MTL108
The t, χ^2 and F tests

Rahul Singh

t-Tests

The Student's t-test is used to compare population means when the true population variance (σ^2) is unknown and must be estimated from the sample data. Because estimating the variance introduces additional uncertainty, we cannot use the Standard Normal (Z) distribution. Instead, we use the t-distribution, which has heavier tails that account for this extra variability. As the sample size (n) increases, the t-distribution converges to the Standard Normal distribution.

There are three primary types of t-tests, each designed for a specific experimental framework.

The One-Sample t-Test

Purpose: To test whether the mean of a single population is equal to a known, specified baseline or theoretical value (μ_0).

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ with σ^2 unknown.

- **Hypotheses:** $H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0$ (Two-sided)
- **Test Statistic:** We standardize the sample mean $\bar{x}$ using the sample standard deviation s:

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- **Degrees of Freedom:** $df = n - 1$

Example 1 (Battery Lifespan)**.** A manufacturer claims their new batteries last exactly 50 hours ($\mu_0 = 50$). An auditor tests a random sample of $n = 16$ batteries and finds a sample mean of $\bar{x} = 48.2$ hours with a sample standard deviation of $s = 3.0$ hours. To test at $\alpha = 0.05$.

$$H_0 : \mu = \mu_0 = 50 \quad vs. \quad H_1 : \mu \neq \mu_0.$$

The test statistic is

$$T_{computed} = \frac{48.2 - 50}{3.0/\sqrt{16}} = \frac{-1.8}{3.0/4} = \frac{-1.8}{0.75} = -2.40$$

The critical value $t_{0.025,\,15}$, i.e., if T follows t-distribution with 15 df then

$$\mathbb{P}(|T| > t_{0.025,\,15}) = 0.05,$$

is obtained from table/calculator as 2.131. Since $|-2.40| > 2.131$, we reject H_0. The batteries do not meet the 50-hour claim.

The Independent Two-Sample t-Test

Purpose: To test whether the means of two entirely independent populations are equal.

Assuming Equal Variances

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent samples from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$.

- **Hypotheses:** $H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 \neq 0$
- **Pooled Variance:** Because we assume $\sigma_1^2 = \sigma_2^2$, we combine both sample variances into a single, more precise "pooled" estimator:

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

- **Test Statistic:**

$$T = \frac{(\bar{x} - \bar{y}) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- **Degrees of Freedom:** $df = n_1 + n_2 - 2$

Remark: If variances cannot be assumed equal, Welch's t-test must be used, which utilizes separate variances and a modified decimal degrees of freedom formula.

Example 2 (Teaching Methods)**.** A school tests two different math curricula A and B on two separate classes. Suppose, method A has mean score μ_1 and B has μ_2. We want to test $H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 \neq 0$ at $\alpha = 0.05$.
Based on an experiment we have following data

- **Method A:** $n_1 = 20, \bar{x} = 85, s_1^2 = 24$
- **Method B:** $n_2 = 20, \bar{y} = 81, s_2^2 = 28$

So, the pooled variance is

$$s_p^2 = \frac{19(24) + 19(28)}{20 + 20 - 2} = \frac{456 + 532}{38} = 26.$$

The test statistics is

$$T_{calculated} = \frac{85 - 81}{\sqrt{26}\sqrt{\frac{1}{20} + \frac{1}{20}}} = \frac{4}{\sqrt{26(0.1)}} = \frac{4}{\sqrt{2.6}} \approx \frac{4}{1.61} \approx 2.48$$

The critical value $t_{0.025,\,38}$, i.e., if T follows t-distribution with 38 df then

$$\mathbb{P}(|T| > t_{0.025,\,38}) = 0.05,$$

is obtained from table/calculator as 2.024. Since $|T_{calculated}| = 2.48 > 2.024$, we reject H_0.

Further, we want to test
$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 > \mu_2$ at $\alpha = 0.05$.
So, we need to obtain critical value $t_{0.05,\,38}$, i.e., if T follows t-distribution with 38 df then

$$\mathbb{P}(T > t_{0.05,\,38}) = 0.05.$$

From table/calculator we obtain $t_{0.05,\,38} = 1.686$. Here, for the alternative hypothesis $H_1 : \mu_1 > \mu_2$ rejection region is

$$C = \{T > t_{0.05,\,38} = 1.686\}.$$

Since

$$T_{calculated} = 2.48 > 1.686,$$

we reject the null hypothesis and conclude that Method A produces significantly higher scores.

The Paired t-Test

Purpose: To test the mean difference between two paired or dependent observations (e.g., measuring the exact same subjects "Before" and "After" an intervention).

Because the two samples are perfectly dependent, we cannot use the independent two-sample formula. Instead, we calculate the pairwise difference for each subject: $d_i = x_i - y_i$. We then treat these differences as a brand new, single dataset. This mathematically reduces the paired test to a standard One-Sample t-test on the differences.

- **Hypotheses:** $H_0 : \mu_d = 0 \quad \text{vs.} \quad H_1 : \mu_d \neq 0$ (where μ_d is the true mean difference).
- **Test Statistic:**
$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$
where $\bar{d}$ is the sample mean of the differences, s_d is the standard deviation of the differences, and n is the number of *pairs*.
- **Degrees of Freedom:** $df = n - 1$

Example 3 (Blood Pressure Medication)**.** A doctor measures the systolic blood pressure of $n = 9$ patients before and after taking a new drug. The calculated differences (Before - After) have a mean of $\bar{d} = 12$ mmHg and a standard deviation of $s_d = 4.5$ mmHg. We want to test $H_0 : \mu_d = 0 \quad \text{vs.} \quad H_1 : \mu_d \neq 0$.
The computed test statistic is

$$T_{calculated} = \frac{12 - 0}{4.5/\sqrt{9}} = \frac{12}{4.5/3} = \frac{12}{1.5} = 8.0.$$

At level of significance $\alpha = 0.05$, the critical value is $t_{0.025,\,8} = 2.306$, i.e., for $T \sim t_8$

$$\mathbb{P}(|T| > t_{0.025,\,8} = 2.306) = 0.05.$$

Since $T_{calculated} > 8.0 \gg 2.306$, we reject H_0.

Further one sided testing can be done to check whether the medication has a statistically significant effect on lowering blood pressure.

Testing Variance!

While much of classical statistics focuses on the mean (μ) as a measure of central tendency, the variance (σ^2) is equally critical. Variance measures volatility, inequality, consistency, and risk. In manufacturing, high variance means defective parts. In finance, high variance means extreme risk. In public health, high variance in outcomes indicates severe systemic inequality.

Therefore, we must be able to statistically test whether a population's variance meets a specific standard, or whether two different populations exhibit the same level of variability.

Testing a Single Variance (χ^2-Test)

Suppose we have a random sample $X_1, X_2, \ldots, X_n$ drawn from a Normal population $N(\mu, \sigma^2)$, where both the mean and variance are unknown. We wish to test if the true population variance σ^2 is equal to a specified baseline value, σ_0^2.

Hypotheses:

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1 : \sigma^2 \neq \sigma_0^2 \text{ (or } > \text{ or } < \text{)}$$

The Test Statistic:

We estimate the population variance using the unbiased sample variance,

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

By statistical theory, if we scale the sum of squared deviations by the true population variance, the resulting variable follows a Chi-Square (χ^2) distribution. Assuming the null hypothesis is true ($\sigma^2 = \sigma_0^2$), our test statistic is

$$\chi^2_{test} = \frac{(n-1)S^2}{\sigma_0^2}$$

Because we had to estimate the unknown mean μ using the sample mean $\bar{X}$, we lost one degree of freedom. Thus, under H_0, the test statistic follows a Chi-Square distribution with $n-1$ degrees of freedom,

$$\chi^2_{test} \sim \chi^2_{n-1}$$

Decision Rules

Unlike the Normal or t-distributions, the χ^2 distribution is strictly positive and right-skewed. Therefore, it is not symmetric, and we must find distinct critical values for the lower and upper tails.

- **Right-Tailed Test ($H_1 : \sigma^2 > \sigma_0^2$):** Reject H_0 if $\chi^2_{test} > \chi^2_{\alpha, n-1}$.

- **Left-Tailed Test ($H_1 : \sigma^2 < \sigma_0^2$):** Reject H_0 if $\chi^2_{test} < \chi^2_{1-\alpha,\, n-1}$.
- **Two-Sided Test ($H_1 : \sigma^2 \neq \sigma_0^2$):** Reject H_0 if $\chi^2_{test} > \chi^2_{\alpha/2,\, n-1}$ OR if $\chi^2_{test} < \chi^2_{1-\alpha/2,\, n-1}$.

Example 4 (NBA Scoring Consistency)**.** In sports analytics, a basketball player's value isn't just their average points per game, but their consistency. Suppose the league average variance for a starting point guard's scoring is $\sigma_0^2 = 25$ (meaning a standard deviation of 5 points). An analyst wants to test if a specific rookie is significantly more volatile (inconsistent) than the league baseline at the $\alpha = 0.05$ level.

- $H_0 : \sigma^2 \leq 25$ (Rookie is as consistent as or more consistent than average).
- $H_1 : \sigma^2 > 25$ (Rookie is highly volatile/inconsistent).

The analyst samples $n = 30$ games for the rookie and finds a sample variance of $S^2 = 38.2$.
Calculation:

$$\chi^2_{test} = \frac{(30-1)(38.2)}{25} = \frac{29 \times 38.2}{25} = 44.31$$

For df = 29 and $\alpha = 0.05$, the critical value from the Chi-Square table is $\chi^2_{0.05,29} = 42.557$.
Conclusion: Since $44.31 > 42.557$, we reject H_0.

Testing Equality of Two Variances (F-Test)

Suppose we have two independent random samples drawn from two Normal populations.

- Sample 1: $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ with sample variance S_1^2.
- Sample 2: $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ with sample variance S_2^2.

We wish to test whether the two populations share the same variance. This is often a crucial preliminary check before conducting a pooled two-sample t-test.

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

The Test Statistic:

If two variances are equal, their ratio should be exactly 1. We construct our test statistic by simply taking the ratio of the two sample variances,

$$F_{test} = \frac{S_1^2}{S_2^2}$$

By definition, an F-random variable is the ratio of two independent Chi-Square variables, each divided by its respective degrees of freedom. Mathematically:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

Under the null hypothesis ($\sigma_1^2 = \sigma_2^2$), the population variances cancel out perfectly, leaving our test statistic $F_{test} = S_1^2/S_2^2$. Therefore, under H_0, the statistic follows an F-distribution with numerator degrees of freedom $\nu_1 = n_1 - 1$ and denominator degrees of freedom $\nu_2 = n_2 - 1$.

The Practical Rule for the Two-Sided F-Test

The F-distribution is heavily skewed and asymmetrical. Finding lower-tail critical values ($F_{1-\alpha/2}$) can be annoying, as many textbook tables only provide upper-tail values. However, we can exploit a mathematical property of the F-distribution:

$$F_{1-\alpha,\nu_1,\nu_2} = \frac{1}{F_{\alpha,\nu_2,\nu_1}}$$

The Shortcut: To avoid dealing with the lower tail entirely when doing a two-sided test by hand, statisticians universally adopt a simple rule: *Always place the larger sample variance in the numerator.* By forcing $F_{test} \geq 1$, we guarantee the statistic will fall on the right side of the distribution. We then simply check if $F_{test} > F_{\alpha/2,\nu_1,\nu_2}$.

Example 5. A public health researcher is analyzing domestic health survey data to evaluate systemic health inequality. They extract data on women's hemoglobin levels (a proxy for anemia and nutritional health) from two different administrative districts. Even if the *average* hemoglobin level is identical, a higher variance in one district indicates severe health inequality (a mix of highly nourished and severely malnourished individuals).
The researcher wants to test if District A has a statistically different level of health inequality compared to District B at the $\alpha = 0.05$ level.

- $H_0 : \sigma_A^2 = \sigma_B^2$ (Health inequality is uniform across districts).
- $H_1 : \sigma_A^2 \neq \sigma_B^2$ (One district has significantly more inequality).

Sample Data:

- **District A:** $n_A = 41$ households, $S_A^2 = 18.5$
- **District B:** $n_B = 31$ households, $S_B^2 = 8.2$

Calculation: Following the practical shortcut, we place the larger variance (District A) in the numerator.

$$F_{test} = \frac{18.5}{8.2} = 2.256$$

The degrees of freedom are $\nu_1 = 41-1 = 40$ (numerator) and $\nu_2 = 31-1 = 30$ (denominator). Because this is a two-sided test with $\alpha = 0.05$, we look up the upper-tail critical value for $\alpha/2 = 0.025$. From the F-table: $F_{0.025,40,30} \approx 1.99$.
Conclusion: Since $F_{test} = 2.256 > 1.99$, we reject the null hypothesis. The researcher concludes with 95% confidence that the variance in hemoglobin levels is structurally unequal between the two districts. District A exhibits a statistically significant wider gap in nutritional health outcomes compared to District B.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

[2] Ross, Sheldon M. (2020). *Introduction to probability and statistics for engineers and scientists.* Academic press.

Disclaimer

This lecture note is prepared solely for teaching and academic purposes. Some parts of the material, including definitions, examples, and explanations, have been adapted or reproduced from the references. These notes are not intended for commercial distribution or publication, and all rights remain with the respective copyright holders.