9.5 The t-test

Tests for samples from a Normal population are one of the most common hypothesis procedures. Here, we focus on the case when both the mean μ and variance σ^2 of the population are unknown.

We could start by looking at the likelihood ratio to find an MLR statistic. But unlike the case when variance is known, this is a non-starter. Think about the hypotheses:

$$H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$$

What is parameter space? We actually omitted the parameter σ^2 from our hypotheses. So we can't actually use MLR to get an efficient test. What alternatives do we have?

Recall that if $\mathbf{X}$ is a sample from $N(\mu, \sigma^2)$, we can form a one-sided γ-level confidence interval for μ as

$$(A, \infty) \quad \text{where } A = \bar{X} - F_{n-1}^{-1}(\gamma)\frac{S}{\sqrt{n}}$$

Since this should give the range of plausible values for μ, we might reject H_0 in favor of H_1 if $\mu_0 < A$. This procedure indeed gives a test of size $\alpha_0 = 1 - \gamma$. Explicitly, we reject H_0 if

$$\mu_0 < \bar{X} - F_{n-1}^{-1}(\gamma)\frac{S}{\sqrt{n}} \quad \iff \quad \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} > F_{n-1}^{-1}(\gamma)$$

This is called the **t-test for a population mean**.

Let $T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$, which has $t(n-1)$ when $\mu = \mu_0$. Thus

$$P(T > F_{n-1}^{-1}(\gamma)|\mu = \mu_0) = P(F_{n-1}(T) > \gamma|\mu = \mu_0) = 1 - \gamma$$

by the Universality of the Uniform. Moreover, if $\mu < \mu_0$, then

$$T = \frac{X - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{X - \mu}{\frac{S}{\sqrt{n}}} - \frac{\mu_0 - \mu}{\frac{S}{\sqrt{n}}} = T^* - W \qquad T^* \sim t(n-1), W > 0$$

So

$$P(T > F_{n-1}^{-1}(\gamma)|\mu) = P(T^* > F_{n-1}^{-1}(\gamma) + W|\mu) < P(T^* > F_{n-1}^{-1}(\gamma)|\mu) = 1 - \gamma$$

Hence, this procedure is indeed a size $\alpha_0 = 1 - \gamma$ test.

Based on the preceding arguments, we can show that the power function $\pi(\mu, \sigma^2|\delta)$ for this procedure δ has the following properties:

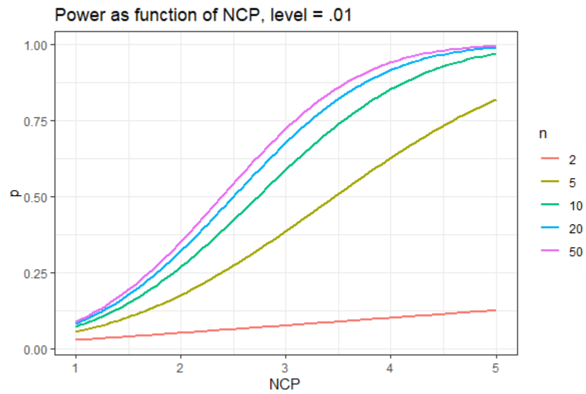
1. $\pi(\mu, \sigma^2|\delta) = \alpha_0$ when $\mu = \mu_0$.
2. $\pi(\mu, \sigma^2|\delta) < \alpha_0$ when $\mu < \mu_0$.
3. $\pi(\mu, \sigma^2|\delta) > \alpha_0$ when $\mu > \mu_0$.
4. $\pi(\mu, \sigma^2|\delta) \to 0$ when $\mu \to -\infty$.
5. $\pi(\mu, \sigma^2|\delta) \to 1$ when $\mu \to \infty$.

However, finding a particular formula for $\pi(\mu, \sigma^2|\delta)$ is more difficult. Note for example when $\mu < \mu_0$,

$$\pi(\mu, \sigma^2|\delta) = P(T > F_{n-1}^{-1}(\gamma)|\mu) = P(T^* > F_{n-1}^{-1}(\gamma) + W|\mu)$$

which involves random variables T^* and W on both sides of the inequality. That is, when $\mu \neq \mu_0$, the variable T is **not** t-distributed. Instead, it is said to have the **noncentral t-distribution** with noncentrality parameter ncp $= \frac{\mu - \mu_0}{\sigma/\sqrt{n}}$.

Note that R can return the PDF, CDF, QF, and random variates for non-central t by adding an `ncp =...` argument inside `dt`, `pt`, `qt`, `rt`.



Nevertheless, we can still calculate p-values of an observation using facts (1) and (2) above. For a test of

$$H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$$

suppose we observe the statistic $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$. Then we would reject H_0 at level α_0 iff

$$t \geq F_{n-1}^{-1}(1-\alpha_0) \iff F_{n-1}(t) \geq 1-\alpha_0 \iff \alpha_0 \geq 1 - F_{n-1}(t)$$

Hence, the p-value for t is $1 - F_{n-1}(t)$.

Two-sided Hypotheses

Suppose instead we consider the two-sided alternative hypothesis:

$$H_0 = \mu = \mu_0 \quad H_1 \mu \neq \mu_0$$

We can derive a α_0 size test from $\gamma = 1 - \alpha_0$ confidence interval, just as we did for the 1-sided test.

In this case, the t-test takes the form. Reject H_0 if

$$|T| \geq F_{n-1}^{-1}(1-\alpha_0/2) \quad \text{where} \quad t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$$

<u>Ex 1:</u> A batch of stout beer is best when it has an original gravity (OG) close to 1.071. Suppose the OG of beer is $N(\mu, \sigma^2)$. We sample 5 OG measurements from a batch of beer and find $\bar{x} = 1.0686$ and $s = 0.0064$. We perform a 2-sided hypothesis test of $H_0 : \mu = 1.071$ vs. $H_1 : \mu \neq 1.071$. Our t-statistic is then

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{1.0686 - 1.071}{0.0064\sqrt{5}} = -0.84$$

At the $\alpha_0 = 0.05$ level, $F_{n-1}(1-\alpha_0/2) =$ `qt(.975, 4)` $= 2.78$. Since $|t| < 2.78$, we do not reject H_0. Moreover, our statistic had p-value of

$$p - \text{value} \; = 2P(T \geq |t|) = \texttt{2*(1-pt(.84, 4))} = 0.448$$

With all that said, if $\mu = 1.0686$ and $s = 0.0064$, then $\text{nc} = \frac{1.0686-1.071}{0.0064/\sqrt{5}}$ and so

$$\begin{aligned} P(|T| > 2.78) =& 1 - P(T < 2.78) + P(T < -2.78) \\ =& \texttt{1-pt(2.78, ncp = nc, df = 4)+pt(-2.78, ncp = nc, df = 4)} \\ =& 0.1 \end{aligned}$$

which shows that this test may not have particularly high power if the true parameter values are close to those actually observed.

9.6 Tests for Difference in Population Mean

Suppose we are interested in determining whether two the means of two distinct populations are equal. If we model the two populations using Normal distributions with common (but unknown) variance, we can modify the t-test in the previous section, we can create a hypothesis test for assessing whether the means of the two Normal distributions are equal.

We assume that $\mathbf{X} = (X_1, \dots, X_{n_x})$ is a random sample from $N(\mu_x, \sigma^2)$ and that $\mathbf{Y} = (Y_1, \dots, Y_{n_y})$ is a random sample from $N(\mu_y, \sigma^2)$. *Note that this model has 3 unknown parameters:* μ_x, μ_y, σ^2. We will first test the hypotheses

$$H_0 : \mu_x \leq \mu_y \qquad H_1 : \mu_x > \mu_y$$

We need to construct a test statistic which can be used to assess they hypotheses. Since $\bar{X}$ and $\bar{Y}$ are MLEs for μ_x and μ_y, it makes sense to incorporate them in our test statistic. Moreover, observing a large positive difference in $\bar{X} - \bar{Y}$ is unlikely, if H_0 is true. Of course, what values count as "large" depends on the standard deviation of $\bar{X} - \bar{Y}$, so it would be helpful to scale by this as well. This leads us to defining the following test statistic T:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)\left(\frac{S_X^2 + S_Y^2}{n_x + n_y - 2}\right)}} \qquad \text{where} \qquad S_x^2 = \sum (X_i - \bar{X})^2 \qquad S_y^2 = \sum (Y_i - \bar{Y})^2$$

Thm: Assuming $\mathbf{X}, \mathbf{Y}$ are independent samples from $N(\mu_x, \sigma^2)$ and $N(\mu_y, \sigma^2)$ of sizes n_x and n_y respectively, and T is defined as above, then T has the t-distribution with $n_x + n_y - 2$ degrees of freedom, when $mu_x = \mu_y$.

Proof. We decompose T into two parts:

$$T = \frac{Z}{\sqrt{\frac{W}{m+n-2}}}$$

where

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}\right)}} \qquad W = \frac{S_X^2 + S_Y^2}{\sigma^2}$$

Since $\bar{X}$ and $\bar{Y}$ are independent $N(\mu_x, \sigma^2/n_x)$ and $N(\mu_y, \sigma^2/n_y)$, respectively, then $\bar{X} - \bar{Y}$ is $N(\mu_x - \mu_y, \sigma^2/n_x + \sigma^2/n_y)$. Moreover, W has the χ^2 distribution with $n_x + n_y - 2$ degrees of freedom, since it is a sum of n_x and n_y squared Normals centered at their sample means. Finally, we note that $\bar{X}, \bar{Y}, S_X^2, S_Y^2$ are all mutually independent. Hence T is t-distributed, as desired. □

Thm: Consider tests δ_c of

$$H_0 : \mu_x \leq \mu_y \qquad H_1 : \mu_x > \mu_y$$

which reject when $T \geq c$. The p-value of an observed sample $\mathbf{X}, \mathbf{Y}$ with $T = t$ is

$$p\text{-value} = P(T \geq t | \mu_x = \mu_y) = 1 - F_{n_x + n_y - 2}(t)$$

where $F_{n_x+n_y-2}$ is the CDF of $t(n_x + n_y - 2)$.

For each value of μ_x, μ_y, σ^2, we can compute the power function for the procedure δ_c using a non-central t-distribution.

Thm: For μ_x, μ_y, σ^2, the statistic T defined above has the non-central t distribution with $n_x + n_y - 2$ degrees of the freedom, and noncentrality parameter:

$$\text{ncp} = \frac{\mu_x - \mu_y}{\sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}}$$

Moreover, if δ_c has size α_0, then the power function $\pi(\mu_x, \mu_y, \sigma^2 | \delta_c)$ has the properties:

1. $\pi(\mu_x, \mu_y, \sigma^2 | \delta_c) = \alpha_0$ when $\mu_x = \mu_y$.

2. $\pi(\mu_x, \mu_y, \sigma^2|\delta_c) < \alpha_0$ when $\mu_x < \mu_y$.
3. $\pi(\mu_x, \mu_y, \sigma^2|\delta_c) > \alpha_0$ when $\mu_x > \mu_y$.

While this t test has nice theoretical properties, it does have some significant drawbacks. The model assumptions that $\mathbf{X}$ and $\mathbf{Y}$ come from Normal populations with equal variance is often inaccurate. And while deviations from Normality do not substantially change the distribution of T when the samples are of large size, deviations from the equal variance assumption can have substantial impact on distribution of T (even when samples are large). As a result, an alternative test can be conducted with the following statistic:

$$V = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n_x}\frac{S_X^2}{n_x-1} + \frac{1}{n_x}\frac{S_X^2}{n_x-1}}}$$

The statistic inside the square-root of the denominator:

$$W = \frac{1}{n_x}\frac{S_X^2}{n_x - 1} + \frac{1}{n_x}\frac{S_X^2}{n_x - 1}$$

is **approximately** Gamma distributed, when n_x, n_y are moderate or large, and so the ratio V is approximately t distributed with degrees of freedom

$$\frac{\left(\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_x}\right)}{\frac{1}{n_x-1}\left(\frac{\sigma_x^2}{n_x}\right) + \frac{1}{n_y-1}\left(\frac{\sigma_y^2}{n_y}\right)}$$

In general, tests of equal means using this statistic are more robust to deviations from the equal variance assumption (even taking into account the approximation).

9.7 The F Distribution

Suppose we are in the situation of the previous section, and are interested in performing a test to assess whether two Normal populations have equal means. In order to use the t-procedure, we need to assume that both populations have equal variance. But how can we assess whether such a claim is valid, based on data? With another hypothesis test, of course!

Before proceeding, we must introduce another named distribution to our repertoire.

<u>**Def:**</u> Let Y and W be independent random variables where $Y \sim \chi^2(m)$ and $W \sim \chi^2(n)$. Let X be the random variable defined as

$$X = \frac{\frac{Y}{m}}{\frac{W}{n}} = \frac{nY}{mW}$$

Then X is said to have the F distribution with (m, n) degrees of freedom.

It is possible to compute the density of an F-distribution directly using the change-of-variables formula (although we will have limited use for the precise formula). This density is included here just for completeness:

<u>**Thm:**</u> Let X have the F distribution with (m, n) degrees of freedom. Then the PDF for X is

$$f(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{x^{m/2-1}}{(mx+n)^{m+n}/2} \quad x \geq 0$$

A few properties of the F distribution are summarized below:

<u>**Thm:**</u> If X has F with (m, n) degrees of freedom, then $1/X$ has the F distribution with (n, m) degrees of freedom. If Y has the t distribution with n degrees of freedom, then Y^2 has the F distribution with $(1, n)$ degrees of freedom.

The pdf, cdf, and quantile function for an F distribution can be accessed in R using the `df`, `pf`, `qf` functions.

The F Test

Suppose $\mathbf{X}$ is a sample of size n_x from $N(\mu_x, \sigma_x^2)$ and $\mathbf{Y}$ is a sample of size n_y from $N(\mu_y, \sigma_y^2)$, and that we are interested in testing the claims:

$$H_0 : \sigma_x^2 \leq \sigma_y^2 \qquad H_1 : \sigma_x^2 > \sigma_y^2$$

Define a test statistic V by

$$V = \frac{S_X^2/(n_x - 1)}{S_Y^2/(n_y - 1)}$$

and supposed δ_c are a collection of tests that reject H_0 when $V \geq c$.

Thm: Let V be defined as above. The variable $\frac{\sigma_y^2}{\sigma_x^2}V$ has the F distribution with n_x and n_y degrees of Freedom. If $\sigma_x^2 = \sigma_y^2$, then V itself has this F distribution.

Proof. We previously showed that S_X^2/σ_x^2 and S_Y^2/σ_y^2 have the χ^2 distribution, with $n_x - 1$ and $n_y - 1$ degrees of freedom, respectively. By assumption, both random variables are derived from independent samples, and so are themselves independent. Then the following variable

$$\frac{S_X^2/[(n_x - 1)\sigma_x^2]}{S_Y^2/[(n_y - 1)\sigma_y^2]} = \frac{\sigma_y^2}{\sigma_x^2}\frac{S_X^2/(n_x - 1)}{S_Y^2/[n_y - 1)}$$

has the $F(n_x - 1, n_y - 1)$ distribution, by definition. □

Thm: Let V be the statistic defined above, and let c be the $1 - \alpha$ quantile of the $F(n_x - 1, n_y - 1)$ distribution, and let G_{n_x-1,n_y-1} be the CDF of this distribution. Consider the test δ_α which rejects H_0 when $V \geq c$. Then the power function for this procedure satisfied:

1. $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2|\delta_c) = 1 - G_{n_x-1,n_y-1}\left(\frac{\sigma_y^2}{\sigma_x^2}c\right)$
2. $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2|\delta_c) = \alpha_0$ when $\sigma_x^2 = \sigma_y^2$.
3. $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2|\delta_c) < \alpha_0$ when $\sigma_x^2 < \sigma_y^2$.
4. $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2|\delta_c) > \alpha_0$ when $\sigma_x^2 > \sigma_y^2$.

Moreover, the test δ_α is has size α and the p-value when $V = v$ is observed is $1 - G_{n_x-1,n_y-1}(v)$.

We can also construct a two-sided F-test of the following hypotheses:

$$H_0 : \sigma_x^2 = \sigma_y^2 \qquad H_1 : \sigma_x^2 \neq \sigma_y^2$$

In this case, for $\alpha > 0$, consider the procedure δ_α which rejects H_0 if either $V \leq c_1$ or $V \geq c_2$, where c_1, c_2 are the $\alpha/2$ and $1 - \alpha/2$ quantiles of the $F(n_x - 1, n_y - 1)$ distribution.

It turns out that this two-sided test is biased (i.e. the minimum value of the power function on the alternative set is smaller than the maximum value on the null set). Additionally, we cannot express the rejection region as $T \geq c$, for some single statistic T. For this reason, the classic 209/310 definition of p-value does not work (but our modified p-value definition does work).

Thm: Let V be defined as above, and consider a collection of tests δ_α defined above. Then the p-value when $V = v$ is observed is

$$2\min\left\{1 - G_{n_x-1,n_y-1}(v), G_{n_x-1,n_y-1}(v)\right\}$$

Proof. Homework. □