

# L16: Statistical Inference with confidence intervals and hypothesis testing

Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

## L16: Statistical Inference with confidence intervals and hypothesis testing

Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Statistics is everywhere

From The New York Times, March 8, 2019

## *The Jobs Report Was the Weakest in Months. Here's Why.*

February's meager gain follows other signs of economic sluggishness. But there were encouraging signs in the data as well.



By Patricia Cohen

March 8, 2019



From The New York Times, March 6, 2020

## *U.S. Added 273,000 Jobs in February Before Coronavirus Spread Widely*

The monthly employment report left unanswered questions about the potential economic impact of the outbreak.



From Fox news, July 2, 2020

## **Trump touts 'historic' June jobs report: 'Our economy is roaring back'**

The unemployment rate fell to 11.1 percent in June

by Tyler Stone | Fox News



L16: Statistical Inference with confidence intervals and hypothesis testing

Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Statistics is everywhere

From the articles:

“The economy’s remarkably steady job-creation machine sputtered in February and produced a mere 20,000 jobs. It was the smallest gain in well over a year and came on top of other signs that the economy was off to a sluggish start in 2019.”

and

“For the second month in a row, the economy churned out a blockbuster number of jobs, the government reported Friday, an impressive performance in an era of slow-and-steady employment growth.”

and

“The United States economy added almost 5 million jobs in the month of June, shattering all expectations,” Trump said. “The stock market is doing extremely well . . . this is the largest monthly jobs gain in the history of our country.”

Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

# Statistics is everywhere

But also later in the 2019 article (not mentioned in either source in 2020):

“Revisions to previous months’ estimates added 12,000 jobs”

Numbers are actually revised twice, once in the month following the first report, and again the month after that.

# Statistics is everywhere

What we rarely see included in the articles talking about the jobs numbers is the margin of error.

For the 2019 estimate, it is 120,000 !

You can read more about this from the article on fivethirtyeight

# Statistical Inference

## Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

## Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

So far in part II we have been talking about probability and underlying probability distributions.

Now we are going to think about how to use these theoretical distributions to put some boundaries around estimates we calculate from samples.



Statistical Inference provides methods for drawing conclusions about a population from sample data. We are using data from a sample to **infer** something about the underlying population.

Today we will talk about

- ▶ Confidence intervals for point estimates
- ▶ Margins of error
- ▶ Testing a hypothesis about your data

## Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Simple conditions for inference about a mean

1. We have a simple random sample from the population of interest. There is no non-response or other systematic bias (i.e., no confounding, no measurement error, no selection bias).
2. The quantitative variable we measure has a perfectly Normal distribution  $N(\mu, \sigma)$
3. We don't know the population mean  $\mu$ , and want to estimate it. But we do know the population standard deviation.

Note that these conditions are idealized and not often realistic, however we will use this idealized version as a base which we will adapt as we move forward and discuss more realistic scenarios.

## Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

## Example 14.1 Baldi and Moore

A recent NHANES reports that the mean height of a sample of 217 eight-year old boys was  $\bar{x} = 132.5$  cm.

We want to use this sample to estimate the mean  $\mu$  in the population of  $> 1$  million American eight-year-old boys.

First, we need to check if the problem description meets the simple conditions required:

Condition 1: sampling

- Is it a SRS?

### Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

## Example 14.1 Baldi and Moore

Condition 2: distribution in the population

- ▶ Assume that the distribution of heights in the total population is Normally distributed

Condition 3: Unknown population mean but known population  $\sigma$

- ▶ We also need to assume a standard deviation. We will assume  $\sigma = 10$  cm. (Note that if you are asked to assume a standard deviation, it will be provided to you by the question.)

### Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Calculating a confidence interval

## Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

- Recall that  $\bar{x}$  is an unbiased estimator of  $\mu$ . Under repeated sampling, the sampling distribution of  $\bar{x}$  is Normally distributed with a mean of  $\mu$  and standard deviation  $\sigma/\sqrt{n} = 10/\sqrt{217} = 0.7$  cm.

## Calculating a confidence interval

- ▶ We can draw the Normal distribution for the sampling distribution, and shade in the middle 95% of the area within 2 standard deviations of the mean. Thus, an  $\bar{x}$  from any random sample has a 95% chance of being within 2 SD of the population mean  $\mu$ . This means that **for 95% of samples, 1.4 cm is the maximum distance separating  $\bar{x}$  and  $\mu$** . Therefore, if we estimate that the value  $\mu$  is somewhere in the interval from  $\bar{x} - 1.4$  to  $\bar{x} + 1.4$ , we'll be right 95% of the times we take a sample.

$$\bar{x} - 1.4 = 132.5 - 1.4 = 131.1$$

to

$$\bar{x} + 1.4 = 132.5 + 1.4 = 133.9$$

# Interpretation of a confidence interval

- ▶ Our best estimate of  $\mu$  is 132.5
- ▶ But, given we only took one sample of size  $n=217$ , this best estimate is imprecise
- ▶ The 95% confidence interval for  $\mu$  is 131.1 to 133.9.
- ▶ If our model assumptions are correct and there is only random error affecting the estimate, **this method for calculating confidence intervals will contain the true value  $\mu$  95% of the time** (19 times out of 20).
- ▶ This means that the interval  $\bar{x} \pm 1.4$  has a 95% success rate in capturing within that interval the mean height  $\mu$  of all eight-year-old American boys.

## Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Interpretation of a confidence interval

## Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

Do not use the textbook's shorthand that “we are 95% confident that  $\mu$  is contained in the CI”. This description is ambiguous and imprecise.



# What would make the CI smaller (and more precise)?

## Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

Remember that we are building the CI based on  $+$  and  $- 2 \times$  standard deviation

$$standarddeviation = \frac{\sigma}{\sqrt{n}}$$

What would make the CI smaller?

# What would make the CI smaller (and more precise)?

## Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions

$$\text{standard deviation} = \frac{\sigma}{\sqrt{n}}$$

- ▶ If we increase the sample size, the confidence interval is more precise
- ▶ If there were less underlying variability in the data (i.e.,  $\sigma$  was smaller), then the CI would be more precise

# Margin of error and confidence level

Form of a confidence interval:

$$\text{estimate} \pm \text{margin of error}$$

The margin of error will differ based on the confidence level (often 90%, 95%, or 99%) that is chosen.

Will a 99% confidence interval be wider or narrower than a 95% confidence interval?

## Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

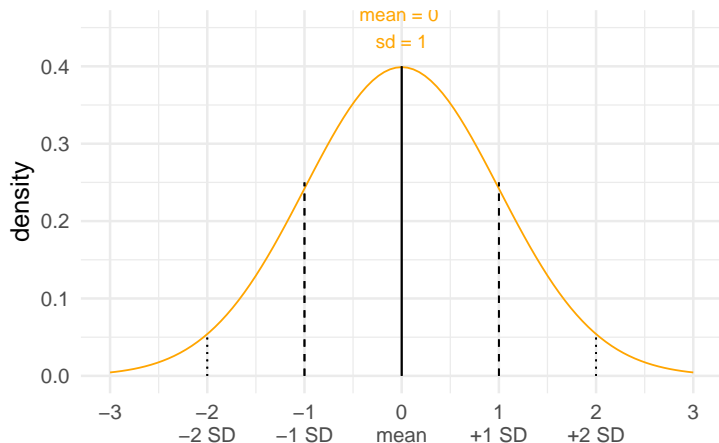
Hypothesis testing using Z tests

Definitions

# The standard Normal distribution

Recall the Standard Normal

- ▶ The standard Normal distribution  $N(0,1)$  has  $\mu = 0$  and  $\sigma = 1$ .
- ▶  $X \sim N(0,1)$ , implies that the random variable  $X$  is Normally distributed.



## Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Standardizing Normally distributed data

- ▶ Any random variable that follows a Normal distribution can be standardized
- ▶ If  $x$  is an observation from a distribution that has a mean  $\mu$  and a standard deviation  $\sigma$ ,

$$z = \frac{x - \mu}{\sigma}$$

###What's the Z

By converting our variable of interest  $X$  to  $Z$  we can use the probabilities of the standard normal probability distribution to estimate the probabilities associated with  $X$ .

- ▶ A standardized value is often called a **z-score**
- ▶ Interpretation:  $z$  is the number of standard deviations that  $x$  is above or below the mean of the data.

## Confidence intervals for the mean $\mu$

## Confidence intervals for the mean $\mu$

Confidence level C	90%	95%	99%
Critical value $z^*$	1.645	1.960 ( $\approx 2$ )	2.576

- These numbers correspond to the value on the x-axis corresponding to having 90%, 95%, or 99% of the area under the Normal density between  $-z$  and  $z$ .

The generic format of a confidence interval is then:

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

# Confidence intervals for the mean $\mu$

- ▶ For example, the middle 90% of the area under the Normal density lies between -1.645 and 1.645.
- ▶ Thus, a 90% confidence interval is of the form:

$$\bar{x} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$



# Confidence interval for the mean of a Normal population

Draw a SRS of size  $n$  from a Normal population having unknown mean  $\mu$  and known standard deviation  $\sigma$ . A level  $C$  confidence interval for  $\mu$  is:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

unbiased estimate  $\pm$  (critical value)  $\times$  (sd of the distribution of the estimate)

unbiased estimate  $\pm$  (critical value)  $\times$  (standard error)

# Steps in finding confidence intervals

1. Problem: Statement of the problem in terms of the parameter you would like to estimate
2. Plan: How will you estimate this parameter? What type of data will you collect?
3. Data: After you plan the study, collect the data you need to answer the problem.
4. Analysis: Evaluate whether the assumptions required to compute a confidence interval are satisfied. Calculate the estimate of the mean and its confidence interval.
5. Conclusion: Return to the practical question to describe your results in this setting.

## Example on IQ scores (pg. 354)

We are interested in the mean IQ scores of 7th grade girls in a Midwest school district. Here are the scores for 31 randomly selected seventh-grade girls. We also know that the standard deviation of IQ scores is 15 points:

```
scores <- c(114, 100, 104, 89, 102, 91, 114, 114, 103, 105,  
            108, 130, 120, 132, 111, 128, 118, 119, 86, 72,  
            111, 103, 74, 112, 107, 103, 98, 96, 112, 112, 93)  
  
iq_data <- data.frame(scores)  
  
known_sigma <- 15
```

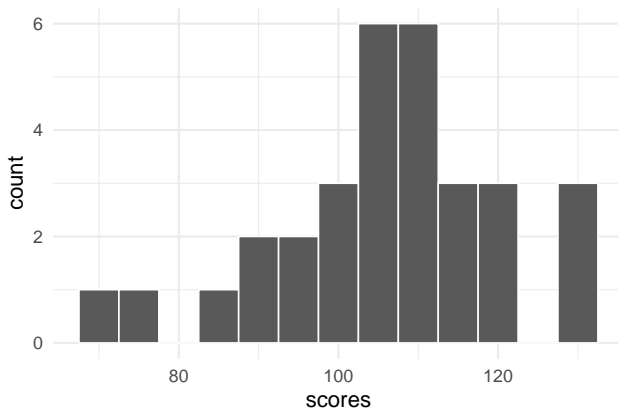
Estimate the mean IQ score  $\mu$  for all seventh grade girls in this Midwest school district by giving a 95% confidence interval.

## Example on IQ scores (pg. 354)

First check the three assumptions:

1. Normality: Can evaluate this using a histogram
2. SRS: Can only use information provided in the problem to assess with an SRS was taken.
3. Known  $\sigma$ : Can use information provided in the problem to determine if  $\sigma$  is known.

# Checking Normality



We can't examine the Normality of the population (because we don't have data on everyone) but we can make a plot for the sample. These data appear slightly left- skewed, but since there is not much data, it may actually follow a Normal distribution.

# Calculating the estimated mean and its confidence interval

Option 1: Perform calculations by hand

By hand:

$$\bar{x} = \frac{114 + 100 + \dots + 93}{31} = 105.8387$$

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{31}} = 2.69408$$

$$\begin{aligned} & \bar{x} \pm 2SE \\ &= 105.8387 \pm 2(2.69408) \\ &= 100.5583 \text{ to } 111.1191. \end{aligned}$$

The average IQ score of the sample is 105.84. The corresponding 95% CI is 100.56 to 111.12. If we were to take samples many times, 95% of the confidence intervals would contain the true population parameter  $\mu$ .

# Calculating the estimated mean and its confidence interval

## Option 2: Perform calculations using R

```
sample_mean <- mean(scores)

standard_error <- known_sigma/sqrt(length(scores))
critical_value <- 1.96

lower_bound <- sample_mean - critical_value*standard_error
upper_bound <- sample_mean + critical_value*standard_error
```

# Calculating the estimated mean and its confidence interval

```
sample_mean
```

```
## [1] 105.8387
```

```
standard_error
```

```
## [1] 2.69408
```

```
lower_bound
```

```
## [1] 100.5583
```

```
upper_bound
```

```
## [1] 111.1191
```



# Calculating the estimated mean and its confidence interval

Thus, our best estimate of the population mean is 105.84.

Its 95% confidence interval is 100.56 to 111.12.

If our model assumptions are correct and there is only random error affecting the estimate, this method for calculating confidence intervals will contain the true value  $\mu$  95% of the time (19 times out of 20).

- ▶ We learned how to create a confidence interval for the mean when the standard deviation for the population is known.
- ▶ We learned about the three required assumptions and how to check the Normality assumption using a histogram.
- ▶ We learned how to interpret the confidence interval and the definitions for the confidence level and the margin of error
- ▶ We introduced our “recipe” for a margin of error

$$\text{unbiased estimate} \pm (\text{critical value}) \times (\text{standard error})$$

Statistical Inference

Confidence intervals for the mean  $\mu$

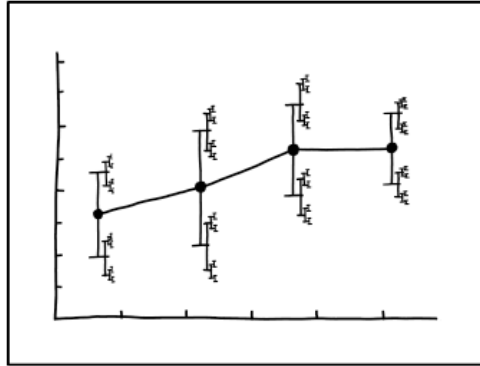
Example: one tailed hypothesis

Two sided testing

Hypothesis testing using Z tests

Definitions

# Central limit theorem



I DON'T KNOW HOW TO PROPAGATE  
ERROR CORRECTLY, SO I JUST PUT  
ERROR BARS ON ALL MY ERROR BARS.

###Hypothesis testing So far we have looked at the theoretical distributions that might give rise to the data we see.

# Hypothesis testing

We start with a hypothesis. This is some specific question about a specific population.

We then state this formally as a **null** and **alternative** hypothesis.

Then, using our knowledge of probability we assess how likely it is to see the data we saw under assumed conditions.

### Define the Hypothesis

A **Null Hypothesis** ( $H_0$ ) is the hypothesis that is assumed to be true and the start of a test. This is often expressed as a statement of equality (ie. mean equal to a certain value or no difference between groups)

An **Alternative Hypothesis** ( $H_A$ ) is usually the inverse of the null hypothesis and is expressed as a statement of difference.

- ▶  $H_A$ : The mean is greater than the Null (one tailed)
- ▶  $H_A$ : The mean is less than the null (one tailed)
- ▶  $H_A$ : The mean is not equal to (greater or less than) the null (two tailed)

# One sample questions

In our lecture on distributions we looked at using the normal distribution to ask how probable it was to observe a value in a population (ie. what proportion of the population would be as tall or taller than a value)

Now we are asking: “How likely is it that we observed a sample mean we did if the true population mean is  $\mu_0$ ”

# One sample questions

Our  $H_0$ : There is no difference between the sample mean and the hypothesized null

Our  $H_A$ : There is a difference between our sample mean and the hypothesized null

Our  $\alpha$  : The level at which we decide an outcome is **not probable**

Our tails: One tailed, or two tailed alternative?

► Let's consider an example.

Statistical Inference

Confidence intervals for the  
mean  $\mu$

**Example: one tailed  
hypothesis**

Two sided testing

Hypothesis testing using Z  
tests

Definitions

## Example: one tailed hypothesis

# Inorganic phosphorus levels in the elderly

Levels of inorganic phosphorus in the blood are known to vary among adults Normally with  $\mu = 1.2$  millimoles per liter and standard deviation 0.1 mmol/l.

A study examined inorganic phosphorus in older individuals to see if it decreases with age. Here are the data from 12 men and women between the ages of 75 and 79:

```
phos <- c(1.26, 1.39, 1, 1, 1, 1,  
          0.87, 1.23, 1.19, 1.29, 1.03, 1.18)  
  
known_sigma <- 0.1
```

\*note that sigma here is given to you

Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

Definitions



The sample mean  $\bar{x}$  is:

```
## [1] 1.12
```

Does the sample mean really differ from the null hypothesis mean?

## The null hypothesis $H_0$ and alternative hypothesis $H_a$

The null hypothesis is a hypothesis of no effect or no difference. For our example:

$$H_0 : \mu = 1.2$$

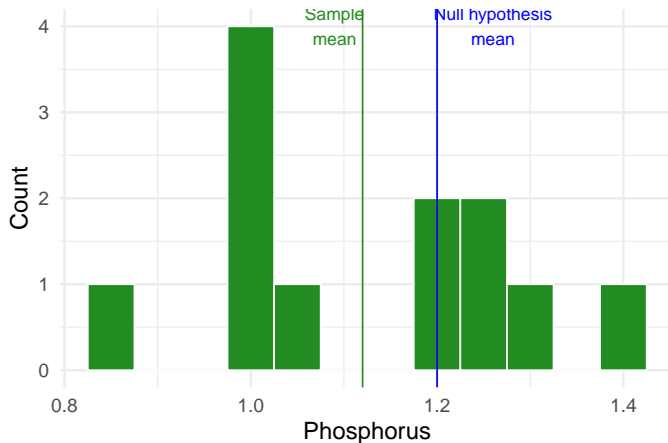
The alternative hypothesis ( $H_a$ ) is the competing hypothesis. In this question, the competing hypothesis is that older individuals have an average phosphorus level lower than that seen in the overall adult population:

$$H_a : \mu < 1.2$$

This is a **one-sided hypothesis**

# Descriptives

We can draw the data, the sample mean, and the hypothesized mean under  $H_0$  all on one figure:



# Performance of the hypothesis test

In our hypothesis test, we assume that the null hypothesis is true and calculate how likely we would observe a sample mean at least as extreme as the mean observed in the sampled data.

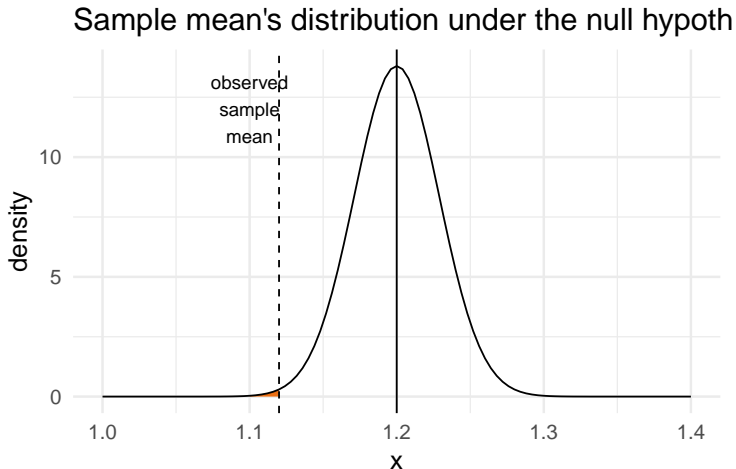
We need to evaluate the evidence *against* the null hypothesis that  $\mu = 1.2$  assuming that the null hypothesis is true.

If it is, then the sampling distribution of  $\bar{x}$  from  $n=12$  individuals has:

- ▶  $\mu = 1.2$  and
- ▶  $sd(\bar{x}) = \frac{\sigma}{\sqrt{n}} = 0.1/\sqrt{12} = 0.0289$ .

## Draw the sampling distribution under the null hypothesis:

What is the probability of observing the sample mean (or lower) we observed if the null hypothesis were true?



## Draw the sampling distribution under the null hypothesis:

What is the probability of observing the sample mean we observed *or lower* if the null hypothesis were true?

```
pnorm(q = 1.12, mean = 1.2, sd = 0.0289)
```

```
## [1] 0.002818636
```

Thus, under the null hypothesis, this is a 0.28% chance of observing a sample mean at least as small as what we saw. This is a very tiny probability, and suggests that the null hypothesis may not be true and that there is evidence in favor of the alternative hypothesis.

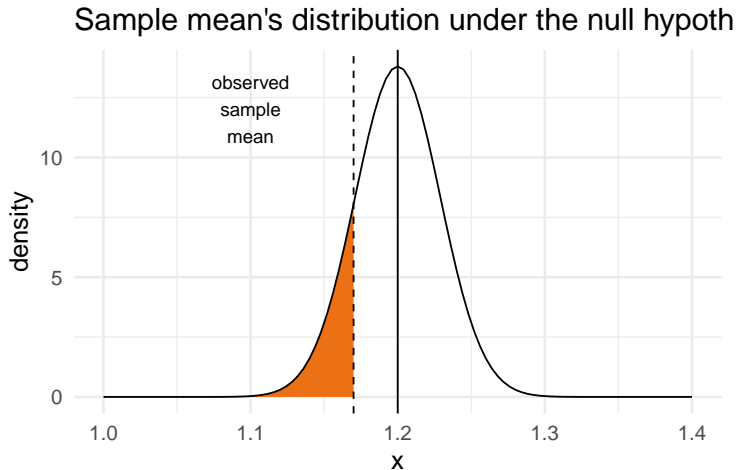
This probability is known as the **p-value**.

## Example, extended.

Suppose instead that for this example, we chose a sample such that  $\bar{x} = 1.17$ .

- ▶ What is the null and alternative hypotheses?
  - ▶  $H_0$ :
  - ▶  $H_a$ :
- ▶ Let's look at the distribution under the null hypothesis and add a line at  $\bar{x}$ .

## Example, extended.





## Example, extended.

- Calculate the probability of observing an observation of 1.17 or lower.

```
pnorm(q = 1.17, mean = 1.2, sd = 0.0289)
```

```
## [1] 0.1496205
```

- Do you think you could observe this  $\bar{x}$  if the null hypothesis were true?

# So far

- ▶ We have considered a one-sided  $H_a$ .
- ▶ We computed the p-value using the Normal distribution of the sampling mean
- ▶ Next: Two-sided alternative hypotheses

Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

**Two sided testing**

Hypothesis testing using Z  
tests

Definitions

## Two sided testing

# Hypothesis testing with a two-sided hypothesis

Suppose you are interested in the Aspirin content in a sample of pills. You have been told that the population of Aspirin tablets is Normally distributed and has a **known standard deviation of 5 mg**. Furthermore, you have a SRS of  $n = 10$  pills. You need to be able to detect if there is evidence that the average Aspirin content in your sample is different from the population average and would be concerned if it appears to be higher or lower. Here:

$$H_0 : \mu = 325mg$$

$$H_a : \mu \neq 325mg$$

This is a two-sided alternative hypothesis because we're interested in knowing if the sample mean appears to be either higher or lower.

Statistical Inference

Confidence intervals for the mean  $\mu$

Example: one tailed hypothesis

Two sided testing

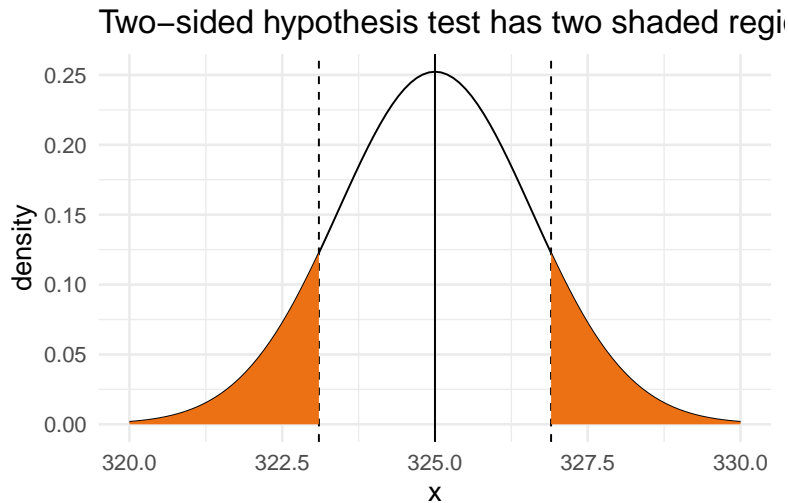
Hypothesis testing using Z tests

Definitions

# Hypothesis testing with a two-sided hypothesis

1. Check the conditions.
2. Calculate the sampling distribution, assuming the null hypothesis is true:
  - ▶  $\mu = 325$
  - ▶  $sd(\bar{x}) = \sigma/\sqrt{n} = 5/\sqrt{10} = 1.581139$
3. Calculate  $\bar{x}$  if provided data. Sketch the sampling distribution and add a vertical line at  $\bar{x}$ . Shade the region corresponding to  $H_a$ . If the hypothesis is two-sided, add another vertical line that is the same distance as  $\bar{x}$  is from  $\mu$  but on the other side of the distribution.

Two-sided alternative: add vertical lines at  $\bar{x}$  and the equivalent distance from the null on the other side of the distribution



# Calculation for the hypothesis testing with a two-sided hypothesis

```
pnorm(q = 326.9, mean = 325, sd = 1.581139, lower.tail = F)
```

```
## [1] 0.1147466
```

```
pnorm(q = 326.9, mean = 325, sd = 1.581139, lower.tail = F)*2
```

```
## [1] 0.2294932
```

- ▶ Why do we need to multiply the probability by 2?
- ▶ Why do we need to set `lower.tail=F`?

Interpret this probability. Does it provide evidence against the null hypothesis or, in the contrary, does this observation seem to follow under the null hypothesis?

Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

**Hypothesis testing using Z  
tests**

Definitions

## Hypothesis testing using Z tests



# Do you remember z-scores?

We briefly considered z-scores of the form:

$$z = \frac{x - \mu}{\sigma}$$

The z-score tells you how far  $x$  is from  $\mu$  in terms of units of standard deviation.

Originally, we considered z-scores to examine how far a baby's birthweight deviated from the average birthweight we expected at a given gestational age.

You could make a statement like: This birthweight is 2.5 standard deviations below the mean, which is very small.

## Do you remember z-scores?

We can generalize the z-score formula to look specifically at z-scores for  $\bar{x}$ , and use its  $\mu$  and the standard deviation of the sampling distribution:

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

Recall that after standardization of a Normal variable,  $z \sim N(0,1)$

## Go back to the earlier example looking at phosphorus

$$\bar{x} = 1.12$$

$$\mu = 1.2$$

$$\sigma/\sqrt{n} = 0.0289 \text{ Thus:}$$

$$z = \frac{1.12 - 1.2}{0.0289} = -2.768166$$

That is,  $z$  is 2.77 standard errors below the mean.

## Go back to the earlier example looking at phosphorus

What is the probability of observing a z-score of -2.768166 (or lower) on the standard Normal distribution?

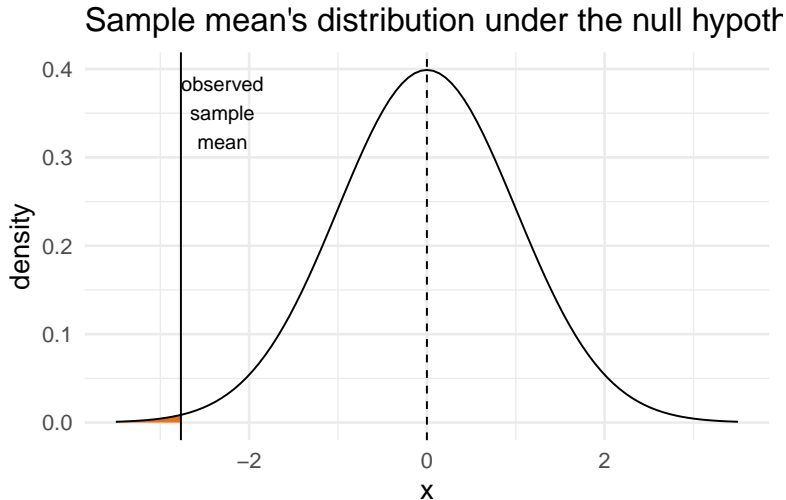
```
pnorm(q = -2.768166, mean = 0, sd = 1)
```

```
## [1] 0.002818637
```

The calculation is the same as before. We just standardized the units!

## Go back to the earlier example looking at phosphorus

And we can see that the graph looks the same as well



## Definitions

Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

Hypothesis testing using Z  
tests

**Definitions**

# Test statistic

The  $Z$  value we calculated in our example is a kind of test statistic.

**Test Statistic:** Measures how far the data diverge from the null hypothesis  $H_0$ . Large values of the statistic show that the data are far from what we would expect if  $H_0$  were true.

The  $Z$  is drawn from comparisons to a standard normal - we will learn about other tests statistics in part III

## Formalizing what we meant by “test statistic”

The Z test for a population mean: Draw a SRS from a Normal( $\mu$ ,  $\sigma$ ) population, where  $\mu$  is unknown, but  $\sigma$  is known. A test statistic and a p-value are obtained to test the null hypothesis that  $\mu$  has a specified value:

$H_0$ :

$$\mu = \mu_0$$

The one-sample z test statistic is:

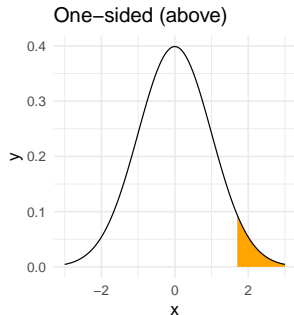
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$



## Z test for a population mean (continued)

As the variable  $Z$  follows the standard Normal distribution (i.e.,  $N(0,1)$ ), the p-value for a test of  $H_0$  against

$$H_a: \mu > \mu_0 \text{ is } P(Z \geq z)$$



Statistical Inference

Confidence intervals for the  
mean  $\mu$

Example: one tailed  
hypothesis

Two sided testing

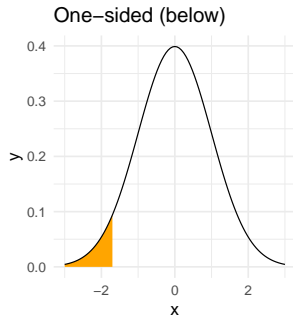
Hypothesis testing using Z  
tests

Definitions

## Z test for a population mean (continued)

As the variable  $Z$  follows the standard Normal distribution (i.e.,  $N(0,1)$ ), the p-value for a test of  $H_0$  against

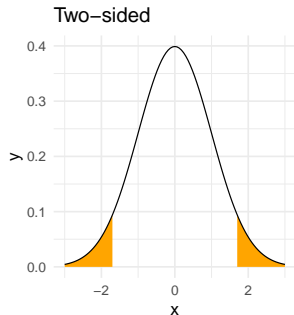
$$H_a: \mu < \mu_0 \text{ is } P(Z \leq z)$$



## Z test for a population mean (continued)

As the variable  $Z$  follows the standard Normal distribution (i.e.,  $N(0,1)$ ), the p-value for a test of  $H_0$  against

$$H_a: \mu \neq \mu_0 \text{ is } 2 \times P(Z \geq |z|)$$



# P-value

**P-value:** The probability, assuming that  $H_0$  is true, that the test statistic would take a value at least as extreme (in the direction of  $H_a$ ) as that actually observed. The smaller the p-value, the stronger the evidence against  $H_0$  provided by the data.

# Interpretation of the p-value

Remember!

We never conclude that the null hypothesis is true, only that there is “no evidence” against the null hypothesis.

We also never conclude that the alternative hypothesis is true, only that there **is** evidence to reject the null hypothesis.

# Definition: Significance level and statistically significant

- ▶ The **significance level**,  $\alpha$ , is an arbitrary threshold that can be used when reporting whether a p-value is “statistically significant”.
- ▶ If the p-value is as small or smaller than  $\alpha$ , people say that the data are statistically significant at level  $\alpha$ .

# Definition: Significance level and statistically significant

- ▶ Many researchers dislike the use of a significance level because it is a completely arbitrary cutpoint. If you are going to report p-values, it is much better to report the magnitude of the p-value than to only report whether the p-value is statistically significant or not.
- ▶ Both  $p\text{-value} = 0.03$  and  $p\text{-value} = 0.004$  are “significant at  $\alpha = 0.05$ ”, but the latter provides more evidence against the null hypothesis.
- ▶ Thus it is more informative to report the p-value you calculated than to only make a statement regarding statistical significance.

## Reminder Statistical significance $\neq$ clinical significance

- ▶ If a finding is “statistically significant” this does not mean it is “clinically significant”, or of a meaningful magnitude.
- ▶ For example, you might find that  $\bar{x} = 1.19$  is statistically different from  $\mu = 1.2$  (say  $\alpha = 0.05$ ), but the estimated difference is only  $1.2 - 1.19 = 0.01$ . This might not be a meaningful difference.
- ▶ Whether the difference is of a meaningful magnitude in practice is not determined by the data, but based on judgement of decision-makers.
- ▶ What is a meaningful reduction in the number of drinks per day for an alcoholic?
- ▶ What is a meaningful reduction in depressive symptoms associated with a new very expensive treatment vs. a current cheaper treatment?



# Relationship between confidence intervals and test statistics

- ▶ A two-sided test statistic at significance level  $\alpha$  can be carried out from a confidence interval with confidence level  $C = 1 - \alpha$ .
- ▶ For example, if a 95% confidence interval does not contain the null value, then this implies that the p-value for the test that  $H_0 : \bar{x} = \mu$  has a p-value  $< \alpha = 0.05$  and is therefore statistically significant at the 5% level.

## Example: Relationship between confidence intervals and test statistics

Recall earlier in the slides that we calculated the 95% CI for the mean height of girls, based on a sample of girls in a Midwestern school district. This CI was from 100.56 to 111.12.

Suppose you wanted to test whether the mean height was different from  $H_0 : \mu = 113$  cm. This mean height is outside of the 95% CI, so we know that the p-value corresponding to the two-sided hypothesis test would be  $< 5\%$ , and we could conclude that it is statistically significant for  $\alpha = 0.05$ .

# Relationship between confidence intervals and test statistic

CIs and p-values provide similar information, because you can deduce directly whether a test will be  $< 0.05$  from a 95% CI.

However, if you only know a p-value you cannot derive the CI.

The CI is better because it puts a range around the magnitude of the value of the parameter.

# Parting Humor

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	