Likelihood and Bayesian Inference

Lecture 10A handout

Name: _____

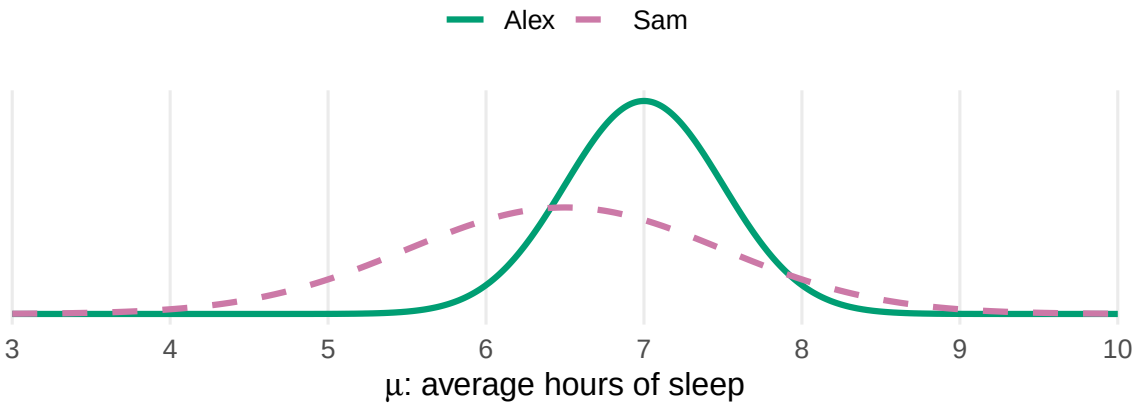
Normal distribution for a population parameter

On Monday, in frequentist inference, a population parameter like π was a _____.

Another way to think about population parameters: we can use a probability distribution to describe _____ about them. Today we will work with a population **mean**, μ, and use a normal distribution to describe what we believe about it.

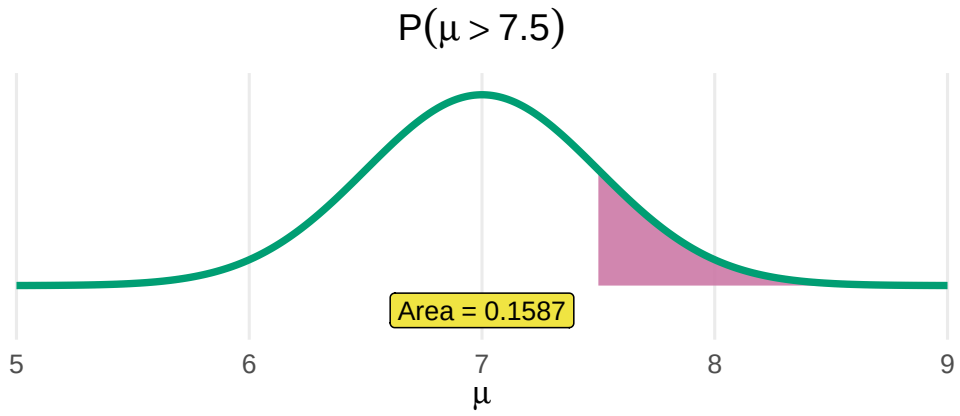
Example. μ: average hours of sleep of Harvard students on a school night.

- Alex believes $\mu \sim N(7, 0.5^2)$: "around 7 hours, fairly sure."
- Sam believes $\mu \sim N(6.5, 1^2)$: "around 6.5 hours, not very sure."



Who is more certain about μ? How can you tell from the plot?

Probability of μ. According to Alex's belief, what is the probability that the average hours of sleep of Harvard students is more than 7.5? The plot below shows Alex's belief, $N(7, 0.5^2)$, with the area above 7.5 shaded. Write the probability in notation and give its value.



Your turn. Use the 68-95-99.7 rule.

a. According to Sam's belief, $N(6.5, 1^2)$, what is the probability that μ is more than 7.5?

b. Who thinks it is more plausible that μ is less than 4.5 hours? Find both probabilities.

c. Write down your own belief about μ as $\mu \sim N(\ ?,\ ?^2)$. Sketch it, and compare with your neighbor.

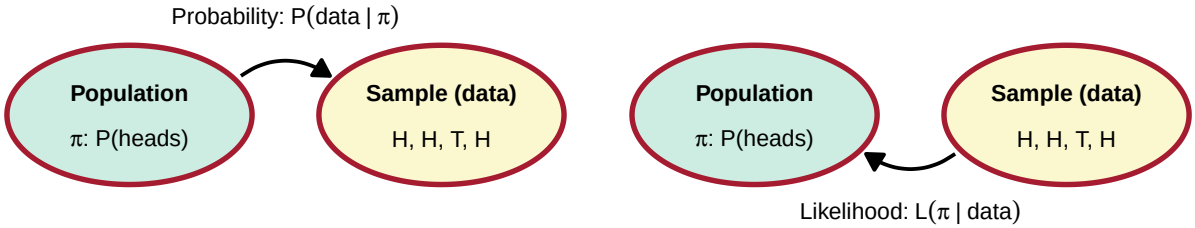
Same notation, different things. Does each normal distribution describe a *random observation*, a *random sample statistic*, or a *random population parameter*? Then use the 68-95-99.7 rule to find the probability in brackets.

Table 1: Is each normal distribution about a random observation, a random sample statistic, or a random population parameter?

Normal distribution	Observation, sample statistic, or parameter?	Probability
$X \sim N(500, 100^2)$: SAT Math score of a randomly selected test taker $[P(X > 600)]$	_____	_____
$\pi \sim N(0.6, 0.1^2)$: our belief about the proportion of all Harvard students who drink coffee daily $[P(\pi > 0.5)]$	_____	_____
$p \mid \pi = 0.5 \sim N(0.5, 0.125^2)$: proportion of 16 Randomville babies who choose the helper $[P(p > 0.75)]$	_____	_____
$X \sim N(70, 10^2)$: resting heart rate of a randomly selected student $[P(X > 80)]$	_____	_____
$\mu \sim N(3.7, 0.1^2)$: our belief about the mean GPA of all Harvard students $[P(\mu < 3.6)]$	_____	_____
$\bar{X} \sim N(7, 0.2^2)$: mean hours of sleep of a random sample of 36 students, if the mean for all students is 7 $[P(\bar{X} < 6.6)]$	_____	_____

Probability vs. likelihood

Flashback: lecture 1. We asked: "If we observe H, H, H, H, T on five coin flips, what is the likelihood that this is a fair coin?" Back then we did not have the words **population**, **sample**, **parameter**, or **likelihood**. Now we do!



- **Probability** goes from the _____: if we knew π, how probable is the data? $P(\text{data} \mid \pi)$
- **Likelihood** goes from the _____: given the data, how well does each value of π explain it? $L(\pi \mid \text{data})$

Is this coin fair? A friend hands you a coin. Let π be the probability that it lands **heads**. You flip it 4 times and get **H, H, T, H**. Is the coin fair? Or is π something else, like 0.75?

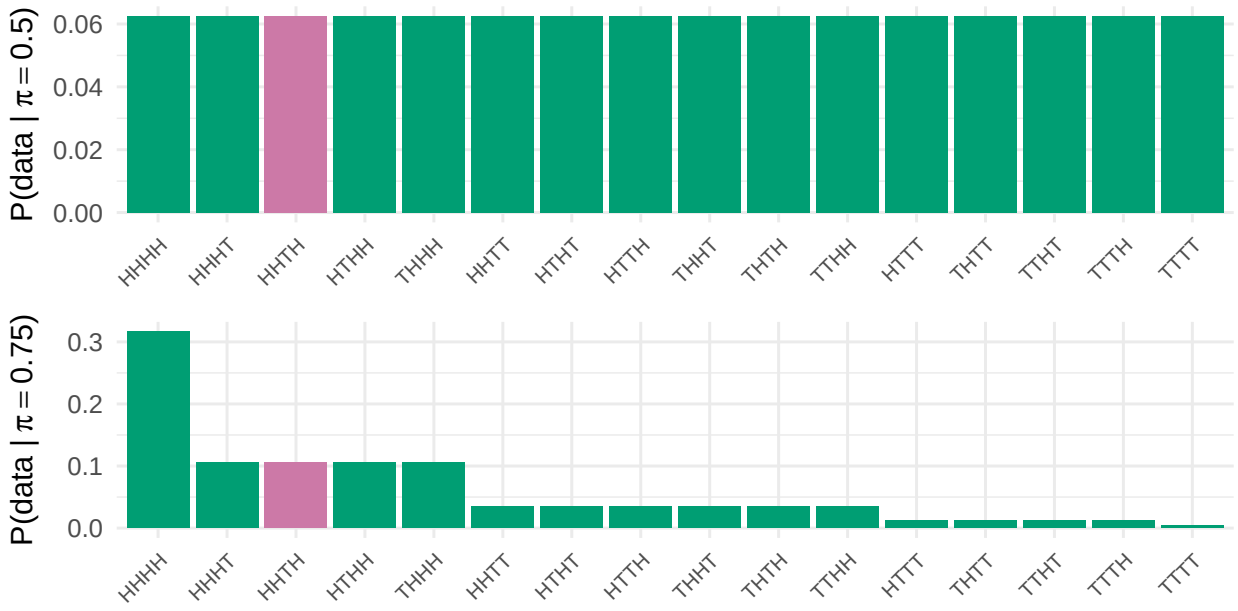
If we knew π: $P(\text{H} \mid \pi) =$ _____ and $P(\text{T} \mid \pi) =$ _____.

The flips are independent, so we multiply. Write $P(\text{H, H, T, H} \mid \pi)$ in terms of π:

Two questions.

- **Probability question:** If the coin is fair ($\pi = 0.5$), how probable is each possible set of 4 flips?
- **Likelihood question:** We **observed** H, H, T, H. How well does each possible value of π (0.5? 0.75? 0.25?) explain these flips?

Probability: fix π, vary the data. If the coin is fair, each of the 16 possible sets of 4 flips has probability $0.5^4 =$ _____. If the coin were biased toward heads, $\pi = 0.75$, each set of 4 flips with k heads has probability $0.75^k \times 0.25^{4-k}$:



In each plot, the 16 probabilities add up to _____. Each is a _____ (over the data).

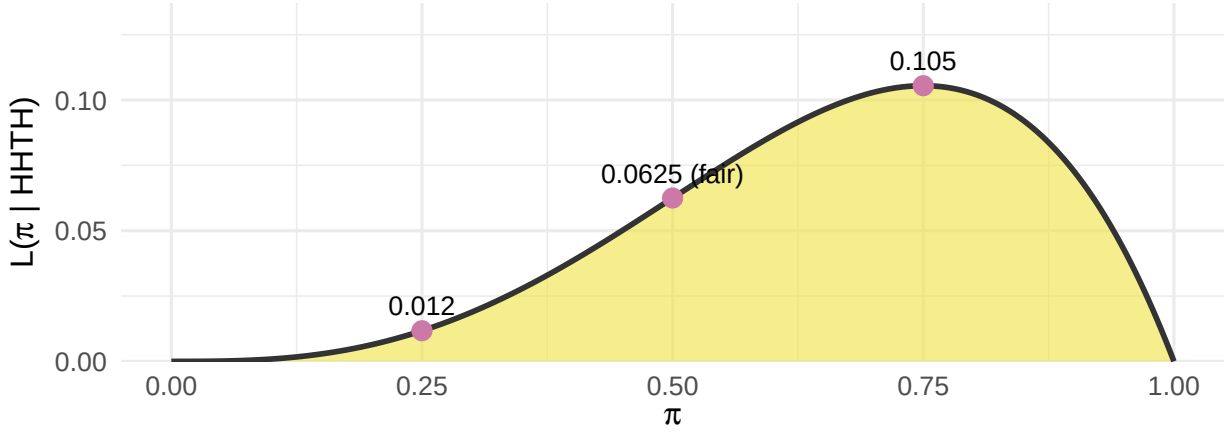
Likelihood: fix the data, vary π. We observed H, H, T, H. Complete the table. Hint: $0.25^3 \approx 0.016$ and $0.75^3 \approx 0.42$.

Table 2: Likelihood of three values of π given the observed flips H, H, T, H

π	$L(\pi \mid \text{HHTH}) = \pi^3(1-\pi)$
0.25 (biased toward tails)	_____
0.5 (fair)	_____
0.75 (biased toward heads)	_____

How many times more likely is $\pi = 0.75$ than a fair coin? _____

Do these three likelihoods add up to 1? _____ This is a _____ (over π). It is **not** a probability distribution.



The value of π with the highest likelihood, _____, is the sample proportion $\frac{3}{4}$. It is called the _____.

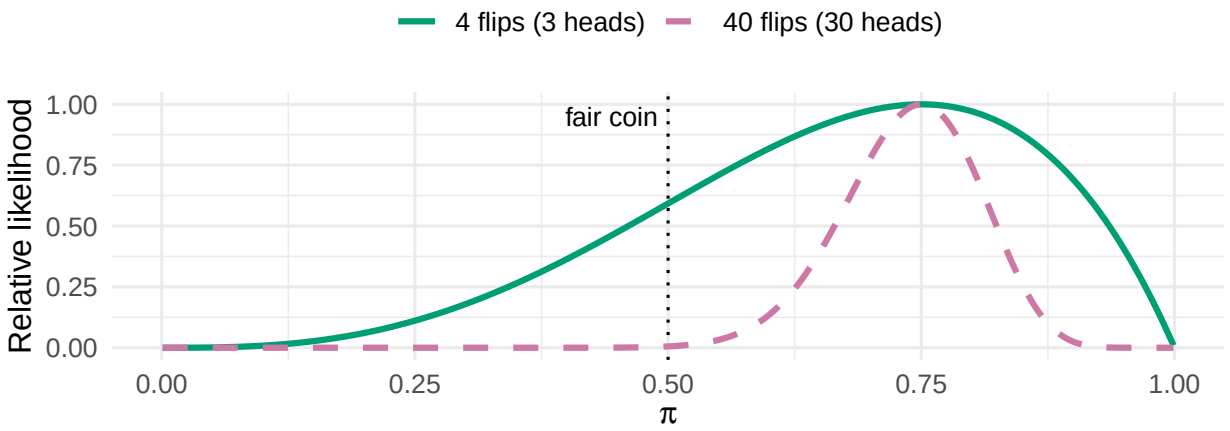
So is the coin fair? $\pi = 0.75$ has the highest likelihood. Can we conclude that the coin is not fair? What would make you more convinced?

The likelihood function. $L(\pi \mid \text{data}) = P(\text{data} \mid \pi)$. The numbers are calculated the same way. What is different is _____.

Table 3: Comparison of probability and likelihood

Feature	Probability	Likelihood
Notation	$P(\text{data} \mid \pi)$	$L(\pi \mid \text{data})$
What is fixed?	_____	_____
What varies?	_____	_____
Adds up to 1?	_____	_____
Question it answers	How probable is this data if π were this value?	How well does this value of π explain the data we saw?

More flips, more focused likelihood. What if we flipped the coin 40 times and got 30 heads? Same proportion, $\frac{30}{40} = 0.75$, more data.



What happens to the likelihood with more data?

Your turn. Your friend has a different coin and suspects it lands tails more often. You flip it 5 times: T, H, T, T, T.

a. Write $P(\text{T, H, T, T, T} \mid \pi)$ in terms of π, the probability of heads.

b. Calculate the likelihood for $\pi = 0.5$ (fair) and for $\pi = 0.25$. Hint: $0.5^5 \approx 0.031$ and $0.75^4 \approx 0.32$.

c. How many times more likely is $\pi = 0.25$ than a fair coin? _____

d. Add the two likelihoods. Should the sum be 1? _____

e. Which value of π has the highest likelihood? _____

What is wrong here? A friend says: "We flipped H, H, T, H. The likelihood of $\pi = 0.5$ is 0.0625, so there is a 6.25% chance that the coin is fair."

Bayesian inference

Remember from last week:

- **Frequentist:** the probability of an event is _____, over many, many repetitions.
- **Bayesian:** the probability of an event is _____. When we get new data, we _____ the probability.

On Monday we did frequentist inference: π was a fixed number, and all of our probabilities were about sample statistics, like $P(\text{data this extreme} \mid \pi = 0.5)$. Today: **Bayesian inference**, where the parameter is a _____.

Table 4: What frequentist and Bayesian inference find

Inference	What we find	Read it as
Frequentist (Monday)	_____	How surprising is the data, if the parameter were this value?
Bayesian (today)	_____	How plausible is each value of the parameter, now that we have seen the data?

The Bayesian recipe.

1. _____: what you believe about the parameter **before** seeing the data, as a probability distribution. $P(\mu)$
2. _____: collect data. How well does each value of the parameter explain the data? $L(\mu \mid \text{data})$
3. _____: the updated belief **after** seeing the data. $P(\mu \mid \text{data})$

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

$\propto$ is read "is proportional to." Values of μ that were plausible before **and** explain the data well become the most plausible after.

How much do college students sleep?

Researchers asked 253 college students to keep a **sleep diary** for two weeks. For each student, we have their average hours of sleep on weekdays (`WeekdaySleep`).

Data: SleepStudy in the `Lock5Data` package. Onyper, S. V., Thacher, P. V., Gilbert, J. W., & Gradess, S. G. (2012). Class start times, sleep, and academic performance in college: A path analysis. *Chronobiology International*, 29(3), 318–335.

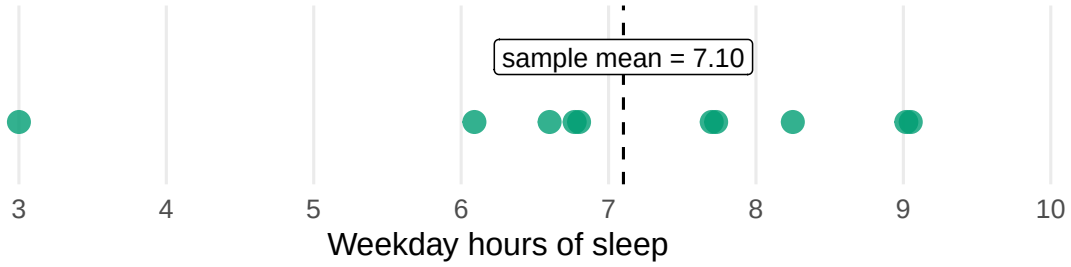
Let μ be the average weekday hours of sleep of **all** college students like the ones in the study. Sleep experts recommend that adults sleep **at least 7 hours** a night (Watson et al., 2015).

Do you think μ is less than 7 hours? Why?

Step 1: prior. Earlier today, Alex described their belief about μ as $\mu \sim N(7, 0.5^2)$. This is Alex's _____ distribution.

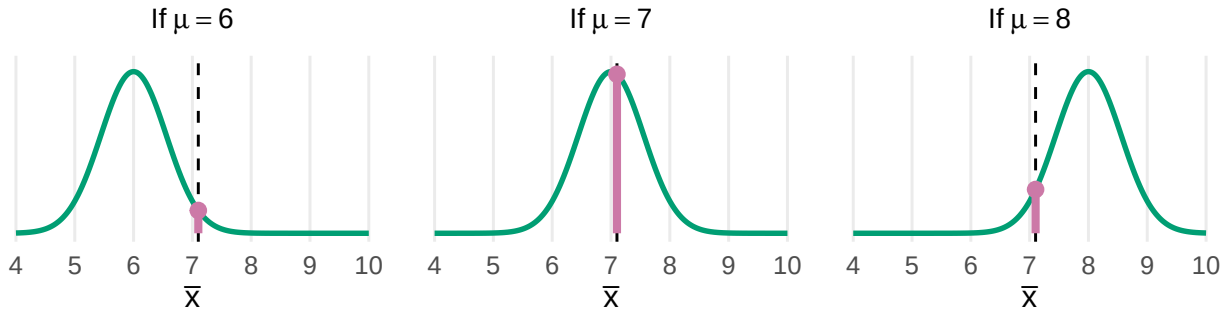
According to Alex's prior, what is the probability that μ is less than 7 hours? _____

Step 2: data. Suppose the researchers had entered only the **first 10** sleep diaries so far.



Sample mean: $\bar{x} = 7.10$ hours. Standard error: 0.56 (given to you, like on Monday).

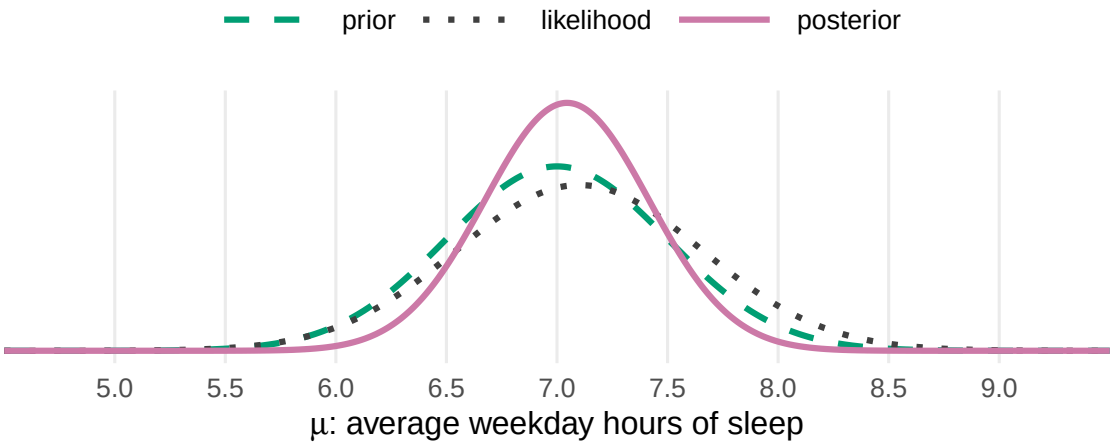
Likelihood of μ. Just like the coin: the data are _____, μ _____. How well does each value of μ explain $\bar{x} = 7.10$? Each curve below is the sampling distribution of $\bar{x}$ if μ were 6, 7, or 8.



The _____ at the observed $\bar{x}$ is the likelihood of that value of μ. Which of the three values of μ explains the data best? _____ Worst? _____

Doing this for every possible value of μ gives the likelihood curve. The likelihood of μ is shaped like a normal curve, centered at _____, with spread equal to _____.

Step 3: posterior.



Alex's **posterior**: $\mu \mid \text{data} \sim N(7.05, 0.37^2)$

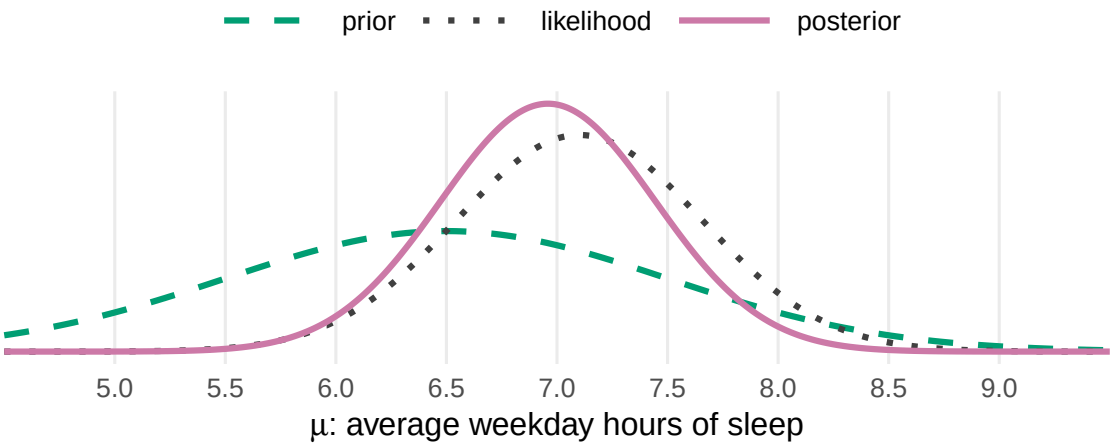
The posterior is a compromise. The posterior mean is a _____ of the prior mean and the sample mean:

$$0.55 \times \underbrace{7}_{\text{prior mean}} + 0.45 \times \underbrace{7.10}_{\text{sample mean}} \approx 7.05$$

- The weights depend on _____. Alex's prior (standard deviation 0.5) and the 10 diaries (standard error 0.56) are about equally certain, so they get about equal weight.
- The posterior is _____ than both the prior and the likelihood: combining two sources of information makes us more certain.

You will not need to calculate the posterior in this course. We will give you its mean and standard deviation.

A different prior. Sam's prior was $\mu \sim N(6.5, 1^2)$: "around 6.5 hours, not very sure." Sam sees the same 10 sleep diaries. Compared to Alex's posterior, will Sam's posterior be centered closer to Sam's prior mean or closer to the sample mean? Why?



Sam's posterior: $\mu \mid \text{data} \sim N(6.96, 0.49^2)$.

Alex and Sam saw the same data but ended up with different posteriors. Is this a problem? What could make their posteriors closer to each other?

All 253 sleep diaries. Now the researchers have entered all 253 diaries: $\bar{x} = 7.87$ hours, standard error 0.073.

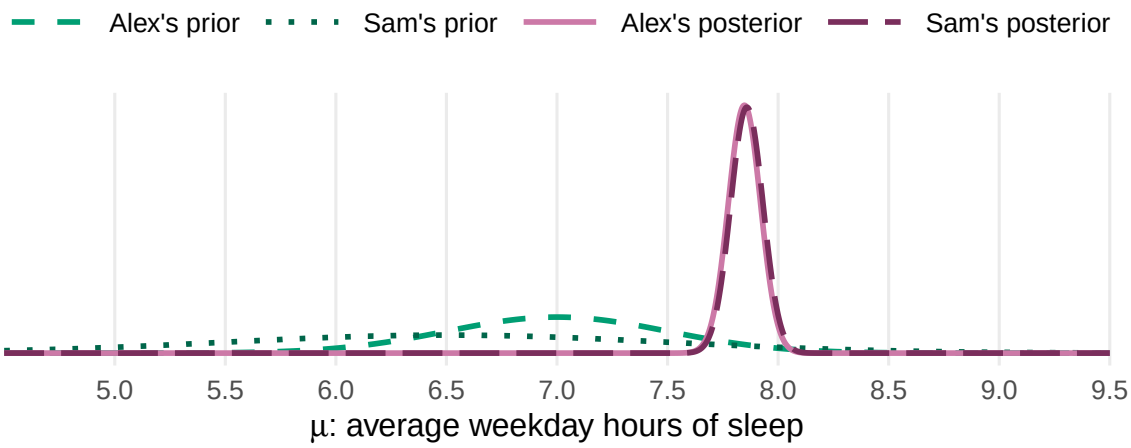


Table 5: Alex's and Sam's priors and posteriors for mu after all 253 sleep diaries

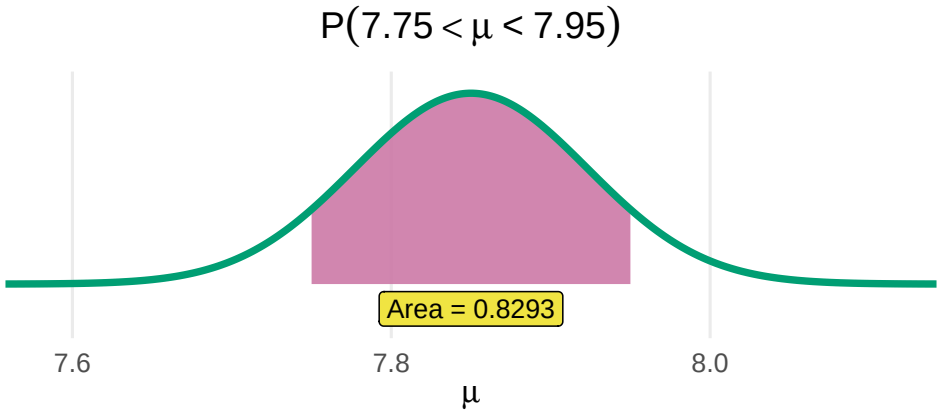
Person	Prior	Posterior after 253 diaries
Alex	$N(7, 0.5^2)$	$N(7.85, 0.073^2)$
Sam	$N(6.5, 1^2)$	$N(7.86, 0.073^2)$

What happened to the two posteriors? Why?

Using the posterior

Probability of a parameter. In Bayesian inference, μ is a random variable, so we can find probabilities **about** μ, given the data, as areas under the posterior. We will use Alex's posterior after all 253 diaries, $\mu \mid \text{data} \sim N(7.85, 0.073^2)$.

What is the probability that μ is between 7.75 and 7.95 hours? The plot below shows the area. Write the probability in notation and give its value.



This is a probability about the _____, the kind of statement we could **not** make on Monday.

Bayesian hypothesis testing

Do college students like these get enough sleep on weekdays, on average?

- **Null hypothesis** H_0: _____. They get enough sleep.
- **Alternative hypothesis** H_A: _____. They do **not** get enough sleep.

On Monday, we _____ and asked how surprising the data would be: $P(\text{data this extreme} \mid H_0)$, the p-value.

In Bayesian hypothesis testing, we do **not** assume either hypothesis is true. μ is a random variable, so _____:

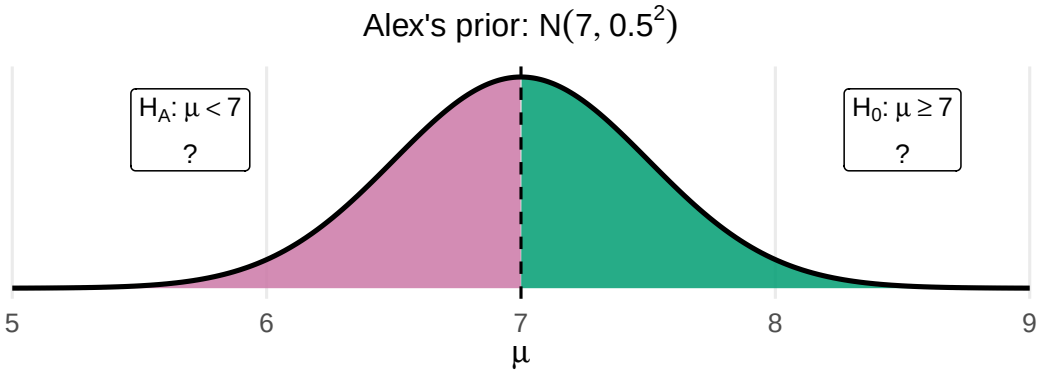
$$P(H_0 \mid \text{data}) = P(\mu \geq 7 \mid \text{data}) \qquad P(H_A \mid \text{data}) = P(\mu < 7 \mid \text{data})$$

The steps.

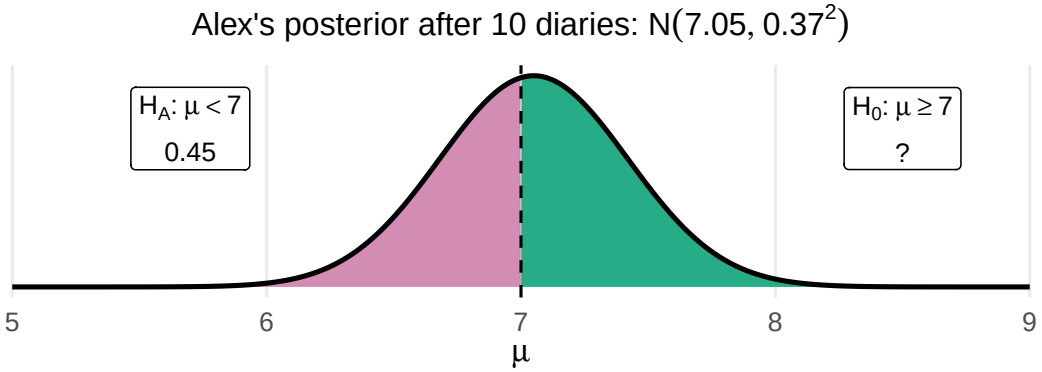
1. State H_0 and H_A.
2. Find the _____ probability of each hypothesis: area under the prior.
3. Find the _____ probability of each hypothesis: area under the posterior.
4. Compare: which hypothesis is more probable, and by how much? $\frac{P(H_0|\text{data})}{P(H_A|\text{data})}$ is how many times as probable H_0 is as H_A.

The two probabilities always add up to _____: μ is either less than 7 or not.

Before the data: prior probabilities. Fill in the "?" in the plot with Alex's prior probabilities $P(H_0)$ and $P(H_A)$. How do you know, without any calculation?



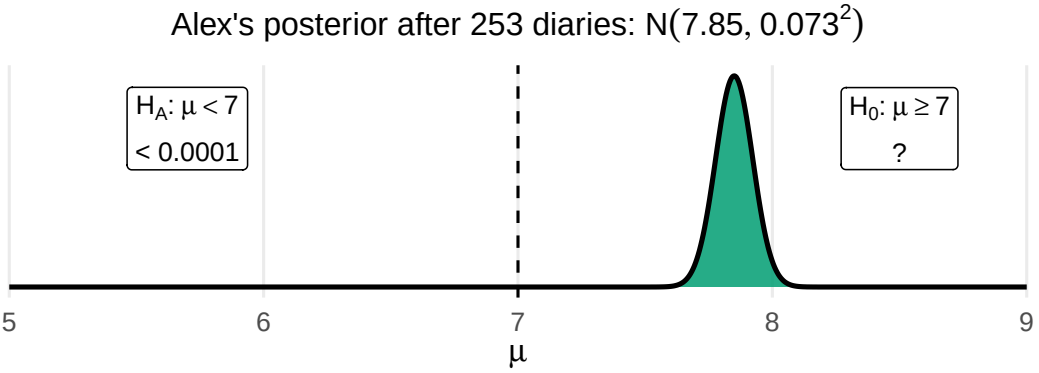
After 10 diaries: posterior probabilities. Alex's posterior is $N(7.05, 0.37^2)$. The plot gives $P(H_A \mid \text{data})$. Use it to find $P(H_0 \mid \text{data})$.



$P(H_0 \mid \text{data}) \approx$ _____ $P(H_A \mid \text{data}) \approx$ _____

H_0 is _____ times as probable as H_A. Would you conclude that college students like these get enough sleep, based on these 10 diaries?

After all 253 diaries. Alex's posterior is $N(7.85, 0.073^2)$. Use the plot to find $P(H_0 \mid \text{data})$.



Complete the table.

Table 6: Alex's prior and posterior probabilities of the null and alternative hypotheses

	$P(H_0)$: $\mu \geq 7$	$P(H_A)$: $\mu < 7$	Which is more probable?
Prior (before any data)	_____	_____	_____
Posterior after 10 diaries	_____	_____	_____
Posterior after 253 diaries	_____	_____	_____

Conclusion after all 253 diaries:

How is this different from Monday?

- On Monday, we could only _____ H_0. Failing to reject H_0 is **not** evidence for H_0.
- Today, both hypotheses get a probability, so the data can give evidence _____, including H_0.

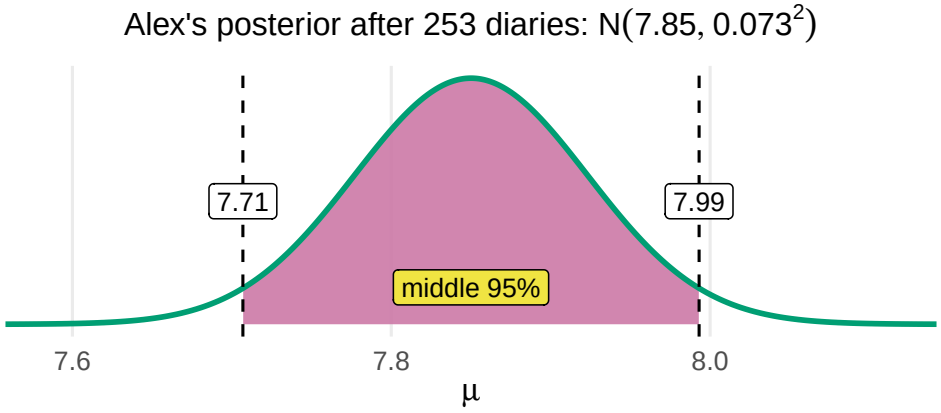
- The p-value, $P(\text{data this extreme} \mid H_0)$, is a probability about the _____. $P(H_0 \mid \text{data})$ is a probability about the _____.

Bayesian statisticians also compare hypotheses with the **Bayes factor**: how much the data changed the odds of one hypothesis against the other. We will not calculate it in this course.

Credible intervals

A **95% credible interval** is the _____: 2.5% of the area below it and 2.5% above it.

The plot shows Alex's posterior after all 253 diaries with the middle 95% shaded. What is Alex's 95% credible interval?

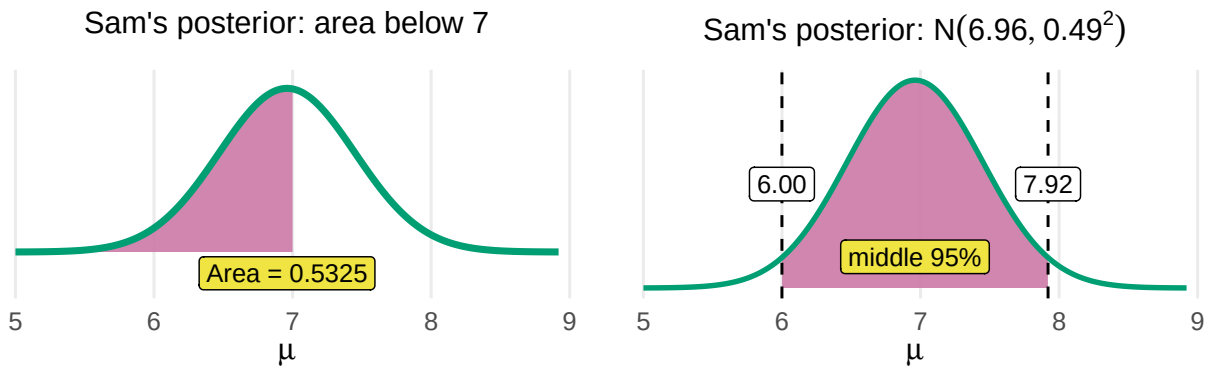


Interpretation:

In Bayesian inference, μ is a random variable, so "95% probability that μ is in the interval" is exactly what the interval means:

$$P(7.71 < \mu < 7.99 \mid \text{data}) = 0.95$$

Your turn. After the first 10 sleep diaries, Sam's posterior was $\mu \mid \text{data} \sim N(6.96, 0.49^2)$.



a. $H_0 : \mu \geq 7$ and $H_A : \mu < 7$. Use the left plot to find Sam's posterior probabilities $P(H_0 \mid \text{data})$ and $P(H_A \mid \text{data})$.

b. Sam's prior probabilities were $P(H_0) = 0.31$ and $P(H_A) = 0.69$. Which hypothesis was more probable before the data, and how many times as probable? What about after the data? What do you conclude?

c. Use the right plot to write Sam's 95% credible interval. _____

d. Interpret the interval in (c) in context.

e. Is Sam's interval wider or narrower than an interval from all 253 diaries? Why?

Frequentist vs. Bayesian

For all 253 diaries:

Table 7: Frequentist and Bayesian 95% intervals for mu from all 253 sleep diaries

Interval	From	To
95% confidence interval (frequentist)	7.72	8.01
95% credible interval (Bayesian, Alex's prior)	7.71	7.99

Almost the same numbers! With a lot of data, the prior matters very little. But they **mean** different things. Interpret each one.

Complete the table.

Table 8: Comparison of frequentist and Bayesian inference

Feature	Frequentist (Monday)	Bayesian (today)
The parameter is	_____	_____
Probability means	long-run proportion over many samples	how plausible something is, given what we know
Uses a prior?	_____	_____
What we find	_____	_____
Hypothesis testing	_____	_____
Interval	_____	_____
"95% probability that the parameter is in the interval"	_____	_____

Correct or not? Write "correct" or "not correct" and explain.

a. Alex's posterior probability that μ is less than 7 hours, after all 253 diaries, is less than 0.0001.

_____

b. In frequentist inference, we can find the probability that μ is less than 7 hours.

_____

c. There is a 95% probability that μ is between 7.71 and 7.99, according to Alex's posterior.

_____

d. The posterior always ends up exactly at the prior mean. _____

e. With more data, the prior has less influence on the posterior. _____

f. The likelihood $L(\mu \mid \text{data})$ is a probability distribution of μ. _____

g. In Bayesian hypothesis testing, we start by assuming that the null hypothesis is true.

_____

h. After 10 diaries, Alex's $P(H_0 \mid \text{data}) \approx 0.55$ and $P(H_A \mid \text{data}) \approx 0.45$, so the data give strong evidence for H_0. _____

Scope of the conclusions. Can we use these results to describe how much **Harvard** students sleep? The students wrote their own sleep diaries. How might this affect the data?

Summary

- Likelihood $L(\pi \mid \text{data})$: _____, does **not** need to add up to 1. Probability $P(\text{data} \mid \pi)$: _____, adds up to 1.
- Bayesian inference: _____ $\times$ _____ $\propto$ _____. The posterior mean is a compromise between the prior mean and the sample mean, weighted by how certain each one is.
- Different priors can give different posteriors. With more data, _____.
- **Bayesian hypothesis testing:** no hypothesis is assumed true. Find _____ and conclude in favor of the much more probable one.
- A 95% **credible interval** has a _____ of containing μ.
- Frequentist: $P(\text{data} \mid \text{parameter})$. Bayesian: $P(\text{parameter} \mid \text{data})$. Remember: $P(A \mid B) \neq P(B \mid A)$!