Date 11-06-2026

PSYCH 10 Tutorial Session · Fall 2026

Week 7 · The Sampling Machine · Answer key

Full name (in CAPITAL LETTERS) *Answer key* **SUNet ID** (the part before @stanford.edu) _____ **Section** (e.g., TUES 9:30) _____

Please write legibly so the teaching team can log your attendance: if we can't read your name, we can't give you credit.

The situation

> **Q. Ledger** has built a sampling machine. "Last year I timed every visit to **Café Variantia**, from joining the line to holding a cup: 10,000 waits in all," Ledger says. "The machine draws random samples from those 10,000 waits. It can print a histogram of all the waits, of one sample, or of the means of 1,000 samples. Jo ran it all morning and forgot to put titles on the printouts."
>
> **Four printouts are missing their titles. Match each printout to its title, then use the printouts in Parts 2 and 3.**

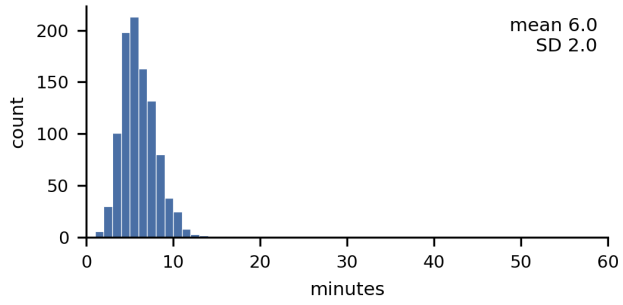
The file

Last year's 10,000 waits had a mean of 6 minutes and an SD of 6 minutes. Most waits were short, but a few were very long, because the espresso machine broke down about once a week.

Part 1 · Match the printouts

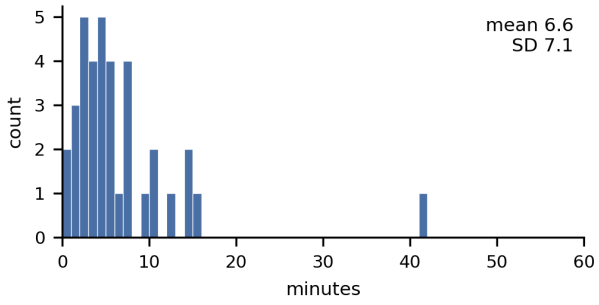
Under each printout, circle its title. Each title belongs to one printout.

Printout A



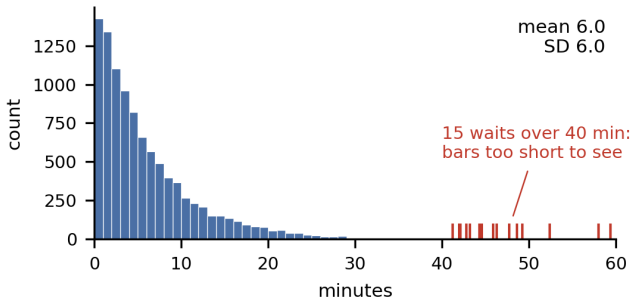
all 10,000 waits · one sample of 36
means of samples of 9 · means of samples of 36

Printout B



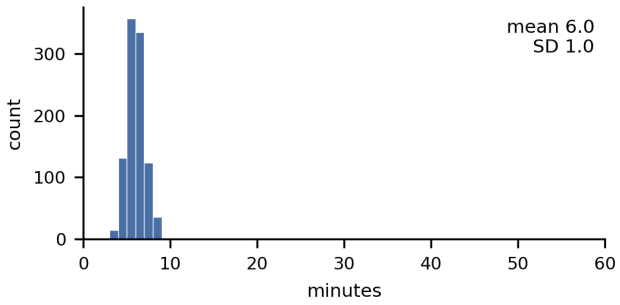
all 10,000 waits · one sample of 36
means of samples of 9 · means of samples of 36

Printout C



all 10,000 waits · one sample of 36
means of samples of 9 · means of samples of 36

Printout D



all 10,000 waits · one sample of 36
means of samples of 9 · means of samples of 36

A: means of samples of 9. B: one sample of 36. C: all 10,000 waits. D: means of samples of 36. The counts settle the matches: the bars on B add up to 36, the bars on A and on D to 1,000 each, and the bars on C to 10,000. B and C are skewed because both are made of waits. A and D are made of means, one mean per sample, so both are centered on 6.0 with an SD far smaller than the waits' SD of 6.0.

How can you tell printout A from printout D?

By their SDs. The SD printed on A is 2.0 minutes and the SD printed on D is 1.0 minute, so D is half as wide as A. Means of samples of 36 vary less from one sample to the next than means of samples of 9, so the narrower printout is the one made from the larger samples. D is also the more symmetric of the two; A keeps a little of the waits' right skew.

Printout B has one wait of about 41 minutes, far from the other waits. Jo thinks the sampling machine made an error and wants to delete this wait. Circle one: delete · keep. Justify your answer.

Keep. The file says that the espresso machine broke down about once a week, so waits of 40 minutes and more are real, and printout C shows 15 of them among the 10,000 (the red ticks). A random sample of 36 that includes one of them is unusual but not an error. The most likely account of the 41 minutes is a real wait during a breakdown. Deleting it would remove a real feature of the café's waits from the sample.

Part 2 · The standard error

Take many random samples of the same size and compute an estimate, such as the mean, from each sample. The distribution of these estimates is a **sampling distribution**, and its SD is called the **standard error**. For the mean of a sample of n waits, the standard error is **the SD of all 10,000 waits divided by $\sqrt{n}$**.

a. For samples of 9, the standard error of the mean is 6 / $\sqrt{9}$ = *2.0* minutes. For samples of 36, it is 6 / $\sqrt{36}$ = *1.0* minutes. The samples of 36 are 4 times as large, and their standard error is *half* as large.

The two values are the SDs printed on A and D, so the formula agrees with the machine's 1,000 samples. Four times the sample size gives half the standard error, because $\sqrt{4} = 2$.

b. The waits are skewed. For samples of 36, circle the shape of the sampling distribution of the mean: **skewed · roughly normal**
roughly normal

The central limit theorem, from Lecture 12: the means of large enough random samples have a distribution close to normal for a population of nearly any shape, including the café's skewed waits. A (samples of 9) still shows a little right skew, and D (samples of 36) is close to symmetric.

c. Jo says: "The standard error for samples of 36 is small, so the 36 waits in my one sample should have an SD much smaller than 6 minutes." Circle one: **right · wrong**. Justify your answer using two printouts. *wrong*

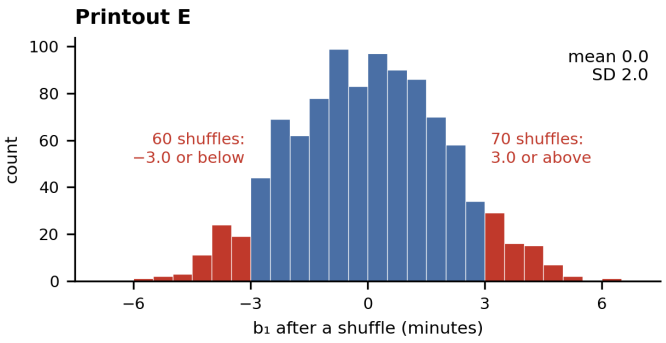
Printout B is one sample of 36 waits, and its SD is 7.1 minutes, about as large as the SD of all 10,000 waits on printout C (6.0). The SD of the waits in a random sample estimates the SD of the population, whatever the sample size. The standard error of 1.0 minute is the SD of printout D, whose values are the means of 1,000 samples of 36, and it measures the variation of those means from one sample to the next. Jo has confused the spread of the waits inside one sample with the spread of the means across samples.

Part 3 · The new espresso machine

The café has a new espresso machine, and the manager, **Percy Lator**, says that it has shortened the wait. To test this claim, Percy timed 36 visits. Half of the visits, chosen at random, were served from the new machine, and the other half from the old machine.

The mean wait was 7.0 minutes with the old machine and 4.0 minutes with the new machine. The difference between the two means is b_1 = 4.0 — 7.0 = —**3.0 minutes**.

Could a difference of this size arise by chance, with no effect of the new machine? Jo set the sampling machine to **shuffle**. In each shuffle, the machine deals the labels "new" and "old" (18 of each) to the 36 waits at random and computes b_1 for the shuffled labels. **Printout E** is a histogram of the b_1s from 1,000 shuffles. It approximates the sampling distribution of b_1 if the new machine made no difference.



a. In one shuffle, b_1 was —2.5 minutes. Circle the source of this difference: **the new machine · the random dealing of the labels**
the random dealing of the labels

After a shuffle, each wait carries a label that was dealt at random, so the label on a wait has nothing to do with the espresso machine that served the visit. Any difference between the two labeled groups is due to chance, and a shuffled b_1 of —2.5 is a chance difference between two random halves of the same 36 waits.

b. The **p-value** is the probability of a b_1 at least as far from 0 as Percy's —3.0, in either direction, if the new machine made no difference (the empty model). Estimate it from the red counts on printout E: p-value = (*60* + *70*) / 1,000 = *.13*

Both red tails count, because a shuffled b_1 of +3.0 is as far from 0 as Percy's —3.0. Counting only the 60 shuffles at —3.0 or below gives .06, half the p-value.

c. Percy says: "The wait was 3 minutes shorter, so the new machine works." By a common convention, researchers reject the empty model when the p-value is below .05. Is Percy's conclusion justified? Circle one: **yes · no**. Justify your answer using the p-value. *no*

The p-value is .13, which is not below .05, so by the convention the empty model is not rejected. A difference at least as far from 0 as Percy's arose in 130 of the 1,000 shuffles, about 1 in 8, with no effect of the new machine, so the random assignment of visits to the two machines is a plausible account of the 3 minutes. Percy's conclusion goes beyond the evidence in these 36 visits. The .13 is computed on the assumption that the new machine made no difference, so it is not the probability that this assumption is true. The p-value also does not show that the new machine has no effect: —3.0 minutes is still the best estimate of the difference from these data, and 36 visits are too few to tell a real difference of 3 minutes from chance.

d. For Percy's 36 visits, the standard error of b_1 is about 2.0 minutes (the SD of printout E). Suppose Percy repeats the study with 4 times as many visits (144, half on each machine), and b_1 is again —3.0 minutes. By the pattern in Part 2a, the new standard error would be about *1.0* minutes, so the shuffled b_1s would vary less around 0. Circle your prediction for the new p-value: **larger than in part b · about the same · smaller than in part b** *smaller than in part b*

Four times the visits gives half the standard error, 2.0 to 1.0, as in Part 2a. A b_1 of —3.0 is then 3 standard errors from 0 in place of 1.5, so far fewer shuffled b_1s reach 3.0: by the empirical rule, about 3 in 1,000 values of a normal distribution fall 3 SDs or more from its center, in either direction. Jo's run of 1,000 shuffles of the 144 visits gave 3 at 3.0 or beyond, so p = .003. The same difference of 3 minutes is weak evidence against the empty model from 36 visits and strong evidence from 144. (A standard error of 0.5, from $6 / \sqrt{144}$, would belong to the mean of one sample of 144 waits. b_1 is a difference between two means, so the halving applies to the 2.0 from printout E.)