Hypothesis Testing/Tests of Significance (about the Mean, Proportion, and Standard Deviation or Variance)

1. The **null hypothesis** is denoted by H_0. It states that the parameter of interest is equal to, greater than or equal to, or less than or equal to a specific, hypothesized value. In the context of a population mean, population proportion, or standard deviation or variance, H_0 has one of the three forms

$$\mathrm{H}_0 : \mu = \mu_0 \text{ (or } \mathrm{H}_0 : \mu \geq \mu_0 \text{ or } \mathrm{H}_0 : \mu \leq \mu_0)$$
$$\mathrm{H}_0 : p = p_0 \text{ (or } \mathrm{H}_0 : p \geq p_0 \text{ or } \mathrm{H}_0 : p \leq p_0)$$
$$\mathrm{H}_0 : \sigma = \sigma_0 \text{ (or } \mathrm{H}_0 : \sigma \geq \sigma_0 \text{ or } \mathrm{H}_0 : \sigma \leq \sigma_0)$$

where μ_0, p_0, and σ_0 stand for the hypothesized value of interest, often matching the TI calculator symbol. The null hypothesis is typically a statement of "no effect" or "no difference". The significance test is designed to assess the **evidence against the null hypothesis**.

2. The **alternative hypothesis** is denoted by H_1. It typically states what the researchers suspect or hope to be true about the parameter of interest. It depends on the purpose of the study and must be specified before seeing the data. The alternative hypothesis can take one of three forms:

$$\mathrm{H}_1 : \mu \neq \mu_0 \text{ (or } \mathrm{H}_1 : \mu < \mu_0 \text{ or } \mathrm{H}_1 : \mu > \mu_0)$$
$$\mathrm{H}_1 : p \neq p_0 \text{ (or } \mathrm{H}_1 : p < p_0 \text{ or } \mathrm{H}_1 : p > p_0)$$
$$\mathrm{H}_1 : \sigma \neq \sigma_0 \text{ (or } \mathrm{H}_1 : \sigma < \sigma_0 \text{ or } \mathrm{H}_1 : \sigma > \sigma_0)$$

The first form is a **two-sided** alternative, while the last two are one-sided alternatives.

3. The **test statistic** is a value computed by standardizing the observed sample statistic on the basis of the hypothesized parameter value. It is used to assess the evidence against the null hypothesis. In the context of a population mean, proportion, or variance or standard deviation, it is denoted by z or t or χ^2 and calculated as follows:

$$\mathrm{z} = \frac{\bar{\mathrm{X}} - \mu}{\frac{\sigma}{\sqrt{\mathrm{n}}}} \quad \text{or} \quad \mathrm{t} = \frac{\bar{\mathrm{X}} - \mu}{\frac{s}{\sqrt{\mathrm{n}}}}$$

Use the z test if the population standard deviation is known (and the population is normally distributed or n > 30), and use the t test if σ is unknown (and the population is normally distributed or n > 30).

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Use the z test if np$\geq$ 5 and nq$\geq$ 5.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Strict requirement: normally distributed population

4. The **p-value** is the probability, assuming the null hypothesis to be true, of obtaining a test statistic as extreme or more extreme than the one actually observed. “Extreme” means “in the direction of the alternative hypothesis,” so the p-value takes one of three forms (corresponding to the appropriate form of H_1).

The p-value of a hypothesis test is the smallest significance level at which the null hypothesis can be rejected. One judges the strength of the evidence that the data provide against the null hypothesis by examining the p-value. The smaller the p-value, the stronger the evidence against H_0 (and thus the stronger the evidence in favor of H_1). For instance, typical evaluations are:

p-value > 0.1	Little or no evidence against H_0
0.05 < p-value < 0.1	Some evidence against H_0
0.01 < p-value < 0.05	Moderate evidence against H_0
0.001 < p-value < 0.01	Strong evidence against H_0
p-value < 0.001	Very strong evidence against H_0

Note: The probability entries in the table above are all p-values and not related to “p-hat”, the sample proportion, $\hat{p}$, or to p, the population proportion.

5. The **significance level**, denoted by α, is an optional “cut-off” level for the p-value that one decides to regard as decisive. The experimenter specifies the significance level in advance. Common values are 0.10, 0.05, and 0.01. The smaller the significance level, the more evidence you require in order to be convinced that the null hypothesis is not true. Remember to consider whether the test involves one or two-tails in the computation.

The probability of a Type I error (rejecting a true null hypothesis) is α.

For a fixed sample size, the smaller we specify the significance level, α, the larger will be the probability of a Type II error, β, (not rejecting a false null hypothesis).

6. If the p-value of the test is less than or equal to the significance level, α, the **test decision** is to reject H_0; otherwise, the decision is to fail to reject H_0. Notice that failing to reject is not the same as affirming its truth; it is simply to admit that the evidence was not convincing enough to reject it. Another very common expression is to say that the data are **statistically significant** at the α level if the p-value is less than or equal to α. Thus, a result is statistically significant if it is unlikely to have occurred by chance.

The value $1 - \alpha$ is the confidence level for a confidence interval. Confidence intervals correspond directly to 2-tailed hypothesis tests and are considered one of the valid methods of hypothesis testing.

The technical assumptions needed to establish the validity of this significance testing procedure are the same as for the confidence interval procedure: that the data are a simple random sample from the population of interest and the population standard deviation, σ, is known and the population is normally distributed or n > 30 (for a z-interval for the mean); σ is unknown and the population is normally distributed or n > 30 (for a t-interval for the mean); that $np \geq 5$ and $nq \geq 5$ (for a 1-prop z-interval for the proportion); and there is no TI confidence interval for the variance or standard deviation.