Homework 1
Advanced Statistics
Zhetessov Nur

Chapter 1 (Probability)
p. 13-17, ex. 5, 13, 19, 22.

5. Suppose we toss a fair coin until we get exactly 2 Heads.
   a) Describe Ω
   b) $P(X=k)$ - ?

a) $\Omega = \{\omega = (\omega_1, \omega_2, \dots \omega_n) : \omega_i \in \{H, T\},$
$|\omega_i = H| = 2\}$

b) Probability of getting the first head along with Tails in k tosses is equal to $\left(\frac{1}{2}\right)^{k-1}$. But the first head can be in different positions (1st, 2nd, 3rd...). There are $(k-1)$ such positions, so

$(k-1)\cdot\left(\frac{1}{2}\right)^{k-1}$

then on the k^{th} toss we get the second head, so probability of 2 heads in k tosses is:

$$(k-1)\cdot\left(\frac{1}{2}\right)^{k-1}\cdot\left(\frac{1}{2}\right)=\frac{(k-1)\cdot 1^{k-1}}{2^{k-1}}\cdot\frac{1}{2}=$$

$$=\frac{k-1}{2^k}$$

13. Suppose that a fair coin is tossed, until a head and a tail appear ≥ 1

a) Describe Ω

b) $P(X=3)-?$

a) $$\Omega=\{\omega=(\omega_1,\omega_2,\dots,\omega_{n-1},\omega_n):\omega_i\in\{H,T\},$$
$$\omega_1=\omega_2=\omega_3=\dots=\omega_{n-1},\ \omega_{n-1}\neq\omega_n\}$$

b) $$P(X=3)=P(\{HHT\})+P(\{TTH\})=$$
$$=\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)+\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)=\frac{1}{8}+\frac{1}{8}=\frac{2}{8}=\frac{1}{4}$$

19.

Mac = 30% $\Rightarrow P(M) = 0,3$

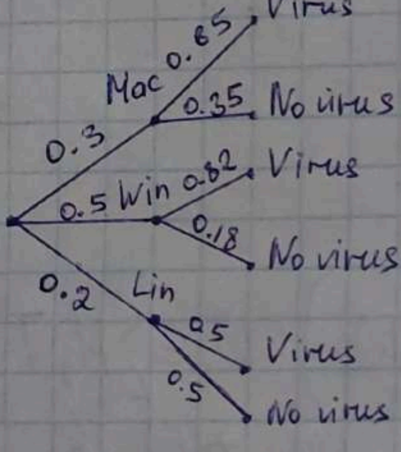
Windows = 50% $\Rightarrow P(W) = 0,5$

Linux = 20% $\Rightarrow P(L) = 0,2$

Virus-Mac = 65% $\Rightarrow P(V\text{-}M) = 0,65$

Virus-Windows = 82% $\Rightarrow P(V\text{-}W) = 0,82$

Virus-Linux = 50% $\Rightarrow P(V\text{-}L) = 0,5$



$P(W|V) - ?$

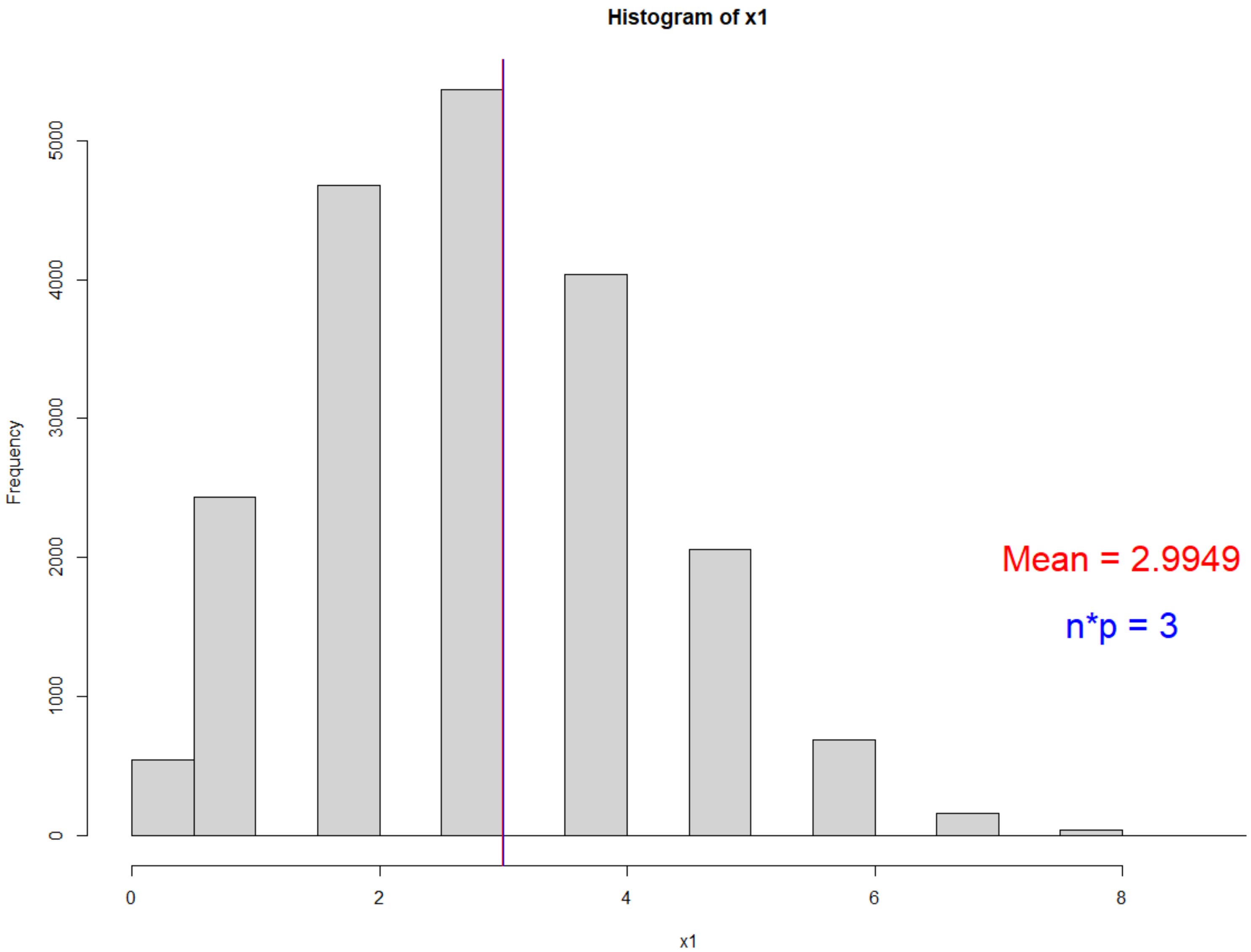
$$P(W|V) = \frac{P(W \cup V)}{P(V)} = \frac{P(V|W)\cdot P(W)}{P(V|M)\cdot P(M) + P(V|L)\cdot P(L) + P(V|W)\cdot P(W)} = \frac{0,5\cdot 0,82}{0,3\cdot 0,65 + 0,5\cdot 0,82 + 0,2\cdot 0,5} =$$

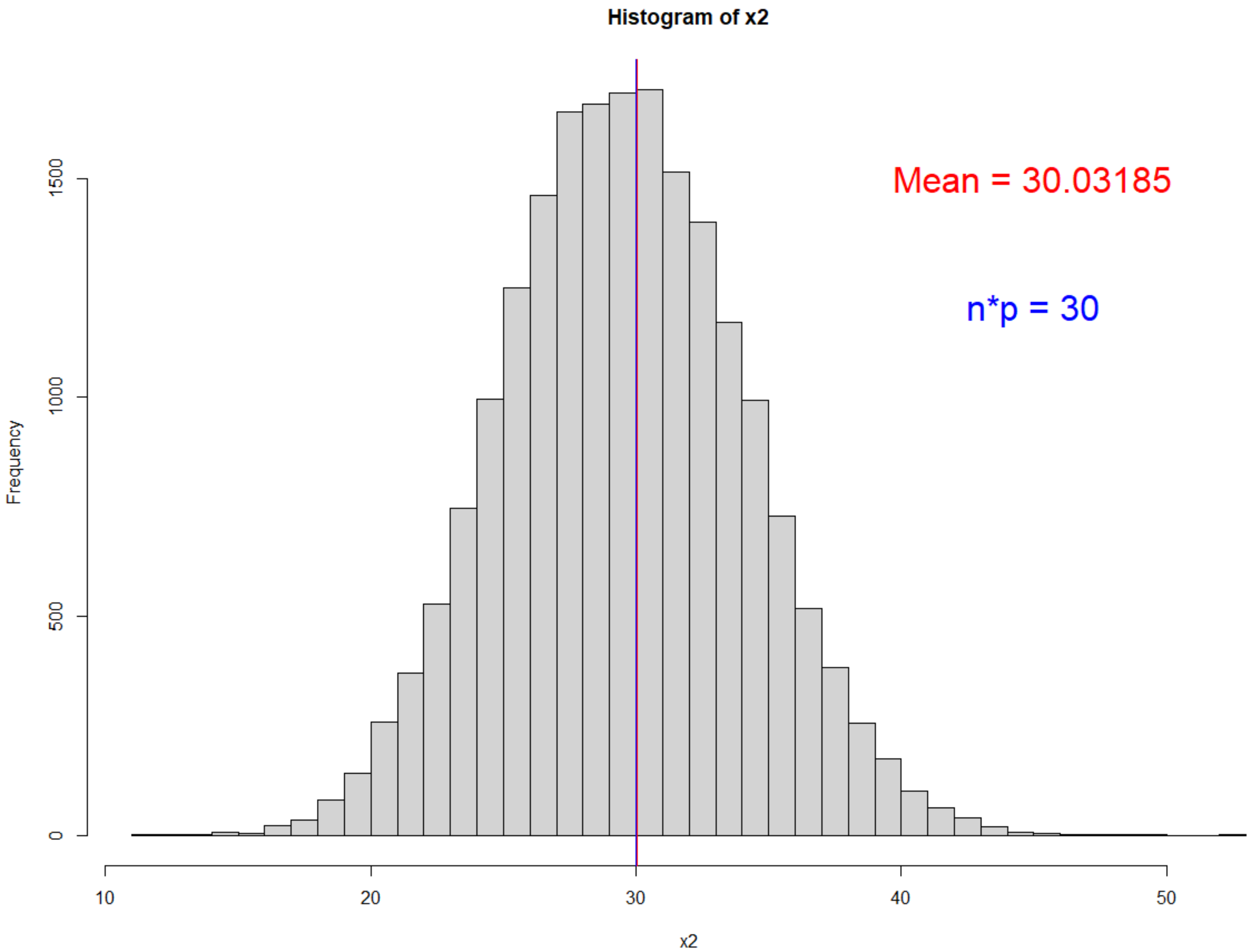
$= 0,58156 \approx 0,5816$

Console Terminal Jobs

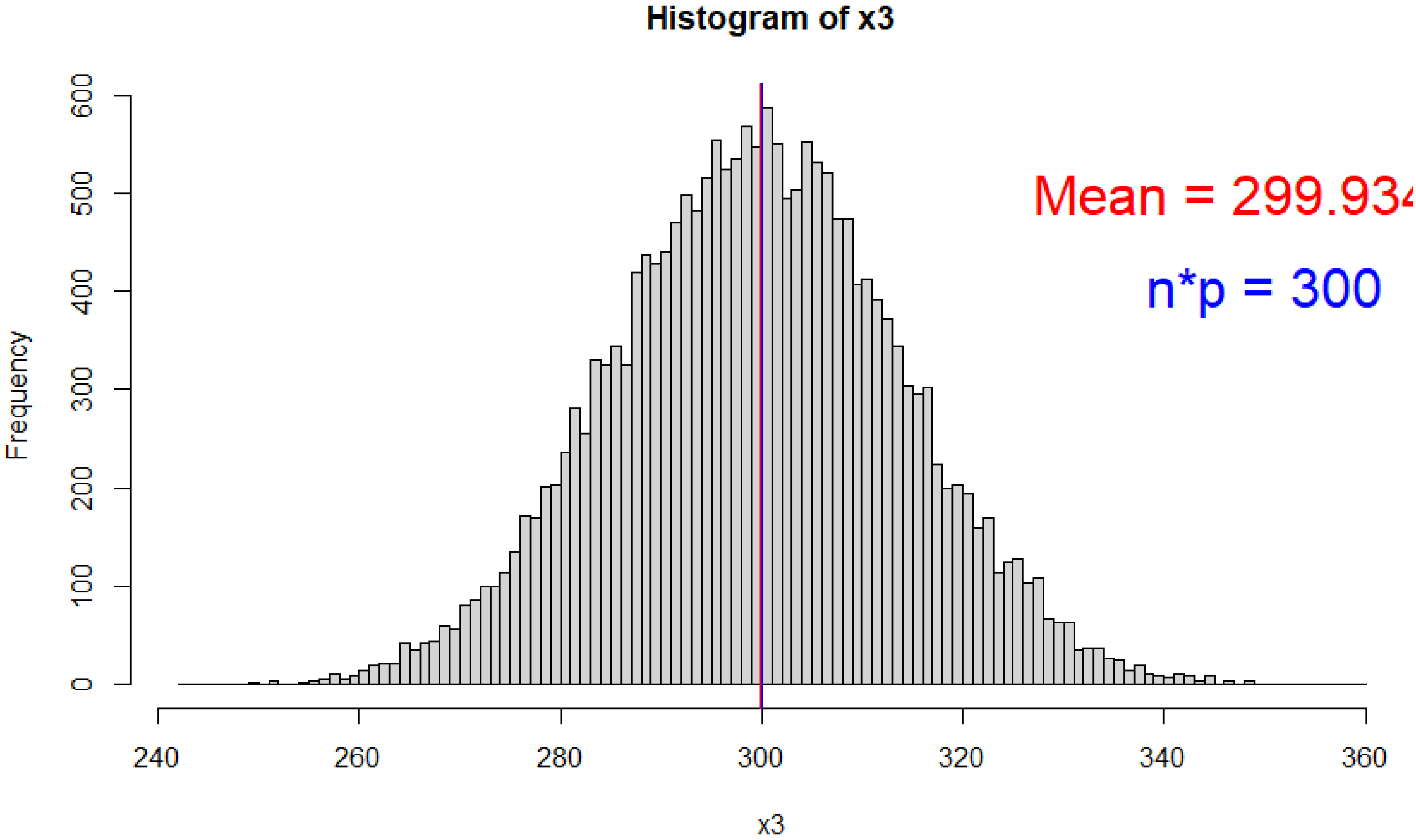
R 4.1.2 · C:/Users/Nur/Desktop/дз/Стат/HW1/

```
> p <- 0.3
> x1 <- rbinom(20000, 10, p)
> x2 <- rbinom(20000, 100, p)
> x3 <- rbinom(20000, 1000, p)
> print(paste("X1 mean:",mean(x1)))
[1] "X1 mean: 2.9949"
> print(paste("X1 np:",p*10))
[1] "X1 np: 3"
> print(paste("X2 mean:",mean(x2)))
[1] "X2 mean: 30.03185"
> print(paste("X2 np:",p*100))
[1] "X2 np: 30"
> print(paste("X3 mean:",mean(x3)))
[1] "X3 mean: 299.93465"
> print(paste("X3 np:",p*1000))
[1] "X3 np: 300"
>
```


Histogram of x1
Frequency
0
1000
2000
3000
4000
5000
0
2
4
6
8
x1
Mean = 2.9949
n*p = 3



Histogram of x2
Mean = 30.03185
n*p = 30
Frequency
0
500
1000
1500
10
20
30
40
50
x2


Histogram of x3
Mean = 299.934
n*p = 300
Frequency
0
100
200
300
400
500
600
240
260
280
300
320
340
360
x3

22.

Computer Experiment.

$p = 0,3$
$n_1 = 10$
$n_2 = 100$
$n_3 = 1000$

Theoretical mean for $n_1 = 10$:

$\mu_1 = n_1 \cdot p = 10 \cdot 0.3 = 3$

Theoretical mean for $n_2 = 100$:

$\mu_2 = n_2 \cdot p = 100 \cdot 0.3 = 30$

Theoretical mean for $n_3 = 1000$:

$\mu_3 = n_3 \cdot p = 1000 \cdot 0.3 = 300$

Let's find out experimental means for n's on computer

Experimental means are:

$\overline{X}_1 = 2.995$ for $n_1 = 10$

$\overline{X}_2 = 30.032$ for $n_2 = 100$

$\bar{X}_3 = 299.935$ for $n_3 = 1000$

Conclusion:

We see that means from computer experiment closely approximates the theoretical means. The distribution of # Heads closely to normal.

Chapter 2

Random variables

p. 43-46, ex. 2, 11a, 13b, 18

2.

$P(X=2) = P(X=3) = \frac{1}{10}$

$P(X=5) = \frac{8}{10}$

1) Plot the CDF:

2) $P(2 < X \leq 4.8)$ - ?

$P(2 \leq X \leq 4.8)$ - ?

CDF is:

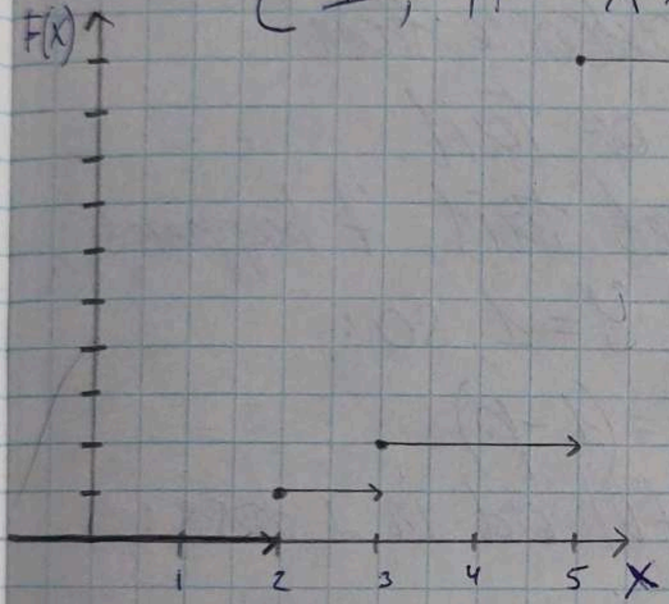
$F(X) = 0$, if $X < 2$

$F(X) = \frac{1}{10}$, if $2 \leq X < 3$

$P(X) = \frac{2}{10}$, if $3 \le X < 5$

$P(X) = \frac{10}{10} = 1$, if $X \ge 5$

$$F(X) = \begin{cases} 0, & \text{if } X < 2 \\ 0{,}1, & \text{if } 2 \le X < 3 \\ 0{,}2, & \text{if } 3 \le X < 5 \\ 1, & \text{if } X \ge 5 \end{cases}$$



2) $P(2 < X \le 4.8) = F(4.8) - F(2) =$

$= 0{,}2 - 0{,}1 = 0{,}1$

$P(2 \le X \le 4.8) = P(X=2) + P(X=3) =$

$= 0{,}1 + 0{,}1 = 0{,}2$

11a.
We toss a coin 1 time. Probability of Head is p. X – # heads. Y – # tails.

a) The events are depended if
$P(A \text{ and } B) \neq P(A) \cdot P(B)$

There is only one outcome from
$\Omega = \{H, T\}$ Head or Tail.
Let's assume we get Tail, it happens only if $X=0$ and $Y=1$, so:
$P(X=0 \text{ and } Y=1) = (1-p)$
Let's get the prob. of 0 head in 1 toss:
$P(X=0) = (1-p)$
Let's get the prob. of 1 tail in 1 toss:
$P(Y=1) = (1-p)$
By the formula and prob of 2 event:
$P(X=0) \cdot P(Y=1) = (1-p) \cdot (1-p) = (1-p)^2$
Therefore,
$P(X=0 \text{ and } Y=1) \neq P(X=0) \cdot P(Y=1)$

Untitled1*

Source on Save

```r
pnorm(7, mean=3, sd=4, log=FALSE) #a) P(X<7)

1 - pnorm(-2, mean=3, sd=4, log=FALSE) #b) P(>-2)

qnorm(0.95, mean=3, sd=4) #c) P(X>x)=0.05

pnorm(4, mean=3, sd=4, log=FALSE) - pnorm(0, mean=3, sd=4, log=FALSE) #d) P(0<=X<4)

qnorm(0.975, mean=3, sd=4) #e) P(|X|>|x|) = 0.05

```

Console Terminal Jobs

R 4.1.2 · ~/

```
> pnorm(7, mean=3, sd=4, log=FALSE) #a) P(X<7)
[1] 0.8413447
> 1 - pnorm(-2, mean=3, sd=4, log=FALSE) #b) P(>-2)
[1] 0.8943502
> qnorm(0.95, mean=3, sd=4) #c) P(X>x)=0.05
[1] 9.579415
> pnorm(4, mean=3, sd=4, log=FALSE) - pnorm(0, mean=3, sd=4, log=FALSE) #d) P(0<=X<4)
[1] 0.372079
> qnorm(0.975, mean=3, sd=4) #e) P(|X|>|x|) = 0.05
[1] 10.83986
>
>
```

the events are depended.

18.

$X \sim N(3, 16)$

a) $P(X<7) = P\left(\frac{X-3}{\sqrt{16}} < \frac{7-3}{\sqrt{16}}\right) =$

$= P(Z < 1) = \Phi(1) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx$

$\approx 0,8413$

b) $P(X > -2) = P\left(\frac{X-3}{\sqrt{16}} > \frac{-2-3}{\sqrt{16}}\right) =$

$= 1 - P\left(Z < -\frac{5}{4}\right) = 1 - \Phi\left(-\frac{5}{4}\right) \approx$

$\approx 1 - \int_{-\infty}^{-1,25} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx 1 - 0,1056 \approx$

$\approx 0,8944$

c) Find x: $P(X > x) = 0,05$

$P(X > x) = 1 - P(X < x) = 0,05$

$P(X < x) = 1 - 0,05$

$P\left(\frac{X-3}{\sqrt{16}} < \frac{x-3}{\sqrt{16}}\right) = 0,95$

According to table, z-score 1.645

corresponds to the area 0.95

So,

$$\frac{X-3}{\sqrt{16}} = 1,645$$

$$X-3 = 4 \cdot 1,645$$

$$X = 6,58+3$$

$$X = 9,58$$

d) $P(0 \le X < 4) = P\left(\frac{0-3}{\sqrt{16}} \le X < \frac{4-3}{\sqrt{16}}\right) =$

$= \Phi\left(\frac{1}{4}\right) - \Phi\left(-\frac{3}{4}\right) = 0,5987 - 0,2266 =$

$= 0,3721$

e) $P(|X| > |x|) = 0,05$

$P(|X| > |x|) = 1 - P(|X| < |x|) =$

$= 1 - P(-x < |X| < x) = P(-x > X) + P(x > X) =$

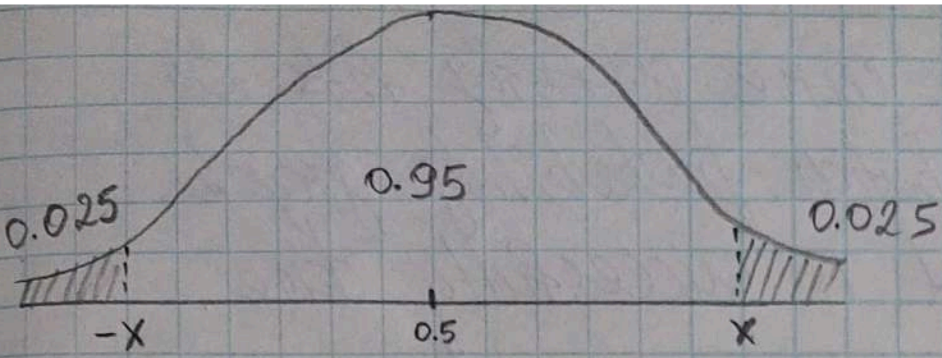
$= 0,05$

$P(-x > X) + P(x > X) = 0.05$

$2P(x > X) = 0.05$ - since normal is symmetric

$P(x > X) = 0.025$

$P(x < X) = 1 - P(x > X) = 1 - 0,025 = 0.975$



z-score for area = 0.975 is 1.96

$$1.96 = \frac{x-3}{\sqrt{16}}$$

$$x-3 = 1.96 \cdot 4$$

$$x-3 = 7.84$$

$$x = 7.84+3$$

$$x = 10.84$$

13 b)

Computer Experiment.

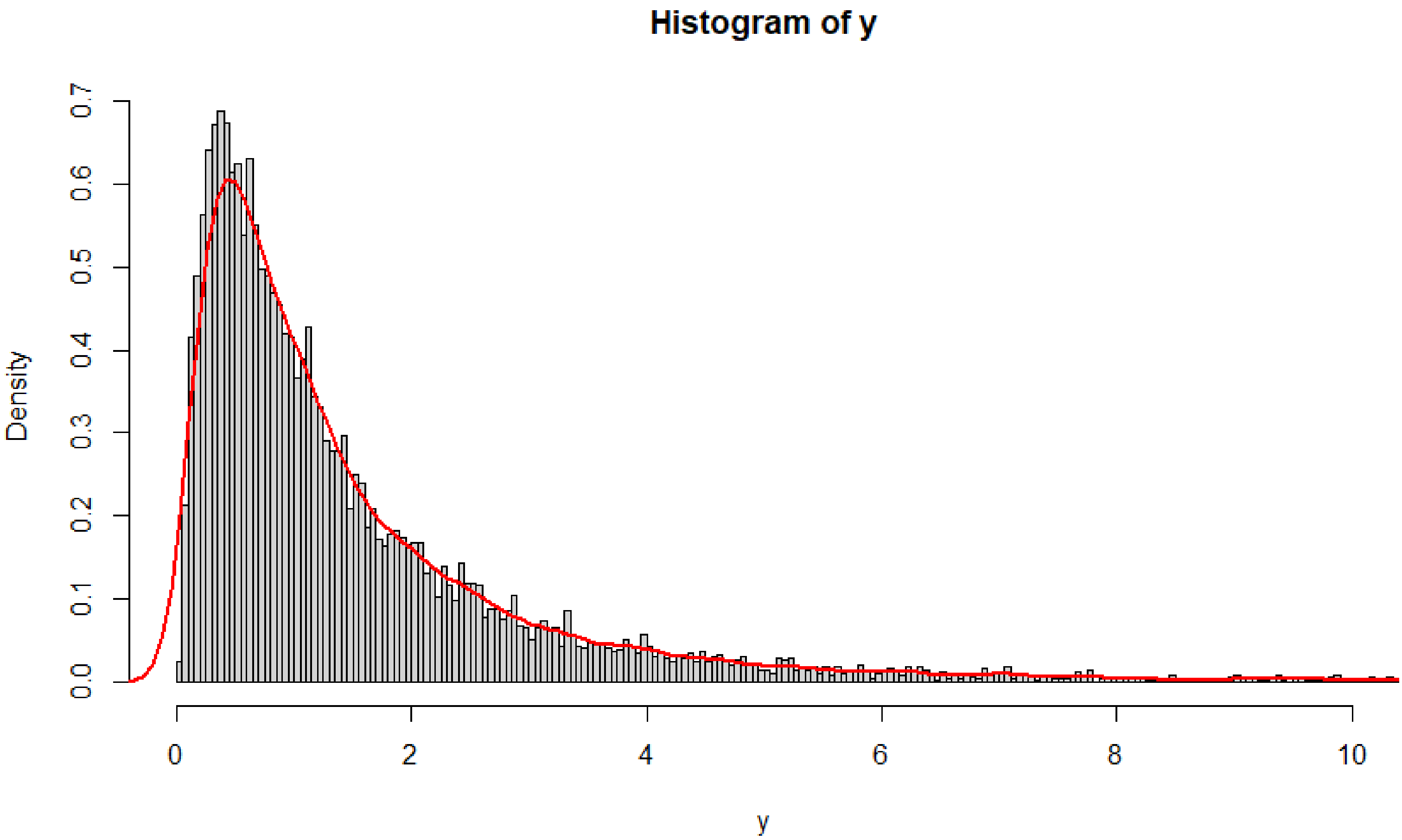
From the experiment I have plotted a histogram where, there were y-values on the x-axis and their probabilities on the y-axis. I got the right-skewed graph, where all the values are positive, since every real number in e^x gives us a positive number.

The density curve perfectly fits on the distribution probability graph. The experiment is successfully done.

13Ex.R

Source on Save

```r
n <- 10000

x <- rnorm(n)

y <- exp(x)

hist(y,breaks = 1000, xlim=c(0,10), prob=TRUE) #Draw a histogram

lines(density(y), col="red", lwd=2) #Compare with the Probability Density Function

#Zhetessov Nur
```


Histogram of y
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
2
4
6
8
10
y