557: MATHEMATICAL STATISTICS II
HYPOTHESIS TESTING: RECAP OF BASIC CONCEPTS

A statistical hypothesis test is a **decision rule** that takes as an input observed sample data and returns an action relating to two mutually exclusive **hypotheses** that reflect two competing hypothetical states of nature. The decision rule partitions the sample space $\mathbb{X}$ into two regions that respectively reflect support for the two hypotheses. The following terminology is used:

- **Hypotheses:**
  - Two **hypotheses** characterize the two possible states of nature. The **null hypothesis** is denoted H_0, the **alternative hypothesis** is denoted H_1.
  - In parametric models, the null and alternative hypotheses define a partition of the 'effective' parameter space Θ (that is, the only possible values of θ that are to be considered). Suppose that disjoint subsets Θ_0, Θ_1 correspond to H_0 and H_1 respectively. We write

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

- **Tests:**
  - A **test**, $\mathcal{T}$, of H_0 versus H_1 defines a partition of sample space $\mathbb{X}$ into two regions. The hypothesis H_0 is **rejected in favour of** H_1 depending on where the data $\mathbf{x}$ (or a suitably chosen statistic $T(\mathbf{x})$) fall within $\mathbb{X}$.
  - A **test statistic**, $T(\mathbf{x})$, is the function of data $\mathbf{x}$ used in a statistical hypothesis test.
  - The **critical region**, $\mathcal{R}$, is the region within which $T(\mathbf{x})$ must lie in order for hypothesis H_0 to be rejected in favour of H_1. The complement of $\mathcal{R}$ will be written $\mathcal{R}'$.
  - A **test function**, $\phi_{\mathcal{R}}(T(\mathbf{x}))$, is an indicator function that reports the result of the test,

$$\phi_{\mathcal{R}}(T(\mathbf{x})) = \begin{cases} 1 & T(\mathbf{x}) \in \mathcal{R} \\ 0 & T(\mathbf{x}) \in \mathcal{R}' \end{cases}$$

- **Errors:**
  - A **Type I error** occurs when the null hypothesis H_0 is **rejected** when it is in fact **true**.
  - A **Type II error** occurs when the null hypothesis H_0 is **accepted** when it is in fact **false**.
  - For test with test statistic T and critical region $\mathcal{R} \subset \mathbb{X}$, and $\theta \in \Theta_0$, define the **Type I error probability** $\xi(\theta)$ by

$$\xi(\theta) = \Pr\left[T \in \mathcal{R}; \theta\right] \qquad \theta \in \Theta_0 \tag{1}$$

    If Θ_0 comprises a single value θ_0, then $\xi = \Pr\left[T \in \mathcal{R}; \theta_0\right]$.

- **Test properties:**
  - The **size** of a statistical test is

$$\overline{\alpha} = \sup_{\theta \in \Theta_0} \xi(\theta)$$

    which is equal to ξ if Θ_0 comprises a single value.
  - Suppose $\alpha \geq \overline{\alpha}$. If $T(\mathbf{x}) \in \mathcal{R}$, then H_0 is rejected at **level** α, and rejected at level $\alpha + \epsilon$ for $\epsilon > 0$.
  - The **power function**, $\beta(\theta)$, is defined by

$$\beta(\theta) = \Pr\left[T \in \mathcal{R}; \theta\right] \qquad \theta \in \Theta$$

    so that $\beta(\theta) = \xi(\theta)$ for $\theta \in \Theta_0$.

    *Note that this notation is not universally used; commonly the **power** of a statistical test is denoted $1 - \beta(\theta)$ and computed for $\theta \in \Theta_1$, whereas the **Type II error probability** is $\beta(\theta)$ for $\theta \in \Theta_1$.*