

# СТАТИСТИКА И ЕМПИРИЧНИ МЕТОДИ

## Домашно 1

Антоан Венциславов Стефанов

Ф.Н.: 61797

гр.5

Софтуерно инженерство

21.11.2016

*# • Прочетете данните и ги запишете в data frame в R;*

```
input = read.csv("C:/Users/Toni/Downloads/pokemon.csv", header=TRUE)
```

*# • Генерирайте си подизвадка от 600 наблюдения. За целта нека f\_nr е  
# вашият факултетен номер. Задайте състояние на генератора на слу-  
# чайни числа в R чрез set.seed(f\_nr). С помощта на подходяща фун-  
# кция генерирайте извадка без връщане на числата от 1 до 705 като  
# не забравяте да я запишете във вектор. Използвайте вектора, за да  
# запишете само редовете със съответните индекси в нов дейтафрейм и  
# работете с него оттук нататък;*

```
set.seed(61797)
```

```
randomIndexes = c(sample({1:705}, 600, replace = FALSE))
```

```
pokemonData = input[randomIndexes, ]
```

*# • Изкарайте на екрана първите няколко (5-6) наблюдения;*

```
head(pokemonData)
```

	Number	Name	Type1	Type2	Attack	Defense	Height	weight
681	681	Clawitzer	Water		73	88	1.30	35.3
507	507	Pansage	Grass		53	48	0.61	10.5
409	409	Shieldon	Rock	Steel	42	118	0.51	57.0
143	143	Snorlax	Normal		110	65	2.11	460.0
636	636	Reshiram	Dragon	Fire	120	100	3.20	330.0
592	592	Ferroseed	Grass	Steel	50	91	0.61	18.8

*# • Какъв вид данни (качествени/количествени, непрекъснати/дискретни)  
# са записани във всяка от променливите?*

*# Name - качествени*

*# Type1 - качествени*

*# Type2 - качествени*

*# Attack - количествени (непрекъснати)*

*# Defence - количествени (непрекъснати)*

*# Height - количествени (непрекъснати)*

*# Weight - количествени (непрекъснати)*

# • Изведете дескриптивни статистики за всяка една от променливите;

**summary(pokemonData)**

Number	Name	Type1	Type2	Attack
Defense	Height			
Min. : 1.0	Abomasnow : 1	Water : 93	:310	Min. : 10.00
Min. : 10.00	Min. :0.100			
1st Qu.:171.8	Abso1 : 1	Normal : 78	Flying : 70	1st Qu.: 53.75
1st Qu.: 50.00	1st Qu.:0.610			
Median :342.5	Accelgor : 1	Bug : 54	Poison : 28	Median : 75.00
Median : 67.00	Median :0.990			
Mean :347.1	Aerodactyl: 1	Grass : 54	Ground : 26	Mean : 75.08
Mean : 71.08	Mean :1.131			
3rd Qu.:523.2	Aggron : 1	Fire : 37	Psychic: 22	3rd Qu.: 95.00
3rd Qu.: 85.25	3rd Qu.:1.500			
Max. :705.0	Aipom : 1	Psychic: 36	Steel : 16	Max. :165.00
Max. :230.00	Max. :9.190			
	(Other) :594	(Other):248	(Other):128	

weight
Min. : 0.10
1st Qu.: 9.00
Median : 28.20
Mean : 56.44
3rd Qu.: 61.62
Max. :950.00

# • Изведете редовете на най-високия и на най-лекия покемон;

**attach(pokemonData)**

**print(pokemonData[Height == max(Height),])**

Number	Name	Type1	Type2	Attack	Defense	Height	weight
208	208 Steelix	Steel	Ground	85	200	9.19	400

**print(pokemonData[Weight == min(Weight),])**

Number	Name	Type1	Type2	Attack	Defense	Height	weight
659	659 FlabÃ©bÃ©	Fairy		38	39	0.1	0.1
93	93 Haunter	Ghost	Poison	50	45	1.6	0.1

# • Изведете редовете на покемоните с общ брой точки за атака и защита над 220;

**print(pokemonData[Attack + Defense > 220,])**

	Number	Name	Type1	Type2	Attack	Defense	Height	weight
	376	Metagross	Steel	Psychic	135	130	1.60	550.0
	522	Gigalith	Rock		135	130	1.70	260.0
	348	Armaldo	Rock	Bug	125	100	1.50	68.2
	669	Doublade	Steel	Ghost	110	150	0.84	4.5
	208	Steelix	Steel	Ground	85	200	9.19	400.0

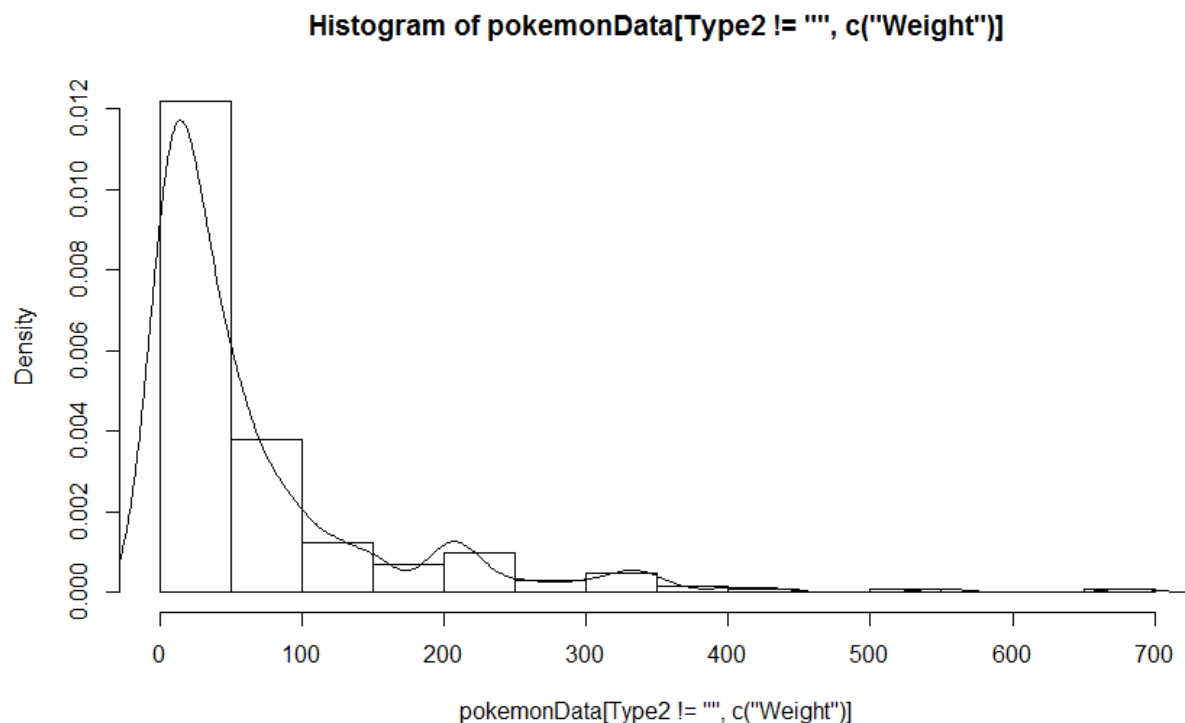
.....

# • Колко на брой покемони имат първичен или вторичен тип "Dragon"или  
# "Flying"и са високи над един метър?

```
nrow(pokemonData[(Type1 == "Dragon" | Type2 == "Dragon" | Type1 == "Flying" | Type2
== "Flying") & Height > 1,])
57
```

# • Направете хистограма на теглото само на покемоните с вторичен тип  
# и нанесете графика на плътността върху нея. Симетрично ли са раз-  
# положени данните?

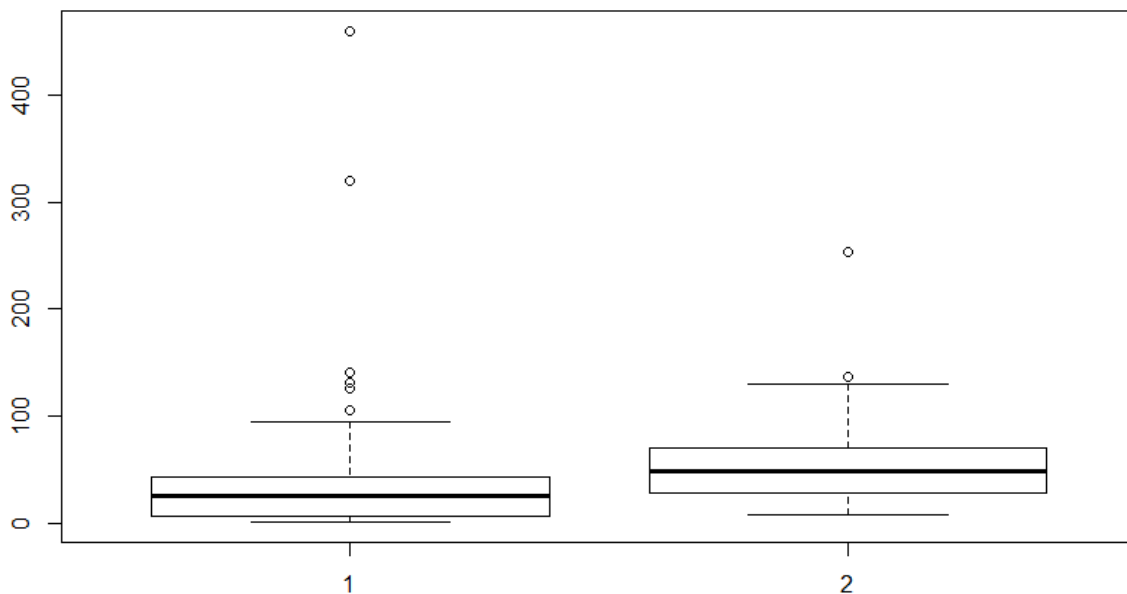
```
hist(pokemonData[Type2 != "", c("Weight")], probability = TRUE)
lines(density(pokemonData[Type2 != "", c("Weight")]))
```



# данните не са разположени симетрично

# • За покемоните с първичен тип "Normal"или "Fighting"изследвайте  
# съвместно променливите Type1 и Height с подходящ графичен метод.  
# Забелязвате ли outlier-и? Сравнете извадковите средни и медианите в  
# двете групи и направете извод;

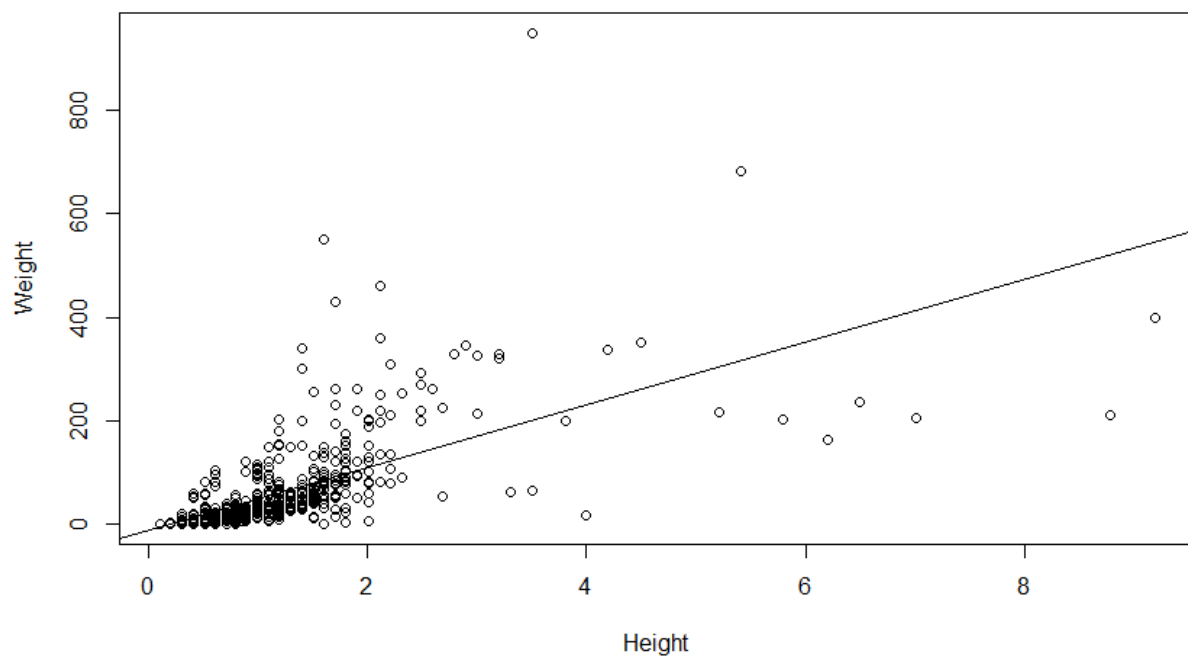
```
boxplot(pokemonData[Type1 == "Normal",c("Weight")], pokemonData[Type1 ==  
"Fighting",c("Weight")])
```



# при покемоните с първичен тип Normal има 7 outlier-и, докато при тези с тип Fighting- само 2, всички от които над съответните медиани.  
# Покемоните с тип Fighting средно тежат повече, от тези с тип Normal

# • Изследвайте съвместно променливите *Height* и *Weight* с подходящ  
# графичен метод. Бихте ли казали, че съществува линейна връзка меж-  
# ду тях? Намерете корелацията между величините и коментирайте  
# стойността ѝ. Начертайте регресионна права (линейната функция, ко-  
# ято най-добре приближава функционалната зависимост). Ако е наб-  
# людаван нов вид покемон с височина 2.1 метра, какво се очаква да е  
# теглото му на базата на линейния модел?

```
plot(Height, Weight)  
abline(lm(Weight ~ Height))
```



# съществува средно силна линейна връзка между височината и теглото на  
покемоните

```
cor(Height, Weight)  
0.6541075
```

```
nums = lm(Weight ~ Height)$coefficients
print( nums[2]*2.1 + nums[1])
Height
115.349
```