# [BIOSTATISTICS]

**Lecture 3 (Hypothesis testing & Tests of significance)**

**Given by: Dr. Nedal Al-Nawaiseh**

**Done by Doctor:**

**Abdullah Daradkh**

**Mahmoud Al-Otoom**

[MUTAH UNIVERSITY]

القلوب أوعية

والشفاه أقفالها

والألسن مفاتيحها

فليحفظ كل إنسان مفتاح سره.

# Hypothesis Testing

**Example: Average intake in children for dietary fat is 70 g of fat per day. Suppose we want to study children who eat a vegetarian diet. Possible hypotheses are**

1. **Average fat intake is 70 g per day**

2. **Average fat intake is less than 70 g per day**

--------------------------------------------------------------------------------------------------------------------------------------

**Def: One-sample problem—a single distribution.**

**Def: Hypothesis—statement about parameters in a population or populations. We want to know how likely this is to be true, given the evidence (data). For example,**

1. **Average number of beds filled per day in the hospital**

2. **Average number of minutes per day the doctor spends with a patient**

3. **Average lead content of water for a housing project**

**Def: Null hypothesis—$H_o$—the hypothesis to be tested. This is usually a statement of no difference.**
 **The population value of the parameter is not different from some specified value.**

**Def: Alternative hypothesis—$H_1$ or $H_A$—This is the statement we will accept if we reject the null hypothesis.**

يعني أنه عندمت نرفض النظرية الصفرية فإننا نقبل النظرية المغايرة

**$H_o$: Mean fat intake in vegetarian children is 70 g per day. $H_o$: $\mu = \mu_0$ or $\mu \geq \mu_0$**

**$H_1$: Mean fat intake in vegetarian children is < 70 g per day. $H_1$: $\mu < \mu_0$**

## ✚ Possible decisions:

1. **Accept $H_o$ (really, fail to reject $H_o$)**

2. **Reject $H_o$**

**Possible Scenarios:**

| Jury Trial | | |
|---|---|---|
| | Truth | |
| **Verdict** | **Innocent** | **Guilty** |
| **Innocent** | Correct decision | Error |
| **Guilty** | Error | Correct decision |

| Test of hypothesis | | |
|---|---|---|
| | Truth | |
| **Results of Test** | $H_o$ | $H_1$ |
| **Accept $H_o$** | Correct decision | Type II error |
| **Reject $H_o$** | Type 1 error | Correct decision |

**Def: <span style="color:red">Type I error</span>** is the probability of <span style="color:red">rejecting $H_o$ when $H_o$ is true.</span>

**Def: <span style="color:red">Type II error</span>** is the probability of <span style="color:red">accepting $H_o$ when $H_1$ is true.</span>

----------------------------------------------------------------------------------------------------------------------------

**Example:** We have developed a new procedure to improve survival of premature infants. If the hospital adopts these procedures, there will have to be new rooms and new equipment purchased. This is very costly.

1. **What does a Type I error imply?**
→when we say : increasing the number of rooms & equipments will improve survival of premature infants , but in reall life there is no improvement.

2. **What does a Type II error implies?**
→ When we say : increasing the number of rooms & equipments will not make a difference , but in reall life it will improve survival of premature infants.

**Def: <span style="color:red">Level of significance</span>:** $\alpha$ = Probability of a Type I error.
This is the area under the curve below (or above) the critical value.
This is the probability of rejecting $H_o$ when $H_o$ is true.

**Def:<span style="color:red">$\beta$</span>: Probability of a Type II error.**

**Def: <span style="color:red">$1$-$\beta$</span>—Power of a test. This is the**
Probability (rejecting $H_o$ when $H_1$ is true).

<span style="color:red">**Goal**</span>: Make $\alpha$, $\beta$ as small as possible. Usually, as $\alpha\uparrow$, $\beta\downarrow$ and as $\alpha\downarrow$, $\beta\uparrow$.

Fix $\alpha$ (0.05 or 0.01). Find a test to minimize $\beta$.

**Best** test for the fat experiment is one based on $\overline{X}$ .

**Def: <u>Acceptance Region</u>**—These are the values of $\overline{X}$ for which H$_o$ is accepted.

**Def: <u>Rejection Region</u>**—These are the values of $\overline{X}$ for which H$_o$ is rejected.

**Note:** For this example, we are conducting a one-sided or one-tailed test. We will only reject H$_o$ for values of $\overline{X}$ that are low.

**Def: <u>One-tailed</u>**—this is a test in which values of parameter under H$_1$ either $>$ or $<$ values under H$_o$ <u>but not both</u>.

## The picture behind the two sided test

0.95

0.025                    0.025

-c            0            c

reject H$_o$    do not reject H$_o$                    reject H$_o$
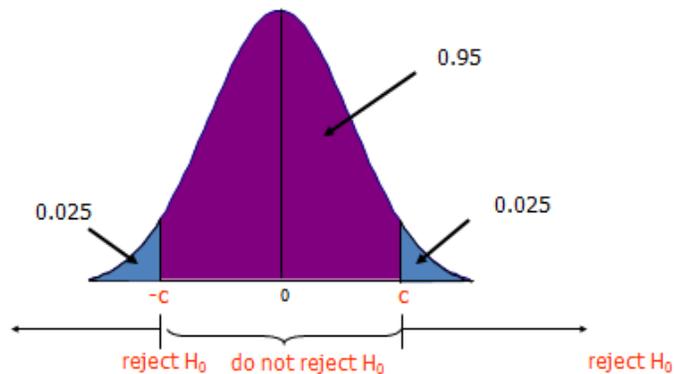
# Tests of Significance

**SUMMARY OF HYPOTHESIS TESTING**

*Basic principles:*

**1. Every time you see a p-value, someone tested a hypothesis**

**2. Your data comes from a sample you select; you hope your sample represents a larger population you'd like to draw conclusions about.**

**3. A statistic is a number (i.e. a proportion, an average or a disease rate) you calculate from your data sample.**

**4. The statistic is an attempt to describe the truth (a parameter) about a larger population using your sample.**

**5. Hypothesis testing allows a researcher to decide whether a difference between numbers is more likely due to random variation or a true difference**

# How to test a hypothesis?

1.  **Write a null hypothesis**

مثلا مستوى ذكاء الطلاب في جامعة مؤته = مستوى ذكاء الطلاب في باقي الجامعات الأردنية

2.  **Write an alternative hypothesis: determines if your test will be one-sided or two-sided (usually two).**

مستوى ذكاء طلاب جامعة مؤته لا يساوي مستوى ذكاء الطلاب في باقي الجامعات الأردنية

3.  **Set alpha (that is, the level of probability you will accept as unlikely related to chance alone), usually=0.05**

4.  **Generate a test statistic: a number you calculate to describe your sample results.**

   **There are many types of test statistics.**

 **The type of test statistic used depends on the type of data you have.**

5.  **Compare your test statistic to a known/published distribution for that type of test statistic in order to find the p-value**

6.  **If the p-value is less than alpha (usually .05), then the numbers in the null hypothesis differed more than expected due to chance alone.**

 **Reject the null hypothesis.**

7. **Conclude that the groups are different**

# Common Statistical Tests

**1.1. Review of the Central Limit Theorem (CLT):**

If we were to take a sample of size n from a population and calculate its **mean**, then sample the same population **again and calculate the mean**, then sample it **again and calculate the mean**, and keep doing this many times...then we graph all of those means on one curve, we would get a "**sampling distribution**" of the mean.

This sampling distribution will be normally distributed, with mean = μ and standard deviation =  **SEM** ( standard error of the mean)= **X-ų/ (s/√n)**


The CLT is very powerful, **but it has two limitations**:

1) it depends on a **large sample size**, and

2) to use it, we **need to know the standard deviation** of the population.

**In reality**, we usually **don't know** the **standard deviation** of the population so we use the standard deviation of our sample (denoted as 's' ) as an estimate.

## Standard error (SE) = X-ų/ (s/√n)

Since we are estimating the standard deviation using our sample, the sampling distribution will not be normal (even though it appears bell-shaped).

It is **a little shorter and wider** than **a normal distribution**,

 and it's called a **t-distribution.**

The t-distribution is **actually a family of distributions** – there is a different distribution for each sample value **of n-1 (degrees of freedom).**

The shape of t depends on the **size of the sample**…the larger the sample size, the more confident we can be that 's' is near sigma , and the **closer** t gets to Z.

Because it is not normal, the t-distribution does not follow **the 68-95-99 rule**, but we can use t-tables or computer programs to estimate the area-under-the-curve (probability) associated with a specific t-score**, t = (X - μ) / SE ,**

 and a specific sample size

**Again**, the t-distribution **approaches the normal distribution as n approaches infinity.**

For qualitative data we make chi-square.

For quantitative data we make 4 types of t-test:

1-one sample t-test

مثلا عندما يكون لدينا عينة واحدة نقارن وسطها الحسابي مع ال national figure .

2-independent sample t-test → اذا كان يوجد لدينا عينتان

3- ANOVA → اذا كان يوجد لدينا ثلاث عينات فأكثر

4-paired t-test

و هو الأكثر شيوعا و هو نفس الشخص قبل و بعد مثل مقارنة الفك العلوي مع السفلي أو الجانب الأيسر من الوجه مع الجانب الأيمن.

----------------------------------------------------------------------------------------------------------------------------

## 2. Statistical tests for Continuous data

### 1-One-sample t-test:

**How do we decide if a continuous measure taken on a sample of people is significantly more extreme than we might find by chance alone?**

**Remember that if we had taken a sample many times, the means collectively would form a curve around the true value that follows the t-distribution. So, whenever we are testing a sample mean, we use a t-statistic with SE in the denominator. Because of this, the larger the sample size, the more values fall into a smaller range.**

**To test a sample of normal continuous data, we need:**

1. **An expected value = the population or true mean**

2. **An observed mean = the average of your sample**

3. **A measure of spread: standard error**

4. **Degrees of freedom (df) = n-1 (number of values used to calculate SD or SE)**

**Then, we can calculate a test statistic to be compared to a known distribution**

**In the case of continuous, normal data, it's the t-statistic and the t-distribution.**

$$\text{T} = \frac{(\overline{X} - \mu)}{\frac{S}{\sqrt{n}}}$$

$$\text{T} = \frac{\text{(observed mean - expected mean)}}{\text{SE}}$$

**\*Notice that t is a measure of the difference between your data and what you expect to see, in units of standard error. This is a common theme of testing continuous variables.**

**<span style="color:red">An example</span>: You would like to see whether your clinic of HIV-positive men has more extreme testosterone levels than you would expect by chance.**

**The lab tells you that, among healthy men,**

**1) Testosterone levels are normally distributed;**

**2) the average population testosterone level is 600 ng/dl.**

**--Null hypothesis: Testosterone levels (your clinic) = Testosterone levels (general population)=600.**

**--Alternative hypothesis: Testosterone levels (clinic) is different (more or less) than general population mean**

**2-sided**

## <span style="color:red">--Set alpha=0.05</span>

**Sample your patients: 25 men who happen to visit in July.**
 **The results return with a mean testosterone = 500 ng/ml in your patients, SD=200 ng/ml. The average seems pretty good to you; it's close to 600. You calculate SE=  200/5 = 40.**

$$-- t = \frac{500 - 600}{40} = -2.5$$

**1.Your results are 2.5 standard error units below the expected value.**

 **The <span style="color:red">degrees of freedom</span>(n-1) are n-1 = 25-1 = 24.**

**2.     You use a computer program or a statistic table to look up the t- distribution with 24 degrees of freedom.**

**3.     A t of 2.5 (positive and negative values are handled the same because the curve is symmetric) has the same area of 0.01 in each tail.**

**4.     Because you're doing a two-tailed test, you need to consider the possibility of both tails, or 2 x 0.0l (again because the normal curve is symmetric). In this case, p=0.02.**

5.    Under the assumption that the true testosterone value of these patients is 600, the likelihood of getting a mean of 500, or more extreme in either direction from 600, by random sampling alone is only 2%.

6.    *The t-value needed to achieve statistical significance* **with a given alpha and**

**a given sample size is called the** *critical value* **P is less than or equal to alpha, so you reject the null hypothesis.**

**You conclude that the average you saw was unlikely to have occurred by chance alone, and that your patients' testosterone levels are lower on average than a healthy population.**

**The picture behind the two sided test**
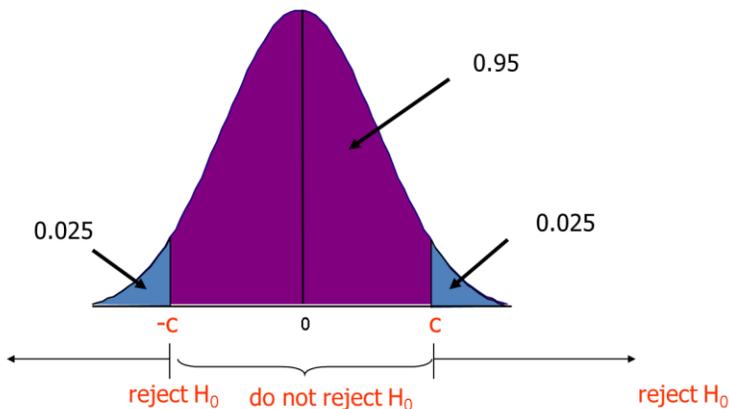


## Acceptable range

**We need to calculate the critical range using the t-distribution:**

**Two-sided:**      **Reject H$_0$ if**   $|T| > t_{n-1,\alpha/2}$

**One-sided upper tailed:**      **Reject H$_0$ if**   $T > t_{n-1,\alpha}$

**One-sided lower tailed:**      **Reject H$_0$ if**   $T < -t_{n-1,\alpha}$

| t-table Degrees of Freedom | Probability, p, level of significance, alpha, $\alpha$ | | | |
|---|---|---|---|---|
| | 0.1 | 0.05 | 0.01 | 0.001 |
| 1 | 6.31 | 12.71 | 63.66 | 636.62 |
| 2 | 2.92 | 4.30 | 9.93 | 31.60 |
| 3 | 2.35 | 3.18 | 5.84 | 12.92 |
| 4 | 2.13 | 2.78 | 4.60 | 8.61 |
| 5 | 2.02 | 2.57 | 4.03 | 6.87 |
| 6 | 1.94 | 2.45 | 3.71 | 5.96 |
| 7 | 1.89 | 2.37 | 3.50 | 5.41 |
| 8 | 1.86 | 2.31 | 3.36 | 5.04 |
| 9 | 1.83 | 2.26 | 3.25 | 4.78 |
| 10 | 1.81 | 2.23 | 3.17 | 4.59 |
| 11 | 1.80 | 2.20 | 3.11 | 4.44 |
| 12 | 1.78 | 2.18 | 3.06 | 4.32 |
| 13 | 1.77 | 2.16 | 3.01 | 4.22 |
| 14 | 1.76 | 2.14 | 2.98 | 4.14 |
| 15 | 1.75 | 2.13 | 2.95 | 4.07 |
| 16 | 1.75 | 2.12 | 2.92 | 4.02 |
| 17 | 1.74 | 2.11 | 2.90 | 3.97 |
| 18 | 1.73 | 2.10 | 2.88 | 3.92 |
| 19 | 1.73 | 2.09 | 2.86 | 3.88 |
| 20 | 1.72 | 2.09 | 2.85 | 3.85 |
| 21 | 1.72 | 2.08 | 2.83 | 3.82 |
| 22 | 1.72 | 2.07 | 2.82 | 3.79 |
| 23 | 1.71 | 2.07 | 2.82 | 3.77 |
| 24 | 1.71 | 2.06 | 2.80 | 3.75 |
| 25 | 1.71 | 2.06 | 2.79 | 3.73 |
| 26 | 1.71 | 2.06 | 2.78 | 3.71 |
| 27 | 1.70 | 2.05 | 2.77 | 3.69 |
| 28 | 1.70 | 2.05 | 2.76 | 3.67 |
| 29 | 1.70 | 2.05 | 2.76 | 3.66 |
| 30 | 1.70 | 2.04 | 2.75 | 3.65 |
| 40 | 1.68 | 2.02 | 2.70 | 3.55 |
| 60 | 1.67 | 2.00 | 2.66 | 3.46 |

| | | | | |
|---|---|---|---|---|
| 120 | 1.66 | <span style="color:red">1.98</span> | 2.62 | 3.37 |
| **infinity** | **1.65** | <span style="color:red">**1.96**</span> | **2.58** | **3.29** |