STAT 423/523 Statistical Methods for Engineers and Scientists

Lecture 5: Hypothesis Testing II

Chencheng Cai

Washington State University

From Z-test to T-test

We discussed Z-test in the last lecture. The Z-test is based on the Z-statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

where σ is the population standard deviation assumed to be known.

In practice, we often do not know the population standard deviation.

T-Statistic

For a population with unknown standard deviation, we consider the **T-statistic**:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

where S is the sample standard deviation:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

T-Statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

When the population is normally distributed, the T-statistic, under $H_0 : \mu = \mu_0$, follows a t-distribution with $n - 1$ degrees of freedom.

$$T \sim t_{n-1}$$

The test based on the T-statistic is called the **T-test**, which is similar to the Z-test.

Two-sided T-Test

Let $X_1, \dots, X_n$ be a random sample from a normal population with unknown mean μ and unknown standard deviation. We want to test

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu \neq \mu_0.$$

The test statistic is

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

The decision rule is

$$\text{reject null if } |T| > t_{\alpha/2, n-1},$$

where $t_{\alpha/2,n-1}$ is the $\alpha/2$ quantile of the t-distribution with $n-1$ degrees of freedom.

Example (Textbook 8.10)

After 24 hours of smoke abstinence, 20 smokers are asked how much time has elapsed during a 45-second period. The sample data is summarized as follows:

$$\bar{x} = 59.30, \quad s = 9.84$$

We consider test the following hypothesis with $\alpha = 0.05$:

$$H_0 : \mu = 45 \quad \text{v.s.} \quad H_a : \mu \neq 45.$$

The test statistic is

$$t = \frac{\bar{x} - 45}{s/\sqrt{n}} = 6.50$$

The quantile of the t-distribution is $t_{0.025,19} = 2.093$. We reject the null hypothesis since $|t| = 6.50 > 2.093$.

P-value of Two-sided T-Test

Similar to the Z-test, the p-value of the two-sided T-test is

$$\text{p-value} = 2(1 - F_{t,n-1}(|T|)),$$

where $F_{t,n-1}(\cdot)$ is the cdf of the t-distribution with $n-1$ degrees of freedom.

We reject the null hypothesis if the p-value is less than the significance level α.

Power of Two-sided T-Test

Consider $\mu_1 \neq \mu_0$ in the alternative hypothesis. The power $(1 - \beta(\mu_1))$ of the two-sided T-test is

$$\begin{aligned}
\text{Power} &= P(\text{reject} \mid H_1) \\
&= P(|T| > t_{\alpha/2,n-1} | \mu = \mu_1) \\
&= F_{t,n-1}\left(-t_{\alpha/2,n-1} + \frac{\mu_1 - \mu_0}{s/\sqrt{n}}\right) + F_{t,n-1}\left(-t_{\alpha/2,n-1} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}}\right)
\end{aligned}$$

Sample Size Determination

In order to have a desired power $1 - \beta^*$, we need to determine the minimum sample size n such that $1 - \beta(\mu_1) \geq 1 - \beta^*$.

Similar to the two-sided Z-test, the approximate formula for the sample size is

$$n^* = \left\lceil \left(\frac{s(t_{\alpha/2,n-1} - t_{1-\beta^*,n-1})}{\mu_1 - \mu_0} \right)^2 \right\rceil$$

Summary of Two-sided T-test

- Hypothesis:

$$H_0: \ \mu = \mu_0 \quad \text{v.s.} \quad H_a: \ \mu \neq \mu_0$$

- Test statistic:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

- Rejection region:

$$\text{reject null if } |T| > t_{\alpha/2, n-1}$$

- P-value:

$$\text{p-value} = 2(1 - F_{t,n-1}(|T|))$$

- Power:

$$1 - \beta(\mu_1) = F_{t,n-1}\left(-t_{\alpha/2,n-1} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}}\right) + F_{t,n-1}\left(-t_{\alpha/2,n-1} + \frac{\mu_1 - \mu_0}{s/\sqrt{n}}\right)$$

- Sample size determination: (approximation)

$$n^* = \left\lceil \left(\frac{s(t_{\alpha/2,n-1} - t_{1-\beta^*,n-1})}{\mu_1 - \mu_0}\right)^2 \right\rceil$$

One-sided T-Test (Summary)

- Hypothesis:

$$H_0: \ \mu = \mu_0 \quad \text{v.s.} \quad H_a: \ \mu > \mu_0$$

- Test statistic:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

- Rejection region:

$$\text{reject null if } T > t_{\alpha,n-1}$$

- P-value:

$$\text{p-value} = 1 - F_{t,n-1}(T)$$

- Power:

$$1 - \beta(\mu_1) = 1 - F_{t,n-1}\left(t_{\alpha/2,n-1} - \frac{\mu_1 - \mu_0}{s/\sqrt{n}}\right)$$

- Sample size determination:

$$n^* = \left\lceil \left(\frac{s(t_{\alpha,n-1} - t_{1-\beta^*,n-1})}{\mu_1 - \mu_0}\right)^2 \right\rceil$$

Example (Textbook 8.9)

The failure stree for 19 carbon nanofibers are measured in MPa. The sample data is summarized as follows:

$$\bar{x} = 562.68, \quad s = 180.874$$

The researchers want to test whether the average failure stress is greater than 500 MPa with $\alpha = 0.05$.

Hypothesis:

$$H_0 : \mu = 500 \quad \text{v.s.} \quad H_a : \mu > 500$$

Test statistic:

$$T = \frac{\bar{x} - 500}{s/\sqrt{n}} = 1.51$$

The p-value is $1 - F_{t,18}(1.51) = 0.074$.
We failed to reject the null hypothesis since the p-value is greater than $\alpha = 0.05$.

Two-sample T-Test

The T-tests we discussed so far are for one sample. We can also use T-test to compare two samples.

Suppose we have two independent random samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ from two normal populations with unknown means μ_1 and μ_2 and unknown standard deviations. We want to test

$$H_0: \ \mu_1 - \mu_2 = \Delta_0.$$

Variations:

- Alternative hypothesis.
- Equal or unequal variances.
- Paired or unpaired samples.

Welch's Two-sample T-Test

We first consider a popular two-sample T-test called Welch's T-test.

- Sampling: $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ are **independent random samples** from two normal populations with unknown means μ_1 and μ_2 and unknown standard deviations σ_1 and σ_2.
  That is X_i and Y_j are independent for all i and j.

- Hypothesis: (two-sided alternative)

$$H_0: \ \mu_1 - \mu_2 = \Delta_0 \quad \text{v.s.} \quad H_a: \ \mu_1 - \mu_2 \neq \Delta_0$$

- Variance assumption: we **do not** assume $\sigma_1 = \sigma_2$.

Welch's Two-sample T-Test

$$H_0: \ \mu_1 - \mu_2 = \Delta_0 \quad \text{v.s.} \quad H_a: \ \mu_1 - \mu_2 \neq \Delta_0$$

- A natural estimator of $\mu_1 - \mu_2$ is

$$\hat{\Delta} = \bar{X} - \bar{Y}.$$

- The standard error of $\bar{X} - \bar{Y}$ is

$$\sqrt{\mathrm{Var}(\bar{X} - \bar{Y})} = \sqrt{\mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- The standard error of $\bar{X} - \bar{Y}$ is estimated by the sample version:

$$S_{\hat{\Delta}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},$$

where S_1 and S_2 are the sample standard deviations.

Welch's Two-sample T-Test

Similar to the one-sample T-test, the test statistic is

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

The T-statistic, under $H_0 : \mu_1 - \mu_2 = \Delta_0$, follows a t-distribution with degrees of freedom ν given by the **Welch-Satterthwaite equation**:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$$

Remark: the degrees of freedom of the t-distribution can be a decimal number.

Welch's Two-sample T-Test

The confidence interval of $\mu_1 - \mu_2$ is

$$\bar{X} - \bar{Y} \pm t_{\alpha/2,\nu} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

The decision rule is reject null if Δ_0 is not in the confidence interval, or equivalently

$$\text{reject null if } |T| > t_{\alpha/2,\nu}.$$

The p-value of the two-sided Welch's T-test is

$$\text{p-value} = 2(1 - F_{t,\nu}(|T|)).$$

The power and sample size determination of Welch's T-test can be done similarly to the one-sample T-test.

Example (Textbook 9.6)

The void volumns within two types of fabrics are measured. The sample data is summarized as follows:

Fabric Type	Sample Size	Sample Mean	Sample Standard Deviation
Cotton	10	51.71	.79
Triacetate	10	136.14	3.59

We want to test whether the void volumns of the two types of fabrics differ by 80 with $\alpha = 0.05$.

The hypothesis:

$$H_0: \ \mu_1 - \mu_2 = -80 \quad \text{v.s.} \quad H_a: \ \mu_1 - \mu_2 \neq 80$$

Example (Textbook 9.6)

The two-sample T-statistic:

$$t = \frac{\bar{x} - \bar{y} - 80}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{51.71 - 136.14 + 80}{\sqrt{\frac{0.79^2}{10} + \frac{3.59^2}{10}}} = -3.811$$

The degrees of freedom:

$$\nu = \frac{\left(\frac{0.79^2}{10} + \frac{3.59^2}{10}\right)^2}{\frac{(0.79^2/10)^2}{9} + \frac{(3.59^2/10)^2}{9}} = 9.87$$

The quantile of the t-distribution is $t_{0.025,9.87} = 2.23$.
We reject the null hypothesis since $|t| = 3.811 > 2.23$.

Or we can compute the p-value as $2(1 - F_{t,9.87}(3.811)) = 0.0035 < 0.05$.
We reject the null hypothesis.

Example (Textbook 9.6)

We can also compute the confidence interval of $\mu_1 - \mu_2$:

$$\bar{x}-\bar{y}\pm t_{0.025,9.87}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 51.71-136.14\pm 2.23\sqrt{\frac{0.79^2}{10}+\frac{3.59^2}{10}} = (-87.02, -81.84).$$

The confidence interval does not contain -80.

So we reject the null hypothesis.

Summary of Welch's Two-sample T-Test

- Hypothesis:

$$H_0 : \ \mu_1 - \mu_2 = \Delta_0 \quad \text{v.s.} \quad H_a : \ \mu_1 - \mu_2 \neq \Delta_0$$

- Test statistic:

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

- Rejection region:

$$\text{reject null if } |T| > t_{\alpha/2,\nu} \quad \text{with } \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

- P-value:

$$\text{p-value} = 2(1 - F_{t,\nu}(|T|))$$

- Confidence interval for $\mu_1 - \mu_2$:

$$\bar{X} - \bar{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Extensions – One-sided

One-sided alternative: $H_a : \mu_1 - \mu_2 > \Delta_0$.

- The rejetion region should be modified to

$$\text{reject null if } T > t_{\alpha,\nu}.$$

- The p-value should be

$$\text{p-value} = 1 - F_{t,\nu}(T).$$

- The confidence interval method does not work for one-sided alternative.

Extensions – Pooled T-Test

Equal variances: we assume $\sigma_1 = \sigma_2$.

- In this case, we should use the pooled variance estimator:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

- The test statistic is modified to

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

- The degrees of freedom is $\nu = n_1 + n_2 - 2$.
- All other procedures are the same.

Paired T-Test

In some cases, we have paired samples. That is, we have two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ such that X_i and Y_i are paired.

The paired T-test is used to test the difference between the means of the two populations.

The test statistic is

$$T = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}},$$

where $D_i = X_i - Y_i$ and S_D is the sample standard deviation of the differences.

Remark: the paired T-test is equivalent to the one-sample T-test on the differences.

Paired T-Test

In a paired t-test, X_i and Y_i are paired, and therefore are usually dependent.

The sample variance of the differences is

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2.$$

Or equivalently,

$$S_D^2 = S_1^2 + S_2^2 - \rho S_1 S_2,$$

where ρ is the correlation coefficient between X_i and Y_i.

Paired T-Test

The other procedures of the paired T-test are similar to the one-sample T-test.

The degrees of freedom of the paired T-test is $n-1$.

The decision rule is

$$\text{reject null if } |T| > t_{\alpha/2,n-1}.$$

The p-value is

$$\text{p-value} = 2(1 - F_{t,n-1}(|T|)).$$

Summary of Paired T-Test

- Hypothesis:

$$H_0: \ \mu_D = \Delta_0 \quad \text{v.s.} \quad H_a: \ \mu_D \neq \Delta_0$$

- Test statistic:

$$T = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}}$$

- Standard error of the differences:

$$S_D = \sqrt{S_1^2 + S_2^2 - \rho S_1 S_2}$$

- Rejection region:

$$\text{reject null if } |T| > t_{\alpha/2, n-1}$$

- P-value:

$$\text{p-value} = 2(1 - F_{t,n-1}(|T|))$$

Exacly the same as the one-sample T-test on the differences.

Example (Textbook 9.9)

To test the neck-and-should disorder, 16 candidates are asked to elevate their arms below 30°. The time of arm elevation is recorded before and after the 18 months of work condition change.

Subject	1	2	3	4	5	6	7	8
Before	81	87	86	82	90	86	96	73
After	78	91	78	78	84	67	92	70
Difference	3	−4	8	4	6	19	4	3
Subject	9	10	11	12	13	14	15	16
Before	74	75	72	80	66	72	56	82
After	58	62	70	58	66	60	65	73
Difference	16	13	2	22	0	12	−9	9

We want to test whether the time of arm elevation is longer after the work condition change with $\alpha = 0.05$.

Example (Textbook 9.9)

We only need to work on the differences d_i. The hypothesis is

$$H_0: \ \mu_D = 0 \quad \text{v.s.} \quad H_a: \ \mu_D > 0.$$

The sample mean and standard deviation of the differences are

$$\bar{d} = 6.75, \quad s_d = 8.234.$$

The T-statistic is

$$t = \frac{6.75 - 0}{8.234/\sqrt{16}} = 3.28.$$

The p-value is

$$\text{p-value} = 1 - F_{t,15}(3.28) = 0.0025 < 0.05$$

We reject the null hypothesis.

Summary of T-tests

- T-tests are hypothesis tests based on the T-statistic:

$$T = \frac{\text{difference between the sample mean and the null}}{\text{estimated standard error of the numerator}}$$

- One-sample T-test: comparing the sample mean with a fixed value.
  - Two-sided, one-sided.
  - Decision based on the confidence interval, T-statistic, or p-value.
- Two-sample T-test: comparing the means of two populations.
  - Welch's T-test: unpaired sampling + unequal variances.
  - Pooled T-test: unpaired sampling + equal variances.
  - Paired T-test: paired sampling. (Same as one-sample T-test on the differences.)