



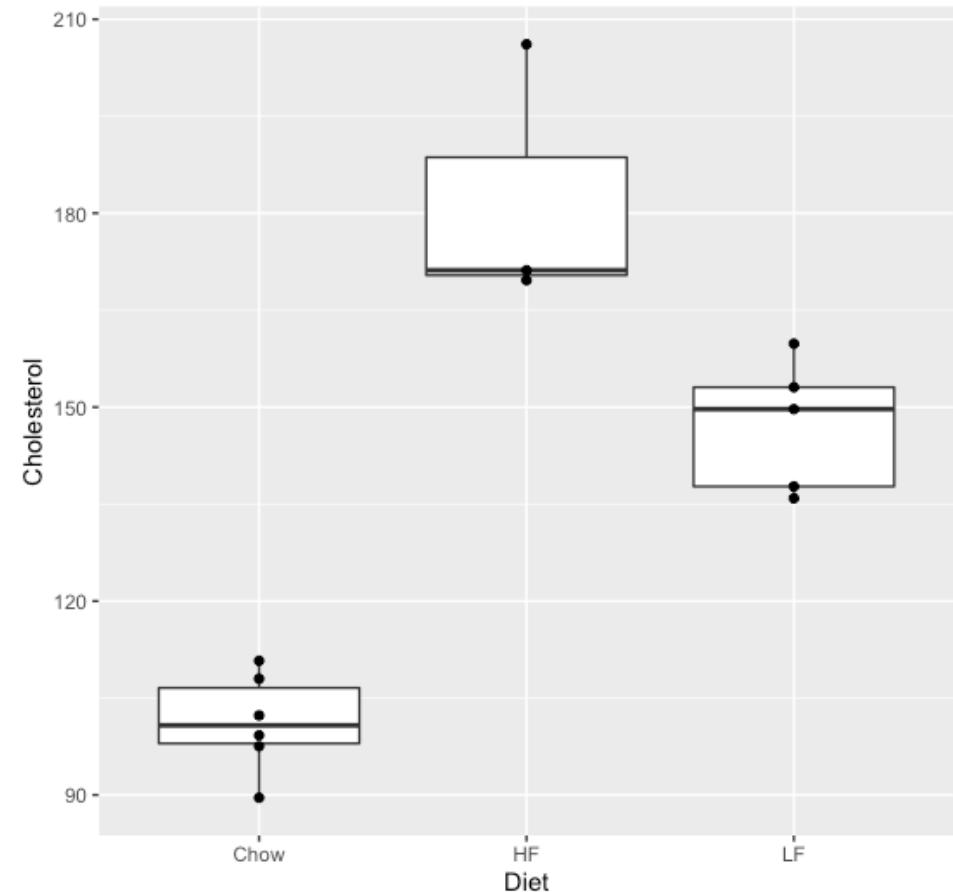
BMR 617

Roles of variables



Introduction

- In many (perhaps all) experiments and studies, we are interested in examining the relationship between two or more variables
- For example, in the data set we looked at last time, we examined the cholesterol level (one variable) of mice fed different diets (another variable)
 - We were examining the relationship between diet and cholesterol





Types of variables

- Recall we already classified variables by *type*
 - Nominal, ordinal, interval, or ratio
 - Nominal and ordinal are categorical
 - Interval and ratio are quantitative
- Diet is ...
- Cholesterol is ...



Types of variables

- Recall we already classified variables by *type*
 - Nominal, ordinal, interval, or ratio
 - Nominal and ordinal are categorical
 - Interval and ratio are quantitative
- Diet is nominal (categorical)
- Cholesterol is ratio (quantitative)



Roles of variables

- For a given experiment or study, we can also classify variables based on their *role* in the experiment
 - Explanatory or response
 - The *response variable* is the particular focus of a question in the study or experiment
 - An *explanatory variable* is a variables that predicts or explains changes in the response variable
- In our previous example:
 - Cholesterol is a ...
 - Diet is a ...



Roles of variables

- For a given experiment or study, we can also classify variables based on their *role* in the experiment
 - Explanatory or response
 - The *response variable* is the particular focus of a question in the study or experiment
 - An *explanatory variable* is a variables that predicts or explains changes in the response variable
- In our previous example:
 - Cholesterol is a response variable
 - Diet is an explanatory variable



Classifying relationships between two variables

- When we look at relationships between two variables, one variable will be the response variable, and the other will be an explanatory variable
- Each variable will either be categorical or quantitative
- We therefore have four possibilities:
 - Categorical explanatory variable, quantitative response: $C \rightarrow Q$
 - Categorical explanatory variable, categorical response: $C \rightarrow C$
 - Quantitative explanatory variable, quantitative response: $Q \rightarrow Q$
 - Quantitative explanatory variable, categorical response: $Q \rightarrow C$
- Understanding which of these cases you are in goes a long way to determining the analysis and presentation of the data



Example: $C \rightarrow Q$

- The example we have already seen: cholesterol levels in the Tallyho mice under three different diets provides an example of a categorical explanatory variable and a quantitative response variable
- We typically present data for this case graphically
- The response variable is plotted on the y-axis
- Graphs are plotted side-by-side using boxplots, column scatter plots, or bar graphs



Recall: boxplots in R

- Remember how we retrieved, wrangled, and filtered our data:

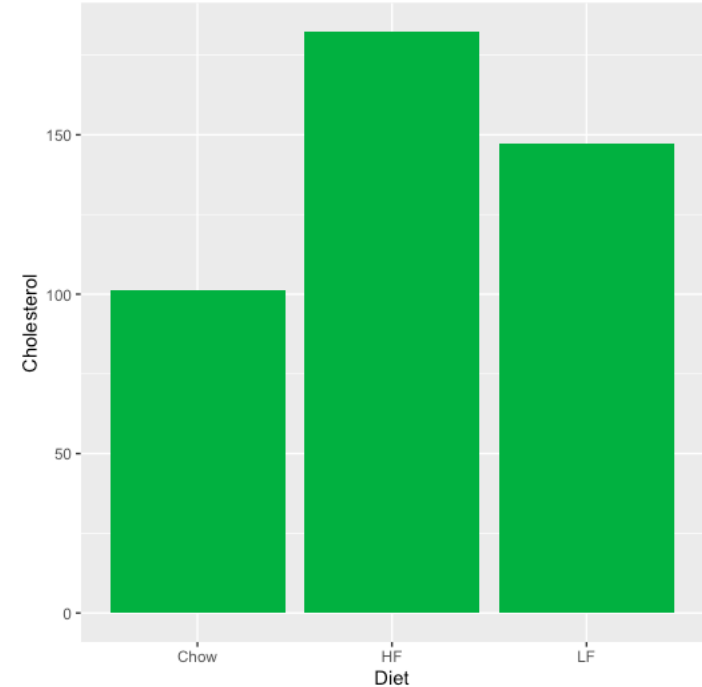
```
library(tidyverse)
met <- read_csv("https://denvirlab.marshall.edu/BMR617-
2021/data/TH-B6-metabolic.csv")
met <- separate(met, MouseID, sep="-", into=c("Strain", "Diet",
"ID"))
th <- filter(met, Strain=="TH")
```

- And we created side-by-side boxplots with
`ggplot(th, aes(x=Diet, y=Cholesterol)) + geom_boxplot()`
- Adding scatter plots on top with
`ggplot(th, aes(x=Diet, y=Cholesterol)) + geom_boxplot() +
geom_point()`



Bar charts in R

- Bar charts in R are a little tricky
 - Do we really like them anyway?
 - ggplot bar graphs are really designed for count data
- We must first *group* and *summarize* the data:
`th_grouped <- group_by(th, Diet)`
- Note this looks no different in the data viewer
- Now we can summarize the data by the group(s), i.e. by Diet:
`th_summ <- summarize(th_grouped, Cholesterol=mean(Cholesterol))`
- And we can plot the bar chart with
`ggplot(th_summ, aes(x=Diet, y=Cholesterol)) +
geom_bar(stat="identity", fill="#00b140")`
- We'll see how to add error bars later





Example: Pfizer SARS-CoV-2 vaccine trial

- Pfizer developed a mRNA vaccine, BNT162b2, for SARS-CoV-2
- Reported the outcomes from a Phase 3 clinical trial to the FDA on December 10, 2020
- Report retrieved from <https://www.fda.gov/media/144246/download>
- In the clinical trial, 18,198 participants received the vaccine, and 18,325 received a placebo
- Two doses in each case, 21 days apart
- Subjects were evaluated to see if they became infected with SARS-CoV-2 within 7 days of the second dose



Variables in the clinical trial

- What is the response variable in this trial? Is it quantitative or categorical?
- What is (are) the explanatory variable(s) in this trial? Is it (are they) quantitative or categorical?



Variables in the clinical trial

- The response variable is SARS-CoV-2 status
 - It is a categorical variable with possible values (“levels”) “infected” and “not infected”
- The explanatory variable is Treatment
 - It is also a categorical variable with possible values “vaccine” and “placebo”
- This is an example of a “ $C \rightarrow C$ ” relationship
- The aim of the study is to see if treatment affects SARS-CoV-2 status



Pfizer BNT162b2 clinical trial design: discussion

- The primary outcome for the trial was SARS-CoV-2 status *seven days after the second dose*
 - Why do you think this outcome (particularly the time period) was chosen?
 - What are the pros and cons of choosing this as the primary outcome?
 - Are there any mitigating factors to any of the cons?



Pfizer BNT162b2 clinical trial results

- Of the 18325 subjects receiving the placebo, 162 were infected with SARS-CoV-2 within a week of the second dose
- Of the 18198 subjects receiving the vaccine, 8 were infected with SARS-CoV-2 within a week of the second dose
- Data for “C → C” type studies are typically presented in a *contingency table*:

		Treatment		
		Placebo	Vaccine	Total
SARS-CoV-2 status	Infected	162	8	170
	Not infected	18163	18190	36353
	Total	18325	18198	36523



Summarizing contingency table data

- There are a number of measures of interest from contingency table data
- For a randomized trial such as this, we can measure the *risk* for each treatment group
- The risk is the number with the “bad” outcome, out of the total
- The risk for the placebo group is $162/18325=0.00884=0.884\%$
- What is the risk for the vaccine group?



Relative risk

- The *relative risk* is the risk in one group as a ratio of the risk in the other group
- The relative risk for the vaccine group (compared to the placebo group) is $0.000440/0.00884 = 0.0497 = 4.97\%$
- So those that receive the vaccine have only about 5% of the risk of those who received the placebo
 - This is essentially the statement that Pfizer made that the vaccine is 95% effective



Attributable risk

- The *attributable risk* is the difference in risk between the two groups
- So the attributable risk for the placebo group is $0.884\% - 0.044\% = 0.84\%$.
- If you are in the placebo group (i.e. if you don't get the vaccine) you have a 0.884% chance of getting SARS-CoV-2 in the four-week timespan represented by the study. Of that 0.884% risk, 0.84% is *attributable to the fact that you did not receive the vaccine*.
- Note that this is not readily interpretable in this case
 - The attributable risk over a more realistic timeline is likely much higher



Number needed to treat

- The *number needed to treat* (NNT) is (in this case) the number of people who need to be vaccinated to prevent one person getting infected with SARS-CoV-2 in the time span of the trial
- NNT is the reciprocal of attributable risk
- In this case, $NNT = 1/0.84\% = 1/0.0084 = 119$
 - Again, this is over a very short time span; the number needed to treat to prevent one person being infected over a more realistic time span is likely much smaller
- To date, approximately 220,000 people in WV have been vaccinated
- Represents about 1845 people saved from infection per four-week time span since vaccination
- At a mortality rate of $\sim 2\%$, this has saved at least 37 lives so far



For next Tuesday

- Produce boxplots with individual points for the *glucose* data for the TH/B6 diet data
- Describe the effect of diet on glucose level for TH, and for B6
- Is the effect of diet the same in both strains?
- Calculate the risk for the placebo groups and vaccine groups, the relative risk, the attributable risk, and the number needed to treat for the Moderna vaccine
 - In the Moderna vaccine trial, 90 out of 13883 of those receiving the placebo, and 5 out of 13934 receiving the vaccine were infected with SARS-CoV-2 within a week of the second dose