Lecture 23: Hypothesis Tests for a Mean with a Unknown Standard Deviation

Chapter 17

Tomer Altman

October 22, 2025

Recap

- For the last few lectures, for instructions purposes, we have assumed that the population standard deviation (σ) was known.
- We conducted the z-test and created CIs using this known σ
- Today, we generalize to the more practical situation where σ is unknown and needs to be estimated by s, the sample standard deviation

Reduced conditions for inference about a mean

- Data is an SRS from a much larger population, say 20 times (really important)
- Observations follow a Normal distribution (leeway thanks to the CLT if sample size large enough)

Estimating the standard error based on the sample

- Previously, we learned that the standard error of the sampling distribution is equal to $\frac{\sigma}{\sqrt{n}}$ and used this in our calculations
- When we don't know σ we can use the sample standard deviation, s, instead to estimate the standard error by

$$\frac{s}{\sqrt{n}}$$

Standard Deviations: Sample vs. the sample mean distribution

s vs. $\frac{s}{\sqrt{n}}$

- Remember, s is our estimate for the population standard deviation. It estimates the variation between *individuals*
- In contrast, $\frac{s}{\sqrt{n}}$ is our estimate for the standard error of the sample mean distribution, $\bar{x}$
- $\frac{s}{\sqrt{n}}$ estimates how much sample *means* vary if we were to take many multiple samples

Recall the z-test!

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

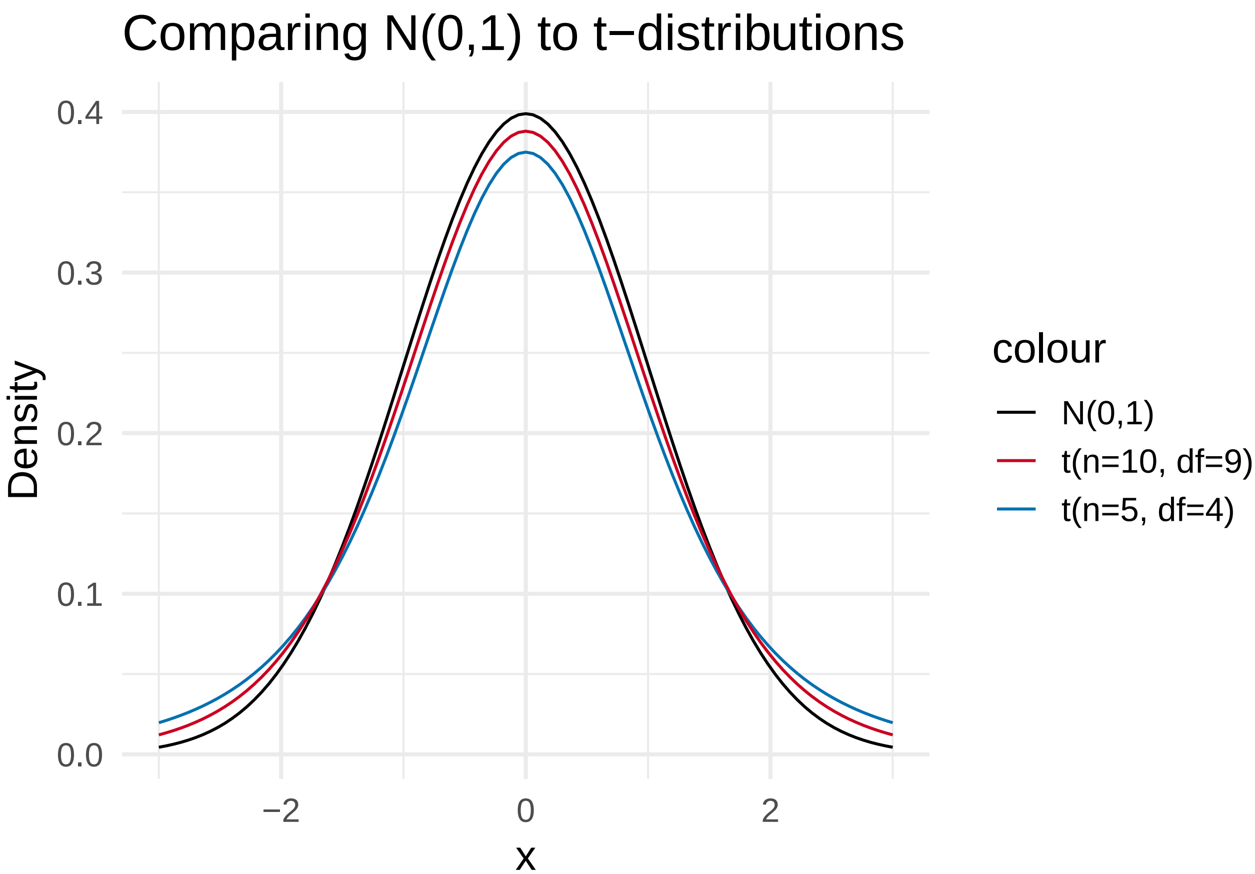
Meet the t-test!

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

- What is the difference between z and t?
- The t-test statistic is more variable, in small sample sizes, than the z-test statistic because we have to estimate σ using s
- Because s is a statistic, it varies across samples
  - Recall σ is a population parameter, which implies it is a constant
- This substitution makes the t-test, in small samples, not follow the standard Normal distribution, *even if the original distribution is normal*
- Its distribution is *more* variable than the standard Normal.
  - For hypothesis testing and confidence intervals, when sample size is relatively small (say $n < 40$) we need a distribution that is like the standard Normal but a little bit wider
  - More area in the tails of the distribution

Introducing the t-distribution

- The t-distribution is like the standard Normal distribution, but wider
- Its width depends on n, the sample size
- As n increases,our estimate s gets better and better, and approaches σ
- As n increases the t-distribution approaches a standard Normal distribution



Meet the t-distribution!

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

The one-sample t statistic (comparing an average to a null value, μ_0) has a t-distribution with $n-1$ **degrees of freedom (df)**

- **What are degrees of freedom?**
  - For this test, the degrees of freedom is equal to $n-1$
  - It represents the number of variables that are free to change
  - If $\bar{x} = \frac{a+b+c}{3}$, and we know the value of $\bar{x}$, and if we know the value of two variables, then we can calculate the value of the third
  - The higher the degrees of freedom, the closer the shape of the t-distribution is to the Normal distribution

Meet the t-test!

Steps to conduct a t-test:

1. Determine whether the assumptions to conduct the t-test are met
2. Calculate the t-test statistic using:
   - $\bar{x}$ and s (estimated from your sample)
   - n which is also a property of your sample
   - μ_0 from the null hypothesis

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

3. Compute the probability of observing a test statistic that is equal to this t statistic or greater under the null hypothesis
   - This is the p-value
   - We use this using the R function `pt(q=t, df=n-1)`
4. Interpret the p-value:
   - Is the probability very small (and shows evidence against the null distribution in favor of the alternative)?
   - Sometimes, you will be asked to compare the p-value to a pre-defined significance level, α
   - $\alpha = 0.05$ most commonly (but we talked about why we don't always do this!)

Guess the R functions

```
pt(q = , df = , lower.tail = TRUE)
qt(p = , df = , lower.tail = TRUE)
```

Which one would we use to calculate the p-value for a hypothesis test after we calculated the t-test statistic? `pt` or `qt`?

Suppose you calculated t = -2 and you know that the sample size was 100. Write the code to calculate the p-value for a two-sided test:

```
pt(-2,df=99)*2
```

```
## [1] 0.04823969
```

Calculating a confidence interval using the t-statistic

Draw an SRS of size n from a large population having unknown mean μ and unknown standard deviation σ. A level C **confidence interval for** μ is:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the critical value for the $t(df = n-1)$ density curve with area C between $-t^*$ and t^*.

Supposing we had $n = 100$, what is t^* for a 95% confidence interval?

Answer to previous question

```
qt(p = 0.975, df = 99)
```

```
## [1] 1.984217
```

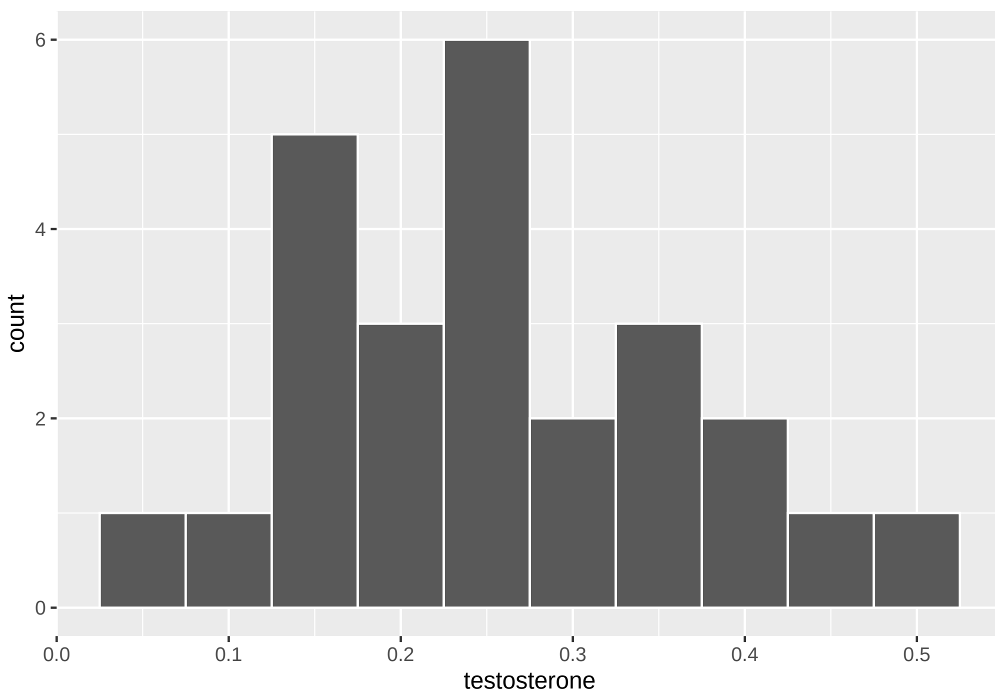
- The above answer should ring some bells!
- What does the t-distribution resemble when n gets large?
- What about the 68%-95%-99.7% rule?
- Wait, where did the 0.975 number come from?

Example: Testosterone and obesity in adolescent males (pg 422 B&M Ed 4)

Here are the data for $n = 25$ adolescent males between the ages of 14 and 20:

```
library(tidyverse)
testosterone <- c(0.30, 0.24, 0.19, 0.17, 0.18, 0.23, 0.24, 0.06, 0.15,
                  0.17, 0.18, 0.17, 0.15, 0.12, 0.25, 0.25, 0.25, 0.32,
                  0.35, 0.37, 0.39, 0.46, 0.49, 0.42, 0.36)
dat_test <- data.frame(testosterone)
```

```
ggplot(dat_test, aes(x = testosterone)) +
  geom_histogram(binwidth = 0.05, col = "white")
```



Example: Testosterone and obesity in adolescent males (pg 422 B&M Ed 4)

Use R to calculate a 95% confidence interval for testosterone. We can do this using `summarize`

```
dat_test %>% summarize(sample_mean = mean(testosterone), #sample mean
                       sample_sd = sd(testosterone), #sample standard dev
                       sample_size = length(testosterone), #sample size n
                       sample_se = sample_sd/sqrt(sample_size)) #standard error of mean
```

```
##   sample_mean sample_sd sample_size  sample_se
## 1      0.2584 0.1115303          25 0.02230605
```

We still need the t^* value:

```
t_star <- qt(p = 0.975, df = 24)
t_star
```

```
## [1] 2.063899
```

```
# Note, just barely bigger than the N(0,1)  97.5% quantile of 1.96
```

Example: Testosterone and obesity in adolescent males (pg 422 B&M Ed 4)

Expand the previous code chunk to calculate the margin of error (which uses the critical t^* value), and then calculate the lower and upper CI

```
dat_test %>% summarize(sample_mean = mean(testosterone),
                       sample_sd = sd(testosterone),
                       sample_size = length(testosterone),
                       sample_se = sample_sd/sqrt(sample_size),
                       margin_of_error = sample_se*t_star,
                       lower_CI = sample_mean - margin_of_error,
                       upper_CI = sample_mean + margin_of_error)
```

```
##   sample_mean sample_sd sample_size  sample_se margin_of_error  lower_CI
## 1      0.2584 0.1115303          25 0.02230605      0.04603743 0.2123626
##    upper_CI
## 1 0.3044374
```

Interpret: The sample mean $\bar{x}$ is 0.26 and its 95% confidence interval is 0.21 to 0.30. Using this method, 95% of the confidence intervals we make will contain the true population mean μ.

The t-test

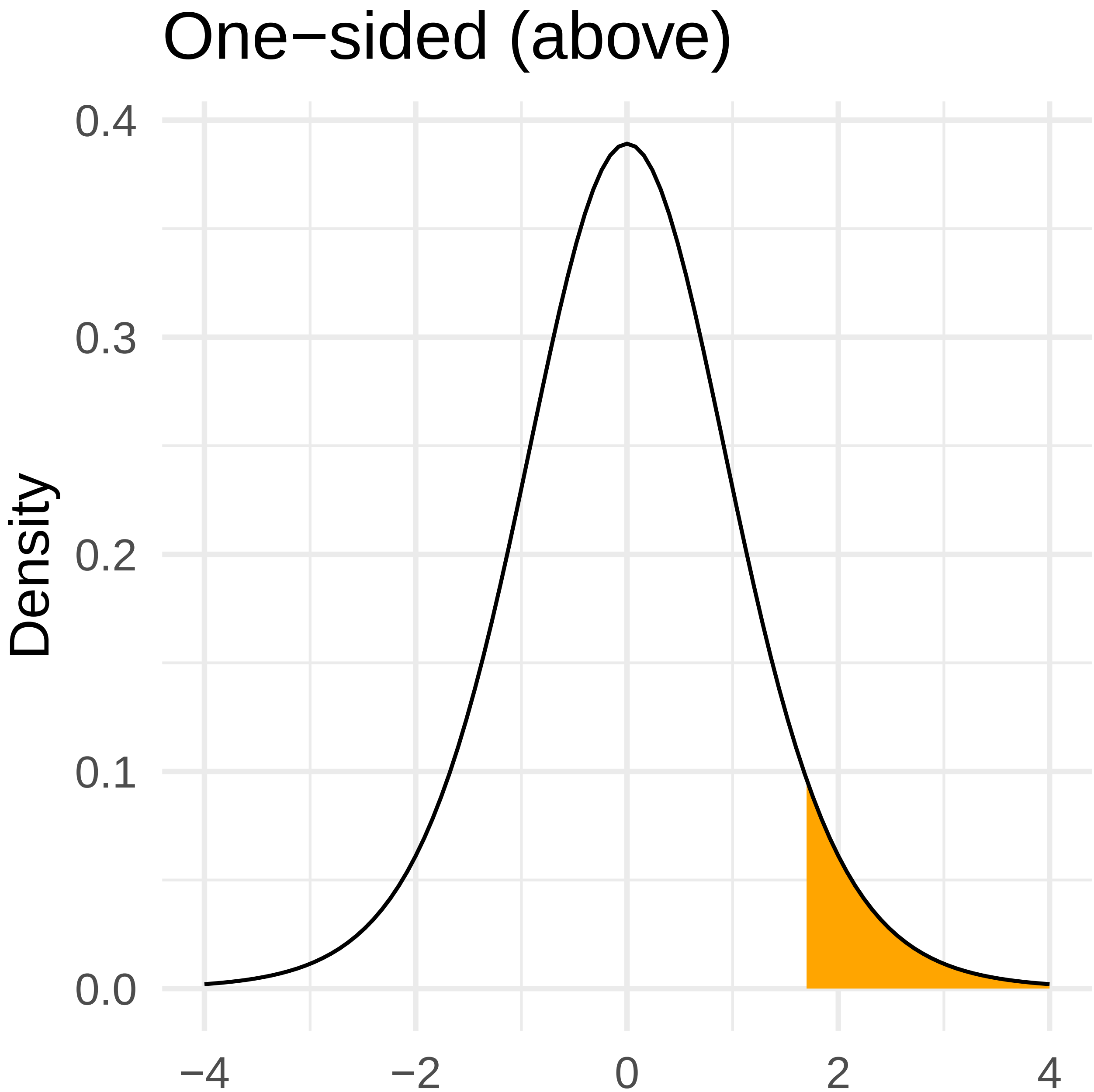
1. Draw an SRS of size n from a large population having unknown mean μ and unknown standard deviation σ
2. To test the hypothesis $H_0 : \mu = \mu_0$, calculate the t statistic from the t-distribution with $n-1$ degrees of freedom:

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

3. Calculate the probability that we would see this t (or a more extreme value) under the null distribution

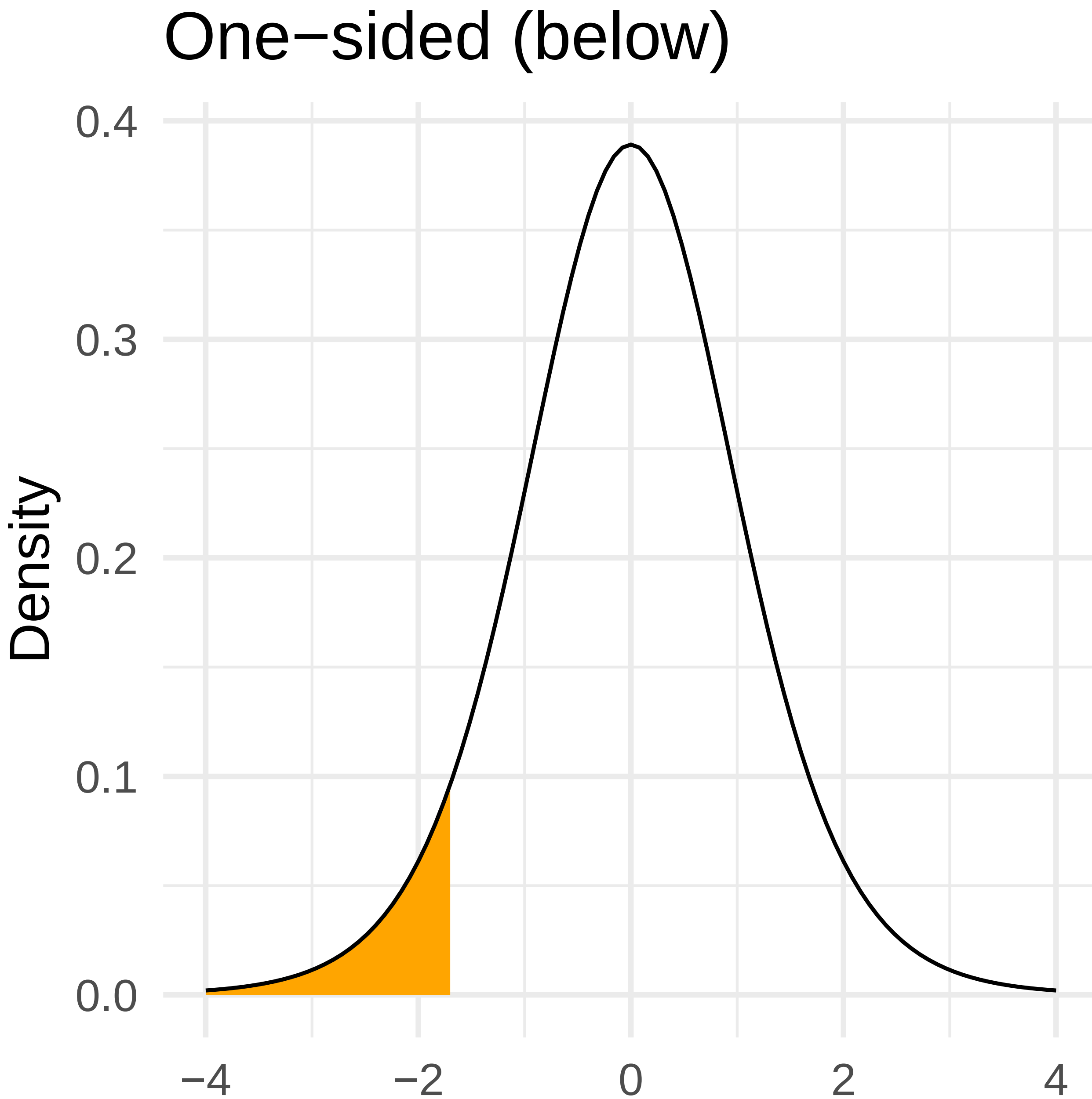
The t-test: $\mu > \mu_0$

- H_a: $\mu > \mu_0$ is $P(T \geq t)$
- `pt(q = t, df = n-1, lower.tail = F)`



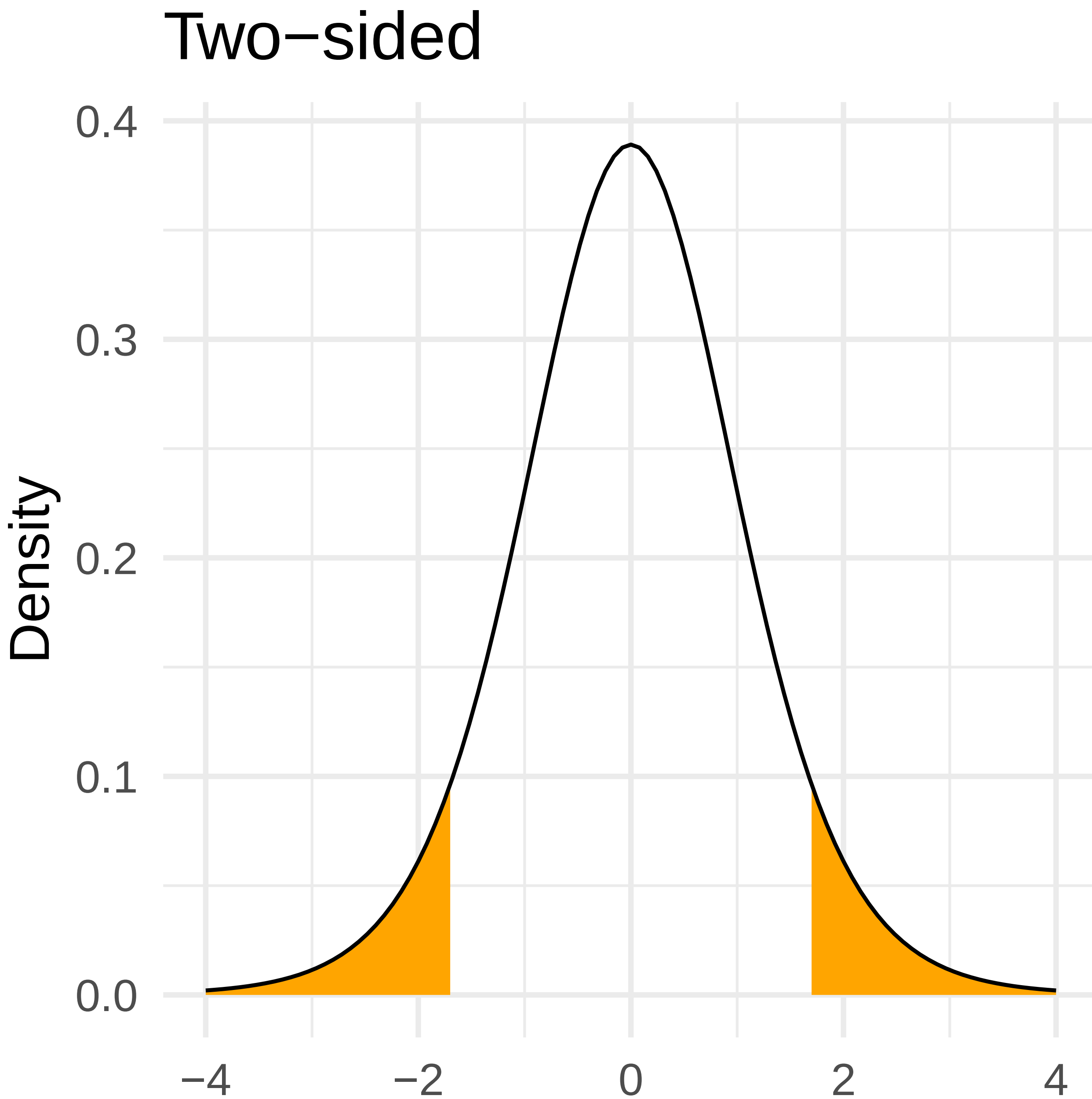
The t-test: $\mu < \mu_0$

- H_a: $\mu < \mu_0$ is $P(T \leq t)$
- `pt(q = t, df = n-1)`



The t-test: $\mu \neq \mu_0$

- H_a: $\mu \neq \mu_0$ is $2 \times P(T \geq |t|)$
- Code:
  - If $t \leq 0$: `pt(q = t, df = n-1)*2`
  - If $t \geq 0$: `pt(q = t, df = n-1, lower.tail = F) * 2`
  - Make sure you understand why



Example of a *t*-test (pg 426 B&M Ed 4)

Here are 18 measures of pulse wave velocity (PWV) from a sample of children diagnosed with progeria, a genetic disorder that produces rapid aging.

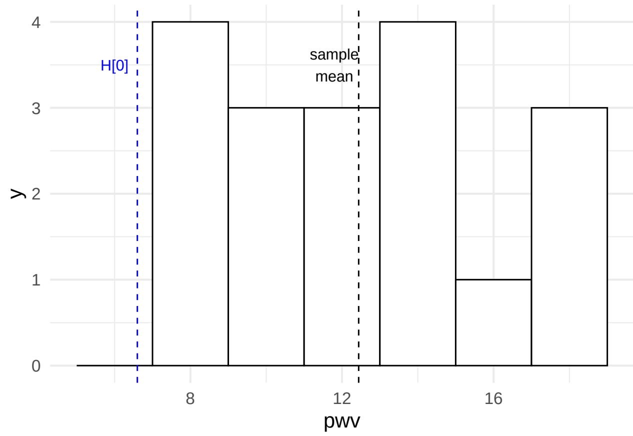
```
pwv <- c(18.8, 17.6, 17.5, 16.0, 14.8, 14.1, 13.7, 13.1, 12.9,
         12.9, 12.4, 10.1, 9.3,  9.1,  8.3,  8.3,  7.9,  7.2)

pwv_dat <- data.frame(pwv)
```

For the general population, `pwv` measures greater than 6.6 are considered abnormally high. We would like to test the hypothesis that the mean for this subset of children is abnormally high.

That is: $H_0 : \mu = 6.6$ and $H_a : \mu > 6.6$

Look at the data and see if there is evidence against the null hypothesis



Calculations using R code

```
pwv_dat %>%
  summarize(sample_mean = mean(pwv),
            sample_sd = sd(pwv),
            sample_size = length(pwv),
            sample_se = sample_sd/sqrt(sample_size),
            t_test = (sample_mean - 6.6)/sample_se,
            p_value = 1 - pt(t_test, df = sample_size - 1))
```

```
##   sample_mean sample_sd sample_size sample_se   t_test      p_value
## 1    12.44444  3.637747          18 0.8574252 6.816273 1.501248e-06
```

- Know also how to do these calculations by hand
- For example, you could be provided with $\bar{x}$ and s for this sample and asked to compute the test statistic
- You cannot compute the p-value by hand, but should know the code required to calculate the p-value and how to interpret it

There's a function for that...

Rather than doing the test using summarize, we could have R do it for us using `t.test`:

```
t.test(x = pwv_dat %>% pull(pwv), alternative = "greater", mu = 6.6)
```

```
##
##  One Sample t-test
##
## data:  pwv_dat %>% pull(pwv)
## t = 6.8163, df = 17, p-value = 1.501e-06
## alternative hypothesis: true mean is greater than 6.6
## 95 percent confidence interval:
##  10.95286      Inf
## sample estimates:
## mean of x
##  12.44444
```

Matched pairs t-test procedures

- Skip this section for now. We will come back to this next week.

Robustness of t-distribution procedures

- Confidence intervals and hypothesis tests ("procedures") are called **robust** if the confidence level or p-value does not change very much when the conditions for use of the procedure are violated
- In particular, how robust are the procedures against non-Normality?
- The t-distribution procedures are quite robust against non-Normality of the population except when outliers or strong skewness are present
- The t-distribution procedures are not robust against a few outliers unless the sample size is sufficiently large

Checking assumptions

- Always plot your data first:
  - Are there any outliers?
  - Is the distribution of the data skewed?

Guidelines for using the t procedures

- The SRS condition is more important that the Normality condition
- If n $<$ 15: Use t procedures if the data appear close to Normal (at least roughly symmetric, single peak, no outliers). If the data are skewed or there are outliers, don't use t.
- Moderate sample size $>$ 15: The t procedures can be used except in the presence of outliers or strong skewness
- Large sample size, roughly $n \geq 40$: The t procedures can be used even for strongly skewed distributions when the sample is large

Example 17.5: Can we use t?

- Good text example. Here you are provided with four datasets and their distributions and sample sizes and are asked whether it is appropriate to use a t-test.
- Pg. 436 of edition 4

Recap

- We use a z-test when the population sd σ is known
- We use a t-test when the population sd has to be estimated by s
- We compare the z-test statistic to a $N(0, 1)$ distribution to calculate the p-value
- We compare the t-test statistic t to a t-distribution with degrees of freedom on $n - 1$
- When n is large, the t-distribution is very close to the $N(0, 1)$ distribution
  - This means that we have some intuition about whether the p-value is going to be small or large when the sample size is large