Untitled

Carsten Ersch

21 December 2016

Statistical interference - Comparing means

Some examples taken from http://onlinestatbook.com/2/tests_of_means/ch10_exercises.html

In genral this is to show how to test for statistical differences between groups

Question 1

The scores of a random sample of 8 students on a physics test are as follows: 60, 62, 67, 69, 70, 72, 75, and 78.

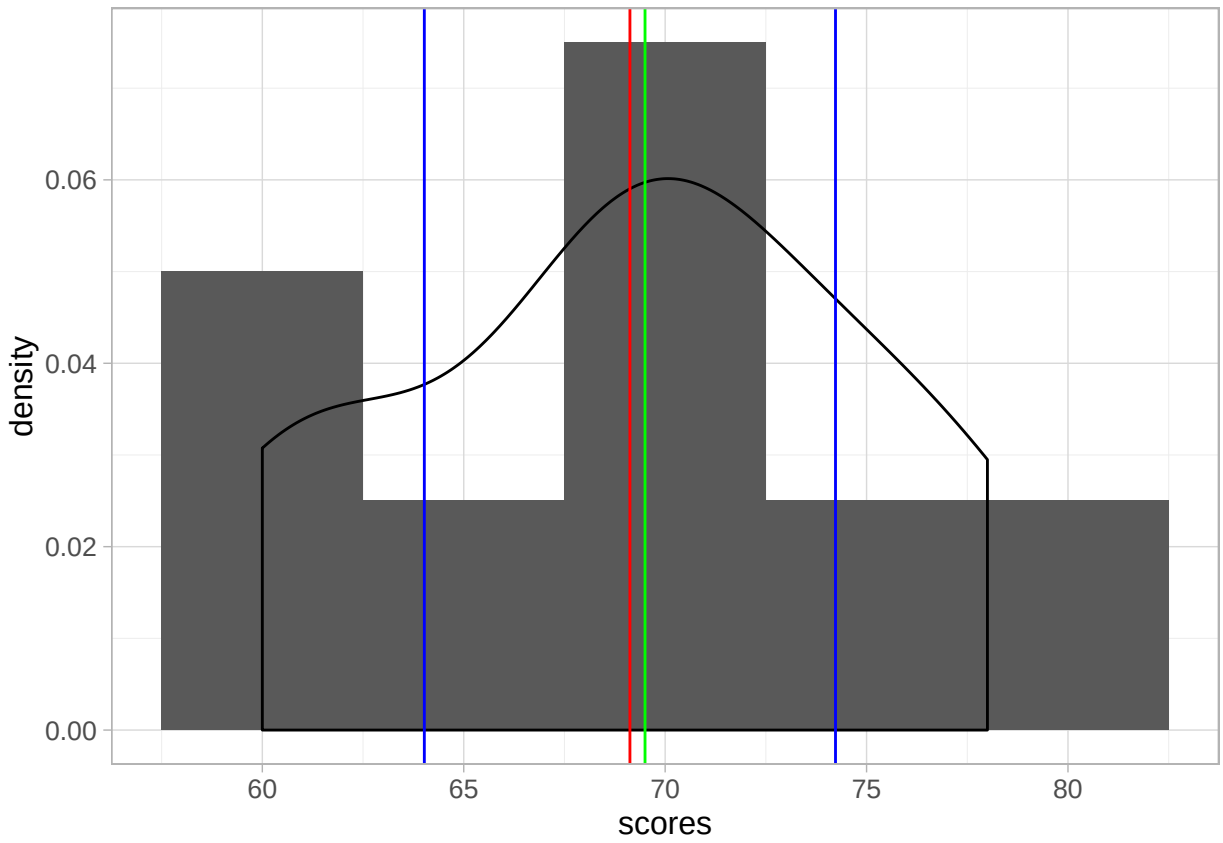
Test to see if the sample mean is significantly different from 65 at the .05 level. Report the t and p values.

```
    scores <- c(60,62,67,69,70,72,75,78)
    resA <- t.test(scores,paired = FALSE,var.equal = TRUE,mu=65)
```

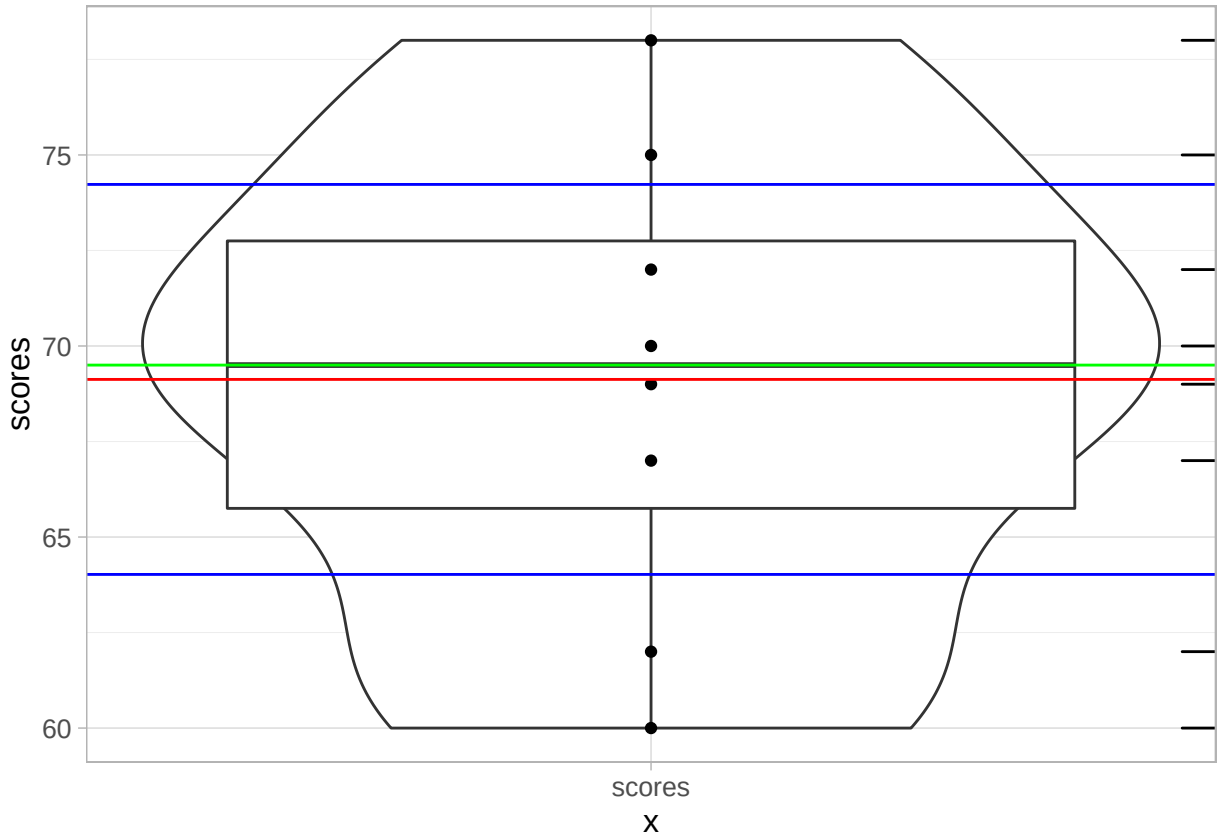
so this seems to e as if the mean is not 65 but the mean is still within the confidence intervals.

When plotting the distribution and data below the relationship between the mean, median and the confidence intervals can be seen.

```
library(ggplot2)
ggplot(data.frame(scores),aes(x=scores)) +geom_histogram(aes(y = ..density..),binwidth = 5) +geom_densi
    geom_vline(xintercept =  median(scores),color="green")+
    geom_vline(xintercept =  resA$estimate,color="red")+
    geom_vline(xintercept =  resA$conf.int,color="blue")
```




```
ggplot(data.frame(scores),aes(x="scores",y=scores))  +geom_violin() +geom_boxplot()+geom_point()+theme_
    geom_hline(yintercept =  median(scores),color="green",show.legend = TRUE)+
    geom_hline(yintercept =  resA$estimate,color="red")+
    geom_hline(yintercept =  resA$conf.int,color="blue")
```



b

The researcher realizes that she accidentally recorded the score that should have been 76 as 67. Are these corrected scores significantly different from 65 at the .05 level?

replacing the number 67 tith 76 and doing this again

```
    scores[3] <- 76
    t.test(scores,paired = FALSE,var.equal = FALSE,mu=65)
```

```
##
##  One Sample t-test
##
## data:  scores
## t = 2.2932, df = 7, p-value = 0.05554
## alternative hypothesis: true mean is not equal to 65
## 95 percent confidence interval:
##  64.83657 75.66343
## sample estimates:
## mean of x
##     70.25
```

Question 2

A (hypothetical) experiment is conducted on the effect of alcohol on perceptual motor ability. Ten subjects are each tested twice, once after having two drinks and once after having two glasses of water. The two tests were on two different days to give the alcohol a chance to wear off. Half of the subjects were given alcohol first and half were given water first. The scores of the 10 subjects are shown below. The first number for each subject is their performance in the "water" condition. Higher scores reflect better performance. Test to see if alcohol had a significant effect. Report the t and p values.

For me this is a paired t-test example

```
library(reshape2)
 water <- c(16,15,11,20,19,14,13,15,14,16)
alcohol <- c(13,13,10,18,17,11,10,15,11,16)
ques2Data <- data.frame(water,alcohol)
ques2Data <- melt(ques2Data)
```

```
## No id variables; using all as measure variables
```

```
res2A <- t.test(water,alcohol,paired = TRUE)
```

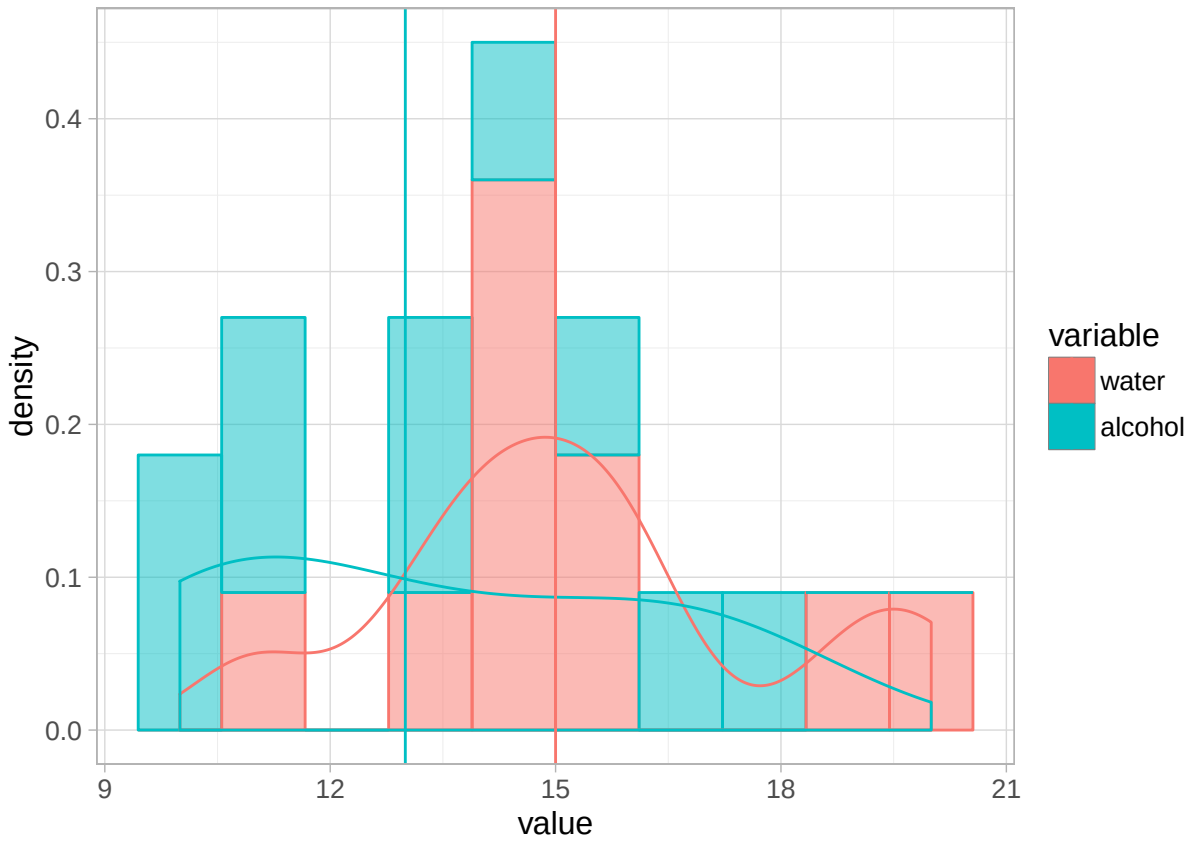
Looking at the results from the t-test it can be seen that there is a difference if 1.9 between the groups

I will use the same plots as before

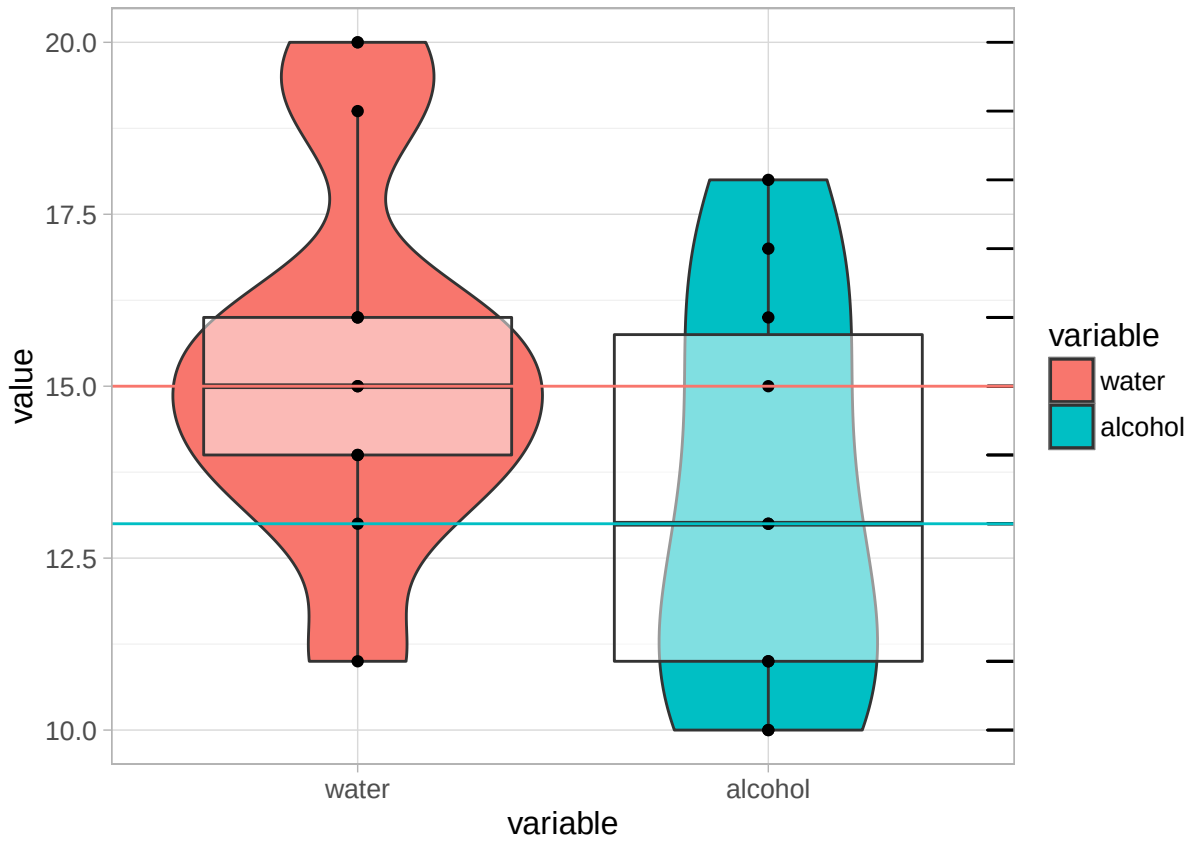
```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
ques2DataMean <- ques2Data %>% group_by(variable) %>% summarize_all(median)
ggplot(ques2Data,aes(x=value,color=variable)) +geom_histogram(aes(y = ..density..,fill=variable),bins =
    geom_vline(aes(xintercept=value,color=variable),data = ques2DataMean)
```




```
ggplot(ques2Data,aes(x=variable,y=value))+geom_violin(aes(fill=variable)) +geom_boxplot(alpha=0.5)+geom_
    geom_hline(aes(yintercept=value,color=variable),data = ques2DataMean,show.legend = TRUE)
```



Question 3

The scores on a (hypothetical) vocabulary test of a group of 20 year olds and a group of 60 year olds are shown below. Test the mean difference for significance using the .05 level.

I will again use the t-test but this time it is not paired

```
Y20 <- c(27,26,21,24,15,18,17,12,13)
Y60 <- c(26,29,29,29,27,16,20,27,NA)
ques3Data <- data.frame(Y20,Y60)
ques3Data <- melt(ques3Data)
```

```
## No id variables; using all as measure variables
```

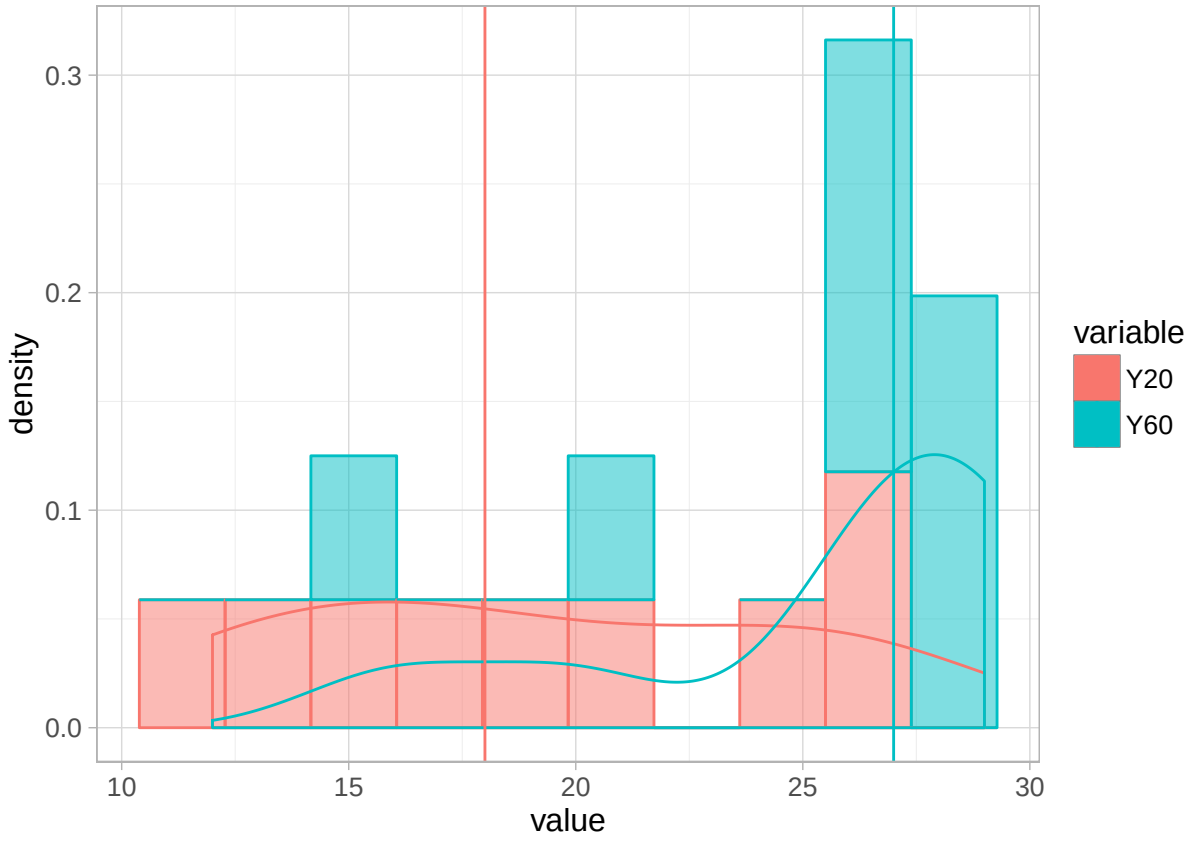
```
res3A <- t.test(Y20,Y60,paired = FALSE,var.equal = FALSE)
```

Again looking at the distribution of the data

```
ques3DataMean <- ques3Data %>% group_by(variable) %>% summarize_all(funs(median(., na.rm = TRUE)))
ggplot(ques3Data,aes(x=value,color=variable)) +geom_histogram(aes(y = ..density..,fill=variable),bins =
    geom_vline(aes(xintercept=value,color=variable),data = ques3DataMean)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

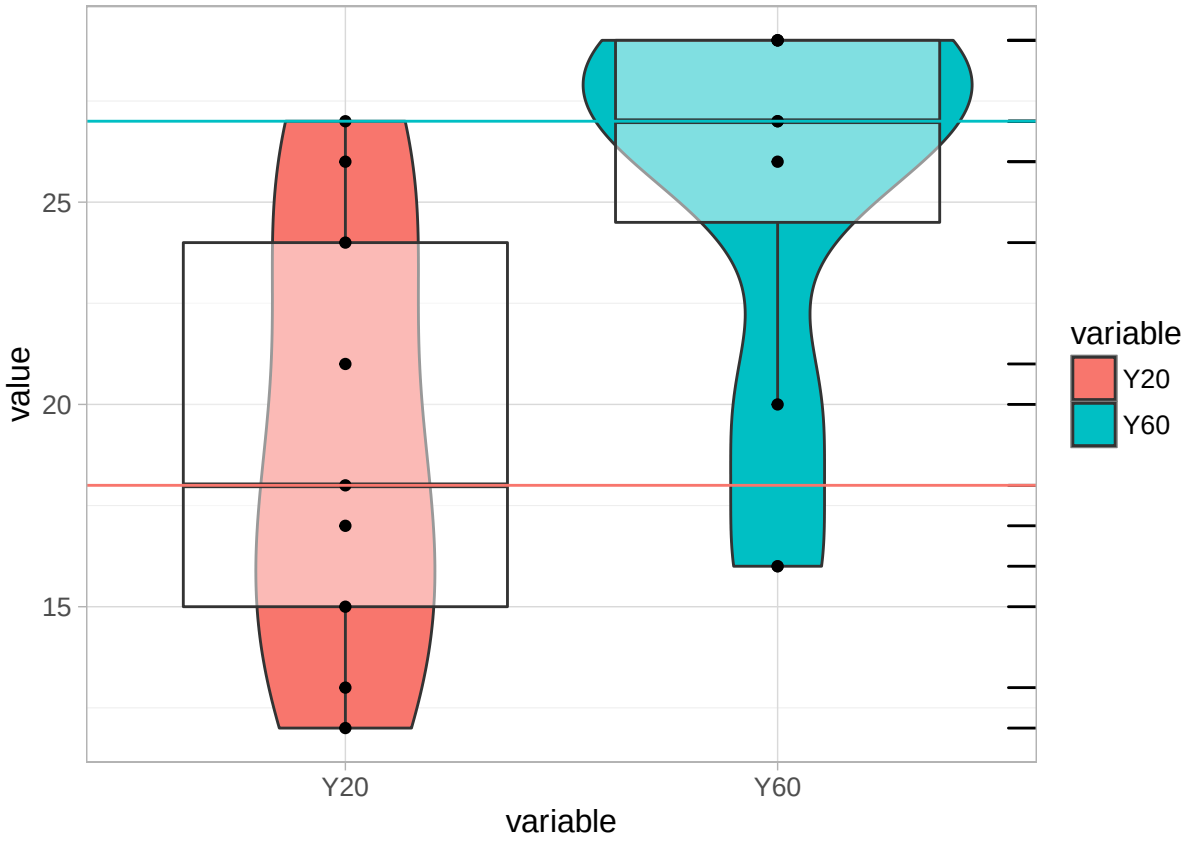



```
ggplot(ques3Data,aes(x=variable,y=value))+geom_violin(aes(fill=variable)) +geom_boxplot(alpha=0.5)+geom_
    geom_hline(aes(yintercept=value,color=variable),data = ques3DataMean,show.legend = TRUE)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_ydensity).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



List the assumptions made in computing your answer.

- The variance is different in the two populations
- the samples were drawn from different populations, (not paired)