

## One-Sided Testing

In a two-sided test, we hypothesize that  $H_0 : \mu = \mu_0$  and look for evidence that it's wrong; such evidence would be a  $t$ -statistic too big in either the positive or negative direction, expressed as  $H_1 : \mu \neq \mu_0$ .

When we do a one-sided test, we are only concerned with whether the true mean is either below or above our guess, but not both. For instance, suppose we think that  $\mu$  is greater than  $\mu_0$  and we want to test this guess. The claim being tested becomes the *alternative* hypothesis. So we test, say at 5% significance,

$$H_0 : \mu \leq \mu_0,$$

$$H_1 : \mu > \mu_0.$$

We again assume that the null is true. We reject the null if we find strong enough evidence against the null, in favor of the alternative. Based on the specification, that evidence would be seen as a value of  $\bar{x}$  that is “far enough” above  $\mu_0$ , in other words, if  $\bar{x} - \mu_0$  is very positive.

We quantify “far enough” by again using the  $t$ -statistic,

$$t \equiv \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim T(n-1).$$

But again, we only reject the null if  $t$  is too far *positive*, and hence we only look at the right-tail of the distribution. Hence we put all 5% of the rejection region into the right-tail. Thus our critical value is  $t_{n-1,0.05} = \text{qt}(1-0.05, n-1)$  in R. We reject the null hypothesis if  $t > t_{n-1,0.05}$ . In other words, the rejection region is  $(t_{n-1,0.05}, \infty)$ .

If instead we think that  $\mu$  is less than  $\mu_0$ , the test becomes

$$H_0 : \mu \geq \mu_0,$$

$$H_1 : \mu < \mu_0.$$

In this setup, evidence against the null is when  $\bar{x}$  is “far enough” below  $\mu_0$ . Thus, if the  $t$ -statistic is too far *negative*, then we reject the null. This means we are only considering the left-tail of the distribution, in which we put all 5% of the test significance. The critical value is therefore  $-t_{n-1,0.05} = \text{qt}(0.05, n-1)$  in R. We reject the null hypothesis if  $t < -t_{n-1,0.05}$ . In other words, the rejection region is  $(-\infty, -t_{n-1,0.05})$ .

**Rule of Thumb:** Put the claim that has the equality as the null hypothesis.

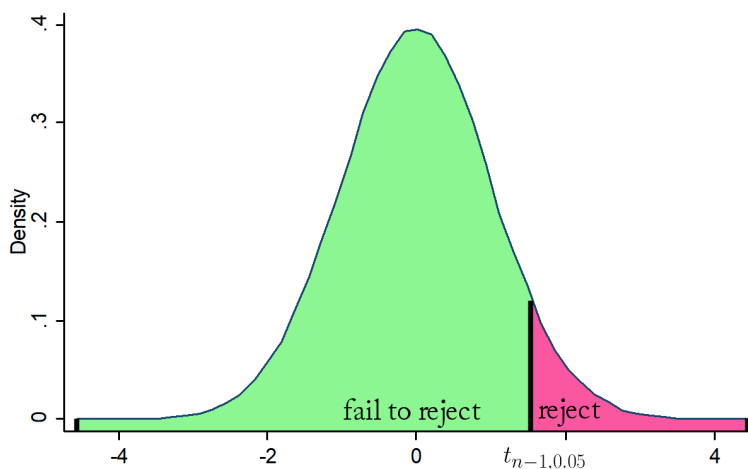


FIGURE 1:  $t_{n-1,0.05}$  is the number such that 5% of the mass of the  $T(n-1)$  distribution falls above it. If  $H_0 : \mu \leq \mu_0$  is true, then it is unlikely that our test statistic will fall above  $t_{n-1,0.05}$ , in which case we reject the null.

## Difference in Means Testing

Suppose we are interested in two groups and how their means,  $\mu_1$  and  $\mu_2$ , differ. We calculate sample means  $\bar{x}_1$  and  $\bar{x}_2$  as well as sample variances  $s_1^2$  and  $s_2^2$ . We hypothesize that the difference in means is  $\mu_0$ ; in practice we will often hypothesize that the difference is  $\mu_0 = 0$ . Thus our null hypothesis is  $H_0 : \mu_1 - \mu_2 = \mu_0$ . Hence we test

$$H_0 : \mu_1 - \mu_2 = \mu_0,$$

$$H_1 : \mu_1 - \mu_2 \neq \mu_0.$$

Suppose group 1 has sample size  $n_1$  and group 2 has sample size  $n_2$ , not necessarily equal. The test statistic is

$$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim T(n_1 + n_2 - 2).$$

The reason we subtract 2 for degrees of freedom is because we are testing with respect to two variables,  $\mu_1$  and  $\mu_2$ . From here, the testing procedure proceeds in the usual way.

## One Proportion Testing

For individual  $i$ , let  $x_i = 1$  for a “successful” event and  $x_i = 0$  for a “failure” event. For example, earning a degree might be the successful event, dropping out would therefore be the

failure event. The sample proportion of individuals who succeeded is the typical mean, now denoted  $p \equiv (\sum_{i=1}^n x_i)/n$ . Think of  $p$  as being an estimate of the true population proportion of successes,  $\pi$ .

Because there are only two possibilities for  $x_i$ , we have to use special techniques and formulas. In particular, the standard error of estimate  $p$  is given by

$$\text{se}(p) = \sqrt{\frac{p(1-p)}{n}}.$$

Furthermore, sample sizes in proportions analysis are typically large. Large enough, in fact, that the standard normal distribution is typically used instead of  $T(n-1)$ . Thus we do not use a  $t$ -statistic but instead the  $z$ -statistic given by

$$z \equiv \frac{p - p^*}{\sqrt{p^*(1-p^*)/n}} \sim \mathcal{N}(0, 1),$$

where  $p^*$  is our hypothesized value for the true proportion of successful events.

A two-sided proportion test would be of the form

$$H_0 : \pi = p^*,$$

$$H_1 : \pi \neq p^*.$$

At 5% significance, we reject the null hypothesis when  $|z| > z_{0.025}$ , where  $z_{0.025} = \text{qnorm}(1-0.025)$  in R or is otherwise found on a normal table.

## Two Proportions Testing

Suppose we have two different population proportions,  $\pi_1$  and  $\pi_2$ . We want to see whether the proportions are the same or not. We sample  $n_1$  times for group 1 and find  $y_1$  successes; we sample  $n_2$  times for group 2 and find  $y_2$  successes. Hence we find estimates

$$p_1 = \frac{y_1}{n_1}, \quad p_2 = \frac{y_2}{n_2},$$

and the total proportion of successes is

$$p^* \equiv \frac{y_1 + y_2}{n_1 + n_2}.$$

Our test is of the form

$$H_0 : \pi_1 - \pi_2 = 0,$$

$$H_1 : \pi_1 - \pi_2 \neq 0,$$

and we use test statistic

$$z \equiv \frac{p_1 - p_2}{\sqrt{p^*(1-p^*)\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{N}(0, 1).$$

## Chi-Squared Distribution

A **chi-squared** random variable, denoted  $\chi^2$ , is a sum of squared standard normal random variables. Many test statistics have chi-squared distribution, so we need to know about it. It has one parameter, the degrees of freedom  $k$ , and as such it is usually denoted  $\chi^2(k)$ . On quizzes and exams, we'll use a chi-squared table (which is available on Canvas).

Suppose we have 20 degrees of freedom. We want to know critical value such that 5% of the area underneath the chi-squared curve lies to the right of it. Then we go to the table, look at the row with 20 degrees of freedom and the column with 0.05, which gives 31.410. Express this number as  $\chi^2_{20,0.05} = \text{qchisq}(1-0.05, 20)$  in R.

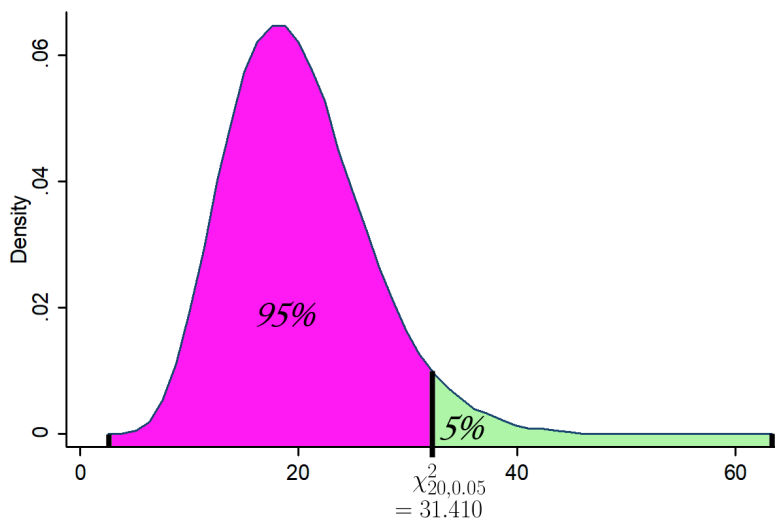


FIGURE 2:  $\chi^2_{20,0.05}$  is the number such that 5% of the mass of  $\chi^2(20)$  distribution falls above it.

## Variance Testing

We usually do not know the true population variance  $\sigma^2$ , so we have to estimate it with

$$s^2 \equiv \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

But this is an estimation using a sample, and hence it has some uncertainty to it. We seek to quantify that uncertainty. For a two-sided test, we perform the test

$$H_0 : \sigma^2 = \sigma_0^2,$$

$$H_1 : \sigma^2 \neq \sigma_0^2,$$

where  $\sigma_0^2$  is the hypothesized value for  $\sigma^2$ . The relevant test statistic is

$$\chi^2 \equiv \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1),$$

where the distribution is valid if either the population is normally distributed or if  $n > 30$ . We reject the null hypothesis if  $\chi^2$  is in the rejection region, which we now define.

Suppose we are testing at the 10% significance level. As usual, we chop the significance level in half for each tail. Problem is, the  $\chi^2(n-1)$  distribution is not symmetric. Thus we must calculate two critical values to determine the rejection region. The values can be found using the  $\chi^2$  table on Canvas. For instance, suppose that  $n = 10$ . Then we have  $n - 1 = 9$  degrees of freedom. We want to find

$\chi_{9,0.05}^2$  = the number such that 0.05 of the area is to the right of it,

$\chi_{9,0.95}^2$  = the number such that 0.95 of the area is to the right of it.

These two numbers are visualized below.

So using the table, go to the 9th row, since that corresponds to 9 degrees of freedom. Then look at the columns for 0.95 and 0.05. Those are the critical values we seek.

Degrees of Freedom	Chi-Square ( $\chi^2$ ) Distribution							
	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578

FIGURE 3:  $\chi_{9,0.05}^2 = 16.919$  is the number such that 5% of the mass of  $\chi^2(9)$  distribution falls above it, and  $\chi_{9,0.95}^2 = 3.325$  is the number such that 95% of the mass of  $\chi^2(9)$  distribution falls above it.

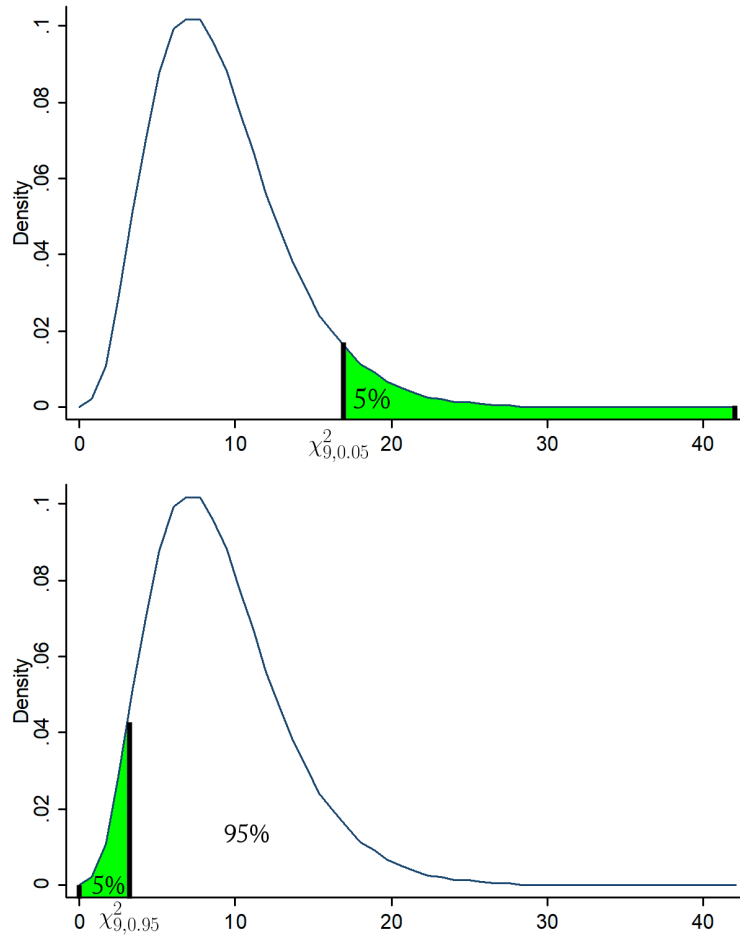


FIGURE 4:  $\chi_{9,0.05}^2$  (top) is the number such that 5% of the mass of the  $\chi^2(9)$  distribution falls above it, and  $\chi_{9,0.95}^2$  (bottom) is the number such that 95% of the mass of the  $\chi^2(9)$  distribution falls above it. Any  $\chi^2$  statistic in the green regions warrants rejecting the null.

These values can be found in R using  $\chi^2_{9,0.05} = \text{qchisq}(0.05, 9, \text{lower.tail} = \text{FALSE})$  and  $\chi^2_{9,0.95} = \text{qchisq}(0.95, 9, \text{lower.tail} = \text{FALSE})$ .

## F-Distribution

The **F-distribution** has two different arguments for two different degrees of freedom, so we denote it  $F(v_1, v_2)$ . What exactly  $v_1$  and  $v_2$  are will become clear once we do regression analysis. Since we have to specify two degrees of freedom, it's difficult to condense a comprehensive  $F$ -distribution onto a single page. Hence we use an  $F$  table that contains only numbers for a 5% right-tail (therefore making it applicable to a one-sided test at 5% significance or a two-sided test with 10% significance, since in the latter we split the 10% into both tails). The table is currently available on Canvas.

Suppose  $v_1 = 3$  and  $v_2 = 15$ , and we want to find the critical value of  $F(3, 15)$  distribution such that 5% of the data falls to the right of it. Then we look at the row corresponding to  $v_2 = 15$  degrees of freedom, and the column corresponding to  $v_1 = 3$  degrees of freedom, which gives 3.287. Express this number as  $F_{3,15,0.05} = \text{qf}(1-0.05, 3, 15)$  in R.

## Difference in Variations Testing

Suppose we have two groups. Group 1 has sample size  $n$  and group 2 has sample size  $m$ . We calculate two different sample variances for each group,  $s_1^2$  and  $s_2^2$ . We want to test if one has true population variance greater than the other. Suppose that the samples give  $s_1^2 > s_2^2$ . Then let us test  $H_0 : \sigma_1^2 \leq \sigma_2^2$  against  $H_1 : \sigma_1^2 > \sigma_2^2$  at 5% significance.

Since our  $F$ -table only gives us right tails, we want to formulate the question in such a way that the test statistic leads to rejection if it is too far in the right tail. So reformulate the test as

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} < 1,$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} \geq 1,$$

where we use test statistic

$$F \equiv \frac{s_1^2}{s_2^2} \sim F(n-1, m-1).$$

Thus, if we have evidence in favor of the alternative, then  $F$  will be greater than 1. If  $F$  is sufficiently greater than 1, then we reject the null.

**Rule of Thumb:** Put the group with the larger sample variance in the numerator. This will ensure that we do a right-tailed test at 5% significance, which is all our  $F$ -table allows.

The critical value we use for the rejection rule is found by looking at the  $F$ -table. The numerator degrees of freedom is  $v_1 = n - 1$ , which is found on the columns; and the denominator degrees of freedom is  $v_2 = m - 1$ , which is found on the rows.

## Testing Correlation Coefficient

We calculate the sample correlation coefficient of  $y$  and  $x$  with  $n$  samples, denoted  $r_{xy}$ . We test that the correlation is significant, i.e. nonzero, by specifying

$$H_0 : \rho = 0,$$

$$H_1 : \rho \neq 0$$

and using test statistic

$$t \equiv \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \sim T(n-2).$$

## Errors in Conclusion

Since we are never 100% confident in our conclusions, it is possible that we reject a null hypothesis even when it is true; and also possible that we fail to reject a null hypothesis even when it is false. We employ the following terminology to discuss such scenarios.

- *Type I Error:* Rejection of a true null hypothesis (false positive).
- *Type II Error:* Failing to reject a false null hypothesis (false negative)

The *size* of a test is the probability of mistakenly rejecting a true null. The *power* of a test is the probability of correctly rejecting a false null. A test is said to have significance level  $\alpha$  if its size is less than or equal to  $\alpha$ . In many cases (and all of our cases), the size and level of a significance level of a test are equal.