Statistics for Data Science - 2

Week 11 Notes

Hypothesis testing

1. **Null hypothesis:**
   The null hypothesis is a kind of hypothesis which explains the population parameter whose purpose is to test the validity of the given experimental data. It is denoted by H_0. The null hypothesis is a default hypothesis that is assumed to remain possibly true.

2. **Alternative hypothesis:**
   The alternative hypothesis is a statement used in statistical inference experiment. It is contradictory to the null hypothesis and denoted by H_A or H_1.

3. **Test statistic:**
   A test statistic is numerical quantity computed from values in a sample used in statistical hypothesis testing.

4. **Type I error:**
   A type I error is a kind of fault that occurs during the hypothesis testing process when a null hypothesis is rejected, even though it is true.

5. **Type II error:**
   A type II error is a kind of fault that occurs during the hypothesis testing process when a null hypothesis is accepted, even though it is not true (H_A is true).

6. **Significance level (Size):**
   Significance level (also called size) of a test, denoted α, is the probability of type I error.
   $$\alpha = P(\text{Type I error})$$

7. $\beta = P(\text{Type II error})$

8. **Power of a test:**
   Power $= 1 - \beta$

9. **Types of hypothesis:**
   (a) **Simple hypothesis:** A hypothesis that completely specifies the distribution of the samples is called a simple hypothesis.
   (b) **Composite hypothesis:** A hypothesis that does not completely specify the distribution of the samples is called a composite hypothesis.

10. **Standard testing method: z-test:**
    Consider a sample $X_1, X_2, \ldots, X_n \sim$ i.i.d. X.

- Test statistic, denoted T, is some function of the samples. For example: sample mean $\overline{X}$
- Acceptance and rejection regions are specified through T.

(a) **Right-tailed z-test:**

- $H_0 : \mu = \mu_0, \quad H_A : \mu > \mu_0$
- Test: reject H_0 if $T > c$.
- Significance level α depends on c and the distribution of $T|H_0$.
- $\alpha = P(T > c|H_0)$
- Fix α and find c.

(b) **Left-tailed z-test:**

- $H_0 : \mu = \mu_0, \quad H_A : \mu < \mu_0$
- Test: reject H_0 if $T < c$.
- Significance level α depends on c and the distribution of $T|H_0$.
- $\alpha = P(T < c|H_0)$
- Fix α and find c.

(c) **two-tailed z-test:**

- $H_0 : \mu = \mu_0, \quad H_A : \mu \neq \mu_0$
- Test: reject H_0 if $|T| > c$.
- Significance level α depends on c and the distribution of $T|H_0$.
- $\alpha = P(|T| > c|H_0)$
- Fix α and find c.

Note: In the test for mean (σ^2 known), $T = \overline{X}$ and when null is true, $\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1)$.

11. ***P*-value:**
Suppose the test statistic $T = t$ in one sampling. The lowest significance level α at which the null will be rejected for $T = t$ is said to be the P-value of the sampling.