# Lab #5 – Power, Sample Size, and Error Bars

Oftentimes our data isn’t readily available to us and we need to take the initiative to collect it ourselves. This could involve surveying people, performing an experiment on groups of participants, or taking measurements. In most of these situations there is a cost associated with each additional case we collect data on, which prompts the question:

*How do we determine the sample size we need for our study?*

Fortunately the formulas we now have at our disposal for constructing confidence intervals and performing hypothesis tests using the normal distribution can be rearranged to help us decide upon the right sample size.

## Part 1: Sample Size and Interval Estimation

In many instances, particularly when we are surveying people on their opinions or beliefs, we aren’t interested in performing a hypothesis test, but instead just want an accurate picture of the population.

For example we might want to estimate the proportion of voters who prefer a particular political candidate. This sort of estimate is only as useful as the margin error attached to it, if the margin of error is too large what do we really learn?

### Question 1:

Recall that we can obtain an approximate P% confidence interval for a single mean using:

Remembering that is the appropriate critical value for P% from the standard normal distribution.

For Question 1, use this formula to solve for sample size as a function of the Margin of Error (MOE), in doing so you should treat and as constants.

### Question 2:

Recall that the t-distribution provides a more accurate representation when we are forced to estimate and the sample size is small. What might be problematic if we tried to use instead of in the above formula? (Hint: think about degrees of freedom)

Degrees of freedom of the t-distribution depends upon n, making it very difficult to find a formula solving for n

### Question 3:

United Airlines Flight 433 is a regular, non-stop flight from Boston’s Logan Airport to San Francisco International Airport. In 2016, Flight 433 was scheduled to depart each day at 6:00 AM (Eastern time) and arrive before 10:00 AM (Pacific) time. Because of the different time zones, the flight is expected to take about 7 hours (420 minutes), but having an accurate estimate is important for schedulers. The airborne time of a sample of n = 31 flights is provided below. Using this preliminary data, determine how large a sample needs to be collected to achieve a 95% confidence interval estimate for the mean airborne time that has a margin of error of 2 minutes.

|  |
| --- |
| Airtime |
| 353 |
| 358 |
| 374 |
| 360 |
| 346 |
| 373 |
| 363 |
| 351 |
| 380 |
| 346 |
| 369 |
| 374 |
| 385 |
| 377 |
| 377 |
| 370 |
| 372 |
| 356 |
| 407 |
| 356 |
| 360 |
| 348 |
| 351 |
| 359 |
| 369 |
| 384 |
| 368 |
| 398 |
| 402 |
| 388 |
| 381 |

So we need approximately 251 flights to estimate the mean airborne time within 2 minutes with 95% confidence

### Question 4:

Recall that we can obtain an approximate P% confidence interval for a single proportion using:

For Question 4, use this formula to solve for sample size as a function of the Margin of Error (MOE), in doing so you should treat and as constants.

### Question 5:

Polling agencies often ask voters which of two candidates they prefer. This data can be expressed as a proportion interested in voting for a particular candidate (ie: 1 sample categorical data). Suppose we have no preliminary idea regarding what the proportion of voters support the candidate. How large of a sample would be needed to estimate the proportion interested in voting for the candidate within 3% at the 95% confidence level?

### Question 6:

For the scenario described in Question 5, suppose an earlier poll estimates support for the candidate at 42%. How large of a sample would be needed to achieve a 3% margin of error in this scenario?

### Question 7:

Consider your answers to questions 5 and 6. Can you find a preliminary estimate that leads to a larger sample size recommendation than the one you found in question 5? Explain.

No, the answer to question 5 is the largest sample size. A proportion of 0.5 corresponds with the maximum uncertainty regarding which candidate is preferred, thus we should except to need the largest sample in this situation.

## Part 2: Sample Size and Power

When we are planning a study that will rely on statistical hypothesis testing an important thing to consider is how likely are we to reject the null hypothesis *if there actually is a difference*

**Power** is a statistical term referring to the probability of rejecting the null hypothesis given that the null hypothesis is false. It is the opposite of the type II error rate. Often you’ll hear: “how powerful is this test?” meaning how likely is the test to identify that the null hypothesis is wrong.

Power depends upon three factors:

1. Effect size (ie: how different is real parameter from the theorized null value)
2. Sample size (ie: how much data do we have to make our assessment)
3. Significance level (ie: what degree of evidence do we need to reject the null hypothesis)

### Question 8:

For question 8-10 use the app: <http://www.statstudio.net/free-tools/power-analysis/>

1. What does blue distribution represent?
2. What do the dashed vertical lines indicate?
3. Why is the light blue region the probability of making a type I error
4. What does the green distribution represent?
5. Why does the proportion of the green curve to the left of the dashed vertical line indicate a type II error?

### Question 9:

Using the power app with default values (, null mean of 0, null sigma of 1) fill out the following table investigating power under different effect sizes

|  |  |
| --- | --- |
| Effect size | Power |
|  | 0.555 |
|  | 0.705 |
|  | 0.851 |
|  | 0.938 |
|  | 0.979 |
|  | 0.994 |

### Question 10:

Using the power app with default values (null mean of 0, null sigma of 1, alternative mean of 3, alternative sigma of 1) fill out the following table investigating power under different effect sizes

|  |  |
| --- | --- |
| Effect size | Power |
|  | 0.664 |
|  | 0.797 |
|  | 0.851 |
|  | 0.912 |
|  | 0.941 |
|  | 0.957 |

### Question 11:

Recall that the standard error of the sampling distribution and the null distribution is given by:

Considering this fact, along with the visual representation of power you’ve seen earlier in the lab. Do you expect a large sample size to lead to higher or lower power? Explain why.

Larger sample size leads to higher power. Each of the normal distributions shown in the app will become more tightly centered on their means, reducing their overlap (potential type II errors)

### Question 12:

Fill out the following table summarizing how various changes to impact the power of a test (assuming all other factors are held constant)

|  |  |
| --- | --- |
| Change | Power (higher or lower) |
| Larger effect size | higher |
| Larger sample size | higher |
| Smaller significance level () | lower |

## Part 3: Error Bars

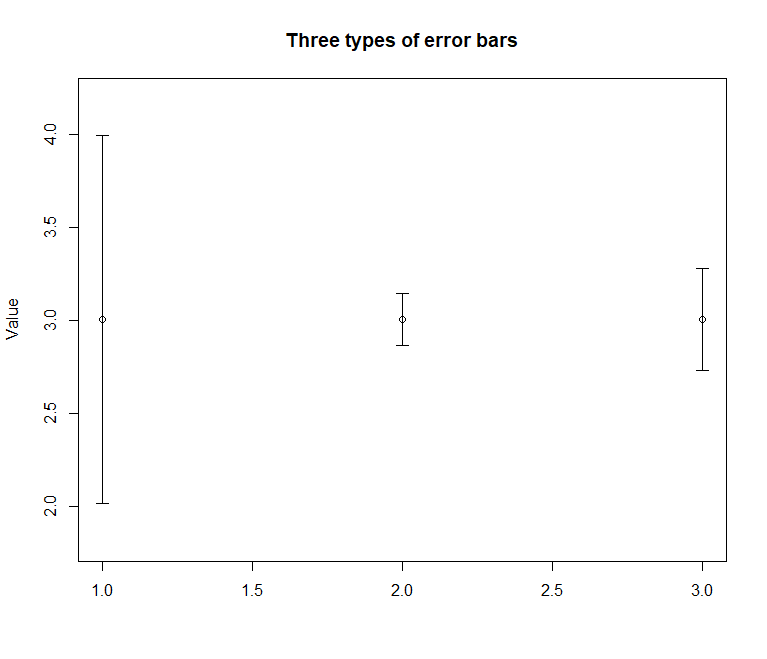
Lately we’ve been studying the standard errors of various statistics, which makes it a good time to look at *error bars*, a commonly included element in many data visualizations. Depending upon the goal of the visualization, error bars can represent very different things, including:

1. The standard error a sample statistic (ie: the sample mean is plotted with bars for +/- 1 standard error). This displays an estimate from the sample and a measure of how reliable that estimate is.
2. The standard deviation of the sample/population (ie: the sample/population mean is plotted with bars for +/- 1 standard deviation). This displays a summary of the sample/population and a measure of how spread out individuals are.
3. The P% confidence interval of a sample statistic (ie: the sample mean is plotted with bars based upon the confidence interval endpoints). This displays an estimate and a measure of how confident we are in that estimate.

### Question 13:

The plot below displays each of these types of error bars for a sample of size n = 50 of a single quantitative variable with a mean of 3. Use the plot to match the following:

|  |  |
| --- | --- |
| Error Bars | Example (left, center, or right) |
| 95% confidence interval | Right |
| 1 standard deviation | Left |
| 1 standard error | Center |



### Question 14:

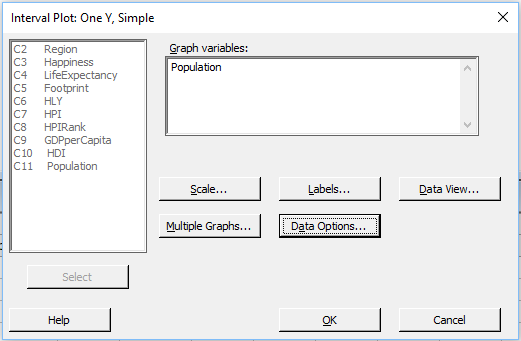
To avoid the ambiguity of different types of error bars, what type of plot could we have used to display our single quantitative variable?

Boxplot!

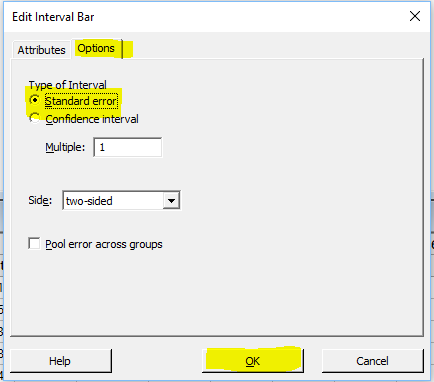
### Question 15:

For the Boston to San Francisco airtime data used in Question 3, construct two plots in Minitab. The first showing the mean airborne time +/- 1 standard error, the second showing the mean airborne time with 95% confidence bounds.

(Hint: use **Graph -> Interval Plot** and follow the screenshots below)



By default, this will produce a plot of the variable’s mean with 95% CI bars. To change to standard error, double click on the interval and use the following menu:



Notice how Minitab’s plot description changes.