Homework due May 13, 2021 23:01 +03

Install the latest version of the **dagdata** packge from the genomicsclass github repository. Load the admissions data from the **dagdata** package:

```
library(dagdata)
data(admissions)
```

Familiarize yourself with this table:

```
print( admissions )
```

You can also obtain this data directly here.

Confounding Exercises #1

1/1 point (graded)

Let's compute the proportion of men who were accepted:

```
index = which(admissions$Gender==1)
accepted= sum(admissions$Number[index] * admissions$Percent
applied = sum(admissions$Number[index])
accepted/applied
```

What is the proportion of women that were accepted?

Note: The code sample above gives the proportion of *men* accepted. Alter the code to find the proportion of *women* instead.

0.30133351

Answer: 0.3033351

0.30133351

Explanation

```
index = which(admissions$Gender==0)
accepted= sum(admissions$Number[index] * admissions$Percent
applied = sum(admissions$Number[index])
accepted/applied
```

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem

Confounding Exercises #2

1/1 point (graded)

Now that we have observed different acceptance rates between genders, test for the significance of this result.

If you perform a chi-square independence test, what is the p-value?

Hint: create a table that has the totals for accepted and not-accepted by gender then use `chisq.test()`.

91.139492e-22

Answer: 9.139492e-22

$91.139492 \times 10^{-22}$

Explanation

```
index = admissions$Gender==1
men = admissions[index,]
women = admissions[!index,]
menYes = sum(men$Number*men$Percent/100)
menNo = sum(men$Number*(1-men$Percent/100))
womenYes = sum(women$Number*women$Percent/100)
womenNo = sum(women$Number*(1-women$Percent/100))
tab = matrix(c(menYes,womenYes,menNo,womenNo),2,2)
chisq.test(tab)$p.value
```

This difference actually led to a lawsuit.

Now notice that looking at the data by major, the differences disappear.

```
index = admissions$Gender==1
men = admissions[index,]
women = admissions[!index,]
print( data.frame( major=admissions[1:6,1],men=men[,3], wom
```

How can this be? This is referred to as Simpson's Paradox.
In the following questions we will try to decipher why this is happening.

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

Confounding Exercises #3

1/1 point (graded)
We can quantify how "hard" a major is using the percent of students that were accepted. Compute the percent that were accepted (regardless of gender) to each major and call this vector `H`.

Which is the hardest major?
Enter your answer as a letter.

F

Answer: F

Explanation

```
major = admissions[1:6,1]
men = admissions[1:6,]
women =admissions[7:12,]
H = (men$Number*men$Percent/100 + women$Number*women$Percen
major[which.min(H)]
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

Confounding Exercises #4

1/1 point (graded)

What proportion of students is admitted for the hardest major from Confounding Exercises #3?

0.06477591

Answer: 0.06477591

0.06477591

Explanation

```
major = admissions[1:6,1]
men = admissions[1:6,]
women =admissions[7:12,]
H = (men$Number*men$Percent/100 + women$Number*women$Percen
min(H)
```

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem

Confounding Exercises #5

1/1 point (graded)

For men, what is the correlation between the number of applications across majors and `H`?

0.7647567

Answer: 0.7647567

0.7647567

Explanation

```
cor(H,men$Number)
```

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem

Confounding Exercises #6

1/1 point (graded)

For women, what is the correlation between the number of applications across majors and `H` ?

-0.6743393

Answer: -0.6743393

-0.6743393

Explanation

```
cor(H,women$Number)
```

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem

Confounding Exercises #7

1/1 point (graded)

Given the answers to Confounding Exercises #5 and #6, which best explains the differences in admission percentages when we combine majors?

- We made a coding mistake when computing the overall admissions percentages.

- ( ) There were more total number of women applications which made the denominator much bigger.
- (•) There is confounding between gender and preference for "hard" majors: females are more likely to apply to harder majors.
- ( ) The sample size for the individual majors was not large enough to draw the correct conclusion.

✔

Submit

You have used 1 of 5 attempts

Answers are displayed within the problem