Homework 1

Rodrigo Petricioli

9/1/2021

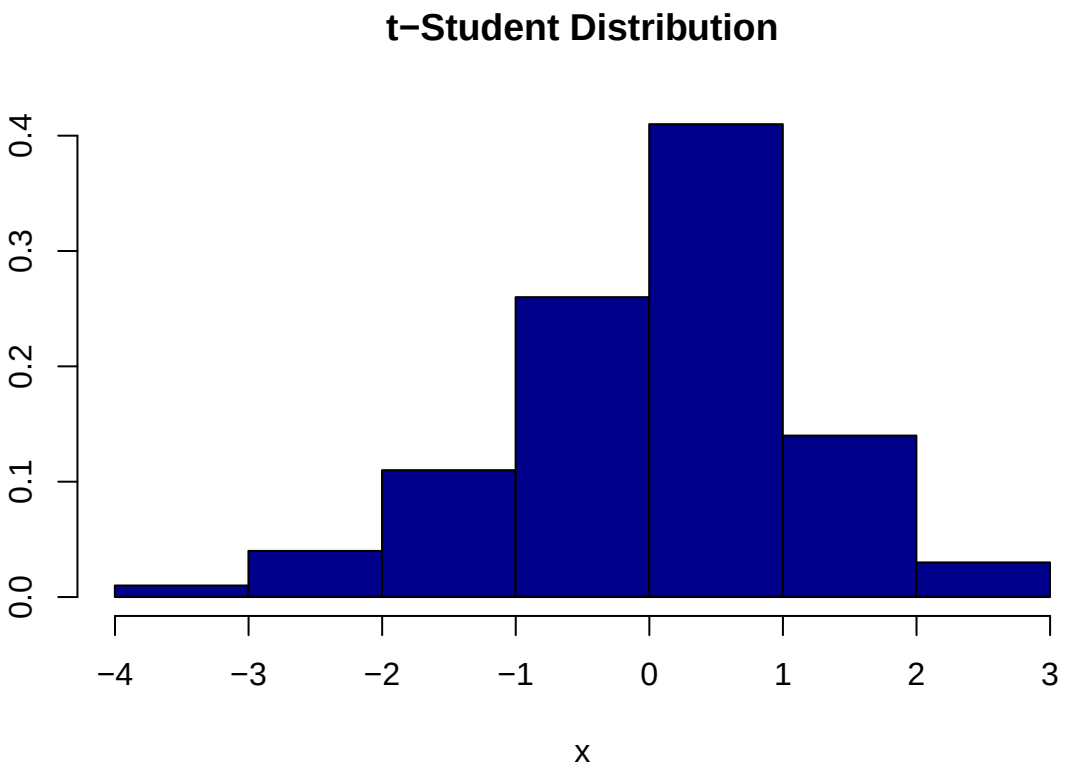
Questions: 1.1, 1.2, 1.4, 1.5, 1.7

Note: Unless otherwise specified, all the numbers are referring to the corresponding exercises in our textbook: Statistical Computing with R (2nd Edition) by Maria L. Rizzo

1.1

Generate a random sample $x_1, ..., x_100$ of data from the $t4(df = 4)$ distribution using the rt function. Use the $MASS :: truehist$ function to display a probability histogram of the sample.

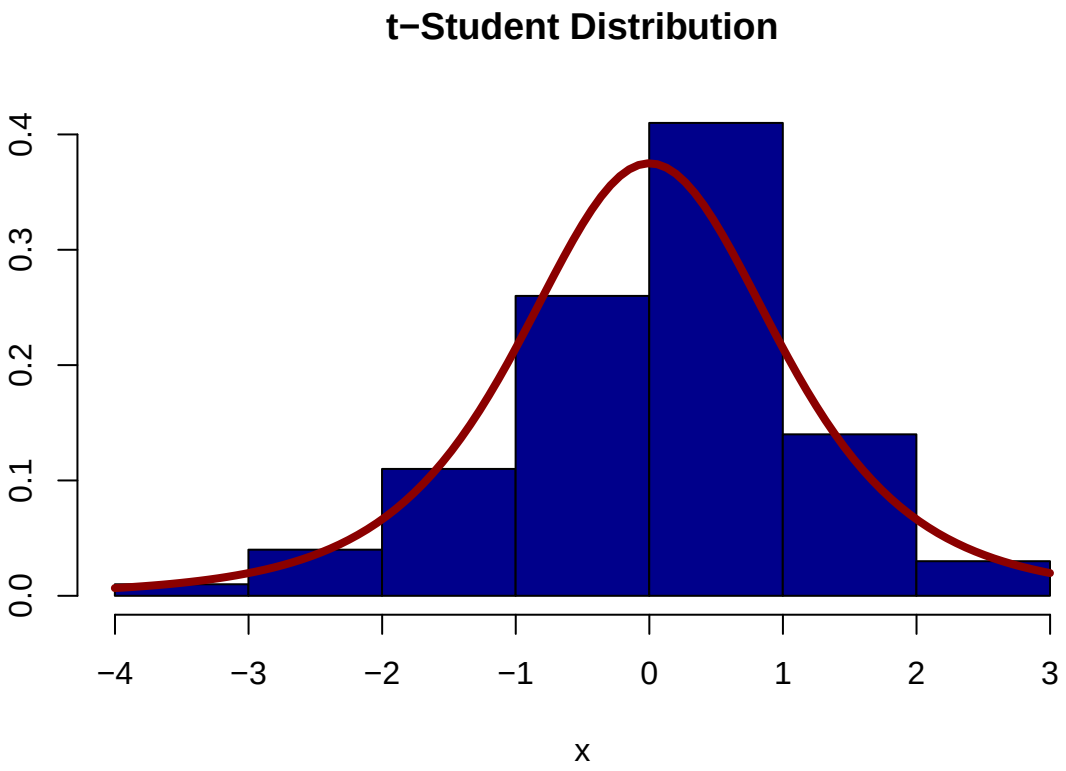
```
set.seed(1234)
x <- rt(100,4)
MASS::truehist(x, col = "blue4", main = "t-Student Distribution")
```



1.2

Add the $t4densitycurve(dt)$ to your histogram in Exercise 1.1 using the curve function with $add = TRUE$.

```
MASS::truehist(x, col = "blue4", main = "t-Student Distribution")
curve(dt(x, 4), add = T, col = "red4", lwd=4)
```



1.4

a. Write an R function f in R to implement the function

$f(x) = \frac{x-a}{b}$

that will transform an input vector x and return the result. The function should take three input arguments: x, a, b.

b. To transform x to the interval $[0, 1]$ we substract the minimum value and divide by the range:

$y < -f(x, a = min(x), b = max(x) - min(x))$

Generate a random sample of $Normal(\mu = 2, \sigma = 2)$ data using $rnorm$ and use your function f to transform this sample to the interval $[0, 1]$. Print a summary of both the sample x and the transformed sample y to check the result.

```
f <- function(x, a, b){
  return (x-a)/b
}
```

```
x_norm <- rnorm(100,2,2)
y <- f(x_norm, min(x_norm), max(x_norm)-min(x_norm))

summary(x_norm)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.6890  0.7048  1.9060  1.9980  3.3351  6.0988
```

```
summary(y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.394   4.595   4.687   6.024   8.788
```

1.5

Refer to Exercise 1.4. Suppose that we want to transform the x sample so that it has mean zero and standard deviation one (studentize the sample). That is, we want:

$z_i = \frac{x_i - \bar{x}}{s}, i = 1, ..., n$

where s is the standard deviation of the sample. Using your function f this is:

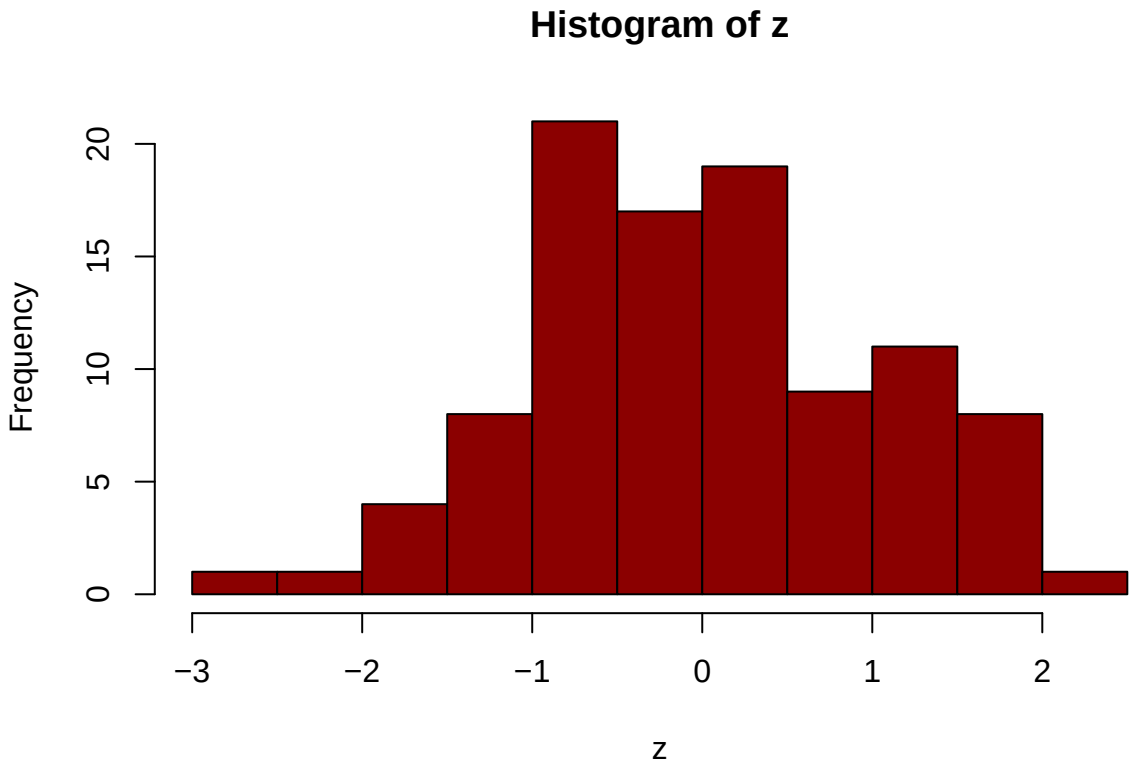
$z_i = f(x, a = mean(x), b = sd(x))$

Display a summary and histogram of the studentized sample z. It should be centered exactly ar zero. Use $sd(x)$ to check that the studentized sample has standard deviation exactly 1.0.

```
s <- f(x_norm, mean(x_norm), sd(x_norm))
z <- (x_norm - mean(x_norm))/sd(x_norm)
summary(z)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.52041 -0.69540 -0.04948  0.00000  0.71905  2.20519
```

```
hist(z, col = "red4")
```




```
sd(z)
```

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

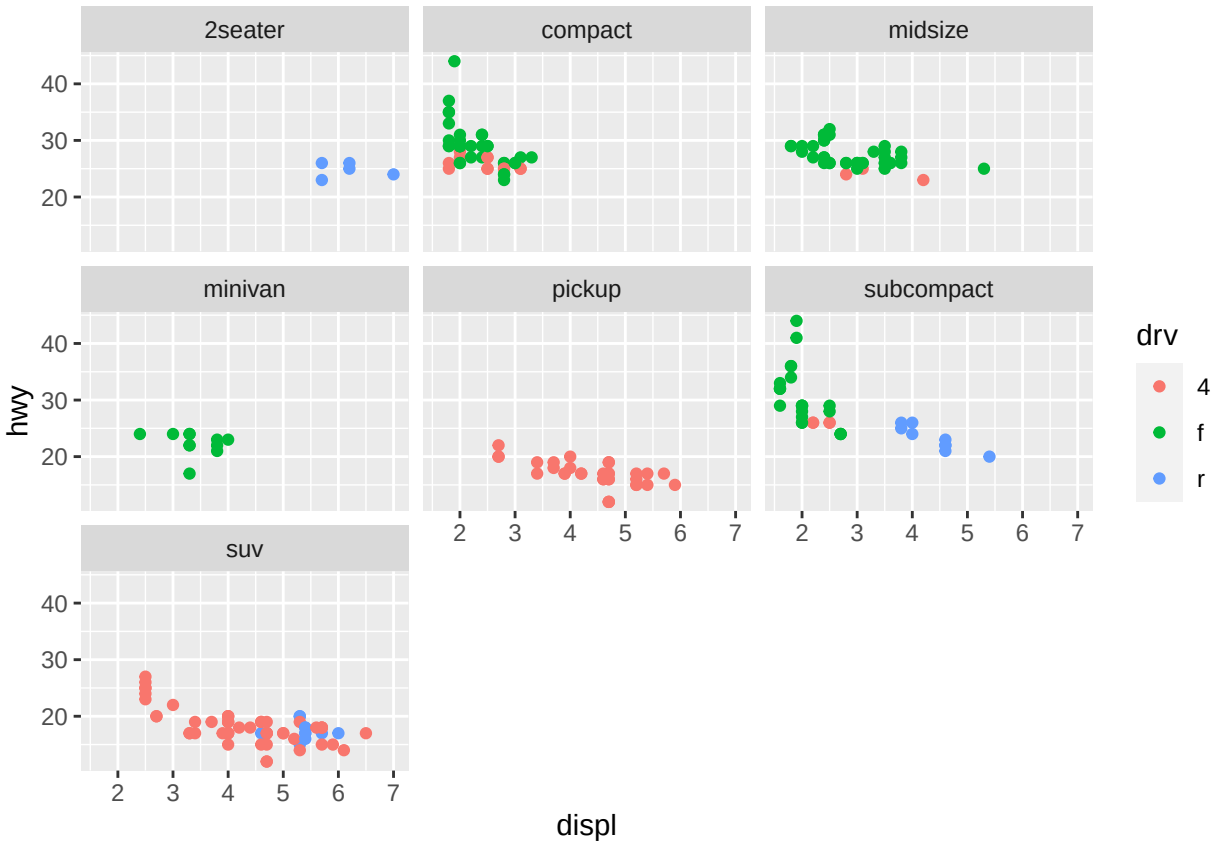
1.7

(ggplot) Refer to Example 1.14 where we displayed an array of scatterplots using *ggplot* with $facet_wrap$. One of the variables in the *mpg* data is *drv*, a character vector indicating whether the vehicle is front-wheel drive, rear-wheel drive, or four-wheel drive. Add $color = drv$ in *aes*:

$aes(displ, hwy, color = drv)$

and display the revised plot. Your scatterplots should now have the three levels of *drv* coded by color and the plot should have automatically generated a legend for *drv* color.

```
ggplot(mpg, aes(displ, hwy, color = drv)) + geom_point() + facet_wrap(~ class)
```


2seater
compact
midsize
minivan
pickup
subcompact
suv
hwy
displ
drv
4
f
r