Chapter 4: Foundations for inference

Yingbo Li

Southern Methodist University

STAT 2331

Outline

1. Variability in estimates and central limit theorem
2. Confidence intervals
3. Hypothesis testing

Parameter estimation

- We are often interested in *population parameters*.

$$\mu,\ \sigma,\ p,\ \ldots$$

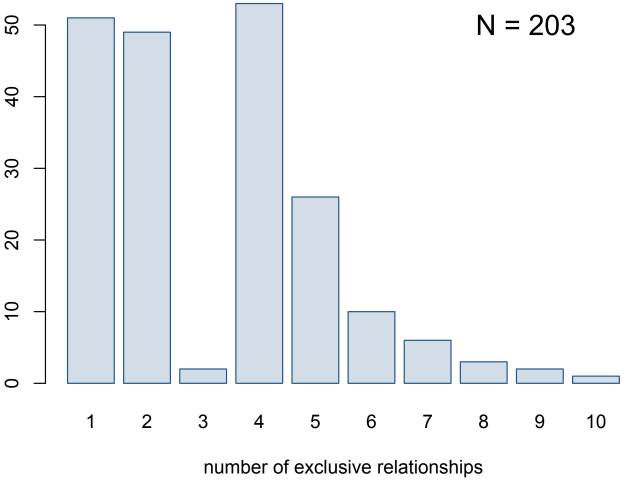
- Since complete populations are difficult (or impossible) to collect data on, we use *point estimates* from samples to estimate parameters.

$$\bar{x},\ s,\ \hat{p},\ \ldots$$

- Point estimates *vary from sample to sample*, and quantifying how they vary gives a way to estimate the *margin of error* associated with our point estimate.
- But before we get to quantifying the variability among samples, let's try to understand how and why point estimates vary from sample to sample.

Estimating number of exclusive relationships

We would like to estimate the number of exclusive relationships stats students have been in, and we actually have the population data:



Sampling scheme

Suppose that you don't have access to the population data. In order to estimate the average number of relationships, you might sample from the population and use your sample mean as the best guess for the unknown population mean.

1. Sample, with replacement, ten students and record the number of exclusive relationships they reported.
2. Find the sample mean and record it.

If we randomly select observations from this data set, which values are most likely to be selected, which are least likely?

3. Repeat Step 1-2, for 100 times, then we get 100 sample means, one for each sample.

Sample means vary from sample to sample

1	2	21	1	41	2	61	2	81	4	101	4	121	7	141	1	161	4	181	2	201	2
2	2	22	5	42	9	62	2	82	4	102	1	122	1	142	4	162	1	182	2	202	2
3	2	23	4	43	1	63	5	83	4	103	5	123	4	143	2	163	1	183	4	203	5
4	7	24	5	44	1	64	7	84	2	104	2	124	1	144	1	164	4	184	4		
5	2	25	4	45	4	65	1	85	6	105	2	125	5	145	2	165	1	185	5		
6	1	26	6	46	4	66	1	86	5	106	7	126	5	146	1	166	5	186	1		
7	2	27	1	47	2	67	1	87	2	107	4	127	4	147	2	167	2	187	2		
8	2	28	1	48	4	68	2	88	1	108	4	128	2	148	2	168	4	188	2		
9	4	29	4	49	4	69	4	89	5	109	4	129	4	149	3	169	4	189	2		
10	2	30	2	50	1	70	1	90	3	110	4	130	1	150	5	170	8	190	2		
11	4	31	4	51	2	71	1	91	2	111	2	131	1	151	9	171	1	191	4		
12	1	32	1	52	1	72	1	92	4	112	8	132	6	152	1	172	2	192	4		
13	5	33	1	53	5	73	4	93	1	113	4	133	5	153	6	173	1	193	1		
14	5	34	5	54	1	74	6	94	1	114	6	134	4	154	2	174	1	194	5		
15	5	35	4	55	1	75	6	95	2	115	2	135	4	155	7	175	2	195	4		
16	4	36	1	56	4	76	7	96	5	116	2	136	4	156	8	176	6	196	2		
17	1	37	4	57	4	77	1	97	10	117	2	137	1	157	5	177	2	197	4		
18	5	38	2	58	4	78	1	98	2	118	5	138	1	158	6	178	2	198	4		
19	4	39	4	59	5	79	1	99	4	119	5	139	4	159	4	179	2	199	1		
20	4	40	2	60	1	80	1	100	5	120	6	140	2	160	1	180	4	200	1		

1st sample: $(9+4+4+2+4+4+1+5+4+4)/10 = 4.1$

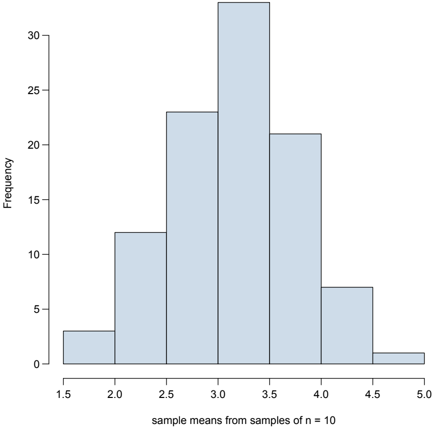
2nd sample: $(1+7+2+5+2+4+4+4+2+5)/10 = 3.6$

$\vdots$

100th sample: $(9+1+2+1+5+1+1+2+1+6)/10 = 2.9$

Sampling distribution

The *Sampling distribution* represents the distribution of the point estimates based on samples of a fixed size.



This histogram is plotted based on the 100 sample means. What is the shape and center of this distribution?

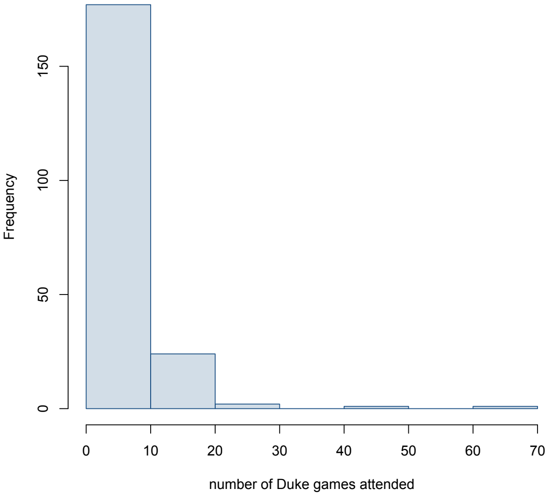
Shape: unimodal, symmetric
Center: between the 3.0 and 3.5

Based on this distribution what do you think is the true population average?

$\mu = 3.207$

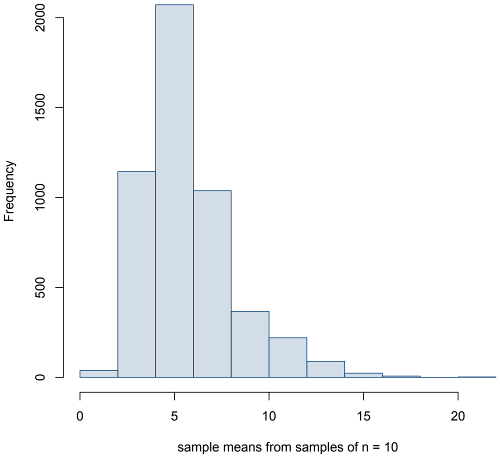
Number of basketball games attended

Next let's look at the population data for the number of basketball games attended:



Number of basketball games attended

Sampling distribution, n = 10:



What does each observation in this distribution represent?

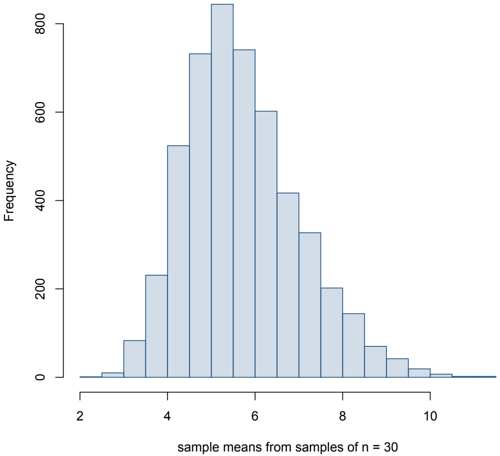
Sample mean, $\bar{x}$, of samples of size $n = 10$.

Is the variability of the sampling distribution smaller or larger than the variability of the population distribution? Why?

Smaller, sample means will vary less than individual observations.

Number of basketball games attended

Sampling distribution, n = 30:

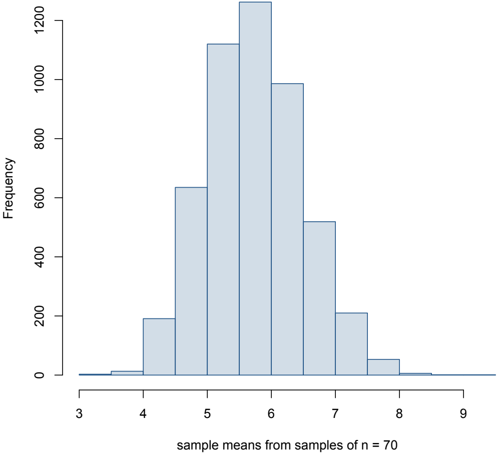


How did the shape, center, and spread of the sampling distribution change going from $n = 10$ to $n = 30$?

Shape is more symmetric, center is about the same, spread is smaller.

Approximately normal!

Sampling distribution, n = 70:



Number of basketball games attended

Question

1. The mean of the sampling distribution is 5.75, and the standard deviation of the sampling distribution (also called the *standard error*) is 0.75. Which of the following is the most reasonable guess for the 95% confidence interval for the true average number of basketball games attended by stats students?

(a) 5.75 ± 0.75

(b) $5.75 \pm 2 \times 0.75 \rightarrow (4.25, 7.25)$

(c) $5.75 \pm 3 \times 0.75$

(d) cannot tell from the information given

Central limit theorem

Central limit theorem

If a sample consists of at least 30 independent observations and the data are not extremely skewed, then the distribution of the sample mean is well approximated by a normal model.

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

- So it wasn't just a coincidence that the sampling distributions we saw earlier were symmetric.
- We won't go into the proving why $SE = \frac{\sigma}{\sqrt{n}}$, but note that as n increases SE decreases.
- This is because as the sample size increases we would expect samples to yield more consistent sample means, hence the variability among the sample means would be lower.

Related video clips

- Bunnies, Dragons and the 'Normal' World
  How to estimate the average weight of bunnies? And how (un)certain is the estimate?
  From CreatureCast.org
  https://www.youtube.com/watch?v=jvoxEYmQHNM
- Wisdom of the crowd
  Ask enough people to estimate something, and the average of all their guesses will get you surprisingly close to the right answer.
  From NOVA scienceNOW.
  http://www.youtube.com/watch?v=r-FonWBEb0o

Confidence intervals

- A plausible range of values for the population parameter is called a *confidence interval*.
- Using only a point estimate to estimate a parameter is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net.

We can throw a spear where we saw a fish but we will probably miss. If we toss a net in that area, we have a good chance of catching the fish.

- If we report a point estimate, we probably will not hit the exact population parameter. If we report a range of plausible values – a confidence interval – we have a good shot at capturing the parameter.

Earlier....

- We built a *sampling distribution* is by drawing many samples from the population, finding their means, and plotting them.
- In reality, we rarely have access to the population.
- So often we need to make inference about population parameters, e.g. estimate the population mean using a confidence interval, using point estimates from just this one sample.

Confidence intervals - a more realistic look

Suppose we sampled 50 students and asked them how many exclusive relationships they've had so far. This sample yielded a mean of 3.2 ($\bar{x} = 3.2$) and a standard deviation of 1.74 ($s = 1.74$). How can we construct a confidence interval for the true average number of exclusive relationships using this sample?

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

For our class,

$\bar{x} = 2.74,$

$s = 3.00.$

Since we don't know σ, $SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25.$

$$\begin{aligned} \bar{x} \pm 2 \times SE &= 3.2 \pm 2 \times 0.25 \\ &= (3.2 - 0.5, 3.2 + 0.5) \\ &= (2.7, 3.7) \end{aligned}$$

Interpret confidence interval

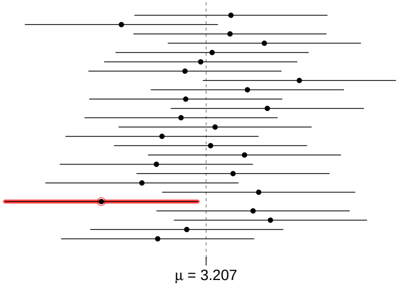
General format:
We are 95% confident that *population parameter* is between *(a, b)*.

In context of this question:
We are 95% confident that *the average number of exclusive relationships among students* is between *2.7 to 3.7*.

What does 95% confident mean?

- Suppose we took many samples and built a confidence interval from each sample using the equation $point\ estimate \pm 2 \times SE$.
- Then about 95% of those intervals would contain the true population mean (μ).

- The figure on the left shows this process with 25 samples, where 24 of the resulting confidence intervals contain the true average number of exclusive relationships, $\mu = 3.207$, and one does not.

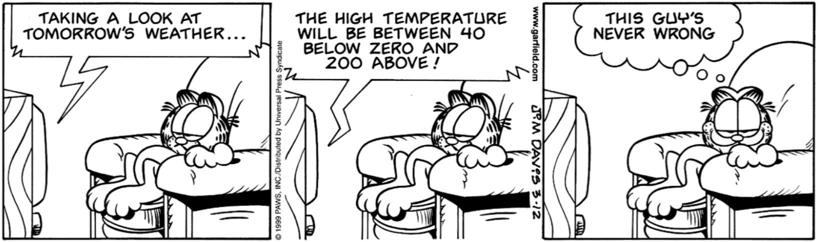


Width of an interval

If we want to be very certain that we capture the population parameter, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?



If the interval is too wide it may not be very informative.

Estimating sleep

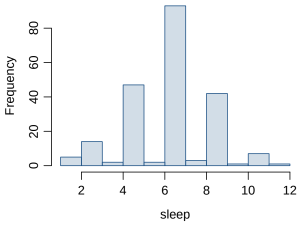
We asked students how many hours of sleep they get per night. A sample of 217 respondents yielded an average of 6.7 hours of sleep with a standard deviation of 2.03 hours. Assuming that this sample is random and representative of all students *(might be leap of faith!)*, construct a 95% confidence interval for the average amount of sleep students get per night.

CLT states that sample means will be nearly normally distributed, and the standard error of the sampling distribution can be estimated by $\frac{s}{\sqrt{n}}$. However there are certain assumptions and conditions that must be verified in order for the CLT to apply.

Assumptions & conditions for inference

1. *Independence Assumption:*
   - *Random sampling condition:* We are assuming that this sample is random.
   - *10% Condition:* $217 < 10\%$ of all college students.

   We can assume that how much sleep one student in this sample gets is independent of another.
2. *Normality:* The sample data has a symmetric distribution, i.e., it is not highly skewed. In addition, $n > 30$, so we can assume that the sampling distribution will be approximately normal.



An approximate interval for sleep

An approximate confidence interval for the average amount of sleep college students get can be calculated as

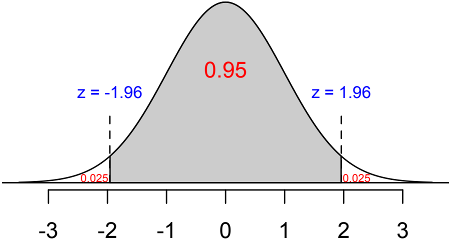
$$6.7 \pm \left(2 \times \frac{2.03}{\sqrt{217}}\right) = 6.7 \pm (2 \times 0.14) = (6.42, 6.98)$$

But we can actually obtain a confidence interval that's a little more accurate.

Question

2. Which of the below Z scores mark the cutoff for the middle 95% of a normal distribution?

(a) $Z = -3.49$ and $Z = 1.65$
(b) $Z = -2.58$ and $Z = 2.58$
(c) $Z = -1.96$ *and* $Z = 1.96$
(d) $Z = -1.65$ and $Z = 1.65$



A more accurate 95% confidence interval

Calculate an exact 95% confidence interval for the average sleep college students get per night.

$$\bar{x} = 6.7, s = 2.03, n = 217$$

$$\begin{aligned} \bar{x} \pm z^{\star} \times \frac{s}{\sqrt{n}} &= 6.7 \pm \left(1.96 \times \frac{2.03}{\sqrt{217}} \right) \\ &= 6.7 \pm 0.27 \\ &= (6.43, 6.97) \end{aligned}$$

Note: We used the approximate confidence interval to introduce this concept and to illustrate how it relates to the 68-95-99.7% rule. When asked for a confidence interval you should calculate it using this more accurate approach.

Changing the confidence level

Confidence interval, a general formula

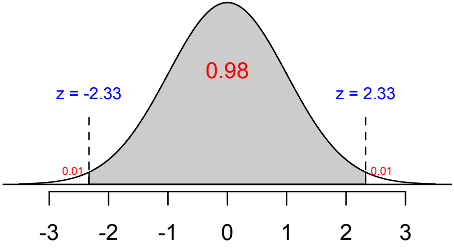
$$point\ estimate \pm z^{\star} \times SE$$

- In order to change the confidence level all we need to do is adjust $z^{\star}$ in the above formula.
- Commonly used confidence levels in practice are 90%, 95%, 98%, and 99%.
- However, using the z table it is possible to find the appropriate $z^{\star}$ for any confidence level.

A 98% confidence interval

Calculate a *98% confidence* interval for the average sleep college students get per night.

$$\bar{x} = 6.7, s = 2.03, n = 217$$



$$= \bar{x} \pm z^{\star} \times SE$$

$$= 6.7 \pm \left(2.33 \times \frac{2.03}{\sqrt{217}} \right)$$

$$= (6.7 - 0.32, 6.7 + 0.32)$$

$$= (6.38, 7.02)$$

Question

3. Which of the following is correct?

(a) 98% of college students sleep between 6.38 and 7.02 hours per night, on average.

(b) *We are 98% confident that college students on average sleep 6.38 to 7.02 hours per night.*

(c) 98% of the time college students sleep 6.38 hours to 7.02 hours per night.

(d) We are 98% confident that the average sleep the 217 college students in this sample get is between 6.38 and 7.02 hours per night.

(e) The standard error is 0.32 hours.

Testing claims based on a confidence interval

Question

4. The 95% confidence interval for the average hours of sleep college students get was (6.43, 6.97). Does this provide convincing evidence that college students do not get the 8 hours of sleep recommended by the CDC?

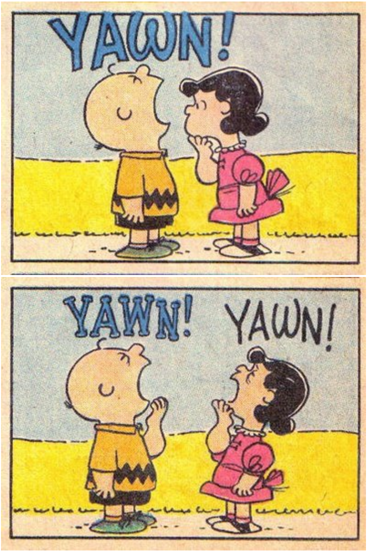
(a) *Yes*

(b) No

Testing claims based on a confidence interval (cont.)

- Using a confidence interval for hypothesis testing might be insufficient in some cases since it gives a yes/no (reject/don't reject) answer, as opposed to quantifying our decision with a probability.
- Formal hypothesis testing allows us to report a probability along with our decision.

Is yawning contagious?



- Do you think yawning is contagious?
- An experiment conducted by the MythBusters tested if a person can be subconsciously influenced into yawning if another person near them yawns.

  http://www.discovery.com/tv-shows/mythbusters/videos/is-yawning-contagious-minimyth/

MythBusters yawning experiment

In this study 50 people were randomly assigned to two groups: 34 to a group where a person near them yawned (treatment) and 16 to a control group where they didn't see someone yawn (control).

	Treatment	Control	Total
Yawn	10	4	14
Not Yawn	24	12	36
Total	34	16	50

$$\hat{p}_{treatment} = \frac{10}{34} = 0.29$$

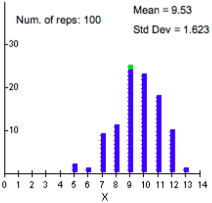
$$\hat{p}_{control} = \frac{4}{16} = 0.25$$

Possible explanations:

- Yawning is *independent* of seeing someone else yawn; therefore, the difference between the proportions of yawners in the control and seeded groups is *due to chance*. $\rightarrow$ *nothing is going on*
- Yawning is *dependent* on seeing someone else yawn; therefore, the difference between the proportions of yawners in the control and seeded groups is *real*. $\rightarrow$ *something is going on*

With a slightly different terminology

- We started with the assumption that yawning is independent of seeing someone else yawn. $\rightarrow$ *null hypothesis*
- We then investigated how the results would look if we simulated the experiment many times assuming the null hypothesis is true. $\rightarrow$ *testing*



- Since the simulation results were similar to the actual data (on average roughly 10 people yawning in the treatment group), we decided not to reject the null hypothesis in favor of the *alternative hypothesis*.

A trial as a hypothesis test

- Hypothesis testing is very much like a court trial.
  - In a trial, the burden of proof is on the prosecution.
  - In a hypothesis test, the burden of proof is on the unusual claim.
- H_0 : Defendant is innocent
  H_A : Defendant is guilty
- Collect data - The null hypothesis is the ordinary state of affairs (the status quo), so it's the alternative hypothesis that we consider unusual (and for which we must gather evidence).
- Then we judge the evidence - "Could these data plausibly have happened by chance if the null hypothesis were true?"
  - If they were very unlikely to have occurred, then the evidence raises more than a reasonable doubt in our minds about the null hypothesis.
- Ultimately we must make a decision. How unlikely is unlikely?

A trial as a hypothesis test (cont.)

- If the evidence is not strong enough to reject the presumption of innocence, the jury returns with a verdict of "not guilty".
  - The jury does not say that the defendant is innocent, just that there is not enough evidence to convict.
  - The defendant may, in fact, be innocent, but the jury has no way of being sure.
- Said statistically, we fail to reject the null hypothesis.
  - We never declare the null hypothesis to be true, because we simply do not know whether it's true or not.
  - Therefore we never "accept the null hypothesis".

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

We'll formally introduce the hypothesis testing framework using an example on testing a claim about a population mean.

Grade inflation?

In 2001 the average GPA of students at a university was 3.37. Recently we asked stats students in this school about their GPA. A sample of 203 respondents yielded an average GPA of 3.59 with a standard deviation of 0.28. Assuming that this sample is random and representative of all students *(another leap of faith!)*, do these data provide convincing evidence that the average GPA has changed?

gradeinflation.com

Setting the hypotheses

- The parameter of interest is the average GPA of current students.
- There may be two explanations why our sample mean is higher than the average GPA from 2001.
  - The true population mean has changed.
  - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.
- We start with the assumption that nothing has changed.

$$H_0 : \mu = 3.37$$

- We test the claim that average GPA has changed.

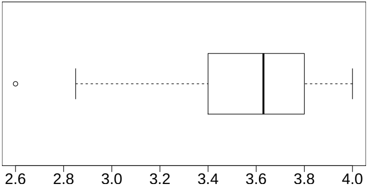
$$H_A : \mu \neq 3.37$$

Assumptions & conditions for inference

Before doing inference using this data set, we must make sure that the assumptions & conditions necessary for inference are satisfied:

1. *Independence Assumption:*
   - *Random sampling condition:* Assuming this sample is random.
   - *10% Condition:* $203 < 10\%$ of all current students in this university.

   We can assume that GPA of one student in this sample is independent of another.
2. *Normality condition:* The distribution appears to be slightly skewed (but not extremely) and $n > 30$ so we can assume that the distribution of the sample means is nearly normal.



Grade inflation in SMU?

In 2007 the average GPA of students at SMU was 3.13. From our class survey, a sample of 19 respondents yielded an average GPA of 3.39 with a standard deviation of 0.49. Assuming that this sample is random and representative of all students, do these data provide convincing evidence that the average GPA has changed over the last decade?

Are all conditions satisfied?

Sample size $n = 19 < 30!$

(We will learn this in Chapter 5.)

gradeinflation.com

Number of college applications - hypotheses

Question

5. The same survey asked how many colleges students applied to, and 206 students responded to this question. This sample yielded an average of 9.7 college applications with a standard deviation of 7. College Board website states that counselors recommend students apply to roughly 8 colleges. Which of the following are the correct set of hypotheses to test if these data provide convincing evidence that the average number of colleges all students in this university apply to is higher than recommended.

(a) $H_0 : \mu = 9.7$
$H_A : \mu > 9.7$

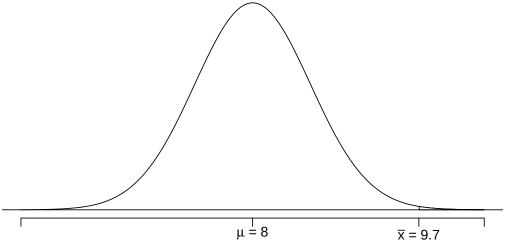
(b) $H_0 : \mu = 8$
$H_A : \mu > 8$

(c) $H_0 : \bar{x} = 8$
$H_A : \bar{x} > 8$

(d) $H_0 : \mu = 8$
$H_A : \mu > 9.7$

http://www.collegeboard.com/student/apply/the-application/151680.html

Number of college applications - test statistic



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

The sample mean is 3.4 standard errors away from the hypothesized value. Is this considered unusually (significantly) high?

Yes, but we can quantify how unusual it is using a p-value.

p-values

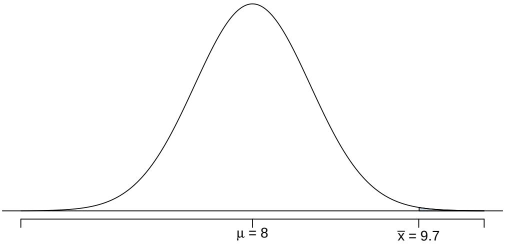
- The *p-value* is the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis is true.

$$\text{p-value} = P(\text{current data or more extreme})$$

- If the p-value is *low* (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject* H_0.
- If the p-value is *high* (higher than α) we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject* H_0.
- We never accept H_0 since we're not in the business of trying to prove it. We simply want to know if the data provide convincing evidence to support H_A.

Number of college applications - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 was true (the true population mean was 8).

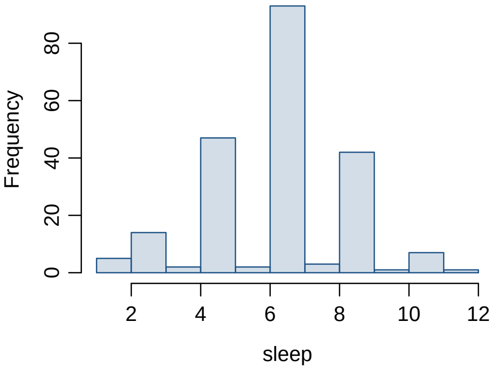


$$P(\bar{x} > 9.7) = P(Z > 3.4) = 0.0003$$

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges students applied to is 8, there is only 0.03% chance of observing a random sample of 206 students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject* H_0.
- The data provide convincing evidence that students average apply to more than 8 schools.
- The difference between the null value of 8 schools and observed sample mean of 9.7 schools is *not due to chance* or sampling variability.

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 217 students yielded an average of 6.7 hours, with a standard deviation of 2.03 hours. Assuming that this is a random sample representative of all college students *(bit of a leap of faith?)*, do these data provide convincing evidence that all students on average sleep less than 7 hours per night?



Setting the hypotheses

A sample of 217 students yielded an average of 6.7 hours, with a standard deviation of 2.03 hours.

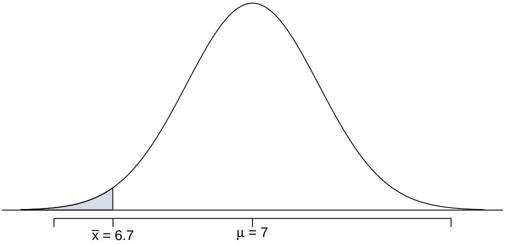
$$H_0 : \mu = 7 \text{ (students sleep 7 hours per night on average)}$$
$$H_A : \mu < 7 \text{ (students sleep less than 7 hours per night on average)}$$

- If in fact the null hypothesis is true, $\bar{x}$ is distributed nearly normally with mean $\mu = 7$ and standard error

$$SE = \frac{s}{\sqrt{n}} = \frac{2.03}{\sqrt{217}} = 0.14$$

- We would like to find out how likely it is to observe a sample mean at least as far from the data as our current sample mean (6.7), if in fact the null hypothesis is true.

Calculating the p-value



$$\bar{x} \sim N\left(\mu = 7, SE = \frac{2.03}{\sqrt{217}} = 0.14\right)$$

$$Z = \frac{6.7 - 7}{0.14} = -2.14$$

p-value $= P(\bar{x} < 6.7) = P(Z < -2.14) = 0.0162$

Question

6. Based on a p-value of 0.0162, which of the following is true?
$H_0 : \mu = 7$ (students sleep 7 hours per night on average)
$H_A : \mu < 7$ (students sleep less than 7 hours per night on average)

(a) Fail to reject H_0, the data provide convincing evidence that students sleep less than 7 hours on average.

(b) *Reject H_0, the data provide convincing evidence that students sleep less than 7 hours on average.*

(c) Reject H_0, the data prove that students sleep more than 7 hours on average.

(d) Fail to reject H_0, the data do not provide convincing evidence that students sleep less than 7 hours on average.

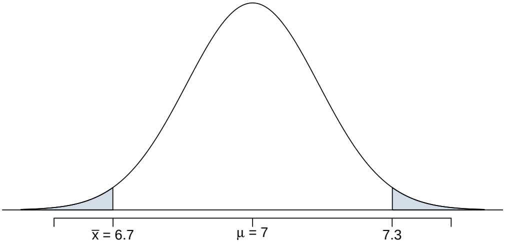
(e) Reject H_0, the data provide convincing evidence that students in this sample sleep less than 7 hours on average.

Two-sided hypothesis testing with p-values

- If the research question was "Do the data provide convincing evidence that the average amount of sleep students get per night is *different* than the national average?", the alternative hypothesis would be different.

$$H_0 : \mu = 7$$
$$H_A : \mu \neq 7$$

- Hence the p-value would change as well:



p-value
$= 0.0162 \times 2$
$= 0.0324$

Recap: Hypothesis testing framework

1. Set the hypotheses.
2. Check assumptions and conditions.
3. Calculate a *test statistic* and a p-value.
4. Make a decision, and interpret it in context of the research question.

Recap: Hypothesis testing for a population mean

1. Set the hypotheses
   - $H_0 : \mu = null\ value$
   - $H_A : \mu <$ or $>$ or $\neq null\ value$
2. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Normality: $n \geq 30$, no extreme skew
3. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{x} - \mu}{SE}, \text{ where } SE = \frac{s}{\sqrt{n}}$$

4. Make a decision, and interpret it in context of the research question
   - If p-value $< \alpha$, reject H_0, data provide evidence for H_A
   - If p-value $> \alpha$, do not reject H_0, data do not provide evidence for H_A

Decision errors

- Hypothesis tests are not flawless.
- In the court system innocent people are sometimes wrongly convicted and the guilty sometimes walk free.
- Similarly, we can make a wrong decision in statistical hypothesis tests as well.
- The difference is that we have the tools necessary to quantify how often we make errors in statistics.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make some decision about which one might be true, but we might choose incorrectly.

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is actually true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is actually true.
- Of course, we (almost) never find out if H_0 or H_A is actually true, but we must entertain the possibility that they might be.

Question

7. Earlier we discussed the example of a trial as a hypothesis test. The hypotheses were

$$H_0: \text{ Defendant is innocent}$$
$$H_A: \text{ Defendant is guilty}$$

Which of the below describes a Type 1 error in this context?

(a) *Deciding that the defendant is guilty when the defendant is actually innocent.*

(b) Deciding that the defendant is not guilty when the defendant is actually guilty.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error.

$$\alpha = P(Type\ 1\ error)$$

- This is why we like keeping α low – increasing α would increase the Type 1 error rate as well.