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
2. Write an alternative hypothesis: determines if your test will be one-sided or two-sided (usually two).
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: **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, $t = (X - \mu) / SE$, and a specific sample size

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

- **2. Statistical tests for Continuous data**

.2.1One-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.

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

$$t = \frac{\textbf{(observed mean - expected mean)}}{\textbf{SE}}$$

- ***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.**
- **An example**: 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/Ml.

1. Null hypothesis: Testosterone levels (your clinic) = Testosterone levels (general population)=600.
2. Alternative hypothesis: Testosterone levels (clinic) is different (**more or less**) than general population mean

2-sided

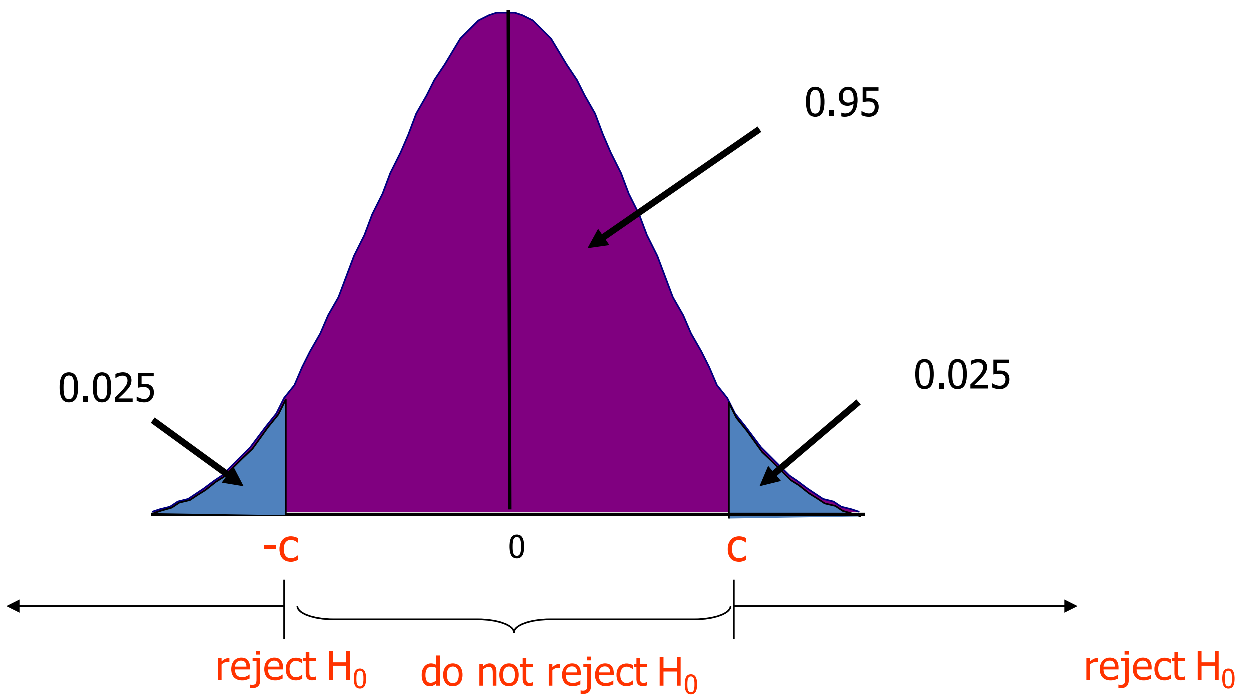
3. Set alpha=0.05
4. Sample your patients: 25 men who happen to visit in July. The results return with a mean testosterone = 500 ng/ml in your patients, SD=200 ng/ml. The average seems pretty good to you; it's close to 600. You calculate SE= 200/5 = 40.

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

- 1.Your results are 2.5 **standard error units** below the expected value. The **degrees of freedom** 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			
	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	1.98	2.62	3.37
infinity	1.65	1.96	2.58	3.29