

# Inference for a Difference in Means

## Learning Objectives.

- Write out the null and alternative hypothesis for one categorical and one quantitative Variable
- Calculate and carry-out simulation based hypothesis test for a difference in means
- Interpret and evaluate a p-value
- Find a bootstrap confidence interval for the difference in means
- Use a confidence interval to determine the conclusion of a hypothesis test

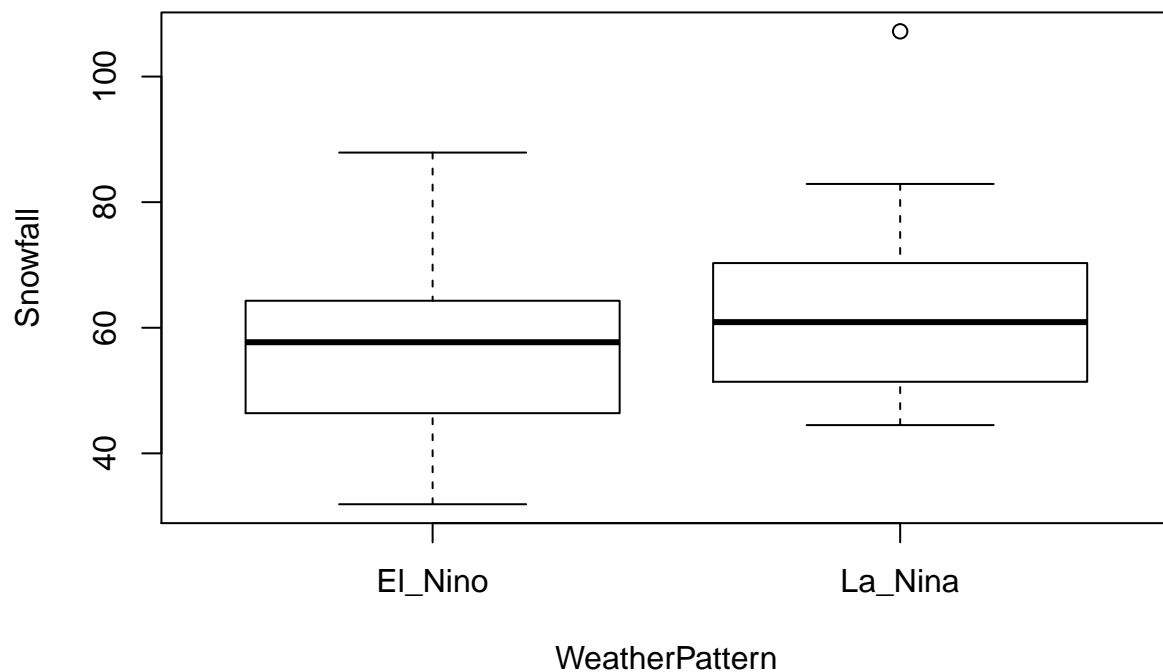
## Background

In the winter of 2018-2019, Bozeman had a record snowfall which resulted in the collapse of two flat-roofed buildings on the MSU campus. A writer for the Washington Post predicted the heavy snowfall for 2018-2019 due to the El Nino weather pattern that occurred in that season. A meteorologist in Montana wanted to see if the weather pattern really was associated with total snowfall. She obtained historical data from 44 years on the weather pattern (El Nino or La Nina) and snowfall (in inches) at the Billings Weather Station.

```
favstats(Snowfall~WeatherPattern, data=Snow)
```

```
##   WeatherPattern  min   Q1 median   Q3   max     mean      sd  n missing
## 1      El_Nino 31.9 46.4   57.7 64.3  87.9 56.23043 13.00823 23      0
## 2      La_Nina 44.5 51.4   60.9 70.3 107.2 63.13333 15.48626 21      0
```

```
boxplot(Snowfall~WeatherPattern, data=Snow)
```



### Quantitative Variables Review

1. The two variables assessed in this study are the type of weather pattern and snowfall. Identify the role for each variable (explanatory, response).
2. Which group (El Nino or La Nina) has the highest center? Explain what measure you are using?
3. Using the side-by-side boxplots, which group has the largest spread? How did you make that choice?

4. Is this an experiment or an observational study? Explain your reasoning.

5. Is this a paired data set or independent groups? Explain your answer.

### **Ask a research question.**

6. Write out the parameter of interest in context of the study. Use proper notation and be sure to define your subscripts. Use El Nino minus La Nina as the order of subtraction.

7. What are the two competing possibilities we will evaluate in this study?

8. Identify which is the null hypothesis and which is the alternative hypothesis.

### **Summarize and Visualize the data.**

9. Calculate the summary statistic. Use El Nino minus La Nina as the order of subtraction. What is the appropriate notation for the statistic?

### **Use statistical inferential methods to draw inferences from the data**

Remember that the null distribution is created based on the assumption the null hypothesis is true. In this study, we assume there is no association between variables. This means that a snowfall value could be in either an El Nino year or a La Nina year.

To demonstrate this your instructor will use cards to represent the sample.

10. How many cards will we start with?

11. What will we write on each card?

12. Next we will mix the cards together and shuffle into two piles. How many cards will go into each pile?  
What should we label the piles?

13. What value is calculated from the cards and plotted on the null distribution?

14. Once we create a null distribution of 1000 simulations, at what value do you expect the distribution to be centered at? Explain your answer.

### **simulation**

15. Load the package CatStats. Using the ....Enter the values for the ....

16. Report the p-value. How much evidence does the p-value provide against the null hypothesis?

17. Using bootstrapping find a 90% confidence interval.

### **bootstrapping simulation**

18. Interpret the interval you calculated in Question 17.

**Communicate the results and answer the research question.**

19. Write a paragraph summarizing the results of the study. Be sure to include:

- Summary statistic
- P-value
- Conclusion in context
- Confidence interval
- Interpretation of the confidence interval
- Scope of inference

**Revisit and Look Forward**

20. Would the results from the theory-based test match the results we saw with the simulation? Explain why or why not.

21. If we had data on 45 La Nina years and 47 El Nino years and found a similar summary statistic, what would happen to the p-value? The width of the confidence interval? The power?