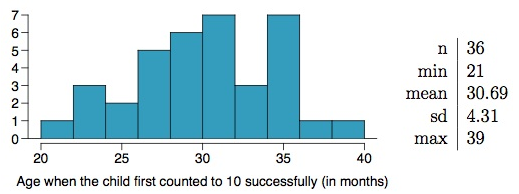
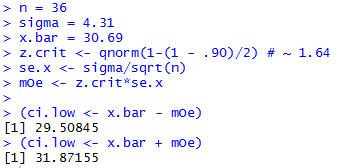
* Suppose we are interested in studying how much chocolate is consumed by Coursera students, measured in g/week. After surveying 500 students, we calculate an average of 175 grams per week with a standard deviation of 195 grams per week. Which of the following is not necessarily true?
* A point estimate for the population standard deviation is 195.
* x¯=175,s=195
* **μ=175, σ=195**
* *Just because the sample statistics are these values doesn't mean the population values will be exactly equal to them, therefore it's not necessarily true that μ=175, σ=195.*
* A histogram of the samples will be skewed to the right.
* Which of the following is true?
* **Standard error computed based on a sample standard deviation will always be lower than the standard deviation of that sample.**
* Standard error measures the variability in means of samples of the same size taken from different populations 🡪 **SAME POPULATION**
* **As the sample size increases, the variability of the sampling distribution decreases.**
* **In order to reduce the standard error by half, sample size should be increased by a factor of 4.**
* The ages of pennies at a particular bank follow a nearly normal distribution w/ mean 10.44 years + SD 9.2 years. Say you take random samples of 30 pennies, find the mean age in each sample, + plot the distribution of these means. Which of the following are the best estimates for the center and spread of this distribution?
* **standard error = 9.2/√=30 = 1.68**



* Which of the following is true about sampling distributions?
* **Shape of the sampling distribution is NOT ALWAYS the same shape as the population distribution 🡪 depends on sample size**
* **Sampling distributions are NOT always nearly normal**
* ~~Sampling distribution of the mean is always right skewed since means cannot be smaller than 0.~~
* **Sampling distributions get closer to normality as the sample size increases.**
* To get an estimate of consumer spending in the U.S. following Thanksgiving, 436 randomly sampled American adults were surveyed. Daily spending for the 6-day period following Thanksgiving averaged $84.71. A 95% CI based on this sample is ($80.31, $89.11). Which are true?
* ~~I. We are 95% confident that the average spending of the 436 American adults in this sample is between $80.31 and $89.11.~~
* **If we collected many random samples of the same size + calculated a CI for daily spending for each sample, then we’d expect 95% of the intervals to contain the true population parameter.**
* **We are 95% confident that average spending of all American adults is between $80.31-$89.11.**

* Which of the following is true about confidence intervals? (All else held constant)
* **As the confidence level increases, the width increases.**
* **As the sample size increases, the margin of error decreases.**
* **As the sample mean increases, the margin of error stays constant.**
* **As the standard deviation of the sample increases, the width increases**
* We want to estimate average coffee intake of Coursera students, measured in cups of coffee. A survey of 1K students yields an average 0.55 cups per day, w/ SD of 1 cup per day. Which of the following is not necessarily true?
* 0.55 is a point estimate for the population mean.
* **μ=0.55, σ=1**
* The sample distribution is right skewed.
* x¯=0.55, s=1
* Which of the following is true?
* **Standard error computed based on a sample standard deviation will always be lower than the standard deviation of that sample.**
* **In order to reduce the standard error by half, sample size should be quadrupled.**
* **Standard error measures the variability in means of samples of the same size taken from the same population.**
* **As the sample size increases, the variability of the sampling distribution decreases.**
* Students are asked to count the # of chocolate chips in 22 cookies for a class activity. They found cookies on average had 14.77 chocolate chips with w/ SD 4.37 chocolate chips. After collecting the data, a student reports the standard error of the mean to be 0.93 chocolate chips. What is the best way to interpret the student’s result?
* **0.93 chocolate chips is a measure of the variability we’d expect in calculations of the mean number of chocolate chips if we took repeated random samples of 22 cookies.**
* Suppose you took a large # of random samples of size n from a large population, calculated the mean of each sample, + plotted the distribution of your sample means in a histogram. Consider the following possible attributes of your collected data + the population from which they were sampled. For which of the following sets of attributes would you NOT expect your histogram of your sample means to follow a nearly normal distribution?
* n = 120. The population distribution is slightly skewed.
* **n = 10. The population distribution is unknown, but the distribution of data in each sample is heavily skewed.**
* n = 20. The population distribution is nearly normal.
* n = 120. The population distribution is unknown, but the distribution of data in each sample is slightly skewed.
* The General Social Survey (GSS) is a sociological survey used to collect data on demographic characteristics + attitudes of residents of the US. In 2010, the survey collected responses from over a 1K US residents via face-to-face, in-person interview of a randomly-selected sample of adults. 1 of the questions on the survey is “For how many days during the past 30 days was your mental health, which includes stress, depression, and problems with emotions, not good?” Based on responses from 1,151 US residents, the survey reported a 95% CI of 3.40 to 4.24 days in 2010. Given this info, which of the following statements would be most appropriate to make regarding the true average number of days of “not good” mental health in 2010 for US residents?
* For these 1,151 residents in 2010, we are 95% confident that the average number of days of “not good” mental health is between 3.40 and 4.24 days.
* **For all US residents in 2010, there is a 95% probability that the true average number of days of “not good” mental health is between 3.40 and 4.24 days.**
* For all US residents in 2010, based on this 95% confidence interval, we would reject a null hypothesis stating that the true average number of days of “not good” mental health is 5 days.
* There is not sufficient information to calculate the margin of error of this confidence interval.
* All but one of the following CI has a margin of error of 0.7. Which is the confidence interval with the different margin of error?
* **(1.6, 4.4)**
* (−4.7, −3.3)
* (−0.5, 0.9)
* (20.3, 21.7)
* Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of 36 children identified as gifted soon after they reached age 4. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics. Calculate a 90% CI for the average age at which gifted children first count to 10 successfully.



* 
* For each of the following situations, state whether the variable is categorical or numerical, and whether the parameter of interest is a mean or a proportion.
* In a survey, college students are asked whether they agree with their parents’ political ideology.
* **Categorical, proportion**
* In a survey, college students are asked what % of their non-class time they spend studying.
* **Numeric, mean**
* Suppose heights of all women in the US have a mean 63.7 inches, + a random sample of 100 women’s heights yield a sample mean 65.2 inches. Which one is the population parameter and which one is the point estimate? Which one is μ and which one is x¯?
* **pop. param mu = 63.7, point estimate x.bar = 62.5**
* Suppose heights of all women in the US have a SD 2.7 inches, + a random sample of 100 women’s heights yields a SD 4 inches. Which is the population parameter and which is the point estimate? Which one is σ and which one is s?
* **Pop. Param sigma = 2.7 point estimate s = 4**
* Conditions necessary for the CLT to hold. Make sure to list alternative conditions for when we know the population distribution is normal vs. when we don’t know what the population distribution is, and the when the sample size is barely over 30 vs. when it’s very large.
* 6. Confirm that z⋆ for a 98% confidence level is 2.33. (Include a sketch of the normal curve in your response.)
* 7. Calculate a 95% confidence interval for the average height of US women using a random sample of 100 women where the sample mean is 63 inches and the sample standard deviation is 3 inches, and interpret this interval in context of the data.
* 8 Explain, in plain English, the difference between standard error and margin of error.
* 9. A little more challenging: Suppose heights of all men in the US have a mean of 69.1 inches and a standard deviation of 2.9 inches. What is the probability that a random sample of 100 men will yield a sample average less than 70 inches? (Hint: First check if we should expect the sample mean to be distributed nearly normally, i.e. if the CLT holds. If so, sketch a normal curve with mean μ and the appropriate standard error. Shade the area you’re interested in, and calculate it using methods we learned in the previous unit.)