

Problem 1. A population has a mean of 50 and a standard deviation of 6. What are the mean and standard deviation of the sampling distribution of the mean for $n = 16$?

Answer 1. By population mean and standard deviation, we are talking about the mean and standard deviation of a single draw, X_i , where $\mu = 50$ and $\sigma = 6$. The sampling distribution refers to \bar{X} . The mean of \bar{X} is also $\mu = 50$, but the standard deviation of \bar{X} is $6/\sqrt{16} = 1.5$. The latter number is called the *standard error*, denoted $se(\bar{X})$.

If we know σ and if $n > 30$, then $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ approximately. If each X_i is normally distributed, or if $n \rightarrow \infty$, then $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ exactly.

Problem 2. Given a test that is normally distributed with mean $\mu = 100$ and standard deviation of $\sigma = 12$. Find the following:

- (a) the probability that a single score drawn at random will be less than 120
- (b) the probability that a single score drawn at random will be greater than 123
- (c) the probability that a sample of 25 scores will have a mean less than 106
- (d) the probability that the mean of a sample of 36 scores will be either less than 95 or greater than 105
- (e) the test score such that the probability of scoring above it is 5%

Answer 2.

- (a) Let X denote a random test score. We want to find $P(X < 120)$. We first need to standardize the test score so that it has mean 0 and standard deviation 1, and accordingly we instead find

$$P\left(\frac{X - 100}{12} < \frac{120 - 100}{12}\right).$$

Let $Z \equiv (X - 100)/12$. Since the test is normally distributed, we also know that Z is normally distributed; and since we've standardized it, it is standard normally distributed. Hence we are to find $P(Z < 1.67)$ for $Z \sim \mathcal{N}(0, 1)$.

To solve this, we need to either appeal to a normal distribution table, or use R. To solve it with R, use the command `pnorm(1.67)`, which gives approximately 0.953. Using the normal table we are provided with, 1.67 is closest to 1.645, so we would use approximately 0.95.

- (b) We set the problem up analogously and arrive at standardized probability $P(Z > 1.92)$. The problem is, `pnorm(1.92)` tells us the probability of Z being *below* 1.92, whereas

we are now trying to find the probability of Z being *above* 1.92. We can exploit the symmetry of the normal distribution to solve this: the probability of being above 1.92 is the same as the probability of being below -1.92 .

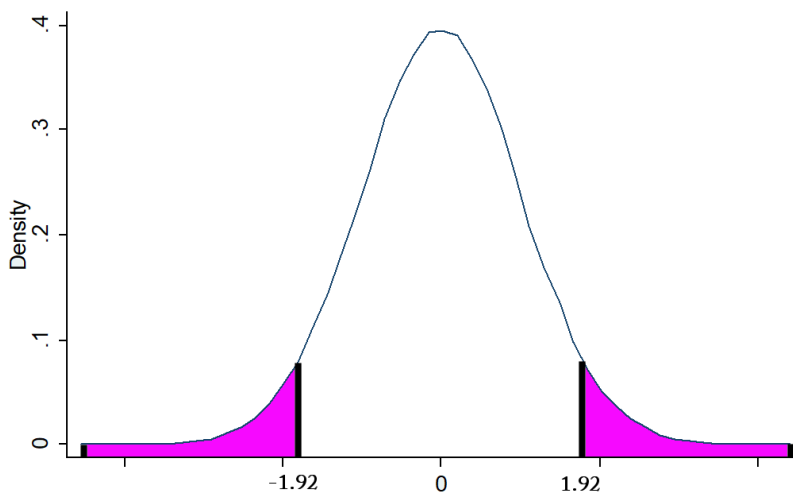


FIGURE 1: The probability of being above 1.92 is the same as the probability of being below -1.92 .

Hence the problem can be solved with `pnorm(-1.92)`, which gives about 0.027.

Another problem: -1.92 is not a number that appears on the normal table. What we can do instead is recognize that the probability of Z being above 1.92 is the complementary probability of Z being below 1.92. That is, $P(Z > 1.92) = 1 - P(Z < 1.92)$.

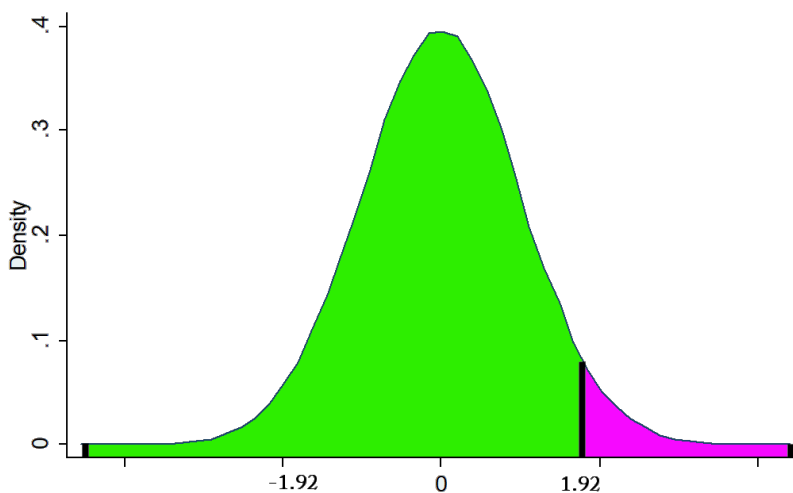


FIGURE 2: The area under the entire curve is 1. Hence, 1 minus the green area gives us the purple area. The green area is $P(Z < 1.92)$. Hence $P(Z > 1.92) = 1 - P(Z < 1.92)$.

Using the normal table, 1.92 is reasonably close to 1.96, so the answer is approximately $1 - 0.9750 = 0.025$.

- (c) Now we are dealing with a sampling distribution, so we appeal to the central limit theorem, which tells us that

$$Z \equiv \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

We want to solve $P(\bar{X} < 106)$. We conform it to central limit theorem form by using

$$\begin{aligned} P(\bar{X} < 106) &= P\left(\frac{\bar{X} - 100}{12/\sqrt{25}} < \frac{106 - 100}{12/\sqrt{25}}\right) \\ &= P(Z < 2.50). \end{aligned}$$

Using R, this gives `pnorm(2.50) \approx 0.994`. Using the normal table, the closest we have is 2.576, which gives probability 0.995.

(Note that if test scores X_i were not normally distributed and we did not know the population standard deviation σ , then we'd have to use estimate s instead of σ , and the t distribution instead of the standard normal distribution since $n < 30$.)

- (d) We want to find $P(\bar{X} < 95) + P(\bar{X} > 105)$. We first standardize each with respect to the central limit theorem, which gives

$$\begin{aligned} P(\bar{X} < 95) + P(\bar{X} > 105) &= P\left(\frac{\bar{X} - 100}{12/\sqrt{36}} < \frac{95 - 100}{12/\sqrt{36}}\right) + P\left(\frac{\bar{X} - 100}{12/\sqrt{36}} > \frac{105 - 100}{12/\sqrt{36}}\right) \\ &= P(Z < -2.50) + P(Z > 2.50). \end{aligned}$$

Since the normal distribution is symmetric, we know $P(Z < -2.50) = P(Z > 2.50)$. Hence we can instead find $2 \times P(Z > 2.50)$. From the argument used in part (b), we know that $P(Z > 2.50) = 1 - P(Z < 2.50)$. From the normal table, 2.50 is close to 2.576, so we can conclude approximately that

$$P(Z > 2.50) = 1 - P(Z < 2.50) = 1 - 0.995 = 0.005.$$

Hence the answer is approximately $2 \times 0.005 = 0.01$. Alternatively, using the R command `2*(1 - pnorm(2.50))` gives 0.012.

- (e) First standardize the test score X into Z in the usual way. We are looking for the value of Z that makes the right tail consist of 5% of the area under the curve. Which is another way of saying, we want the value of Z such that the area below that number

is 0.95. According to our normal table, that number is 1.645. Using R, we find the number by using command `qnorm(0.95)`, which gives the same number.

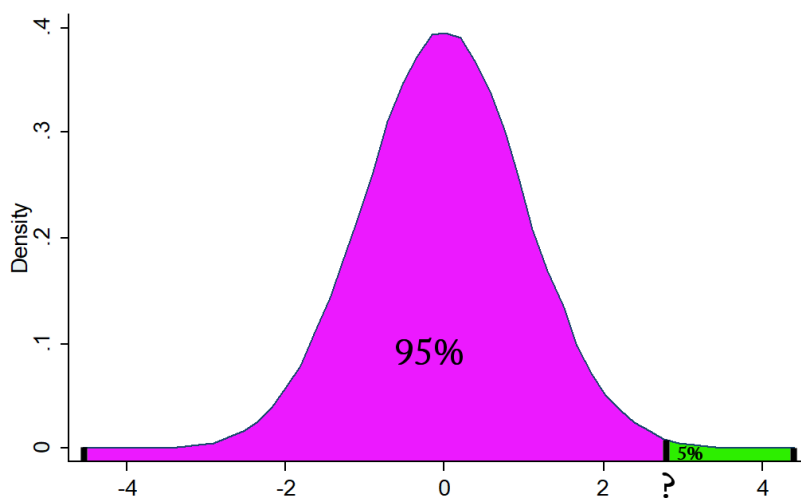


FIGURE 3: We want to find the value of Z such that the area underneath the curve above the value is 0.05.

But this is not a test score. To convert it back into a test score, we have to un-standardize it. So multiply it by the standard deviation and then add the mean back, and you get

$$1.645 \times 12 + 100 \approx 120.$$

Thus we conclude there is a 5% chance that someone receives a score above 120. Note that this is completely consistent with part (a), where we found the probability of being below a score of 120 is 0.95.

Problem 3. In Wisconsin, the mean donut consumption in a week is 48 donuts per person, and the standard deviation of weekly donut consumption is 12 donuts.¹ This week, Jiminy Glick has a weekly donut Z-score of 1.5. How many donuts did Jiminy Glick eat this week?

Answer 3. The Z-score tells us how many standard deviations from the mean. Since Jiminy's Z-score is 1.5, that means he ate $1.5 \times 12 = 18$ donuts more than the mean. Hence, he ate $48 + 18 = 66$ donuts this week.

Problem 4. On average, I eat 7 pizzas per week, with a standard deviation of 1 pizza, and my pizza consumption is normally distributed. What is the probability that I eat less than 5 pizzas in a given week? Don't use R or a normal table.

¹I miss Wisconsin.

Answer 4. For any normal distribution, it is approximately true that

- 68% of the data lies within one standard deviation of the mean
- 95% of the data lies within two standard deviations of the mean
- 99.7% of the data lies within three standard deviations of the mean.

5 pizzas is two standard deviations less than the mean of 7, so let's consider the second bullet more closely. Since 95% of the data is within two standard deviations of the mean, that means the two tails outside of that must comprise the remaining 5% of the data. Since the normal distribution is symmetric, that means each tail gets 2.5%. Thus the probability of me consuming fewer than 5 pizzas in a week is approximately 2.5%.

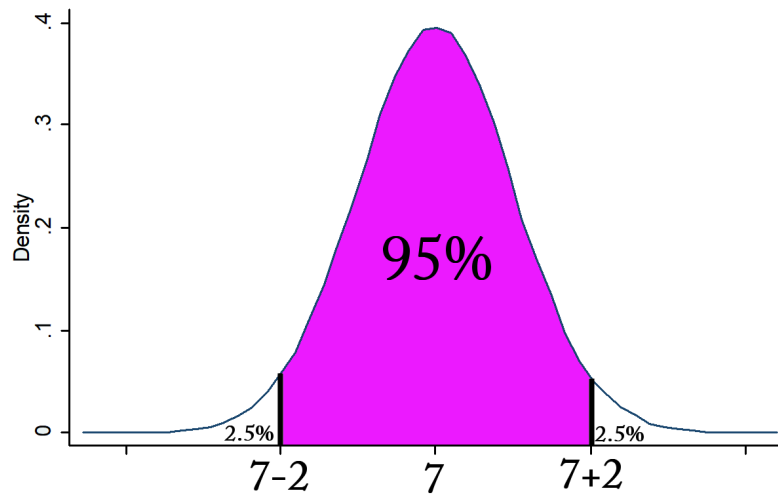


FIGURE 4: 95% of the data is found within ± 2 standard deviations of the mean.