
Weather Patterns and Record Snowfall

9.1 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 a difference in means
- Interpret a confidence interval
- Use a confidence interval to determine the conclusion of a hypothesis test

9.2 Terminology review

In today's activity, we will use simulation-based methods to analyze one categorical and one quantitative variable, where the groups formed by the categorical variable are independent. Some terms covered in this activity are:

- Independent groups
- Difference in means

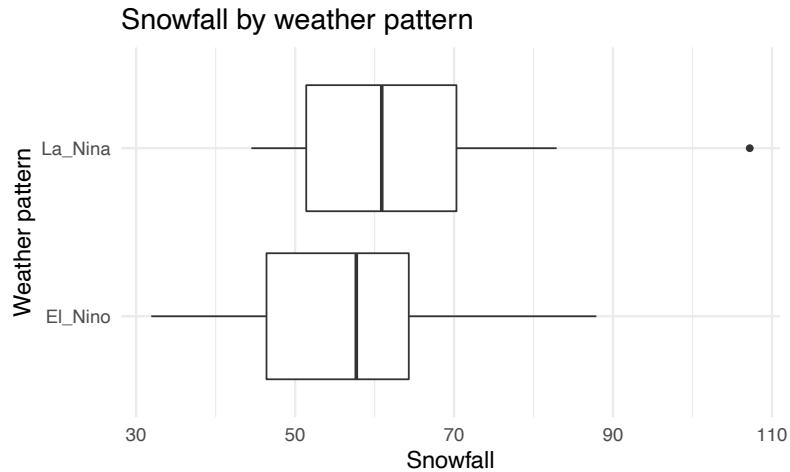
To review these concepts, see Section 6.3 in the textbook.

9.3 Weather patterns and record snowfall

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 Niño 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 Niño or La Niña) and snowfall (in inches) at the Billings Weather Station.

```
Snow <- read.csv("data/SnowfallbyWeatherPattern.csv") # Read in data set
# Code categorical variables as factors
Snow <- # Write over original data with the following
  Snow %>% # Pipe data set into
  mutate(WeatherPattern = factor(WeatherPattern)) # Convert to factor
```

```
# Side-by-side box plots
Snow %>%
ggplot(aes(x = WeatherPattern, y = Snowfall)) +
  geom_boxplot() +
  labs(title = "Snowfall by weather pattern",
       x = "Weather pattern") +
  coord_flip()
```



```
# Summary statistics
Snow %>%
  group_by(WeatherPattern) %>%
  summarise(favstats(Snowfall))
```

```
#> # A tibble: 2 x 10
#>   WeatherPattern   min     Q1 median     Q3    max  mean    sd    n missing
#>   <fct>          <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <int>    <int>
#> 1 El_Nino       31.9  46.4  57.7  64.3  87.9  56.2  13.0    23      0
#> 2 La_Nina       44.5  51.4  60.9  70.3 107.   63.1  15.5    21      0
```

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 Niño or La Niña) has the highest center? Explain which 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? Justify your answer.

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

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 Niño minus La Niña as the order of subtraction.

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

8. Identify which of your answers in question 7 is the null hypothesis and which is the alternative hypothesis.

Summarize and visualize the data

9. Calculate the summary statistic. Use El Niño minus La Niña 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 Niño year or a La Niña 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? Explain your reasoning.

Simulation method

15. Using the provided R markdown file, enter the values for the variables, data set, first in subtraction, number of simulations, observed statistic, and direction of the alternative hypothesis.

```
two_mean_test(RESPONSE~PREDICTOR, data = DATASET, #Variables and data
               first_in_subtraction = "VALUE", #First value in order of subtraction
               number_repetitions = ###, #Number of simulations
               as_extreme_as = ###, #Observed statistic
               direction = "?") #Direction of alternative: "greater", "less", or "two-sided"
```

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. Use the provided R markdown file. Enter the variables, first in subtraction, number of repetitions, and the confidence level.

```
two_mean_bootstrap_CI(RESPONSE~EXPLANATORY, data = DATASET, #Variables and data
  first_in_subtraction = "VALUE", #First value in order of subtraction
  number_repetitions = ###, #Number of simulations
  confidence_level = ##)
```

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 describe:

- Summary statistic
- P-value and interpretation
- Conclusion (written to answer the research question)
- Confidence interval and interpretation
- 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 Niña years and 47 El Niño years and found a similar summary statistic, what would happen to the p-value? The width of the confidence interval? The power of the test?

9.4 Additional notes

Use this space to summarize your thoughts and take additional notes on today's activity.