

Side comment: In this example, we don't actually *know* that $p = 0.68$ is the correct value for the whole population — that's only an estimate based on the sample. But the value of the standard error is not all that sensitive to p — for example, it would be 0.0136 at $p = 0.75$ and 0.0157 at $p = 0.5$ (compared with 0.0147 at $p = 0.68$) — these numbers are not very much different. Strictly speaking, the standard error is only an estimate of the standard deviation of the proportion, but this point is usually ignored in practice.

Example (Problems 7.5 and 7.6, p. 320).

An exit poll was taken of 2705 voters for the governorship of California. Among *all* voters, 55.9% voted for Schwarzenegger

- (a) Define a binary random variable x taking values 0 or 1 (1=vote for Schwarzenegger). State its probability distribution.
- (b) Find the mean and standard error of the sampling distribution of the *proportion* of the 2705 people in the sample who voted for Schwarzenegger.
- (c) If the sampling distribution of the sample proportion had mean 0.559 and standard error 0.0095, given an interval of values within which the sample proportion will almost certainly fall. (Hint: normal distribution)
- (d) Based on (c), if you take an exit poll and observe a sample proportion of 0.60, would this be unusual? Why?
- (e) The actual exit poll of 2705 voters had sample proportion 0.565. Given $p = 0.559$, was this unusual? Answer by finding out how many standard errors 0.565 falls from 0.559.

Brief answers:

(a) $P(x = 1) = 0.559$, $P(x = 0) = 1 - 0.559 = 0.441$.

(b) Mean 0.559, standard error $\sqrt{\frac{0.559 \times (1 - 0.559)}{2705}} = 0.0095$.

(c) $0.559 \pm 3 \times 0.0095 = (.5305, .5875)$.

(d) Yes, because the result lies more than three standard errors from the mean.

(e) $\frac{.565 - .559}{.0095} = 0.63$. This is not unusual, because the result is less than 1 standard error from the mean.

How close are sample means to population means?

Let's go back to the "women's heights" example from Chapter 2.

In our class, I took a survey in which I asked students their heights. Of the 69 female students who responded, the mean was 65.71 and the standard deviation 2.81. The question now is: how much variability is there in these numbers? In particular, how close is that 65.71 likely to be to the true mean of all female students at UNC?

We do a simulation experiment. Individual samples have means like 65.7, 65.9, 64.9, 65.8...

Now let's do 1000 samples. The mean of the sample means is 65.70 and the standard deviation 0.333.

Try increasing the size of an individual sample, say to 1001 (a typical size for an opinion survey)

Now the individual samples have means like 65.68, 65.66, 65.70, 65.62

In a repetition of 1000 samples of size 1001, the mean is 65.714 and the standard deviation .087.

Conclusions:

1. The standard deviation of the sample mean (resp. 0.333, 0.087 for the two sample sizes) is much smaller than the standard deviation for the observations themselves
2. The distribution of the sample mean is also much closer to normal as indicated by the shape of the histogram.

Property 2 is called the **Central Limit Theorem**.

Formula: The standard deviation of \bar{x} (the sample mean of a sample of size n is

$$\frac{\sigma}{\sqrt{n}}.$$

where σ is the standard deviation of a single observation.

The quantity is known as the *standard error* (the same word as we used with sample proportions, but this is more general, because we're not necessarily estimating the mean of binary random variables — they can be anything).

With $\sigma = 2.81$, $n = 69$ we have $\frac{\sigma}{\sqrt{n}} = 0.338$, very close to the 0.333 we saw in our simulation.

With $\sigma = 2.81$, $n = 1001$ we have $\frac{\sigma}{\sqrt{n}} = 0.089$, very close to the 0.087 we saw in our simulation.

Moreover, it's very close to a normal distribution so we can use that to calculate probabilities.

Problem 7.16, page 332.

A roulette wheel in Las Vegas has 38 slots. If you bet \$1 on a particular number, you win \$35 if the ball ends up in that slot and \$0 otherwise. All outcomes are equally likely.

- (a) Let X be your winnings when you play once. What is the probability distribution of X ? (It has mean 0.921 and standard deviation 5.603.)
- (b) You play 5040 times. Show that the sampling distribution of your sample mean winnings has mean 0.921 and standard deviation 0.079.
- (c) Using the central limit theorem, find the probability that your mean winnings is at least \$1, so that you don't lose money overall.

Solution

(a) The table is

x	$P(x)$	$xP(x)$
\$35	$\frac{1}{38}$	$\frac{35}{38}$
\$0	$\frac{37}{38}$	0

Mean is $\frac{35}{38} = 0.921$; SD is 5.603 (given)

(b) $5.603/\sqrt{5040} = .079$.

(c) With $x = 1$, $\mu = 0.921$, $\sigma = .079$, we have $z = \frac{1-0.921}{0.079} = 1$; left tail probability is 0.84; right tail probability is $1 - 0.84 = 0.16$ (the answer).

Problem 7.43, page 340.

A survey of physicians' assistants reported a mean annual income of \$84,396 and a standard deviation of \$21,975.

- (a) Suppose they had randomly sampled 100 physicians' assistants instead of collecting data for all of them. Describe the mean, standard error and shape of the sampling distribution.
- (b) Using this sampling distribution, find the z score for a sample mean of \$80,000.
- (c) Using (a) and (b), find the probability that the sample mean would fall within approximately \$4,000 of the population mean.

Solution

(a) The mean is \$84,396, the standard deviation is \$ $\frac{21,975}{\sqrt{100}} =$ \$2,197.5, and the distribution is approximately normal by the central limit theorem.

(b) $z = \frac{80000 - 84396}{2197.5} = -2.00.$

(c) If $x - \mu = \pm 4000$ then $z = \frac{x - \mu}{\sigma} = \pm \frac{4000}{2197.5} = \pm 1.82.$ The probabilities associated with $z = \pm 1.82$ are .9656 and .0344. Therefore the required probability is $.9656 - .0344 = .9312.$