

## Rule of sample proportions (p. 359)

- IF:
1. There is a population proportion of interest
  2. We have a random sample from the population
  3. The sample is large enough so that we will see at least five of both possible outcomes

THEN: If numerous samples of the same size are taken and the sample proportion is computed every time, the resulting histogram will:

1. be roughly bell-shaped
2. have mean equal to the true population proportion
3. have standard deviation equal to:

$$\sqrt{\frac{(\text{population proportion}) \times (1 - \text{population proportion})}{\text{sample size}}}$$

## Sample proportions: Suppose the true proportion is known

Recall that in craps,  $P(\text{win}) = 244/495 = 0.493$  for each game.

Q: If you play 100 games of craps, what proportion will you win?

A: Dunno.

Q: Okay, I guess that was obvious. But can we at least give a range of possible proportions that will be valid most (say, 95%) of the time?

A: Sure. Use the Rule of Sample Proportions.

## Sample proportions (cont'd): Suppose the true proportion is known

Recall that in craps,  $P(\text{win}) = 244/495 = 0.493$  for each game.

The RoSP says that the sampling distribution of the proportion of wins in 100 games will be:

- roughly normal,
- with mean 0.493
- and standard deviation  $\sqrt{\frac{(.493) \times (.507)}{100}} = 0.05$ .

Therefore, the 68-95-99.7 rule says that 95% of the time, the proportion of wins in 100 games will be between  $0.493 - 2 \times 0.05$  and  $0.493 + 2 \times 0.05$ .

(0.393 and 0.593)

## True proportion unknown

Next, suppose we do not know the true population proportion value. This is far more common in reality!

How can we use information from the sample to estimate the true population proportion?

Suppose we have a sample of 200 students in STAT 100 and find that 28 of them are left handed.

Our sample proportion is: 0.14

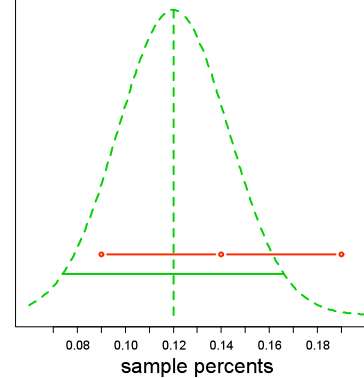
We can now *estimate* the standard deviation of the sample proportion based on a sample of size 200:

$$\sqrt{\frac{(.14) \times (1 - .14)}{200}} = 0.025$$

Hence, 2 standard deviations =  $2 \times .025 = .05$

*On the following two slides, we'll pretend that the true population proportion is 0.12.*

**Normal curve of sample proportions based on sample size 200**



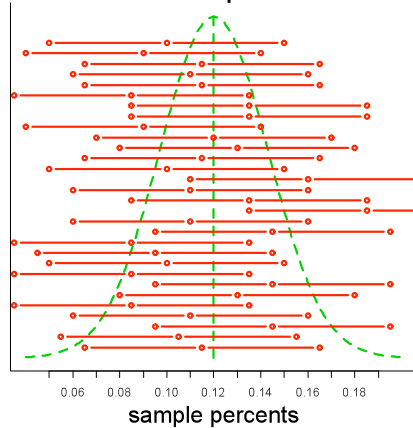
Note the green curve, which is the truth.

Of course, ordinarily we don't know where it lies, but at least we know its standard deviation.

Thus, we can build a *confidence interval* around our 14% estimate (in red).

If we take another sample, *the red line will move but the green curve will not!*

**30 confidence intervals based on sample size 200**



If we repeat the sampling over and over, 95% of our confidence intervals will contain the true proportion of 0.12.

This is why we use the term "95% confidence interval".

**Definition of "95% confidence interval for the true population proportion":**

An interval of values computed from the sample that is almost certain (95% certain in this case) to cover the true but unknown population proportion.

The plan:

1. Take a sample
2. Compute the sample proportion
3. Compute the estimate of the standard deviation of the sample proportion:  $\sqrt{\frac{(\text{proportion}) \times (1 - \text{proportion})}{\text{sample size}}}$
4. 95% confidence interval for the true population proportion: sample proportion  $\pm 2 \times \text{SD}$

## Other confidence coefficients

The confidence intervals we've been constructing (using  $\pm 2 \times \text{stdev}$ ) are called 95% confidence intervals.

The confidence coefficient is 95%, or .95.

It means that we cover the middle 95% of the normal curve associated with sample proportions, and this requires 2 standard deviations.

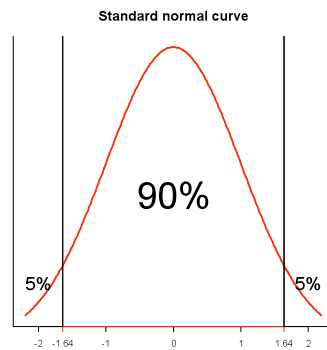
**We can change the confidence coefficient by using the normal table (p. 157) to determine the appropriate number of standard deviations.**

## Other confidence coefficients: An example

Suppose we want a 90% confidence interval instead of 95%.

How many standard deviations span the middle 90% of the normal curve?

## 90% confidence interval (see p. 157)



Since 90% is in the middle, there is 5% in either end.

So find z for .05 and z for .95.

We get  $z = \pm 1.64$

90% confidence interval: sample proportion  $\pm 1.64 \times \text{stdev}$

Back to the original example of sample size 200, 14% left-handed in sample.

We found the sample stdev to be .025.

Hence, to construct a 90% CI we have

$$.140 - 1.64 \times .025 = .140 - .041 = .099$$

$$.140 + 1.64 \times .025 = .140 + .041 = .181$$

90% confidence interval: .099 to .181

(95% confidence interval was: .090 to .190)

So the 90% CI is **shorter** (a more precise estimate) but **less accurate**. It has a 10% chance of missing the true population proportion.