

Practice test 1 solutions

Zack Treisman

9/22/2021

Monte Vista Birds

There are many possible answers. Here is one.

Variables: Age, sex, wingspan

Individuals/ observations are individual birds

Age and wingspan are numerical, sex is categorical

Is average wingspan different for male and female birds?

Sex is explanatory, wingspan is response

Cocaine Treatment

I'll write R for relapse, D for desipramine, L for lithium and C for control (placebo).

$P(R|D) = 10/24$, $P(R|L) = 18/24$, $P(R|C) = 20/24$

At first glance, it appears to me that desipramine is probably more effective than lithium or placebo, and lithium may be more effective than a placebo, though that is certainly debatable.

It is not possible to conclude significant differences in efficacy from sample proportions alone. We will need to carefully combine these differences with sample sizes to determine if the differences are significant.

Empathy study

```
empathy <- c(38,53,41,55,56,61,62,48,43,47,56,65,19,61,32,105)
mean(empathy)
```

```
## [1] 52.625
```

```
median(empathy)
```

```
## [1] 54
```

```
sd(empathy)
```

```
## [1] 18.6257
```

```
quantile(empathy, 0.25)
```

```
## 25%
```

```
## 42.5
```

```
quantile(empathy, 0.75)
```

```
## 75%
```

```
## 61
```

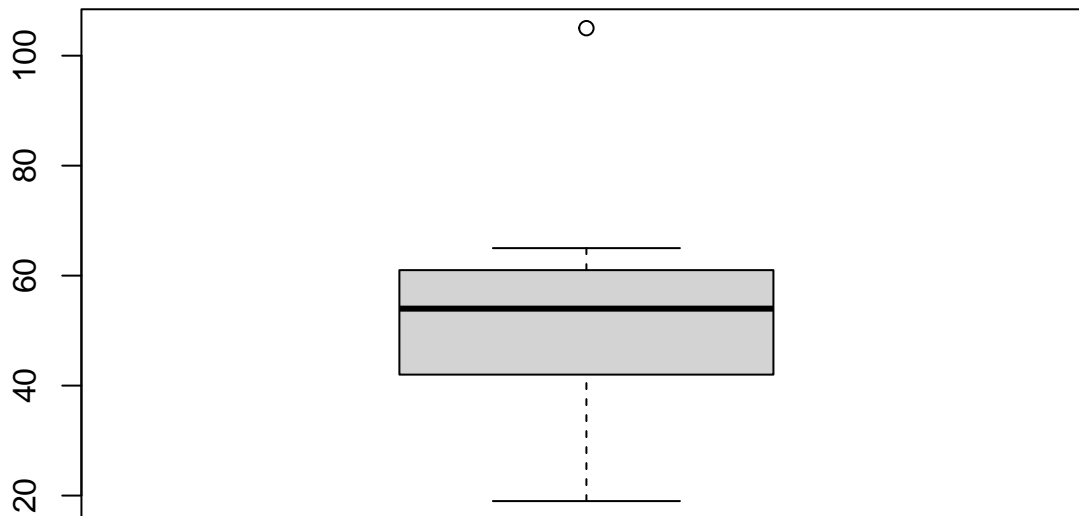
```
quantile(empathy) #for the whole five number summary
```

```
##      0%    25%    50%    75%   100%  
##  19.0  42.5  54.0  61.0 105.0
```

```
IQR(empathy)
```

```
## [1] 18.5
```

```
boxplot(empathy)
```



Summary stats

Use the mean and standard deviation for symmetric data without severe outliers. Use median and quartiles when there is skew or outliers.

Summary of histogram

The data are roughly normal, so use the mean for center and sd spread. There is very small amount of left skew, but probably not enough to make median and IQR preferable.

Movie characters

The population of interest is characters in popular American movies. The sample is the 4342 characters surveyed. If the sample was randomly collected, then the results can be generalized. Since this is an observational study, a causal relationship can not be demonstrated.

Prairie dogs

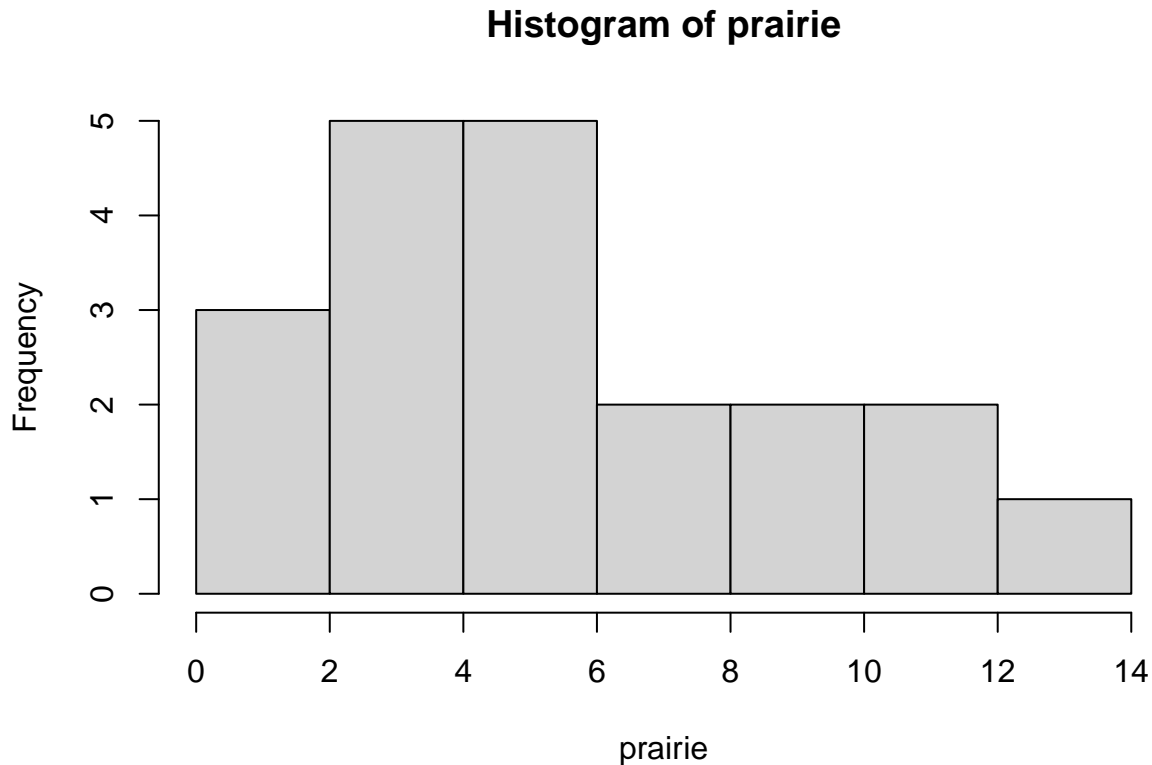
```
prairie<-c(7, 5, 3, 10, 2, 5, 13, 3, 12, 4, 4, 1, 9, 7, 5, 6, 3, 5, 12, 1)  
sum(prairie<8)/length(prairie) #P(fewer than 8)=0.75
```

```
## [1] 0.75
```

```
sum(prairie>=10)/length(prairie) #P(10 or more)=0.2
```

```
## [1] 0.2
```

```
hist(prairie)
```



The distribution is unimodal with a peak around 4, and somewhat right skewed.

College drug use

parameter; statistic

Demographic data

Income is categorical in these data

scatter plot

The minimum value of 30 would be identified as outlier in a box plot. Note there is a typo in this problem because the data show values of score less than 30. But a value of 30 is more than 1.5 IQR below the median ($30 < 69.5 - 1.5(77 - 57)$).

Government and economy

Observational - stratified

Phones and sleep

The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can say the response and explanatory variables are not independent.

Cars and transit

0.125

34/160 vs 94/111

(iv)

Breakfast and cognition

Block on typical breakfast habits.

Sleeping flies

mean = 24, median = 18

Disease probabilities

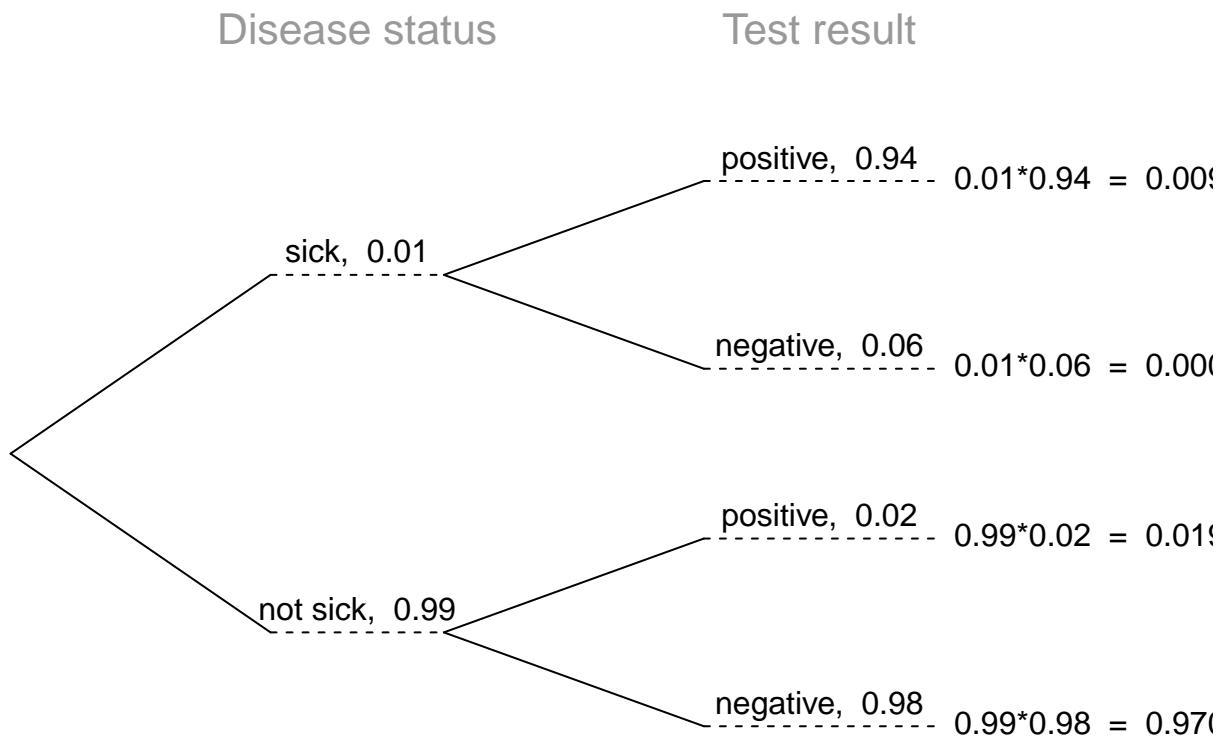
```
library(openintro)
```

```
## Loading required package: airports
```

```
## Loading required package: cherryblossom
```

```
## Loading required package: usdata
```

```
treeDiag(c('Disease status', 'Test result'),  
  out1 = c("sick", "not sick"),  
  out2 = c("positive", "negative"),  
  c(0.01, 0.99),  
  list(c(0.94, 0.06), c(0.02, 0.98)),  
  showWork = TRUE)
```



```
0.009/(0.009+0.019) #P(sick/test+)=0.3214
```

```
## [1] 0.3214286
```

Gambler

```
(0-10)*0.5+(10-10)*0.25+(20-10)*0.20+(100-10)*0.05
```

```
## [1] 1.5
```

She should expect to win \$1.50 per game, so if she plays 100 games she can expect to have won \$150.