Summary of Hypothesis Testing

Note: The notations and terminology can vary slightly from book to book.

General Definitions

A **hypothesis** is a statement about the population distribution that may or may not be true. In other words, a hypothesis is a question with a "yes" or "no" answer; think of the "yes" answer as accepting, or not rejecting, the null hypothesis H_0. The question is answered by performing an experiment, that is, by sampling from the population. The decision about accepting or rejecting H_0 is made using the **test statistic**, a function φ of the sample that takes large values if the null hypothesis does not hold. Mathematically, the null hypothesis is *rejected* when $\varphi_0 > c$, where $c > 0$ is a number called **critical value** of the test and φ_0 is the value of the test statistic on the sample.

When the null hypothesis H_0 is chosen, at least one one of the following applies :
1. The null hypothesis represents "status quo", "things as usual", "what is supposed to be", "no change", and so on. In particular, the null hypothesis is the *opposite* of what you **suspect**.
2. You hope or expect to reject H_0.
3. Erroneous rejection (that is, rejecting H_0 when it is true) is worse than erroneous acceptance (that is, not rejecting H_0 when the alternative H_1 is true). As an example, think of the prosecution in court testing the null hypothesis "the defendant is innocent" versus the alternative "the defendant is guilty".
4. You express the null hypothesis mathematically as $\theta = \theta_0$, where θ is the unknown parameter of the population distribution, and θ_0 is a fixed value of that parameter.

Accepting or rejecting the null hypothesis is a *random event*: the outcome depends on the random sample you collect from the population. On the other hand, whether the null hypothesis is indeed true or not is *not a random event* but a property of the population, even if you might never know that property.

The probability of erroneous rejection (the probability to reject H_0 when H_0 holds) is denoted by α and is called the **probability of type I error** or **the size of the test** or **the level of significance of the test**. The probability of erroneous acceptance is denoted by β, and $1 - \beta$ is called **the power of the test**. The power of the test is the probability of *correct rejection*, that is, the probability to reject H_0 when it is not true. You cannot have the ideal situation, when $\alpha = \beta = 0$; in fact, if $\alpha = 0$ then $\beta = 1$ and vice versa. Instead, you decide on the value of α in advance and then try to minimize β (the *Neyman-Pearson* approach that often results in the *most powerful test*). This approach allows you to control the (more serious) type I error.

Whether you reject H_0 or not depends on two things: the sample you got and the level of significance of the test you chose. The smaller the level of significance, the less likely you are to reject the null hypothesis. As a result, the same sample can lead to either acceptance or rejection depending on the level of significance used. For example, you may reject at the level of significance 10%, but not at 5%. To avoid possible misunderstanding, the results are often reported as the **p-value** of the experiment, that is, the smallest level of significance at which the null hypothesis can be rejected on the basis of that particular sample. You compute the p-value by computing the probability that, when H_0 is true, the test statistic φ will be bigger than the value φ_0 you got on the sample. Note that p-value is the size of the test that uses φ_0 as the critical value. **You reject the null hypothesis if and only if $\alpha >$p-value.**

The steps in hypothesis testing.

(1) Choose the null hypothesis H_0 and the alternative H_1. If necessary, choose the level of significance α.
(2) Choose the test statistic φ so that
- the random variable φ is more likely to be large when *the alternative* is true, and
- the distribution of φ is known when the *null hypothesis* is true.

(3) Determine the critical value $c = c_\alpha$ from the condition
$$P(\varphi > c_\alpha \text{ when } H_0 \text{ is true}) = \alpha.$$
(note that the critical value is a function of α, hence the notation c_α)
(4) Compute φ_0, the value of the test statistic on the sample.
(5) If $\varphi_0 > c_\alpha$, then reject H_0; otherwise, do not reject H_0 (note that in all our models the test statistic is a continuous random variable, so $P(\varphi_0 = c_\alpha) = 0$).
(6) Compute the p-value of the test, that is, compute the probability
$$P(\varphi > \varphi_0 \text{ when } H_0 \text{ is true}).$$
(7) Note that the larger the φ_0 (the value of the test statistic on the sample), the smaller the p-value. Small p-value means a strong evidence against the null hypothesis. Large p-value (usually bigger than 0.1) means that there is not enough evidence against the null hypothesis.

Main Models and Questions

Model / Question	$\mathcal{N}(\mu,\sigma^2)$, σ known $\theta=\mu$	$\mathcal{N}(\mu,\sigma^2)$, σ unknown $\theta=\mu$	$\mathcal{B}(1,p)$, p unknown $\theta=p,\ np_0(1-p_0)>5$
Two-sided. $H_0:\theta=\theta_0$ $H_1:\theta\neq\theta_0$	$\varphi=\dfrac{\sqrt{n}\lvert\bar X_n-\mu_0\rvert}{\sigma}$ $\varphi=\lvert Z\rvert$ $Z\sim\mathcal{N}(0,1)$ $c_\alpha=z_{\alpha/2}$ p-value$=2P(Z>\varphi_0)$	$\varphi=\dfrac{\sqrt{n}\lvert\bar X_n-\mu_0\rvert}{s_n}$ $\varphi=\lvert t_{n-1}\rvert$ $c_\alpha=t_{n-1,\alpha/2}$ p-value$=2P(t_{n-1}>\varphi_0)$	$\varphi=\dfrac{\sqrt{n}\lvert\hat p-p_0\rvert}{\sqrt{p_0(1-p_0)}}$ $\varphi\approx\lvert Z\rvert$ $Z\sim\mathcal{N}(0,1)$ $c_\alpha=z_{\alpha/2}$ p-value$=2P(Z>\varphi_0)$
One-sided. $H_0:\theta=\theta_0$ $H_1:\theta<\theta_0$	$\varphi=\dfrac{\sqrt{n}(\mu_0-\bar X_n)}{\sigma}$ $\varphi=Z$ $Z\sim\mathcal{N}(0,1)$ $c_\alpha=z_\alpha$ p-value$=P(Z>\varphi_0)$	$\varphi=\dfrac{\sqrt{n}(\mu_0-\bar X_n)}{s_n}$ $\varphi=t_{n-1}$ $c_\alpha=t_{n-1,\alpha}$ p-value$=P(t_{n-1}>\varphi_0)$	$\varphi=\dfrac{\sqrt{n}(p_0-\hat p)}{\sqrt{p_0(1-p_0)}}$ $\varphi\approx Z$ $Z\sim\mathcal{N}(0,1)$ $c_\alpha=z_\alpha$ p-value$=P(Z>\varphi_0)$
One-sided. $H_0:\theta=\theta_0$ $H_1:\theta>\theta_0$	$\varphi=\dfrac{\sqrt{n}(\bar X_n-\mu_0)}{\sigma}$ $\varphi=Z$ $Z\sim\mathcal{N}(0,1)$ $c_\alpha=z_\alpha$ p-value$=P(Z>\varphi_0)$	$\varphi=\dfrac{\sqrt{n}(\bar X_n-\mu_0)}{s_n}$ $\varphi=t_{n-1}$ $c_\alpha=t_{n-1,\alpha}$ p-value$=P(t_{n-1}>\varphi_0)$	$\varphi=\dfrac{\sqrt{n}(\hat p-p_0)}{\sqrt{p_0(1-p_0)}}$ $\varphi\approx Z$ $Z\sim\mathcal{N}(0,1)$ $c_\alpha=z_\alpha$ p-value$=P(Z>\varphi_0)$

Notation $\varphi=t_{n-1}$ means that, under the null hypothesis, the test statistics has the t distribution with $n-1$ degrees of freedom. Notation $\varphi=|Z|$ means that the distribution of φ is the same as the distribution of the absolute value of the standard normal random variable.

The values of $t_{n,\alpha}$ and z_α are taken from the tables. Remember that $P(t_n>t_{n,\alpha})=\alpha$ and $P(Z>z_\alpha)=\alpha$.

Keep in mind that $\bar X_n$ is the sample mean; for proportions, $\bar X_n=\hat p=\frac{\text{number of objects with the given property}}{n}$; s_n is the sample **standard deviation**.

If $t_{n-1,\alpha_1}<\varphi_0<t_{n-1,\alpha_2}$, then $\alpha_2<\alpha_1$, and $\alpha_2<P(t_{n-1}>\varphi_0)<\alpha_1$.

A real-life message from a student
Hi professor,
I am going home for Thanksgiving and will not attend the class on Nov 24, but I noticed that the schedule says "Hypothesis test for 2 means" on that day. Just to make sure, are we having a test on that topic that day? If we are, can I take the test the week before I leave?