STAT 515 fa 2023 Lec 15

p-values

Karl Gregory

The p-value

So far we have tested hypotheses in the following two steps:

1. By computing a *test statistic*, for example the quantity

$$Z_{\text{test}} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \quad \text{or} \quad T_{\text{test}} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$

   for hypotheses concerning a mean μ or the quantity

$$Z_{\text{test}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

   for hypotheses concerning a proportion p.

2. Then choosing to reject or not to reject H_0 at significance level α according to whether the test statistic falls in the *rejection region* defined by the *critical value* obtained from the z- or the t-table.

Based on the comparison of our test statistic with the critical value, our conclusion will be to reject H_0 at significance level α or not to reject H_0 at significance level α. For example, if $\alpha = 0.05$, our conclusion will to reject H_0 at significance level 0.05 or not to reject H_0 at significance level 0.05.

If this is all that we report to our audience, it may leave them wondering what our decision would have been for a smaller α. Would we also have rejected H_0 at the $\alpha = 0.01$ significance level? What about at the $\alpha = 0.001$ significance level? Remember that when we choose a smaller α, we require stronger evidence against H_0 in order to reject it. For $\alpha = 0.01$, we need very strong evidence against H_0 in order to reject it, but for $\alpha = 0.10$ we do not need very strong evidence against H_0 in order to reject it.

In order to express exactly how much evidence the data carries against H_0, it is customary to report *the smallest significance level α at which the data lead us to reject H_0*. This is precisely what we refer to as the *p-value*. Reporting the p-value is more informative than saying simply whether H_0 is rejected at one particular value of α. For example, if we report a p-value of 0.034, it communicates that we would reject H_0 at every significance level α which is greater than 0.034 and fail to reject H_0 at every significance level α which is less than or equal to 0.034. The value 0.034 thus conveys exactly how strong the evidence is against H_0.

Definition of the p-value

We here give the formal definition of the p-value. The two definitions are equivalent.

Definition 1: p-value

The p-value of a random sample with respect to a null hypothesis H_0 is

1. the smallest significance level at which the sample leads to a rejection of H_0.

2. the (maximum) probability, assuming that H_0 is true, of drawing a sample carrying as much or more evidence against H_0 than the sample which was drawn.

Intuitively, we can think of the p-value as a *measure of the plausibility of H_0 in light of the data.* Smaller p-values cast greater doubt on H_0; larger p-values cast little doubt on H_0.

How to use the p-value

We can use the p-value to decide whether to reject H_0 at a given significance level α. We simply use the following rule:

$$\text{Reject } H_0 \text{ if } p\text{-value} < \alpha$$
$$\text{Fail to reject } H_0 \text{ if } p\text{-value} \geq \alpha$$

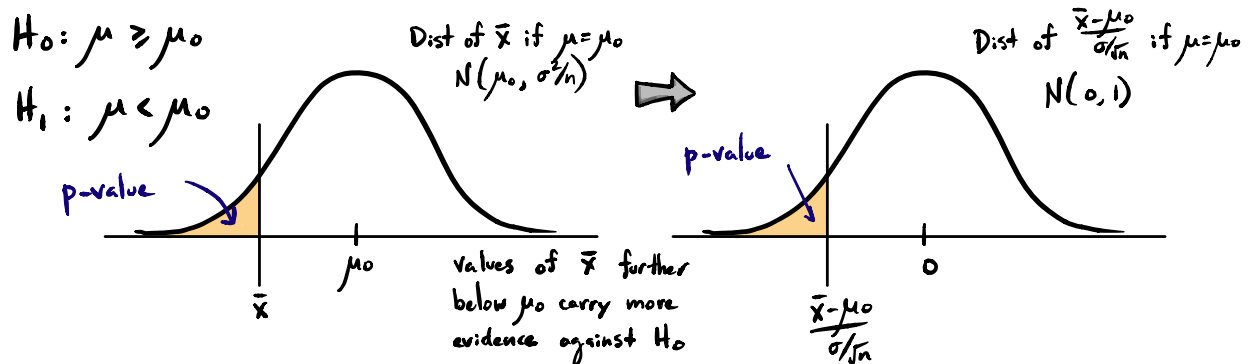This will lead to the same decisions as the decision rules previously discussed.

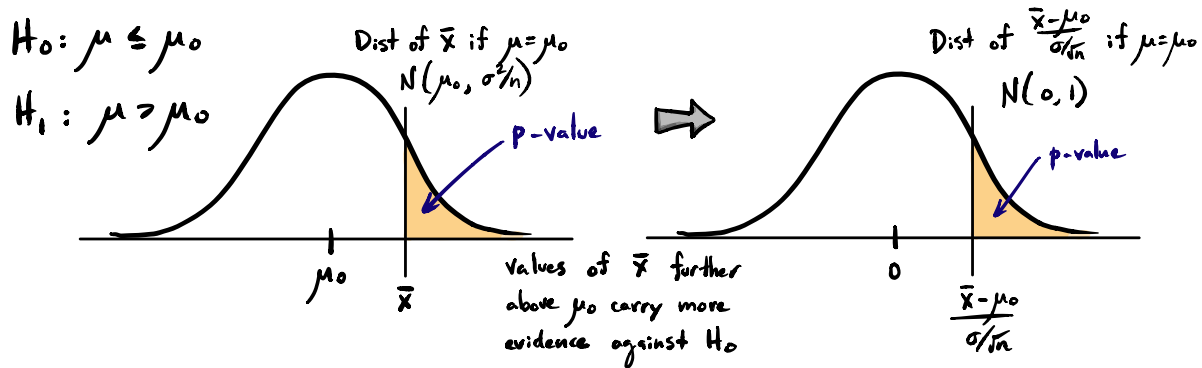## Computing the p-value for hypotheses about μ (σ known)

Using the second definition of the p-value, which says that the p-value is the probability, under H_0, of getting data carrying more evidence against H_0 than the observed data, we can

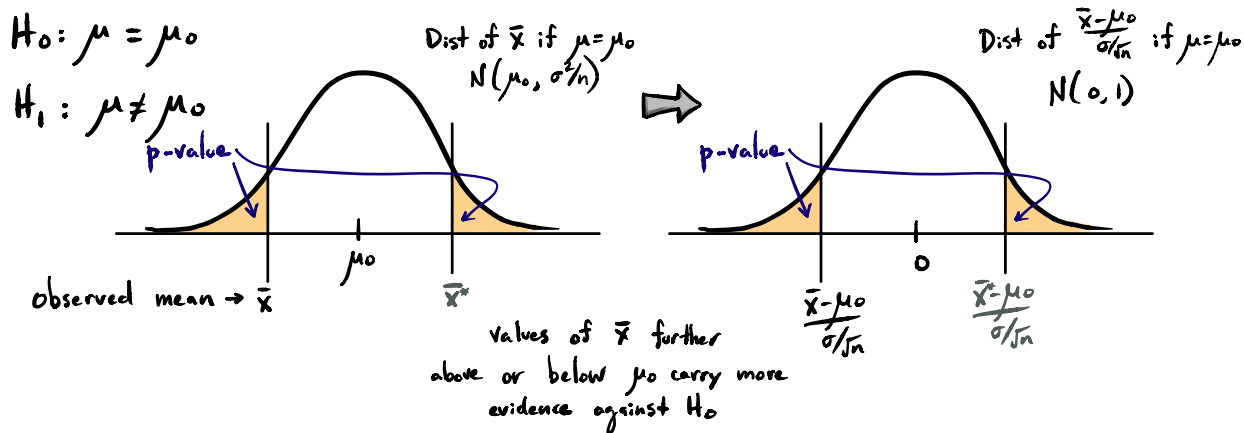come up with the following depictions of the p-values for the sets of hypotheses

$$H_0:\ \mu \geq \mu_0 \qquad H_0:\ \mu = \mu_0 \qquad H_0:\ \mu \leq \mu_0$$
$$H_1:\ \mu < \mu_0 \qquad H_1:\ \mu \neq \mu_0 \qquad H_1:\ \mu > \mu_0.$$

If we may assume that the distribution of $\bar{X}_n$ is Normal, the p-values for the one-sided tests admit the depictions

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

Dist of $\bar{x}$ if $\mu = \mu_0$
$N(\mu_0, \sigma^2/n)$
p-value
μ_0
$\bar{x}$
Values of $\bar{x}$ further above μ_0 carry more evidence against H_0

Dist of $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ if $\mu = \mu_0$
$N(0,1)$
p-value
0
$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

$H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$

Dist of $\bar{x}$ if $\mu = \mu_0$
$N(\mu_0, \sigma^2/n)$
p-value
μ_0
$\bar{x}$
Values of $\bar{x}$ further below μ_0 carry more evidence against H_0

Dist of $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ if $\mu = \mu_0$
$N(0,1)$
p-value
0
$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

In the one-sided case we see that the probability of getting data more in favor of the alternative than the sample at hand is given by the area under the curve beyond $\bar{X}_n$ in the direction of the alternate hypothesis. This area can be found by taking $\bar{X}_n$ into the Z world and looking up the area in the z-table. Note that taking $\bar{X}_n$ into the Z world is the same as computing the test statistic; so the p-value is simply the area under the standard Normal pdf beyond the test statistic in the direction of the alternate hypothesis.

In the two-sided case, our depiction of the p-value changes due to the fact that we will reject H_0 for values of $\bar{X}_n$ which are far away from μ_0 in either direction. That is, as $\bar{X}_n$ moves further from μ_0 to the left or to the right, it carries greater evidence against H_0. If we may assume that the distribution of $\bar{X}_n$ is Normal, the p-value for the two-sided test admits the depiction

3

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

p-value

observed mean → $\bar{x}$

μ_0

$\bar{x}^*$

Values of $\bar{x}$ further above or below μ_0 carry more evidence against H_0

Dist of $\bar{x}$ if $\mu = \mu_0$

$N(\mu_0, \sigma^2/n)$

Dist of $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ if $\mu = \mu_0$

$N(0,1)$

p-value

0

$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

$\frac{\bar{x}^*-\mu_0}{\sigma/\sqrt{n}}$

where the value $\bar{X}_n^*$ lies at the same distance away from μ_0 as $\bar{X}_n$, but in the other direction. We must consider all possible values of $\bar{X}_n$ which would carry as much or more evidence against the null as the observed $\bar{X}_n$; sample means greater than or equal to $\bar{X}_n^*$ would carry as much evidence against H_0 as sample means less than or equal to $\bar{X}_n$, so the p-value is the sum of two shaded areas. Since the distribution of the sample mean is symmetric, the p-value is twice the area below the curve beyond $\bar{X}_n$ away from μ_0.

We can compute this area by taking $\bar{X}_n$ into the Z world, which is the same as computing the test statistic Z_{test}. So the p-value is twice the area under the standard Normal pdf beyond the value of the test statistic away from 0.

Exercise. Refer to 6.92 of the textbook. Suppose a bottler of soft-drinks claims that its bottling process results in an internal pressure of 157 psi with a standard deviation $\sigma = 3$ psi. You are interested in contracting testing the hypotheses

$$H_0: \mu \geq 157 \text{ versus } H_1: \mu < 157.$$

Suppose you collect a sample of size $n = 40$ and get a sample mean of $\bar{X}_n = 155.7$. What is the p-value for testing the above hypotheses?

Answer: To find the p-value we first compute the test statistic, obtaining

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{155.7 - 157}{3/\sqrt{40}} = -2.741.$$

The p-value is the area under the Normal$(0,1)$ pdf to the left (the direction of the alternate hypothesis) of -2.741. This area is equal to 0.003, so the p-value is 0.003. We would therefore reject H_0 at every significance level α which is greater than 0.003.

Exercise. Refer to 6.84 of the textbook. A machine should produce ball bearings such that the standard deviation of the diameters is $\sigma = 0.001$ inches. The mean diameter should be 0.5 inches. You are interested in testing the hypotheses

$$H_1: \mu \neq 0.5 \text{ versus } H_0: \mu = 0.5.$$

Suppose that the diameters are Normally distributed. You take a random sample of 5 ball bearings and compute a mean diameter of 0.499. What is the p-value for testing the above hypotheses?

Answer: To find the p-value, we first compute the test statistic, obtaining

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{0.499 - 0.5}{0.001/\sqrt{5}} = -2.24.$$

Since this is a two-sided test, the p-value is twice the area under the standard Normal pdf to the left of -2.24. This is $2(0.0125) = 0.025$, so the p-value is 0.025. We would therefore reject H_0 at every significance level α which is greater than 0.025.

Exercise. Suppose you have a random sample of size $n = 35$ with sample mean $\bar{X}_n = 25$ from a right-skewed population with unknown mean μ and variance $\sigma^2 = 10$.

1. Compute the p-value for testing the hypotheses

$$H_0: \mu \geq 27 \text{ versus } H_1: \mu < 27.$$

   Answer: We first compute the test statistic

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{25 - 27}{10/\sqrt{35}} = -1.18.$$

   The p-value is the area under the standard Normal pdf to the left of -1.18. This is equal to 0.119, so the p-value is 0.119.

2. Compute the p-value for testing the hypotheses

$$H_0: \mu = 27 \text{ versus } H_1: \mu \neq 27.$$

   Answer: The value of the test statistic is the same, but since the alternate hypothesis is two-sided, the p-value is now two times the area under the standard Normal pdf to the left of -1.18. This is equal to $2(0.119) = 0.238$.

3. Compute the p-value for testing the hypotheses

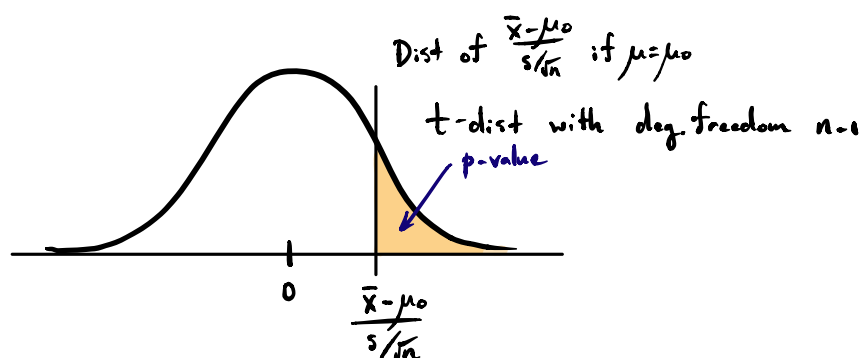$$H_0: \mu \leq 27 \text{ versus } H_1: \mu > 27.$$

   Answer: The value of the test statistic is the same. The p-value is the area under the standard Normal pdf pdf to the *right* of -1.18. This is equal to $1 - 0.119 = 0.881$, so the p-value is 0.881. Note that if the $\bar{X}_n$ lies in the region specified for μ under H_0, the p-value will be greater than or equal to 0.5.

Computing the p-value for hypotheses about μ (σ unknown)

In the σ-unknown case, we depict the p-values as areas under the t-distribution with degrees of freedom $n-1$ beyond the test statistic $T_{\text{test}} = (\bar{X}_n - \mu_0)/(S_n/\sqrt{n})$ in the direction of the alternate hypothesis. We arrive at this depiction using the same reasoning as in the σ-known case. For the one-sided tests, the p-values admit the depictions
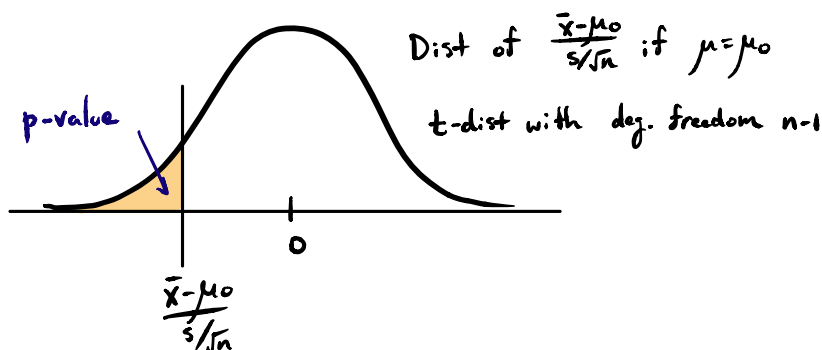
$H_0: \mu \leq \mu_0$

$H_1: \mu > \mu_0$

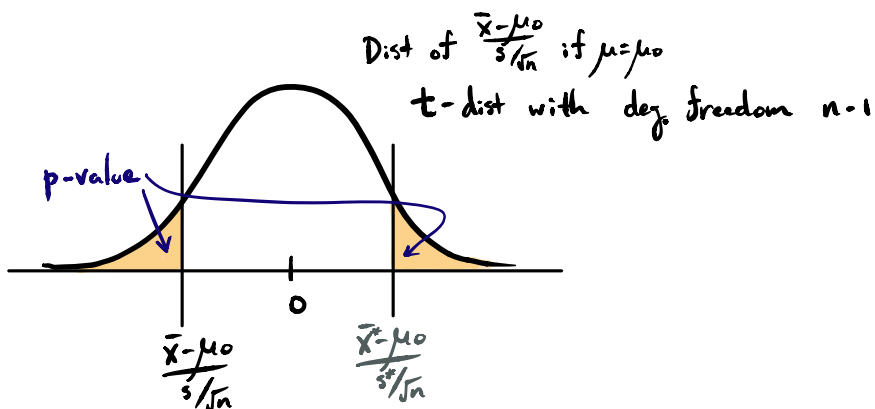Dist of $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ if $\mu = \mu_0$

t-dist with deg. freedom $n-1$

p-value

0

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

$H_0: \mu \geq \mu_0$

$H_1: \mu \leq \mu_0$

p-value

Dist of $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ if $\mu = \mu_0$

t-dist with deg. freedom $n-1$

0

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

The two-sided p-value admits the depiction

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

p-value

Dist of $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ if $\mu = \mu_0$

t-dist with deg. freedom $n-1$

0

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

$\frac{\bar{x}^*-\mu_0}{s^*/\sqrt{n}}$

where $\bar{X}_n^*$ and S_n^* represent the mean and standard deviation of a sample for which the test statistic would lie the same distance away from 0 as the observed test statistic but in the

opposite direction. Thus, $\bar{X}_n^*$ and S_n^* may be considered as coming from a sample which carries the same amount of evidence against H_0 as the observed sample, but for which the sample mean falls on the other side of μ_0. The p-value is thus twice the area beyond the test statistic away from 0 under the density of the t-distribution with degrees of freedom $n - 1$.

We cannot in general get the exact p-value from the t-table, but we can determine a range within which it must lie. We see this in the following example.

Example. Suppose you wish to test whether the LDL (bad cholesterol) level of South Carolinians exceeds the nationwide mean of 150 mg/dl. You take a random sample of size $n = 20$ and get a mean of $\bar{X}_n = 162.5$ mg/dl and a standard deviation of $S_n = 27.6$ mg/dl. What is the p-value for the hypotheses H_0: $\mu \le 150$ versus H_1: $\mu > 150$?
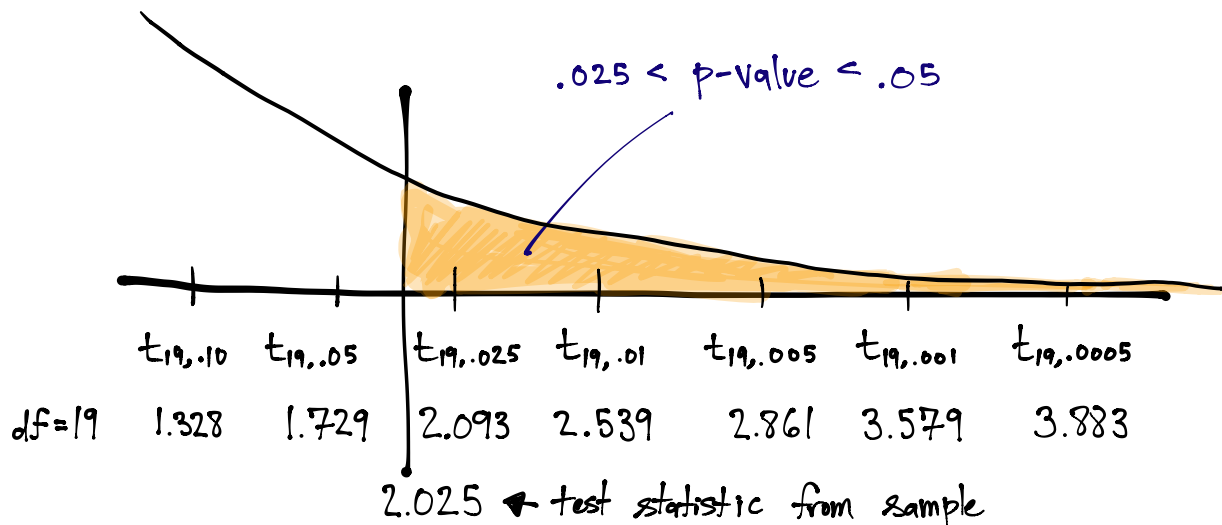
Answer: Let us begin by computing the test statistic and drawing a picture. We get

$$2.025 = \frac{162.5 - 150}{27.6/\sqrt{20}},$$

so we may depict the p-value as the area
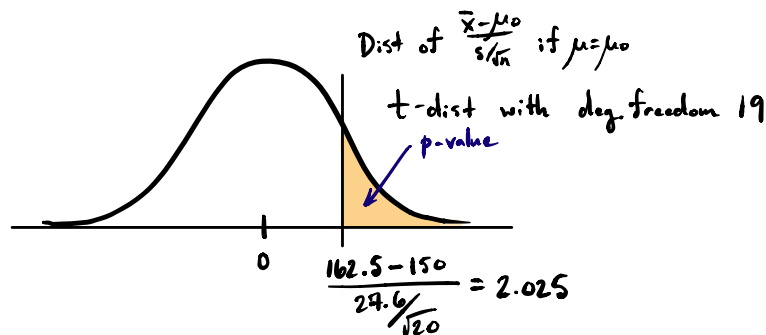
On the t-table we cannot find the exact p-value, as only a small number of quantiles of are tabulated for each t distribution. We can, however, determine a range within which the p-value must lie. Zooming in on the right tail of the t-distribution with degrees of freedom 19, we find that the p-value must lie between 0.025 and 0.05.

7

.025 < p-value < .05

	$t_{19,.10}$	$t_{19,.05}$	$t_{19,.025}$	$t_{19,.01}$	$t_{19,.005}$	$t_{19,.001}$	$t_{19,.0005}$
df=19	1.328	1.729	2.093	2.539	2.861	3.579	3.883

2.025 ← test statistic from sample

It is not very nice to use the t-table for computing p-values, as we cannot find the exact p-value. To get an exact p-value, we may use the R function pt(), where pt(q,ν) returns $P(T_\nu < q)$, where T_ν is a random variable having the t-distribution with degrees of freedom equal to ν. The function pt(q,ν) thus returns the area of the shaded region depicted below:

pdf of t_ν-distribution

Area = pt(q,ν)
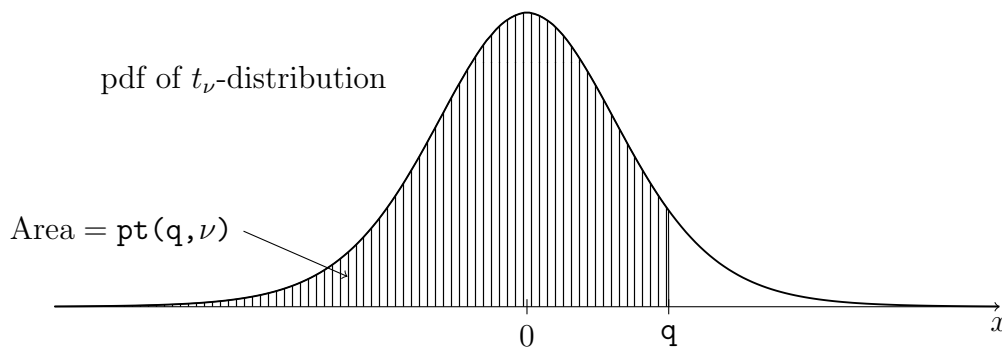
So, returing to the example, we can get the exact p-value from R by entering 1 - pt(2.025,19), which gives 0.02857553.

Computing the p-value for hypotheses about p

We have analogous depictions of the p-value when we are concerning with the following hypotheses about the proportion:

$$
\begin{array}{ccc}
H_0\colon p \ge p_0 & H_0\colon p = p_0 & H_0\colon p \le p_0 \\
H_1\colon p < p_0 & H_1\colon p \ne p_0 & H_1\colon p > p_0.
\end{array}
$$

We will assume in this section that the condition $\min\{np_0, n(1 - p_0)\} \ge 15$ is satisfied, so that $\hat{p}$ can be assumed to have a Normal distribution under $p = p_0$.

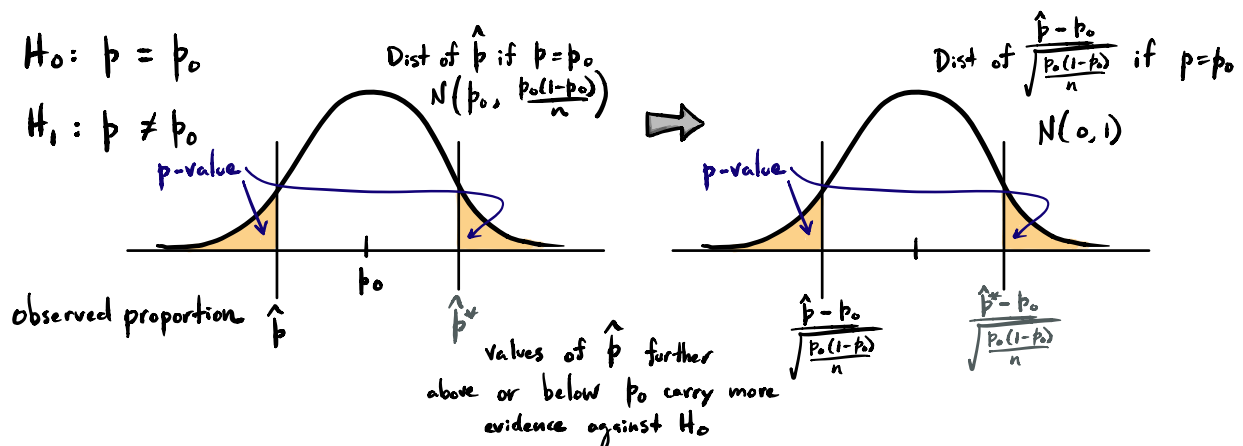Just as before, we consider all sample proportions which carry as much or more evidence against the null hypothesis as the observed $\hat{p}$, and the p-value is the probability, under $p = p_0$, of drawing any of these sample proportions. To get the areas of the shaded regions, we must take $\hat{p}$ into the Z world, which is done by computing the familiar test statistic.

$H_0: p \le p_0$

$H_1: p > p_0$

Dist of $\hat{p}$ if $p = p_0$
$N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

p-value

Values of $\hat{p}$ further above p_0 carry more evidence against H_0

Dist of $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ if $p = p_0$
$N(0,1)$

p-value

$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

$H_0: p \ge p_0$

$H_1: p < p_0$

Dist of $\hat{p}$ if $p = p_0$
$N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

p-value

Values of $\hat{p}$ further below p_0 carry more evidence against H_0

Dist of $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ if $p = p_0$
$N(0,1)$

p-value

$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
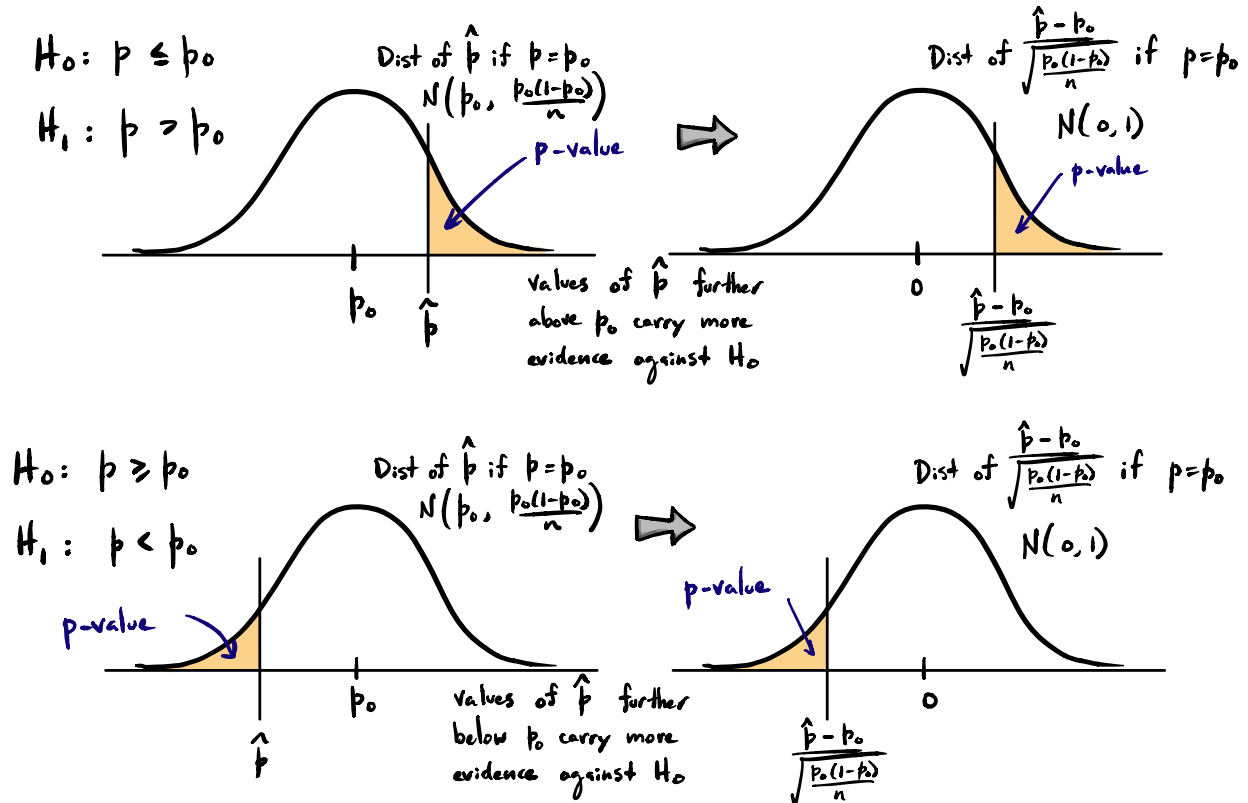
We see that the p-value for a sample proportion is the area under the standard Normal pdf beyond the test statistic in the direction of the alternate hypothesis.

The two-sided p-value admits the depiction

$H_0: p = p_0$

$H_1: p \ne p_0$

Dist of $\hat{p}$ if $p = p_0$
$N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

p-value

observed proportion $\hat{p}$

$\hat{p}^*$

Values of $\hat{p}$ further above or below p_0 carry more evidence against H_0

Dist of $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ if $p = p_0$
$N(0,1)$

p-value

$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

$\frac{\hat{p}^* - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

where $\hat{p}^*$ denotes a sample proportion carrying the same amount of evidence against H_0 as the observed sample proportion $\hat{p}$, but which lies on the other side of p_0. In the end, we see that for the two-sided case, the p-value is twice the area beyond the test statistic away from 0 under the standard Normal pdf.

Exercise. Suppose that an invasive species of fish will begin to dominate an inland system of lakes if it makes up more than 10% of the population of fish in those lakes. In a random sample of 527 fish drawn from the system of lakes, 70 of the fish were of the invasive species. Suppose that the Department of Natural Resources will take action if it is concluded that the invasive species comprises more than 10% of the fish population in the system of lakes.

1. What are the appropriate null and alternate hypotheses?

   Answer: Let p denote the proportion of fish belonging to the invasive species and test

   $$H_0\colon p \leq 0.10 \quad \text{versus} \quad H_1\colon p > 0.10.$$

2. What is the p-value?

   Answer: We compute the p-value by first computing the test statistic

   $$\frac{70/527 - 0.10}{\sqrt{\frac{0.10(0.90)}{527}}} = 2.51.$$

   The p-value is the area under the standard Normal pdf to the right (the direction of the alternative) of this value, which is 0.006.