PS5: Compare Mean Values

Pooja Sadarangani

2022-11-12

Collaborato: Pranali Oza

1 Descriptive Analysis (20pt)

1. (3pt) Load the data. Do the basic descriptive work on it: what is the number of observations? Are there any missings?

```
## The number of observations are: 1078
```

```
## [1] "There are no missing values in the dataset"
```

2. (7pt) Describe fathers and sons: compute the mean, median, standard deviation, and range of their heights. According to these figures, who are taller: fathers or sons?

```
## [1] "Father's Description"
```

```
## The mean of the fathers' height is: 171.9252
```

```
## The median of the fathers' height is: 172.1
```

```
## The standard deviation of the fathers' height is: 6.972346
```

```
## The range of the fathers' height is: 149.9 191.6
```

```
## [1] "Son's Description"
```

```
## The mean of the sons' height is: 174.4572
```

```
## The median of the sons' height is: 174.3
```

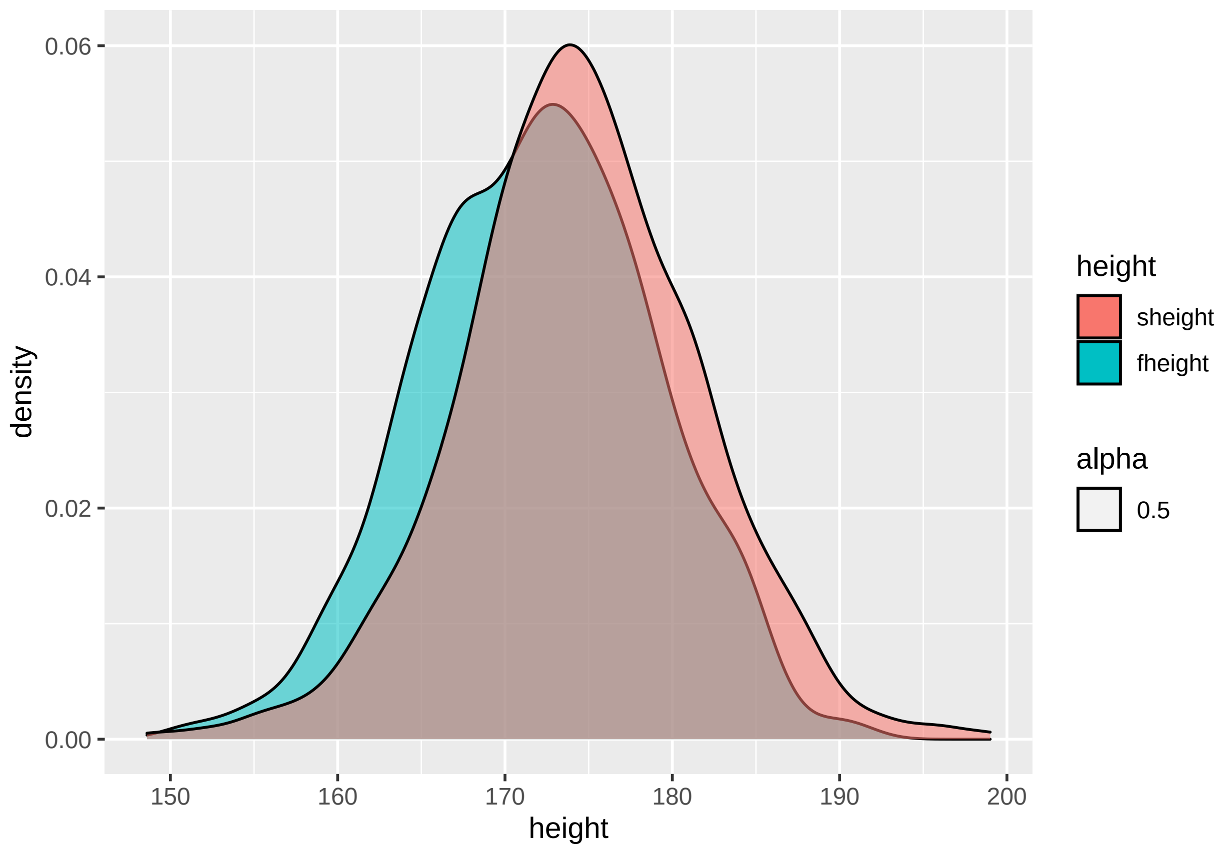
```
## The standard deviation of the sons' height is: 7.150713
```

```
## The range of the sonss' height is: 148.6 199
```

Ans. According to the above figures, sons are taller than fathers

3. (7pt) Lets add a graphical comparison. Create density plots of both heights on the same figure. Comment the density plots. Which distribution do these resemble? Do they agree with the conclusion above that sons are taller?

```
## Warning: Ignoring unknown aesthetics: stat
## Ignoring unknown aesthetics: stat
```



Ans. These distributions resemble normal distribution.But the mean of the sonss' heights is greater than that of the fathers' heights. Thus, they agree with the above conclusion that sons are taller than fathers.

4. (3pt) Finally, for the further reference, compute how much taller are sons on average.

```
## Sons are taller than fathers by 2.532004 on an average
```

2 Simulations

1. (5pt) Compute the overall mean and standard deviation of pooled fathers' and sons' heights. (That is, combine all heights, and compute just one mean and one standard deviation for this combined 2156 heights.)

```
## The mean of the pooled fathers' and sons' height is:  173.1912
```

```
## The standard deviation of the pooled fathers' and sons' height is:  7.173111
```

2. (5pt) Now create two sets of random normals, both with the same mean and the same standard deviation that you just computed above. Call one of these "fake fathers" and the others "fake sons". What is the fake father-fake son mean difference? Compare the result with that you found in the previous problem.

```
## The fake father-fake son mean difference is: -0.4149475
```

The result is much lesser than that found previously which was 2.532. It is also much closer to zero.

3. (8pt) Now repeat the previous question a large number of times R (1000 or more). Each time store the difference, so you end up with R different values for the difference.

```
## [1] 0.0001540755
```

4. (2pt) What is the mean of the differences? Explain why do you get what you get.

```
## The mean of the difference is:  0.0001540755
```

Ans. From question 2.2, we can see that the difference in the mean is very close to zero. We get the above result because the data for fake fathers and fake sons has been created normally using same mean and standard deviation.

5. (2pt) What is the standard deviation of the differences?

```
## The standard deviation of the differences is  0.3072647
```

6. (2pt) What is the largest difference you got (in absolute value)? How does it compare to the actual sons/fathers difference of 2.5?

```
## The largest difference that I got is:  1.027759
```

Ans. It is lesser than the actual sons' and fathers' height's mean difference

7. (8pt) find 95% confidence intervals for the differences you computed. Does the actual difference fall inside or outside of the CI?

```
##         0%       2.5%      97.5%
## -1.0277586 -0.5778669  0.6045579
```

Ans. The actual difference falls outside of the CI

8. (8pt) What is your conclusion? Can you confidently say that sons are taller than fathers? Why?

Ans. From the above result, we can reject the null hypothesis (H0) i.e son's and father's heights are the same, because the actual difference falls outside the CI. After comparing the means of the fathers' and sons' heights', we can conclude that sons are taller than the fathers.

3 t-test

1. (5pt) Compute the standard error for the difference of means.

```
## The standard error for the difference of means is:  0.3158264
```

Compute the t-value (OIS denotes it by T). Here we ask to compute it yourself, not use any pre-existing t-test functions! What is the t-value you find?

```
## The t-value is:  -8.017075
```

Look up the t-distribution table (there is one in OIS appendices, and a web search can find you a ton of those tables). What is the likelihood that such a t-value happens just by random chance?

The t-value is very large (around 8). This means that there is a large difference between the two samples. Since we have a large sample size, the likelihood of getting such a large t-value is very low.

4. (5pt) Finally, state clearly your conclusion. Is the actual difference you see compatible with H0 that fathers and sons are of similar height?

In conclusion, the actual difference that we see i.e 2.532 is not compatible with H0 that fathers and sons are of similar heights.

4 Canned t-test software

1. (6 pt) Use t.test function to compare fathers and sons.

```
## By using the t.test function, we get the t-value as:

##
##  Welch Two Sample t-test
##
## data:  fatherson$fheight and fatherson$sheight
## t = -8.3239, df = 2152.6, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.128532 -1.935475
## sample estimates:
## mean of x mean of y
##  171.9252  174.4572
```

2. (4 pt) Do all three methods agree whether sons are taller than fathers?

Ans. Yes, all three methods agree that the sons are taller than the fathers.