Moodle Questions

Homer White

Monday, June 16, 2014

# Seating and the Liberal Arts at UC-Davis I

The ucdavis1 data frame gives information on a random sample of students from the University of California at Davis. One variable, **Seat**, records where the student preferred to sit in a classroom (Front, Middle or Back), and another variable, **Sex**, says whether or not the student was planning to major in one of the liberal arts disciplines.

We are interested in the following Research Question:

*Do male and female students at UC-Davis differ in their seating preferences, and if so, how?*

Which of the two variables in the Research Question should be considered the explanatory variable, and which should be considered the response variable?

* **Sex** is explanatory, and **Seat** is the response.
* **Seat** is explanatory, and **Sex** is the response.
* Neither variable can be considered to be an explanatory variable.
* Both of the variables are explanatory.

# Seating and the Liberal Arts at UC-Davis II

The ucdavis1 data frame gives information on a random sample of students from the University of California at Davis. One variable, **Seat**, records where the student preferred to sit in a classroom (Front, Middle or Back), and another variable, **Sex**, says whether or not the student was planning to major in one of the liberal arts disciplines.

We are interested in the following Research Question:

*Do male and female students at UC-Davis differ in their seating preferences, and if so, how?*

Should the data we gathered to study this question be considered an observational study or an experiment? Select the best answer.

* It's an observational study, because we took a random sample from the population.
* It's an observational study, because the explanatory variable (**sex**) was not assigned to subjects by the researchers who collected the data. The researchers simply recorded the sex of each subject.
* It's an experiment, because we took a random sample from the population.
* It's an experiment because we manipulated something about the situation, and then recorded the value of the response variable **Seat** for each subject.

# Pseudoscorpions: Experiment or Observational Study?

The data for this project is in data frame Pseudoscorpions, from package abd:

require(abd)  
data(Pseudoscorpions)  
View(Pseudoscorpions)  
help(Pseudoscorpions)

From the text *Analysis of Biological Data*, by Michael C. Whitlock and Dolph Schluter:

Pseudoscorpions of the species *Chordylochernes scorpioides* live in tropical forests where they ride on the backs of harlequin beetles to reach rotting figs where they feed. Females of the species are promiscuous and mate with multiple males over short lifetimes. It is unclear what advantages there are for a female to mate multiple times, because the males don't help to care for her young, and mating just once provides all the sperm she needs to fertilize her eggs.

One possible advantage is that the sperm of some males is genetically incompatible with a given female and, by mating multiple times, a female increases the chances of mating with at least one male whose sperm is compatible with her. To investigate this idea, Newcomer et al. (1999) recorded the number of successful broods by female pseudoscorpions randomly assigned to one of two treatments. One group of females was each mated to two different males (DM), whereas females in the other group were each mated twice to the same male (SM). By mating each female twice, the same amount of sperm was provided in both treatments, but the DM females received genetically more diverse sperm than the SM females.

The Pseudoscorpions data frame gives the results of this study.

From the above description, it appears that the investigators were interested in the following Research Question:

*On average, do female pseudoscorpions who mate with two males produce a larger number of successful broods than do female pseudoscorpions who mate with only one male?*

Study the data frame and the above information, and decide which of the following options best explains whether the Pseudoscorpions study was an observational study or a randomized experiment.

* **Observational study**, because the response variable was **successful.broods**, and the researchers merely observed how many successful broods each female in the study produced: they did not randomly assign the number of broods to each female.
* **Randomized experiment**, because the explanatory variable was **treatment**, and the researchers randomly assigned the value of this variable to the females in the study: some females were assigned to mate with one male (value "SM") and the others were assigned to mate with two males (value "DM").
* **Randomized experiment**, because the number of males you mate with is believed to affect the genetic diversity of the sperm you store, which in turn affects your chance of a producing a successful brood.
* **Observational study**, because for a randomized experiment you would have to take a random sample of female Psuedoscorpions from the population, and nowhere in the above information did we learn that this was done.

# Pseudoscorpions: Investigating the Research Question

The data for this project is in data frame Pseudoscorpions, from package abd:

require(abd)  
data(Pseudoscorpions)  
View(Pseudoscorpions)  
help(Pseudoscorpions)

From the text *Analysis of Biological Data*, by Michael C. Whitlock and Dolph Schluter:

Pseudoscorpions of the species *Chordylochernes scorpioides* live in tropical forests where they ride on the backs of harlequin beetles to reach rotting figs where they feed. Females of the species are promiscuous and mate with multiple males over short lifetimes. It is unclear what advantages there are for a female to mate multiple times, because the males don't help to care for her young, and mating just once provides all the sperm she needs to fertilize her eggs.

One possible advantage is that the sperm of some males is genetically incompatible with a given female and, by mating multiple times, a female increases the chances of mating with at least one male whose sperm is compatible with her. To investigate this idea, Newcomer et al. (1999) recorded the number of successful broods by female pseudoscorpions randomly assigned to one of two treatments. One group of females was each mated to two different males (DM), whereas females in the other group were each mated twice to the same male (SM). By mating each female twice, the same amount of sperm was provided in both treatments, but the DM females received genetically more diverse sperm than the SM females.

The Pseudoscorpions data frame gives the results of this study.

From the above description, it appears that the investigators were interested in the following Research Question:

*On average, do female pseudoscorpions who mate with two males produce a larger number of successful broods than do female pseudoscorpions who mate with only one male?*

Which of the following bits of R-code would be appropriate for investigating the above Research Question? (There is more than one correct answer.)

Option A:

favstats(successful.broods~treatment,data=Pseudoscorpions)

Option B:

densityplot(~successful.broods,data=Pseudoscorpions,  
 groups=treatment,  
 auto.key=TRUE,  
 xlab="Number of successful broods",  
 main="Pseudoscorpions Study")

Option C:

densityplot(~successful.broods,data=Pseudoscorpions,  
 xlab="Number of successful broods",  
 main="Pseudoscorpions Study")

Option D:

barchartGC(~treatment,data=Pseudoscorpions,  
 type="frequency",  
 main="Pseudoscorpions Study")

Option E:

favstats(~successful.broods,data=Pseudoscorpions)

# A Small Experiment: Completely Randomized Design

Recall the imaginary list of subjects, stored in the data frame SmallExp, discussed in Chapter 6:

data(SmallExp)  
View(SmallExp)  
help(SmallExp)

For your convenience, here's the entire data frame:

SmallExp

## name sex athlete  
## 1 Linda female yes  
## 2 Deb female yes  
## 3 Caitlin female yes  
## 4 Yolanda female yes  
## 5 Valerie female no  
## 6 Anne female no  
## 7 Christine female no  
## 8 Priya female no  
## 9 Jorge male yes  
## 10 Jeff male yes  
## 11 Charles male yes  
## 12 Mike male yes  
## 13 Marcus male no  
## 14 Kumar male no  
## 15 Ed male no  
## 16 Alexei male no

George plans to perform an experiment that involves assigning the subjects to two treatment groups. One treatment (called "music") involves listening to headphone-music while going through a stationary bike workout, and the other treatment (called "control") involves just going through the workout, without any music. George will measure how many calories each person burns during the workout.

Which of the following assignments of subjects to treatment group is possible if George plans to use a completely randomized design to divide subjects into two groups of eight subjects each? (More than one answer may be correct.)

Option A:

set.seed(2020)  
opta <- RandomExp(SmallExp,sizes=c(8,8),  
 groups=c("control","music"))  
opta

## name sex athlete treat.grp  
## 2 Deb female yes control  
## 4 Yolanda female yes control  
## 5 Valerie female no control  
## 6 Anne female no control  
## 8 Priya female no control  
## 10 Jeff male yes control  
## 15 Ed male no control  
## 16 Alexei male no control  
## 1 Linda female yes music  
## 3 Caitlin female yes music  
## 7 Christine female no music  
## 9 Jorge male yes music  
## 11 Charles male yes music  
## 12 Mike male yes music  
## 13 Marcus male no music  
## 14 Kumar male no music

Option B:

set.seed(2610)  
optb <- RandomExp(SmallExp,sizes=c(8,8),  
 groups=c("control","music"))  
optb

## name sex athlete treat.grp  
## 3 Caitlin female yes control  
## 4 Yolanda female yes control  
## 5 Valerie female no control  
## 7 Christine female no control  
## 8 Priya female no control  
## 11 Charles male yes control  
## 14 Kumar male no control  
## 16 Alexei male no control  
## 1 Linda female yes music  
## 2 Deb female yes music  
## 6 Anne female no music  
## 9 Jorge male yes music  
## 10 Jeff male yes music  
## 12 Mike male yes music  
## 13 Marcus male no music  
## 15 Ed male no music

Option C:

set.seed(3320)  
optc <- RandomExp(SmallExp,sizes=c(8,8),  
 block="sex",  
 groups=c("control","music"))  
optc

## name sex athlete treat.grp  
## 2 Deb female yes control  
## 5 Valerie female no control  
## 6 Anne female no control  
## 7 Christine female no control  
## 9 Jorge male yes control  
## 10 Jeff male yes control  
## 14 Kumar male no control  
## 15 Ed male no control  
## 1 Linda female yes music  
## 3 Caitlin female yes music  
## 4 Yolanda female yes music  
## 8 Priya female no music  
## 11 Charles male yes music  
## 12 Mike male yes music  
## 13 Marcus male no music  
## 16 Alexei male no music

Option D:

set.seed(5555)  
optd <- RandomExp(SmallExp,sizes=c(8,8),  
 block="sex",  
 groups=c("control","music"))  
optd

## name sex athlete treat.grp  
## 1 Linda female yes control  
## 2 Deb female yes control  
## 6 Anne female no control  
## 8 Priya female no control  
## 11 Charles male yes control  
## 12 Mike male yes control  
## 13 Marcus male no control  
## 14 Kumar male no control  
## 3 Caitlin female yes music  
## 4 Yolanda female yes music  
## 5 Valerie female no music  
## 7 Christine female no music  
## 9 Jorge male yes music  
## 10 Jeff male yes music  
## 15 Ed male no music  
## 16 Alexei male no music

# A Small Experiment: Randomized Block Design

Recall the imaginary list of subjects, stored in the data frame SmallExp, discussed in Chapter 6:

data(SmallExp)  
View(SmallExp)  
help(SmallExp)

For your convenience, here's the entire data frame:

SmallExp

## name sex athlete  
## 1 Linda female yes  
## 2 Deb female yes  
## 3 Caitlin female yes  
## 4 Yolanda female yes  
## 5 Valerie female no  
## 6 Anne female no  
## 7 Christine female no  
## 8 Priya female no  
## 9 Jorge male yes  
## 10 Jeff male yes  
## 11 Charles male yes  
## 12 Mike male yes  
## 13 Marcus male no  
## 14 Kumar male no  
## 15 Ed male no  
## 16 Alexei male no

George plans to perform an experiment that involves assigning the subjects to two treatment groups. One treatment (called "music") involves listening to headphone-music while going through a stationary bike workout, and the other treatment (called "control") involves just going through the workout, without any music. George will measure how many calories each person burns during the workout.

Which of the following assignments of subjects to treatment group is possible if George plans to use a randomized block design, blocking by the possible confounder variable **sex**, to divide subjects into two groups of eight subjects each, in such a way that the percentage of males in each group is the same? (More than one answer may be correct.)

(Present same set of options as in the last question.)

# The Beans Experiment: What is the Design?

In the tigerstats package we find the beans data frame:

data(beans)  
View(beans)  
help(beans)

This was an experiment performed at UC-Davis; fifteen students participated. Each student was asked to place as many beans into a cup as he/she could, in 15 seconds. Each student performed this task once with the dominant hand, and once with the non-dominant hand, but the order of performance was randomized. The purpose of the study was to see whether manual dexterity was better for the dominant hand. Assume that the subjects in the experiment are the fifteen students. (*Reminder about terminology*: your dominant hand is the hand you use the most.)

What was the design of this experiment? Choose the most accurate option below.

Option A: Completely Randomized Design.

Feedback: This is tempting response, because researchers randomized the order in which the subjects performed the work. However, in a a randomized design you assign subjects randomly to the various treatment groups, and this time there weren't even treatment groups. This study is actually a repeated measures design, with the randomization being done to correct for bias that could arise if practice moving beans makes the subject better at moving them. (For a similar example, see the labels experiment in the Chapter Six course notes (Repeated Measures section) and in the Chapter Six Part 2 slides.)

Option B: Randomized Block Design

Feedback: There was no blocking variable.

Option C: Matched Pairs Design.

Feedback: Close, but no cigar. There weren't any blocks, so there weren't any blocks of size two each to make the "matched pairs". If you thought of the subjects in the experiment as the 30 hands (two for each person), then you could say that we have a matched pairs design where each pair is the pair of hands of a person, but in the directions we said to think of the subjects as the fifteen people.

Option D: Repeated Measures Design.

Feedback: Correct! We made two measurements on each subject: one measurement is the number of beans moved with the dominant hand, and the other is the number of beans moved with the non-dominant hand.

# Experiment Terminology

Researchers wanted to know which treatment is more effective for relieving joint pain: an ibuprofen pill, or a newly-developed topical ointment rubbed on the affected joint. Fifty subjects agreed to participate in the experiment. Researchers had the following supplies on hand:

* real ibuprofen pills;
* pills that looked and tasted like ibuprofen pills but which were actually made of an inert substance;
* the real topical ointment;
* a topical ointment that looked and felt like the real ointment but which actually has no affect on the body.

Twenty-five of the subjects were randomly selected to take the real pills and use the fake ointment. The remaining twenty-five used the fake pills and the real ointment. The subjects did not knew that they were using one real and one fake item, but they did not know which was which.

At the end of the period of study, each patient reported their decrease in pain on a numerical scale form 0 to 10, where 0 stands for "no decrease in pain at all" and 10 stands for "my pain is all gone!"

Which of the following terms apply to this experiment? Select all correct answers.

Option A: Completely Randomized Design

Feedback: Yes, there two treatments (real pills vs. real ointment) and the subject were divided randomly into these two groups.

Option B: Randomized Block Design

Feedback: There was no blocking variable.

Option C; Matched Pairs Design.

Feedback: Treatments were give in pairs (one real and one fake), but the subjects were not blocked into matched pairs, so this is not a matched pairs design.

Option D: Repeated Measures Design

Feedback: No each subject reported their decrease in pain just once.

Option E: Double-Dummy Design

Feedback: Yes, there were two dummy treatments: the fake pills and the fake ointment.

Option F: Blinding

Feedback: Yes, the double-dummy design prevented the subjects from knowing which group they were in.

Option G: Replication

Feedback: Yes, there was more than one subject in each treatment group (there were 25 in each group, in fact).

# A favstats() Review: The Attitudes Experiment

Review the attitudes data frame (as discussed in Chapter 6 Course Notes and slides):

data(attitudes)  
View(attitudes)  
help(attitudes)

Match the question with the favstats() R-code that will help to answer it.

What was the mean sentence recommended by all the subjects?

favstats(~sentence,data=attitudes)

What was the median sentence recommended by all the subjects?

favstats(~sentence,data=attitudes)

What was the mean sentence reommended by the subjects who were looking at a form that suggested the race of the defendant was White?

favstats(sentence~def.race,data=attitudes)

What was the mean sentence reommended by the subjects who were looking at a form that suggested the race of the defendant was Black?

favstats(sentence~def.race,data=attitudes)

What was the mean sentence recommended by the subjects who were looking at a form that suggested the race of the victim was Black?

favstats(sentence~vic.race,data=attitudes)

# The Difference of Two Sample Proportions: I

Belinda plans to take a simple random sample of 2500 U.S. adults and to compute the proportion of them who believe in the theory of evolution. (Call this sample proportion p1-hat.)

She also intends to take an independent simple random sample of 1600 adults from France and to compute the the proportion of them who believe in the theory of evolution. (Call this sample proportion p2-hat.)

Unknown to Belinda, the proportion of all U.S. adults who believe in the theory of evolution is 0.50 (or 50%), whereas the proportion of all French adults who believe in the theory of evolution is 0.75, or 75%.

What is the expected value of the difference in sample proportions, p1-hat minus p2-hat? Give your answer as a proportion (a number between 0 and 1), not as a percentage. Also, make sure your answer is correct to two decimal places.

# The Difference of Two Sample Proportions: II

Belinda plans to take a simple random sample of 2500 U.S. adults and to compute the proportion of them who believe in the theory of evolution. (Call this sample proportion p1-hat.)

She also intends to take an independent simple random sample of 1600 adults from France and to compute the the proportion of them who believe in the theory of evolution. (Call this sample proportion p2-hat.)

Unknown to Belinda, the proportion of all U.S. adults who believe in the theory of evolution is 0.50 (or 50%), whereas the proportion of all French adults who believe in the theory of evolution is 0.75, or 75%.

What is the standard deviation of the difference in sample proportions, p1-hat minus p2-hat? Give your answer as a proportion (a number between 0 and 1), not as a percentage. Also, make sure your answer is correct to two decimal places.

sqrt(0.5\*0.5/2500+0.75\*0.25/1600)

## [1] 0.01474

# The Difference of Two Sample Proportions: III

Belinda plans to take a simple random sample of 2500 U.S. adults and to compute the proportion of them who believe in the theory of evolution. (Call this sample proportion p1-hat.)

She also intends to take an independent simple random sample of 1600 adults from France and to compute the the proportion of them who believe in the theory of evolution. (Call this sample proportion p2-hat.)

Unknown to Belinda, the proportion of all U.S. adults who believe in the theory of evolution is 0.50 (or 50%), whereas the proportion of all French adults who believe in the theory of evolution is 0.75, or 75%.

Using the expected value and the SD of p1-hat minus p2-hat, which of the following makes the most sense?

Option A

The percentage of adults in the U.S. sample who believe in evolution should turn out to be about 25% less than the percentage of adults in the french sample who believe in evolution, give or take about 1.5% or so for chance error in the sampling process.

Feedback: Right! You expect a random variable to turn out to be about its EV, give or take an SD or so.

Option B

The percentage of adults in the U.S. sample who believe in evolution should turn out to be about 1.5% less than the percentage of adults in the French sample who believe in evolution, give or take about 25% or so for chance error in the sampling process.

Feedback: In this statement the EV and SD are switched around from where they should be.

Option C

The percentage of adults in the U.S. sample who believe in evolution should turn out to be about 25% more than the percentage of adults in the French sample who believe in evolution, give or take about 1.5% or so for chance error in the sampling process.

Feedback: It seems that in this statement the sample two populations got switched around.

Option D

The percentage of adults in the U.S. sample who believe in evolution should turn out to be about 25% less than the percentage of adults in the French sample who believe in evolution, give or take about 3% or so for chance error in the sampling process.

Feedback: The 3% figure is actually about TWO SDs, so by the 68-95 rule there is about a 95% chance that the difference will be within 3% of the 25% figure. The "or so" in the expression doesn't sound right, since the chance of being more than 30% away is rather small.

# The Difference of Two Sample Proportions: IV

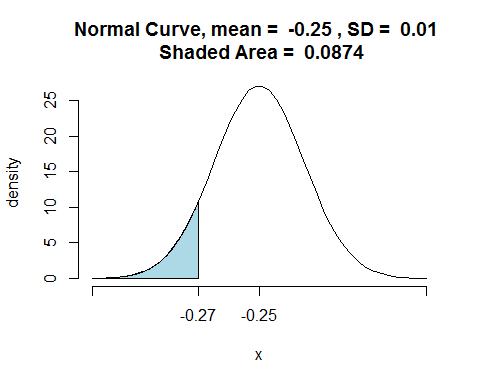
Belinda plans to take a simple random sample of 2500 U.S. adults and to compute the proportion of them who believe in the theory of evolution. (Call this sample proportion p1-hat.)

She also intends to take an independent simple random sample of 1600 adults from France and to compute the the proportion of them who believe in the theory of evolution. (Call this sample proportion p2-hat.)

Unknown to Belinda, the proportion of all U.S. adults who believe in the theory of evolution is 0.50 (or 50%), whereas the proportion of all French adults who believe in the theory of evolution is 0.75, or 75%.

Using the normal approximation: approximately what is the probability that p1-hat minus p2-hat will turn out to be less than or equal to -0.27 (minus 27%)? Write your answer as a decimal (between 0 and 1), not as a percentage. It should be correct to at least two decimal places.

myEV <- -0.25  
mySD <- sqrt(0.5\*0.5/2500+0.75\*0.25/1600)  
pnormGC(-0.27, mean=myEV, sd=mySD, graph=TRUE)



## [1] 0.08737

# The 68-95 Rule for Probability: I

Belinda plans to take a simple random sample of size 64 from a population, and to take the mean x-bar of heights of the people in the sample. Unknown to Belinda, the SD of the population happens to be 16 inches, so the SD of x-bar is

16/sqrt(64)

## [1] 2

Match each chance with its approximate value.

The chance that x-bar will be 2 or more inches below the mean of the population.

The chance that x-bar will be within 4 inches of the mean of the population.

The chance that x-bar will be more than 6 inches above the mean of the population.

The chance that x-bar will be within 6 inches of the mean of the population.

The chance that x-bar will be within 4 inches of the mean of the sample.

# The 68-95 Rule for Probability: II

Harriet plans to take a simple random sample of size 36 from a population, and to take the mean x-bar of the weights of the people in the sample. Unknown to Harriet, the mean of the population is 150 pounds and the SD of the population is 6 pounds. Which of the following are true?

There is about a 68% chance that x-bar will be somewhere between 149 and 151 pounds.

There is about a 2.5% chance that x-bar will be more than 152 pounds.

There is about an 84% chance that x-bar will be less than 151 pounds.

There is about a 68% chance that x-bar will be somewhere between 144 and 156 pounds.

There is about a 2.5% chance that x-bar will be more than 162 pounds.

There is about an 84% chance that x-bar will be less than 156 pounds.

Feedback for wrong answers: Wrong SD. You need to work with the SD of x-bar, not the SD of the population. The SD of x-bar is 6/sqrt(36), which works out to 1 pound.

# One Sample Mean: I

George plans to take simple random sample of size 50 from a population, and to compute the mean x-bar of the weights of the people in the sample. Unknown to George, the mean of the population is 145 and the SD of the population is 23.

Using the normal-curve approximation from Central Limit Theorem, compute the approximate probability that x-bar will turn out to be between 141 and 149 pounds.

Write your answer as a decimal between 0 and 1, not as a percentage, and make sure that it is correct to at least two decimal places.

mySD <- 23/sqrt(50)  
pnormGC(c(141,149),region="between",mean=145,sd=mySD)

## [1] 0.7812

# One Sample Mean: II

George plans to take simple random sample of size 50 from a population, and to compute the mean x-bar of the weights of the people in the sample. Unknown to George, the mean of the population is 145 and the SD of the population is 23.

Using the normal-curve approximation from Central Limit Theorem, compute the approximate probability that x-bar will turn out to be more than 147 pounds.

Write your answer as a decimal between 0 and 1, not as a percentage, and make sure that it is correct to at least two decimal places.

mySD <- 23/sqrt(50)  
pnormGC(147,region="above",mean=145,sd=mySD)

## [1] 0.2693

# One Sample Mean: III

George plans to take simple random sample of size 50 from a population, and to compute the mean x-bar of the weights of the people in the sample. Unknown to George, the mean of the population is 145 and the SD of the population is 23.

Using the normal-curve approximation from Central Limit Theorem, compute the approximate probability that x-bar will turn out to be less than 148 pounds.

Write your answer as a decimal between 0 and 1, not as a percentage, and make sure that it is correct to at least two decimal places.

mySD <- 23/sqrt(50)  
pnormGC(148,region="below",mean=145,sd=mySD)

## [1] 0.8218