Introduction to Statistics Note

2024 Spring Semester

21 CST H3Art

Chapter 7: Inference for Distributions

7.1 Inference for the Mean of a Population

When the sampling distribution of $\bar{x}$ is close to Normal, we can find probabilities involving $\bar{x}$ by standardizing:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

When we don’t know σ, we can estimate it using the sample standard deviation s_x, our statistic has a new distribution called a **t distribution**.

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$$

There is a different t distribution for each sample size, specified by its **degrees of freedom （自由度）(df)**, the one-sample t statistic has the **t distribution** with degrees of freedom df = $n – 1$.

The One-Sample t Interval for a Population Mean:

Choose an SRS of size n from a population having unknown mean . A level C **confidence interval** for μ is:

$$\bar{x} \pm t \times \frac{s_x}{\sqrt{n}}$$

where t is the **critical value** for the $t(n – 1)$ distribution.

The **margin of error** is:

$$t \times \frac{s_x}{\sqrt{n}}$$

The One-sample t Test:

Choose an SRS of size n from a large population that contains an unknown mean μ. To test the hypothesis $H_0 : \mu = \mu_0$, compute the one-sample t statistic:

Matched Pairs t Procedures:

To compare the responses to the two treatments in a matchedpairs design, find the difference between the responses within each pair. Then apply the one-sample t procedures to these differences.

7.2 Comparing Two Means

When data come from two random samples or two groups in a randomized experiment, the statistic $\bar{x_1} - \bar{x_2}$ is our best guess for the value of $\mu_1 - \mu_2$.

When the two samples are independent of each other, the **standard deviation** of the statistic $\bar{x_1} - \bar{x_2}$ is:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

We standardize the observed difference to obtain a t statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

When the Random, Normal, and Independent conditions are met, a level C confidence interval for $(\mu_1 - \mu_2)$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm t \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where t is the critical value at confidence level C for the t distribution with degrees of freedom either gotten from technology or equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

Approximate Distribution of the Two-Sample t Statistic:

The distribution of the two-sample t statistic is very close to the t distribution with degrees of freedom given by:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1} \times \frac{s_1^2}{n_1}\right)^2 + \left(\frac{1}{n_2-1} \times \frac{s_2^2}{n_2}\right)^2}$$

This approximation is accurate when both sample sizes are 5 or **larger**.

Pooled Two-Sample Procedures:

degrees of freedom: $n_1 + n_2 - 2$

Suppose both populations are Normal and they have the same standard deviations. The pooled estimator of σ^2 is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

A level C confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm t \times s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where the degrees of freedom for t are $n_1 + n_2 - 2$

To test the hypothesis $H_0 : \mu_1 = \mu_2$ against a **one-sided** or a **two-sided** alternative, compute the pooled two-sample t statistic for the $t(n_1 + n_2 - 2)$ distribution.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$