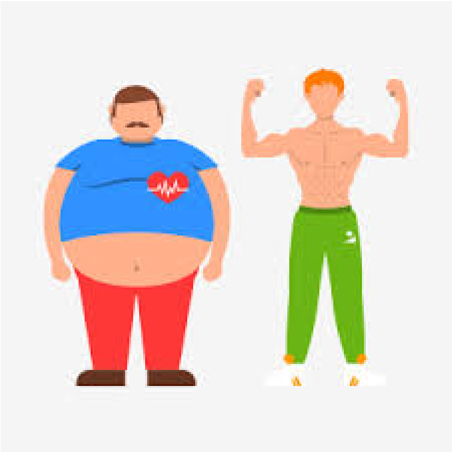
- For a **binary variable** (such as cure, relapse, hypertension, diabetes, etc.,) which is either **yes** or **no** for a subject.

- If we take **1** for **yes** and **0** for **no** and consider this as a **score** attached to each **study subject**.
- Then the **sample proportion (p)** and the **sample mean** would be the **same.**
- Therefore, the **approach of t-test** for mean can be used for **proportion** as well.

T-test for one sample:

For example

- In a random sample of 30 **hypertensive males**, the observed mean **body mass index** (BMI)is 27.0 kg/m2 and the **standard deviation** is 4.0.



- Also, suppose it is known that the mean **BMI in non-hypertensive males** is 25 kg/m2.
- If the question is to **know whether or not** these 30 observations could have **come from a population** with a **mean** of 25 kg/m2.
- To determine this, **one sample t-test** is used with the **null hypothesis** H~0~: Mean = 25, against alternate hypothesis of H~1~: Mean ≠ 25.
- Since the **standard deviation** of the **hypothesized population** is not known.
  - Therefore,**t-test** would be **appropriate**.
  - Otherwise, **Z-test** would have been used.

T-test for two related samples:

- Two samples can be regarded as related in a **pre-** and **post-design** (self-pairing).
- It can also be **in two groups** where the subjects have been matched on a third factor a **known confounder** (artificial pairing).
- In a **pre** and **post–design**,each subject is used as **his** or **her** own control.

For example:

- An **investigator** wants to assess effect of an intervention in **reducing systolic blood pressure**(SBP) in a pre- and post-design.

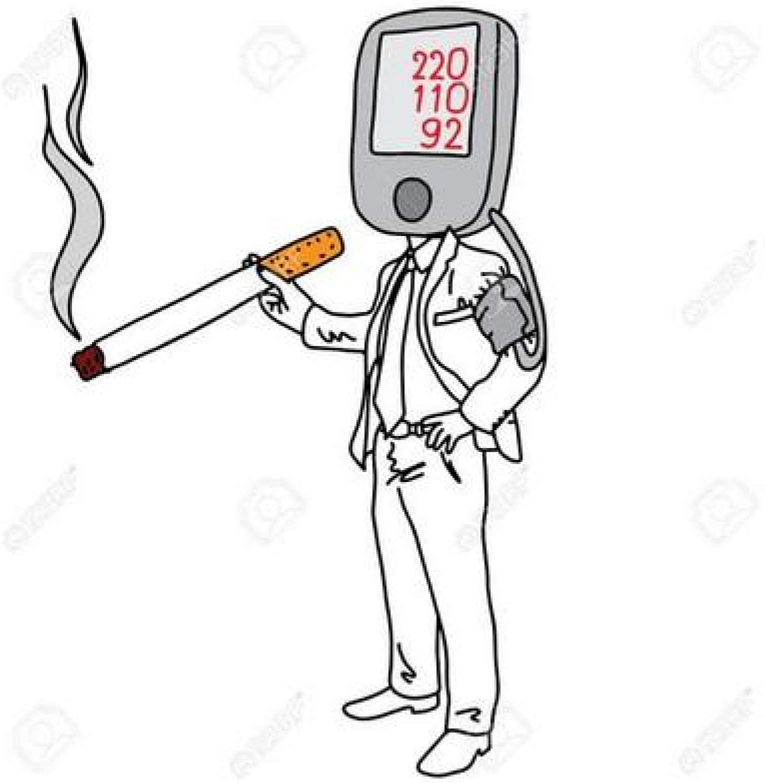
- Here, for each patient,there would be **two observations of SBP**, that is, before and after.

- Here, instead of individual observations, **difference between pairs of observations** would be of interest.
- The problem reduces to **one-sample situation** where the null hypothesis would be to test the **mean difference in SBP equal to zero**
- It is used to **test against** the alternate hypothesis of **mean SBP being not equal to zero**.
- The underlying assumption for **using paired t-test** is that under the null hypothesis the population of difference in **normally distributed**.
- This can be judged using the **sample values**.
- Using the **mean difference** and the **standard error** of the mean difference, **95% confidence interval** can be computed.

- The other situation of the two sample being related is the **two-group matched design**.

For example:

- In a **case–control study** to assess association between **smoking** and **hypertension.**
- Both **hypertensive** and **non-hypertensive** are matched on some third factor, say **obesity**, in a pair-wise manner.

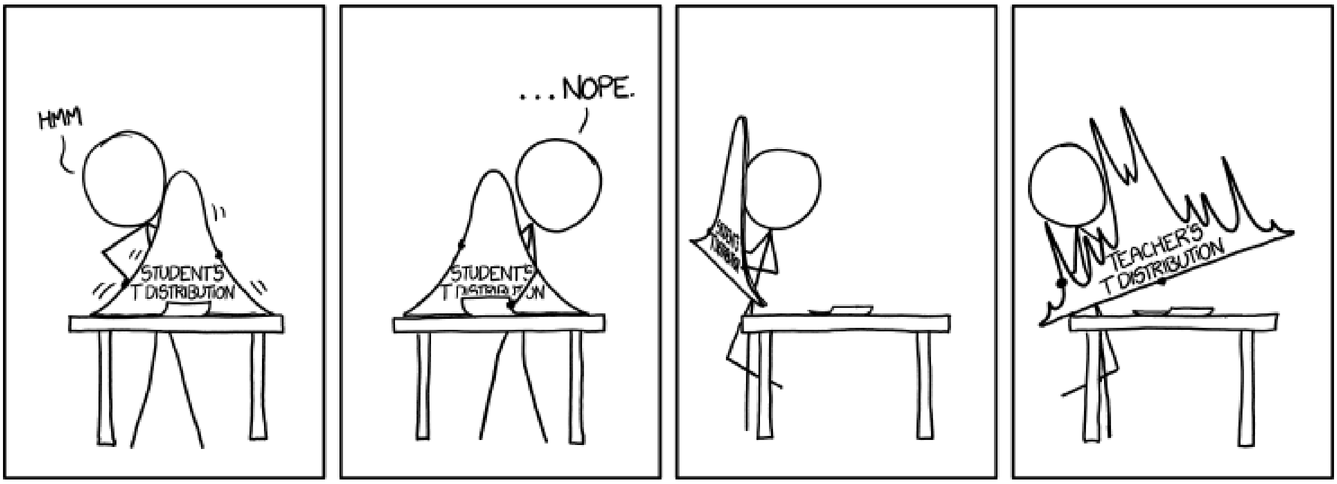


- Same approach of **paired analysis** would be used.
- In this situation, **cases and controls** are different subjects. However, they are related by the factor.

T-test for two independent samples:

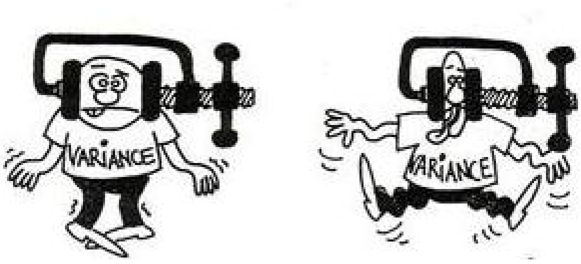
- To test the null hypothesis that the means of **two populations are equal**.
- **Student's t-test** is used provided the **variances** of the **two populations are equal.**

!



- The **two samples** are assumed to be **random sample**.

- When this assumption of **equality of variance is not fulfilled**, the form of the test used is a **modified t-test**.
- These tests are also known as **two-sample independent t-tests** with equal variance or unequal variance, respectively.



- The only difference in the **two statistical tests** lies in the denominator, that is, in determining the **pooled variance**.
- Prior to choosing **t-test for equal** or **unequal variance**, very often a test of variance is carried out to compare the two variances.
- It is **recommended** that this should be **avoided.**
- Using a **modified t-test** even in a situation when the **variances are equal**, has **high power.**
- Therefore, to compare the means in the two unrelated groups, using a **modified t-test is sufficient.**
- When there are **more than two groups**, use of **multiple t-test**(for each pair of groups) is **incorrect** because it may give **false-positive result**.
- Hence, in such situations, **one-way analysis of variance (ANOVA)**,followed by correction in P value for multiple comparisons (post-hoc ANOVA), if required, is used.

- It is used to **test the equality** of more than two means as the **null hypothesis**
- It ensures that the total **P value** of all the pair-wise **does not exceed 0.05**.

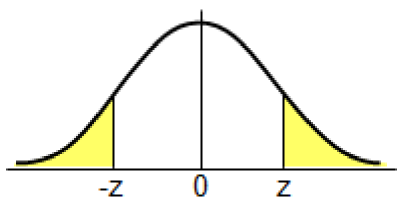
For Example:

- Do men spend the **same amount of money** on clothing as women?
- We can't reasonably ask the **entire population of men** and **women** how much they spend.
- So we'll draw a **sample of men** and **women**.
- These **samples are independent** because they **don't overlap**: everybody is either man or woman, never both.

- If the average amount spent is precisely equal for **all men and women.**
- We'll probably still see **slightly different means** between our samples.
- However, very different sample means suggest that the **population means weren't equal** after all.
- A **t test** tells us if a **sample difference is big enough** to draw this conclusion.

Commonly used Z-tests

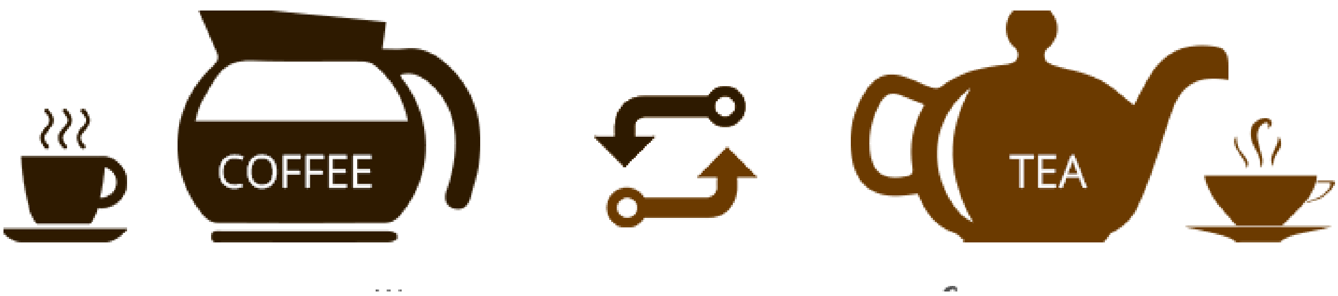
- It is a **statistical test** where **normal distribution** is applied.



- It is basically used for **dealing with problems** relating to **large samples when n ≥ 30** (n = sample size).

For example:

- Suppose we want to test if both **tea** and **coffee** are **equally popular** in a particular city.



- We can take a sample of size say **500 from the city** out of which suppose **280 are tea drinkers**.
- To **test** the **hypothesis**, we can use **Z-test**.

Z-test's for different purposes:

z-test for single proportion:

- It is used to **test a hypothesis** on a **specific value** of the **population proportion**.
- Statistically speaking, we **test the null hypothesis** H~o~: P = P~o~ against the alternative hypothesis H~a~: P >< P~o~
    - Where **P** is the **population proportion**
    - **P~0~** is a **specific value of the population proportion** we would like to test for acceptance.

The example on **tea drinkers** explained above requires this test.

- In that example, **P~0~ = 0.5**.
- Notice that in this particular example, proportion refers to the **proportion of tea drinkers**.

z-test for difference of proportions:

- It is used to **test** the **hypothesis** that **two populations** have the **same proportion**.

For example:

- Suppose one is interested to test if there is any significant difference in the **habit of tea drinking** between **male** and **female citizens** of a city.

- In such a situation,**z-test** for **difference of proportions** can be applied.
- One would have to obtain **two independent samples** from the city, **one from males** and the **other from females**.
- Determine the **proportion of tea drinkers** in each sample in order to perform this test.

z-test for single mean:

- It is used to test a hypothesis on a **specific value** of the **population mean**.
- Statistically speaking, we test the null hypothesis Ho: μ = μ~o~ against the alternative hypothesis Ha: μ >< μ~0~
    - Where **μ** is the **population mean**
    - **μ~0~** is a **specific value of the population** that we would like to test for acceptance.
- Unlike the **t-test** for **single mean**, this test is used if **n ≥ 30** and population **standard deviation** is known.

z-test for single variance:

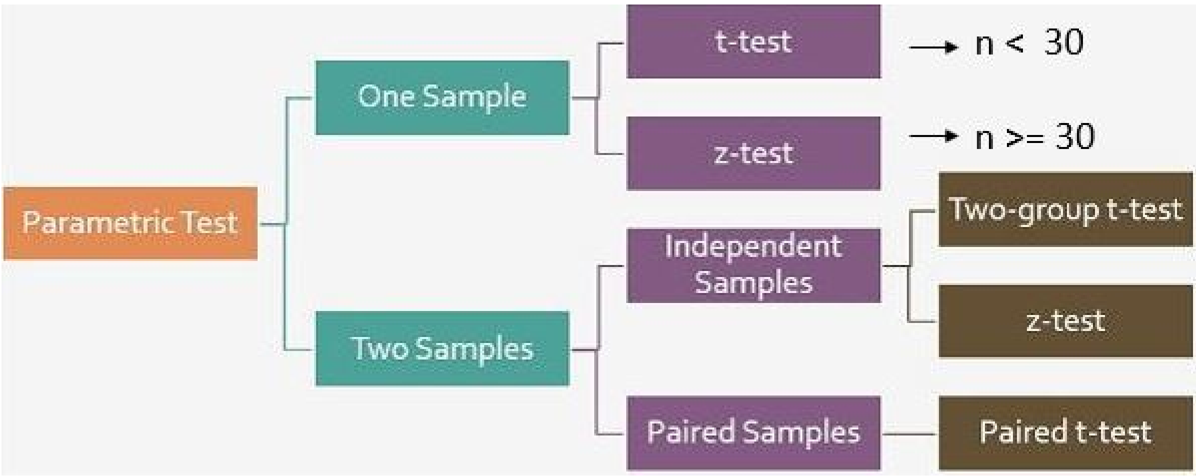
- It is used to test a hypothesis on a **specific value** of the **population variance.**
- Statistically speaking, we test the null hypothesis Ho: σ = σ~0~ against Ha: σ >< σ~0~
    - Where **σ** is the **population mean.**
    - **σ~0~** is a specific value of the **population variance** that we would like to test for acceptance.
- In other words, this test enables us to test if the given sample has been drawn from a population with **specific variance σ~0~**.
- Unlike the **chi square test** for **single variance**, this test is used if **n ≥ 30**.

z-test for testing equality of variance:

- It is used to test the hypothesis of equality of **two population variances** when the **sample size** of each sample is **30 or larger**.


TRUST ME KID
YOU DON'T WANT TO TEST ME

Flowchart


Parametric Test
One Sample
Two Samples
t-test
z-test
n < 30
n >= 30
Independent Samples
Paired Samples
Two-group t-test
z-test
Paired t-test