

## 8.4: Tails of a test

### Introduction

The basics of **statistical inference** is to establish the **null** and **alternative hypotheses**. Starting with the simplest cases, where there is one sample of observations and the comparison is against a population (theory) mean, how many possible comparisons can be made? The next simplest is the two-sample case, where we have two sets of observations and the comparison is against the two groups. Again, how many total comparisons may be made?

Let  $\bar{X}$ , “X bar”, equal the sample mean and  $\mu$ , “mu”, represent the population mean. For sample means, designate groups by a subscript, 1 or 2. We then have Table 8.4.1.

Table 8.4.1. Possible hypothesis involving one or two groups.

Comparison	One-sample	Two-sample
1.	$\bar{X} = \mu$	$\bar{X}_1 = \bar{X}_2$
2.	$\bar{X} \neq \mu$	$\bar{X}_1 \neq \bar{X}_2$
3.	$\bar{X} \geq \mu$	$\bar{X}_1 \geq \bar{X}_2$
4.	$\bar{X} \leq \mu$	$\bar{X}_1 \leq \bar{X}_2$
5.	$\bar{X} > \mu$	$\bar{X}_1 > \bar{X}_2$
6.	$\bar{X} < \mu$	$\bar{X}_1 < \bar{X}_2$

Classical statistics classifies inference into null hypothesis,  $H_O$ , vs. alternate hypotheses,  $H_A$ , and specifies that we test null hypotheses based on the value of the estimated test statistic (see discussion about critical value and **p-value**, [Chapter 8.2](#)). From the list of six possible comparisons we can divide them into **one-tailed** and **two-tailed** differences (Table 8.4.1). By “tail” we are referring to the ends or tails of a distribution (Figure 8.4.1, Figure 8.4.2); where do our results fall on the distribution?

Two-tailed hypotheses: Comparison 1 and comparison 2 in the table above are two-tailed hypotheses. We don’t ask about the direction of any difference (less than or greater than).

Figure 8.4.1 shows the “two-tailed” distribution — if our results fall to the left ( $\leq -1.96$ ) or to the right ( $\geq +1.96$ ) we reject the null hypothesis (blue regions in the curve). We divide the type I error into two equal halves.

#### Note:

It’s a nice trick to shade in **regions of the curve**. A package `tigerstats` includes the function `pnormGC` that simplifies this task.

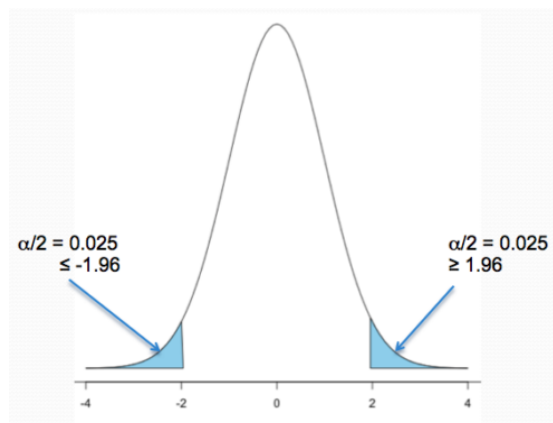


Figure 8.4.1: Two-tailed distribution.

```
library(tigerstats)
pnormGC(c(-1.96,1.96), region="outside", mean=0, sd=1, graph=TRUE)
```

Figure 8.4.2 shows the “one-tailed” distribution — if our alternate hypothesis was that the sample mean was less than the population mean, then our fall to the left ( $\leq -1.645$ ) for the “lower tail” of the distribution. If, however, our alternate hypothesis was that the sample mean was

greater than the population mean, then our region of interest falls to the right ( $\geq +1.645$ ). Again, we reject the null hypothesis (blue regions in the curve). Note for one-tailed hypothesis, all **Type I error** occurs in the one area, not both, so  $\alpha$  (alpha) remains 0.05 over the entire rejection region.

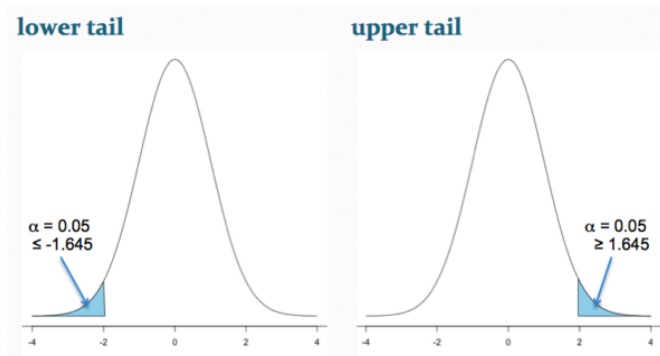


Figure 8.4.2: One-tailed distribution, lower tail (left) and upper tail (right).

```
library(tigerstats)
pnormGC(1.645, region="above", mean=0, sd=1, graph=TRUE)
pnormGC(-1.645, region="below", mean=0, sd=1, graph=TRUE)
```

One-tailed hypotheses: Comparison 3 through comparison 6 in the table are one-tailed hypotheses. The direction of the difference matters.

#### Note:

A simple trick to writing one-tailed hypotheses: first write the alternate hypothesis because the null hypothesis includes all of the other possible outcomes of the test.

### Examples

Let's consider some examples. We learn best by working through cases.

Chemotherapy as an approach to treat cancers owes its origins to the work of Dr. Sidney Farber, among others in the 1930s and '40s (DeVita and Chu 2008; Mukherjee 2011). Following up on the observations of others that folic acid (vitamin B<sub>9</sub>) improved anemia, Dr Farber believed that folic acid might reverse the course of leukemia (Mukherjee 2011). In 1946 he recruited several children with acute lymphoblastic leukemia and injected them with folic acid. Instead of ameliorating their symptoms (e.g., white blood cell counts and percentage of abnormal immature white blood cells, called blast cells), treatments accelerated progression of the disease. That's a scientific euphemism for the reality — the children died sooner in Dr. Faber's trial than patients not enrolled in his study. He stopped the trials. Clearly, adding folic acid was not a treatment against this leukemia.

**Question 1.** Do you think these experiments are one-sample or two-sample? Hint: Is there mention of a control group?

Answer: There's no mention of a control group, but instead, Dr. Faber would have had plenty of information about the progression of this disease in children. This was a one-sample test.

**Question 2.** What would be a reasonable interpretation of Dr. Faber's alternate hypothesis with respect to percentage of blast cells in patients given folic acid treatment? Your options are

1. Folic acid supplementation has an effect on blast counts.
2. Folic acid supplementation reduces blast counts.
3. Folic acid supplementation increases blast counts.
4. Folic acid supplementation has no effect on blast counts.

Answer: At the start of the trials, it is pretty clear that the alternate hypothesis was intended to be a one-tailed test (option 2). Dr. Faber's alternative hypothesis clearly was that he believed that addition of folic acid would reduce blast cell counts. However, that they stopped the trials shows that they recognized that the converse had occurred, that blast counts increased; this means that, from a statistician's point of view, Dr Faber's team was testing a two-sided hypothesis (option 1).

Here's another example.

Dr. Farber reasoned that if folic acid accelerated leukemia progression, perhaps anti-folic compounds might inhibit leukemia progression. Dr Farber's team recruited patients with acute lymphoblastic leukemia and injected them with a folic acid agonist called aminopterin. Again, he predicted that blast counts would reduce following administration of the chemical. This time, and for the first time in recorded medicine, blast

counts of many patients drastically reduced to normal levels and the patients experienced remissions. The remissions were not long-lasting and all patients eventually succumbed to leukemia. Nevertheless, these were landmark findings — for the first time a chemical treatment was shown to significantly reduce blast cell counts, even leading to remission, if however brief (Mukherjee 2011).

Try Question 3 and Question 4 yourself.

**Question 3.** Do you think these experiments are one-sample or two-sample? Hint: Is there mention of a control group?

**Question 4.** What would be a reasonable interpretation of Dr Faber’s alternate hypothesis with respect to percentage of blast cells in patients given aminopterin treatment? Your options are

1. Aminopterin supplementation has an effect on blast counts.
2. Aminopterin supplementation reduces blast counts.
3. Aminopterin supplementation increases blast counts.
4. Aminopterin supplementation has no effect on blast counts.

### Pros and Cons to One-sided testing

Here’s something to consider: why not restrict yourself to one-tailed hypotheses? Here’s the pro-argument. Strictly speaking you gain statistical power to test the null hypothesis. For example, look up the t-test distribution for degrees of freedom equal to 20 and compare  $\alpha_{(1)}$  (one tail) vs.  $\alpha_{(2)}$  (two-tail). You will find that for the one-tailed test, the critical value of the t-distribution with  $df = 20$  is 1.725, whereas for the two-tailed test, the critical value of the t-distribution with the same  $df$  numbers is 2.086. Thus, the difference between means can be much smaller in the one-tailed test and prove to be “statistically significant.” Put simply, with the same data, we will reject the Null Hypothesis more often with one-tailed tests.

The con-argument. If you use a one-tailed test you **MUST CLEARLY** justify its use and be aware that a deviation in the opposite direction **MUST** be ignored! More specifically, you interpret a one-tailed result in the opposite direction as acceptance of the null — you cannot, after the fact, change your mind and start speaking about “statistically significant differences” if you had specified a one-tailed hypothesis and the results showed differences in the opposite direction.

#### Note:

Recall also that, by itself, statistical significance judged by the p-value against a specified cut-off critical value is not enough to say there is evidence for or against the hypothesis. For that we need to consider effect size, see Power analysis in [Chapter 11](#).

### Questions

1. For a Type I error rate of 5% and the following degrees of freedom, compare the critical values for one tail test and a two tailed test of the null hypothesis.
  - $df = 5$
  - $df = 10$
  - $df = 15$
  - $df = 20$
  - $df = 25$
  - $df = 30$
2. Using your findings from Question 1, make a scatterplot with degrees of freedom on the horizontal axis and critical values on the vertical axis. What trend do you see for the difference between one- and two-tailed tests as degrees of freedom increase?
3. A clinical nutrition researcher wishes to test the hypothesis that a vegan diet lowers total serum cholesterol levels compared to an omnivorous diet. What kind of hypothesis should he use, one-tailed or two-tailed? Justify your choice.
4. Spironolactone, introduced in 1953, is used to block aldosterone in hypertensive patients. A newer drug eplerenone, approved by the FDA in 2002, is reported to have the same benefits as spironolactone (reduced mortality, fewer hospitalization events), but with fewer side effects compared with spironolactone. Does this sentence suggest a one-tailed test or a two-tailed test?
5. Write out the appropriate null and alternative hypothesis statements for the spironolactone and eplerenone scenario.
6. You open up a bag of Original Skittles and count the number of green, orange, purple, red, and yellow candies in the bag. What kind of hypothesis should be used, one-tailed or two-tailed? Justify your choice.
7. Verify the probability values from the table of standard normal distribution for  $Z$  equal to  $-1.96$ ,  $-1.645$ ,  $1.645$ , and  $1.96$ .

This page titled [8.4: Tails of a test](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.