

---

---

# MODULE 17

---

## 1-SAMPLE T-TEST

### Contents

17.1 t-distribution . . . . .	138
17.2 1-Sample t-Test Specifics . . . . .	140
17.3 1-Sample t-Test in R . . . . .	141

PRIOR TO THIS MODULE, hypothesis testing methods required knowing  $\sigma$ , which is a parameter that is seldom known. When  $\sigma$  is replaced by its estimator,  $s$ , the test statistic follows a Student's  $t$  rather than a standard normal ( $Z$ ) distribution. In this module, the  $t$ -distribution is described and a 1-Sample  $t$ -Test for testing that the mean from one population equals a specific value is discussed.

### 17.1 t-distribution

A  $t$ -distribution is similar to a standard normal distribution (i.e.,  $N(0,1)$ ) in that it is centered on 0 and is bell shaped (Figure 17.1). The  $t$ -distribution differs from the standard normal distribution in that it is heavier in the tails, flatter near the center, and its exact dispersion is dictated by a quantity called the degrees-of-freedom ( $df$ ). The  $t$ -distribution is “flatter and fatter” because of the uncertainty surrounding the use of  $s$  rather than  $\sigma$  in the standard error calculation.<sup>1</sup> The degrees-of-freedom are related to  $n$  and generally come from the denominator in the standard deviation calculation. As the degrees-of-freedom increase, the  $t$ -distribution becomes narrower, taller, and approaches the standard normal distribution (Figure 17.1).

---

<sup>1</sup>Recall that the sample standard deviation is a statistic and is thus subject to sampling variability.

Figure 17.1. Standard normal (black) and t-distributions (red) with varying degrees-of-freedom.

Proportional areas on a t-distribution are computed using `distrib()` similar to what was described for a normal distribution in Modules 8 and 13. The major exceptions for using `distrib()` with a t-distribution is that `distrib="t"` must be used and the degrees-of-freedom must be given in `df=` (how to find `df` is discussed in subsequent sections). For example, the area right of  $t = -1.456$  on a t-distribution with 9 df is 0.9103 (Figure 17.2).

```
> ( distrib(-1.456,distrib="t",df=9,lower.tail=FALSE) )  
[1] 0.9103137
```

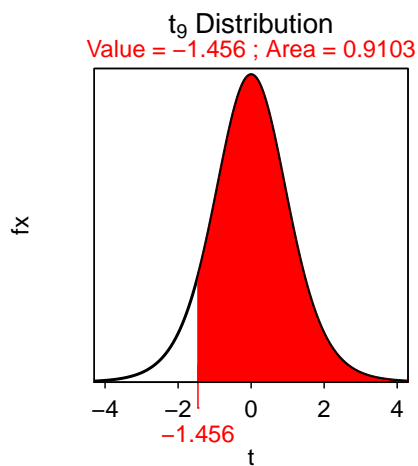


Figure 17.2. Depiction of the area to the right of  $t = -1.456$  on a t-distribution with 9 df.

Similarly, the  $t$  with an upper-tail area of 0.95 on a  $t$ -distribution with 19 df is -1.729 (Figure 17.3).<sup>2</sup>

```
> ( distrib(0.95,distrib="t",type="q",df=19,lower.tail=FALSE) )
[1] -1.729133
```

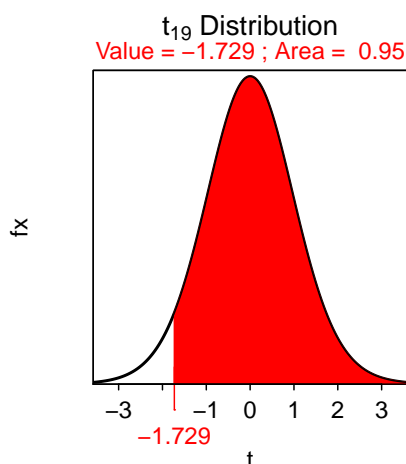


Figure 17.3. Depiction of the value of  $t$  with an area to the right of 0.95 on a  $t$ -distribution with 19 df.

## 17.2 1-Sample t-Test Specifics

A 1-Sample  $t$ -Test is similar to a 1-Sample  $Z$ -test in that both test the same  $H_0$ . The difference, as discussed above, is that when  $\sigma$  is replaced by  $s$ , the test statistic becomes  $t$  and the scaling factor for confidence regions becomes a  $t^*$ . Other aspects are similar between the two tests as shown in Table 17.1.<sup>3</sup>

Table 17.1. Characteristics of a 1-Sample  $t$ -Test.

- **Hypothesis:**  $H_0 : \mu = \mu_0$
- **Statistic:**  $\bar{x}$
- **Test Statistic:**  $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$
- **Confidence Region:**  $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$
- **df:**  $n - 1$
- **Assumptions:**
  1.  $\sigma$  is UNKNOWN
  2.  $n \geq 40$ ,  $n \geq 15$  and the **sample** (i.e., histogram) is not strongly skewed, OR the **sample** is normally distributed.
- **When to Use:** Quantitative response, one population,  $\sigma$  is UNKNOWN.

<sup>2</sup>This “reverse” calculation would be  $t^*$  for a 95% lower confidence bound.

<sup>3</sup>Compare Table 17.1 to Table 16.1.

### 17.2.1 Example - Purchase Catch of Salmon?

Below are the 11-steps (Section 16.1) for completing a full hypothesis test for the following situation:

*A prospective buyer will buy a catch of several thousand salmon if the mean weight of all salmon in the catch is at least 19.9 lbs. A random selection of 50 salmon had a mean of 20.1 and a standard deviation of 0.76 lbs. Should the buyer accept the catch at the 5% level?*

1.  $\alpha=0.05$ .
2.  $H_0 : \mu = 19.9$  lbs vs.  $H_A : \mu > 19.9$  lbs where  $\mu$  is the mean weight of ALL salmon in the catch.
3. A 1-Sample t-Test is required because (i) a quantitative variable (weight) was measured, (ii) individuals from one population were sampled (this catch of salmon), and (iii)  $\sigma$  is **UN**known.<sup>4</sup>
4. The data appear to be part of an observational study with random selection.
5. (i)  $n=50 \geq 40$  and (ii)  $\sigma$  is unknown.
6.  $\bar{x} = 20.1$  lbs (and  $s = 0.76$  lbs).
7.  $t = \frac{20.1-19.9}{\frac{0.76}{\sqrt{50}}} = \frac{0.2}{0.107} = 1.87$  with  $df = 50-1 = 49$ .
8. p-value = 0.0337.
9.  $H_0$  is rejected because the p-value  $< \alpha$ .
10. The average weight of ALL salmon in this catch appears to be greater than 19.9 lbs; thus, the buyer should accept this catch of salmon.
11. I am 95% confident that the mean weight of ALL salmon in the catch is greater than 19.92 lbs (i.e.,  $20.1 - 1.677 \frac{0.76}{\sqrt{50}} = 20.1 - 0.18 = 19.92$ ).

#### R Appendix:

```
( pval <- distrib(1.87,distrib="t",df=49,lower.tail=FALSE) )
( zstar <- distrib(0.95,distrib="t",type="q",df=49,lower.tail=FALSE) )
```

## 17.3 1-Sample t-Test in R

If raw data exist, the calculations for a 1-Sample t-test can be efficiently computed with `t.test()`. The arguments to `t.test()` are the same as those for `z.test()`, with the exception that `sd=` is not used with `t.test()`. Thus, `t.test()` requires the vector of quantitative data as the first argument, the null hypothesized value for  $\mu$  in `mu=`, the type of alternative hypothesis in `alt=` (again, can be `alt="two.sided"` (the default), `alt="less"`, or `alt="greater"`), and the level of confidence as a proportion in `conf.level=` (defaults to 0.95). The use of `t.test()` is illustrated in the following example.

### 17.3.1 Example - Crab Body Temperature

Below are the 11-steps (Section 16.1) for completing a full hypothesis test for the following situation:

*A marine biologist wants to determine if the body temperature of crabs exposed to ambient air temperature is different than the ambient air temperature. The biologist exposed a sample of 25 crabs to an air temperature of 24.3°C for several minutes and then measured the body temperature of each crab (shown below). Test the biologist's question at the 5% level.*

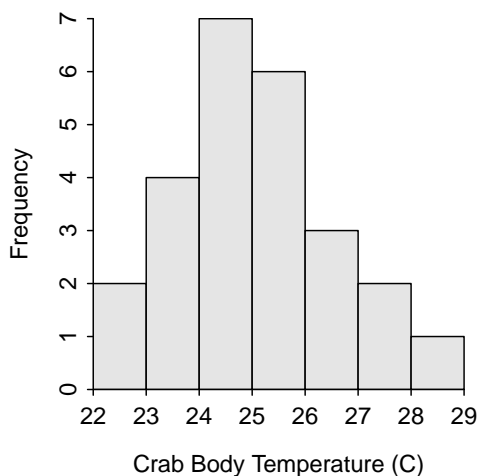
22.9, 22.9, 23.3, 23.5, 23.9, 23.9, 24.0, 24.3, 24.5, 24.6, 24.6, 24.8, 24.8,  
25.1, 25.4, 25.4, 25.5, 25.5, 25.8, 26.1, 26.2, 26.3, 27.0, 27.3, 28.1

<sup>4</sup>If  $\sigma$  is given, then it will appear in the background information to the question and will be in a sentence that uses the words "population", "assume that", or "suppose that."

1.  $\alpha = 0.05$ .
2.  $H_0 : \mu = 24.3^\circ\text{C}$  vs.  $H_A : \mu \neq 24.3^\circ\text{C}$ , where  $\mu$  is the mean body temperature of ALL crabs.
3. A 1-Sample t-Test is required because (1) a quantitative variable (temperature) was measured, (ii) individuals from one population were sampled (an ill-defined population of crabs), and (iii)  $\sigma$  is **UN**known.
4. The data appear to be part of an experimental study (the temperature was controlled) with no suggestion of random selection of individuals.
5. (i)  $n = 25 \geq 15$  and the sample distribution of crab temperatures appears to be only slightly right-skewed (Figure 17.4) and (ii)  $\sigma$  is **UN**known.
6.  $\bar{x} = 25.0^\circ\text{C}$  (Table 17.2).
7.  $t = 2.713$  with 24 df (Table 17.2).
8. p-value = 0.0121 (Table 17.2).
9.  $H_0$  is rejected because the p-value  $< \alpha$ .
10. It appears that the average body temperature of ALL crabs is greater than the ambient temperature of  $24.3^\circ\text{C}$ .
11. I am 95% confident that the mean body temperature of ALL crabs is between  $24.5^\circ\text{C}$  and  $25.6^\circ\text{C}$  (Table 17.2).

Table 17.2. Results from 1-Sample t-Test for body temperature of crabs.

```
t = 2.7128, df = 24, p-value = 0.01215
95 percent confidence interval:
 24.47413 25.58187
sample estimates:
mean of x
 25.028
```

Figure 17.4. Histogram of the body temperatures of crabs exposed to an ambient temperature of  $24.3^\circ\text{C}$ .**R Appendix:**

```
df <- read.csv("data/CrabTemps.csv")
hist(~ct,data=df,xlab="Crab Body Temp (C)")
( ct.t <- t.test(df$ct,mu=24.3,conf.level=0.95) )
```