Packet 6: Hypothesis Testing

Chap 8.1 Test about One Mean

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. $\sim N(\mu, \sigma^2)$. $\bar{X}$ is the sample mean and s^2 is the sample variance.

If σ^2 is known, the test statistic is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

If σ^2 is unknown, the test statistic is

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}.$$

They are also the pivotal quantities in the construction of confidence intervals.

To test the value of μ at significance level α, *Example 1*: It is well known that a new-born baby sleeps about 15 hours per day. New patients were curious about whether it was the case for their baby. So they recorded their baby's sleeping times for 10 days ($n = 10$), and assumed that the data $X_1, \ldots, X_{10}$ are i.i.d. $N(\mu, \sigma^2)$ r.v.s, where μ denote their baby's true sleeping time. They want to test $H_0 : \mu = 15$ vs $H_1 : \mu \neq 15$.

The average sleeping time for 10 days ($n = 10$) is $\bar{x} = 17$ hours with $s^2 = 1.5^2$. What are intuitive test statistic and rejection region for this hypothesis testing at significance level $\alpha = 0.05$?

1. Specify the null and alternative hypotheses:

2. Set up a test statistic for μ, and find its distribution under H_0:

3. Determine the rejection region when $\alpha = 0.05$:

Chap 8.2 Test about Two Means

We are interested in comparing two populations denoted by X and Y. We independently collect two sets of random samples:

$X_1, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$, $Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$.

Case I: If σ_X^2 and σ_Y^2 are both known.

Case II: If σ_X^2 and σ_Y^2 are both unknown, but we may assume $\sigma_X^2 = \sigma_Y^2$.

Case III: If σ_X^2 and σ_Y^2 are both unknown and $\sigma_X^2 \neq \sigma_Y^2$, $n \geq 30$, $m \geq 30$.

Case IV: paired samples Suppose X and Y are always collected in pairs (X_1, Y_1), (X_2, Y_2),..., (X_n, Y_n).

$X_i \sim N(\mu_X, \sigma_X^2)$ – blood pressure before treatment for patient i.

$Y_i \sim N(\mu_Y, \sigma_Y^2)$ – blood pressure after treatment for patient i.

Let $D_i = X_i - Y_i \sim N(\mu_X - \mu_Y, \sigma_D^2)$. It is reduce to a one-population problem with $D = X - Y$, and the parameter of interest being $\mu_D = \mu_X - \mu_Y$.

Textbook Examples: 8.1-1, 8.1-2, 8.1-3, 8.1-4, 8.1-5 (paired t-test), 8.2-1, 8.2-2, 8.2-3.

Example 2: Air pollution levels are measured in two cities, 13 samples from Melbourne $X_i \sim N(\mu_X, \sigma_X^2)$, and 16 samples from Huston$Y_i \sim N(\mu_Y, \sigma_Y^2)$.

We observe $\bar{x} = 72.9$, $\bar{y} = 81.7$, $s_x^2 = 25.6^2$, $s_y^2 = 28.2^2$.

Test $H_0 : \mu_X = \mu_Y$ vs $H_1 : \mu_X < \mu_Y$ at $\alpha = 0.05$.