03 Worksheet. Crash Course in Statistics (Summer 2025)

Neuroscience Center Zurich, University of Zurich

Zofia Baranczuk

2025-08-25

1. Example IQ

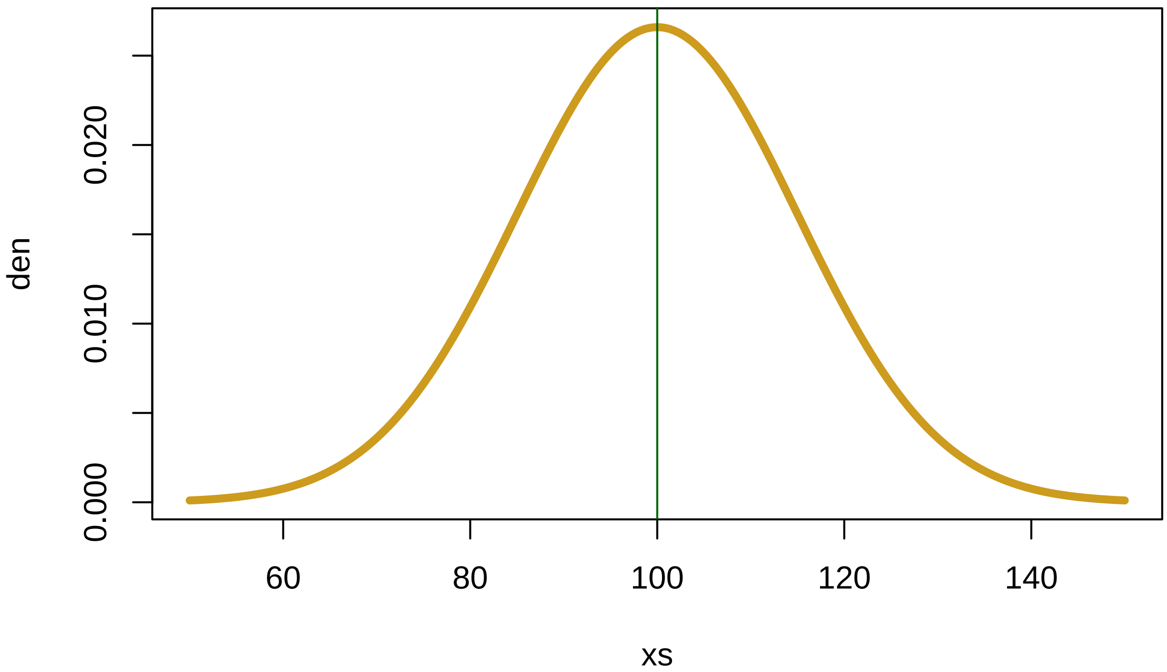
IQ scores are designed to follow a normal distribution with a mean of 100 and a standard deviation of 15. Using this, compute the following. For each of these points, add a corresponding plot. (It doesn't have to be aesthetically pleasing, but it should help you get the intuition.) a) P(IQ < 100) b) P(IQ = 100) c) P(IQ > 130) d) P(95 < IQ < 120) e) P(IQ < X) = 0.8. Compute X f) P(|IQ-100| > X) = 0.2. Compute X

a)

```
xs <- seq(50,150, length = 200)

mu <- 100
sd <- 15
pnorm(100, mu, sd)
```

```
## [1] 0.5
```

```
den <- dnorm(xs, mu, sd)
plot(xs, den, col="goldenrod3", type = "l", lwd=4)
abline(v = 100, col ="darkgreen")
```



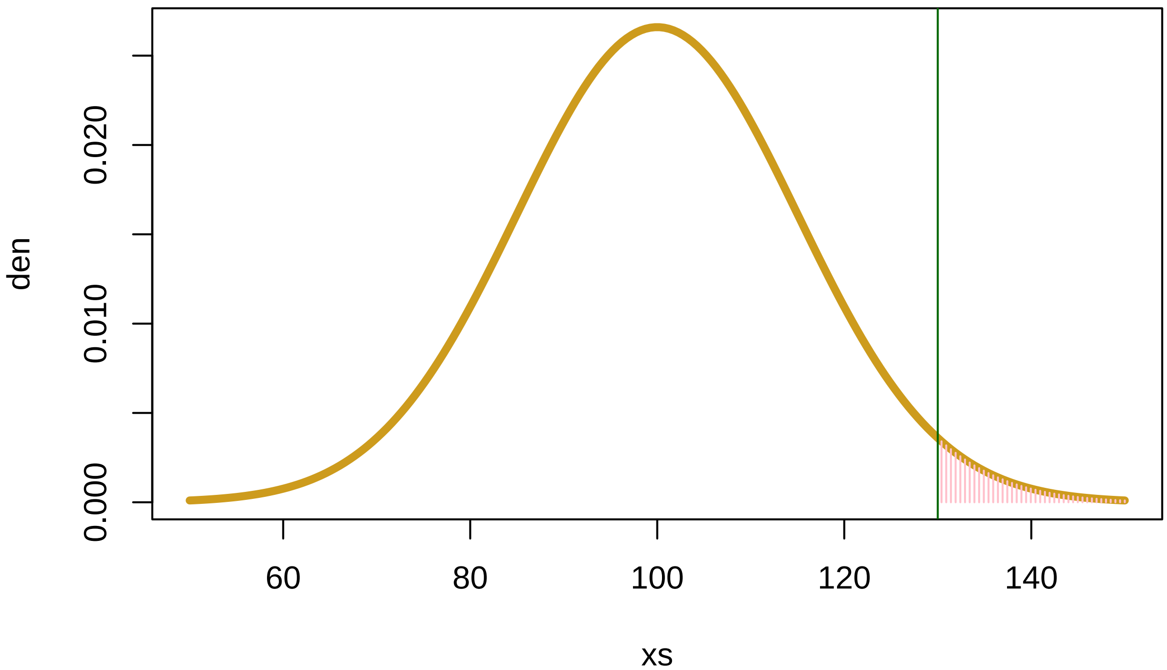
b) P(IQ = 100)

c)

```
pnorm(130, mean = 100, sd = 15, lower.tail = FALSE)
```

```
## [1] 0.02275013
```

```
plot(xs, den, col="goldenrod3", type = "l", lwd=4)
segments(x0 = xs[xs>130], y0 = 0, x1 = xs[xs>130], y1 = den[xs>130],col = "pink")
# that's not elegant, but maybe the easiest way for marking the area of interest
abline(v = 130, col ="darkgreen")
```

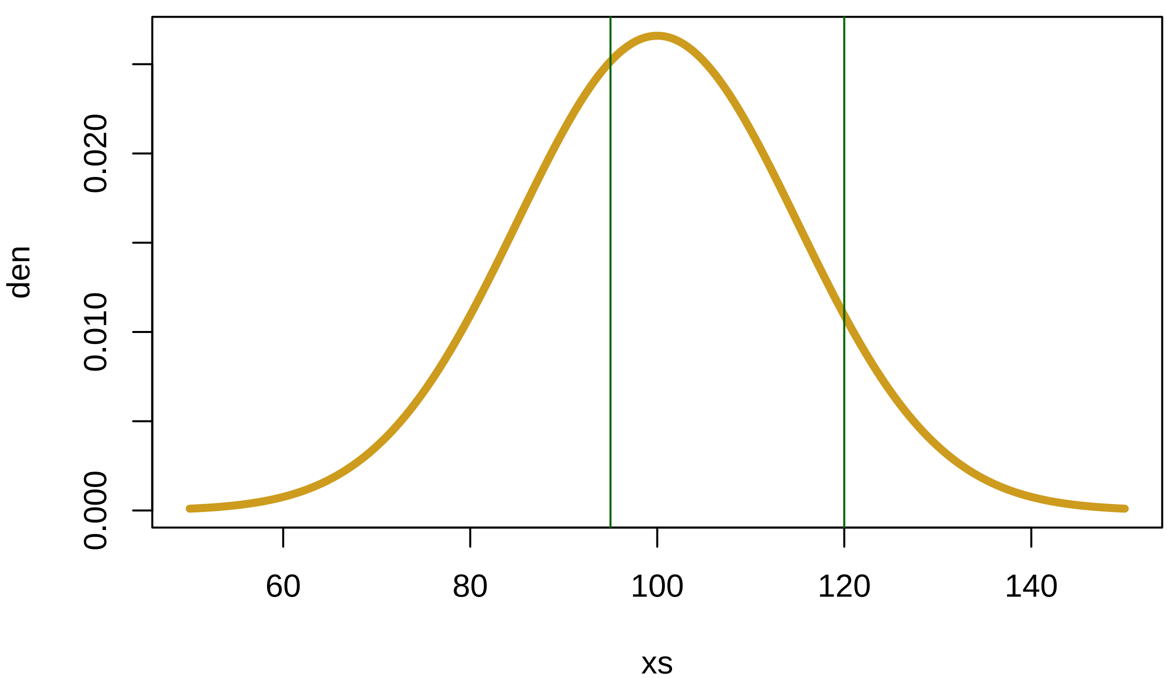


d)

```
pnorm(120, mean = 100, sd = 15)-pnorm(95, mean = 100, sd = 15)
```

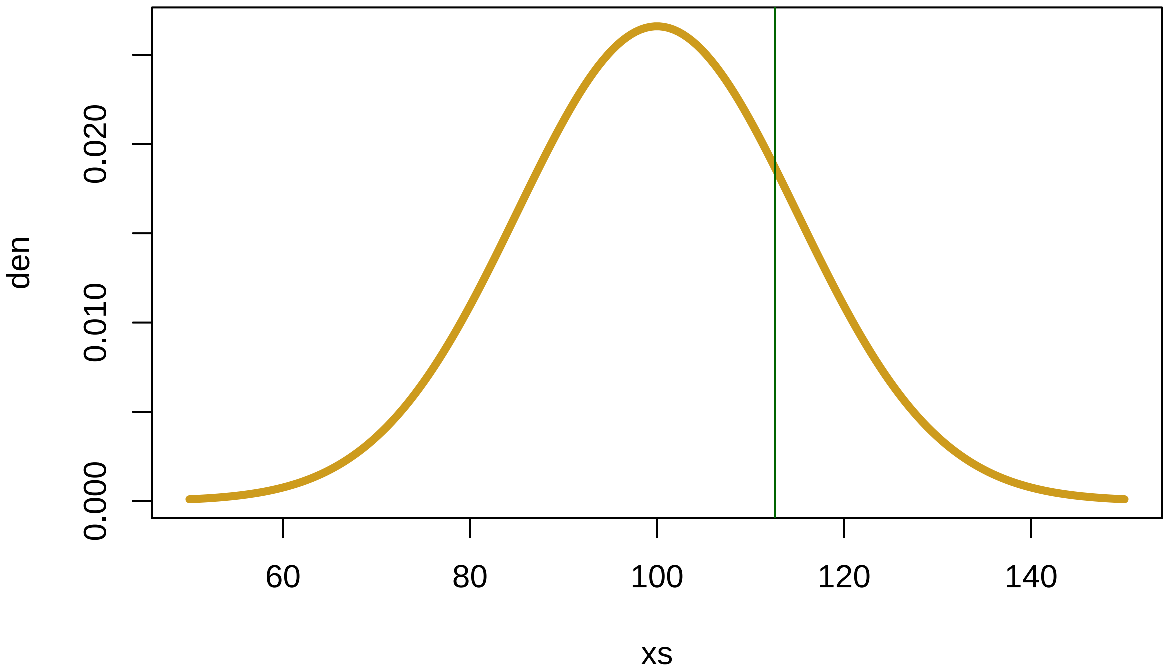
```
## [1] 0.5393474
```

```
plot(xs, den, col="goldenrod3", type = "l", lwd=4)
abline(v = c(95,120), col ="darkgreen")
```



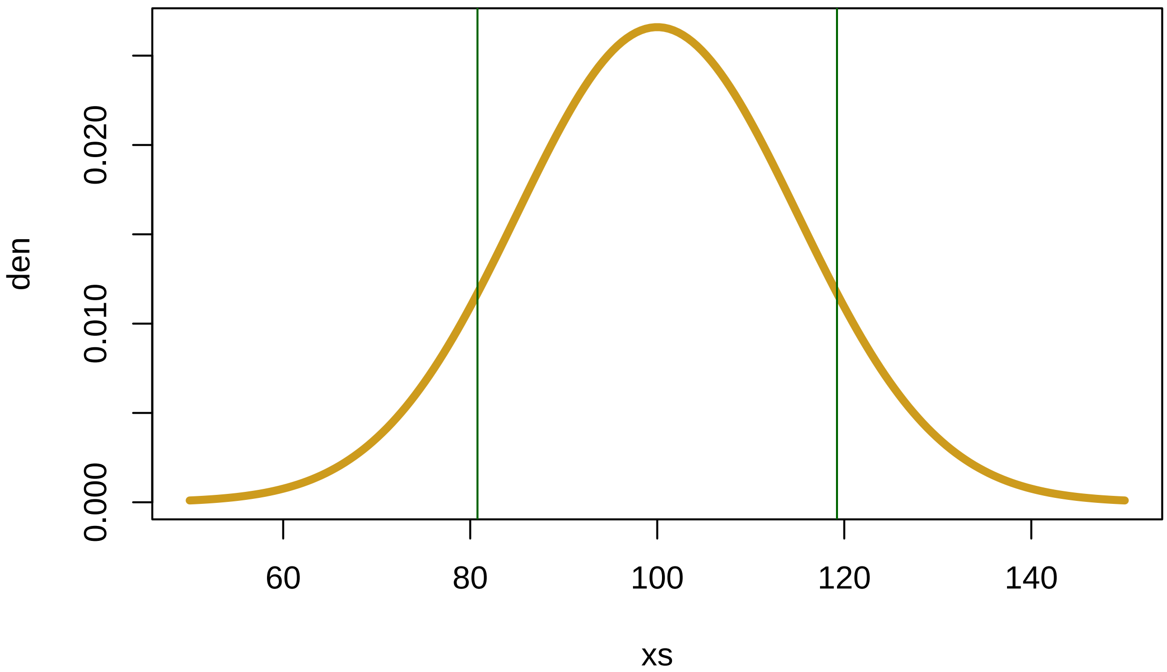
e)

```r
q <- qnorm(0.8, mean = 100, sd = 15)

plot(xs, den, col="goldenrod3", type = "l", lwd=4)
abline(v = q , col ="darkgreen")
```



f)

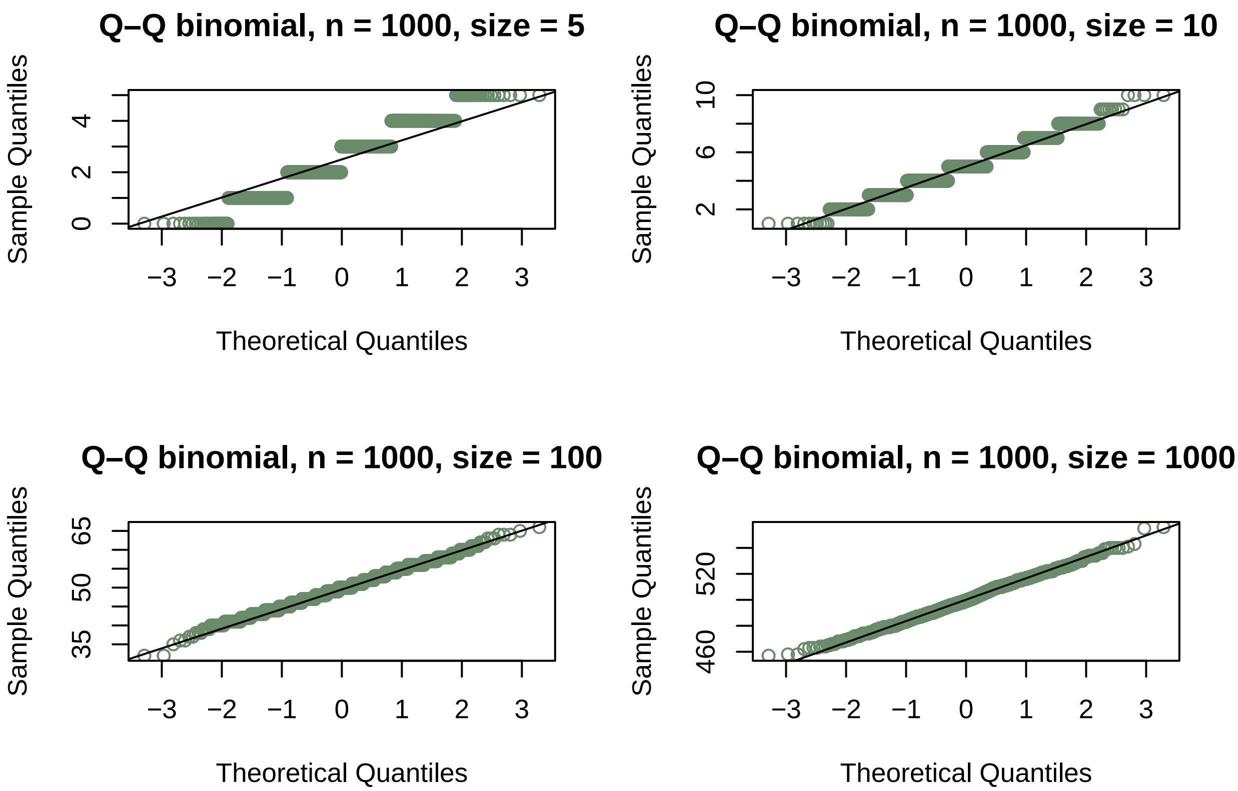
```r
# P(|IQ - 100 |>X) = 0.2
# P (two-sided tail = 0.2 , so each tail = 0.1).

lower <- qnorm(0.1, mean = 100, sd = 15)
upper <- qnorm(0.1, mean = 100, sd = 15, lower.tail = FALSE)

X <- upper - 100 # = 100 - lower

plot(xs, den, col="goldenrod3", type = "l", lwd=4)
abline(v = c(lower, upper) , col ="darkgreen")
```



Q-Q plots.

For large enough sample size, we can approximate binomial distribution using the normal distribution. For different values of size, sample from a binomial distribution and plot a q-q plot for the sampled values.

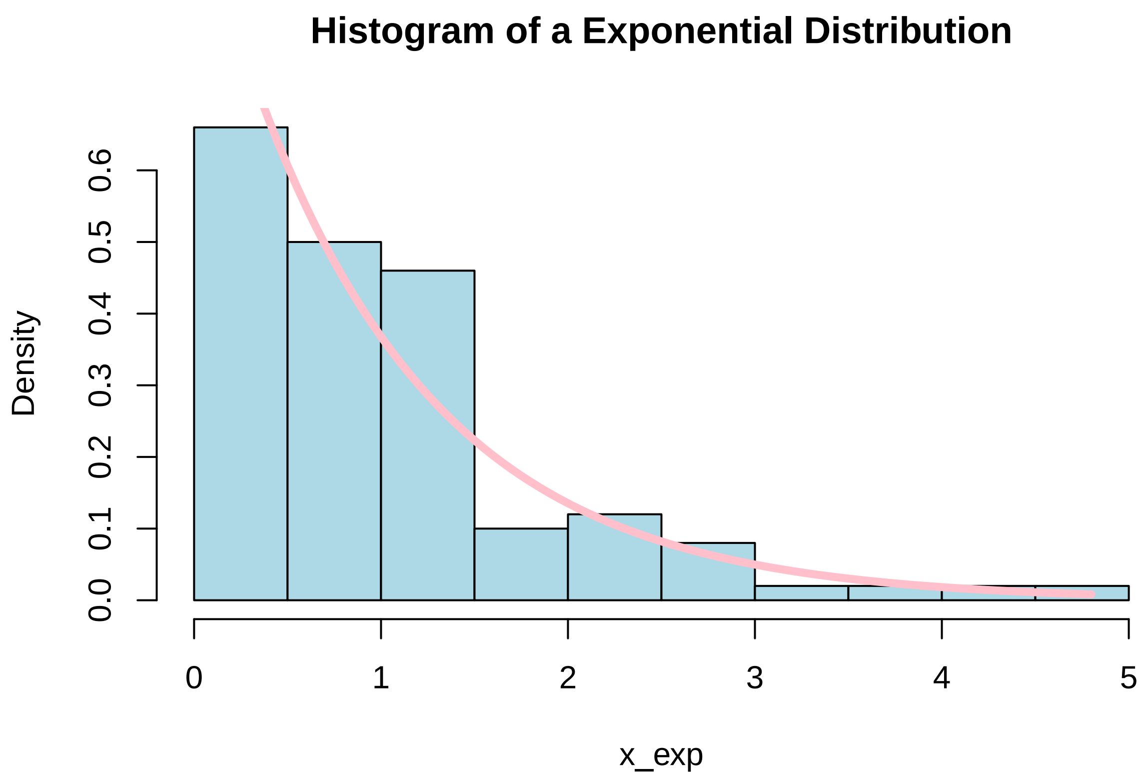
```
# to discuss loop
set.seed(2025)
n <- 1000 # to look at different n!
size = c(5,10, 100, 1000)
prob <- 0.5

par(mfrow = c(2, 2))
for (i in size){
  bin <- rbinom(n=n, size = i, prob = prob)
  qqnorm(bin, col = "darkseagreen4", main = sprintf("Q-Q binomial, n = %d, size = %d", n, i))
  qqline(bin)
}
```



Exponential distribution

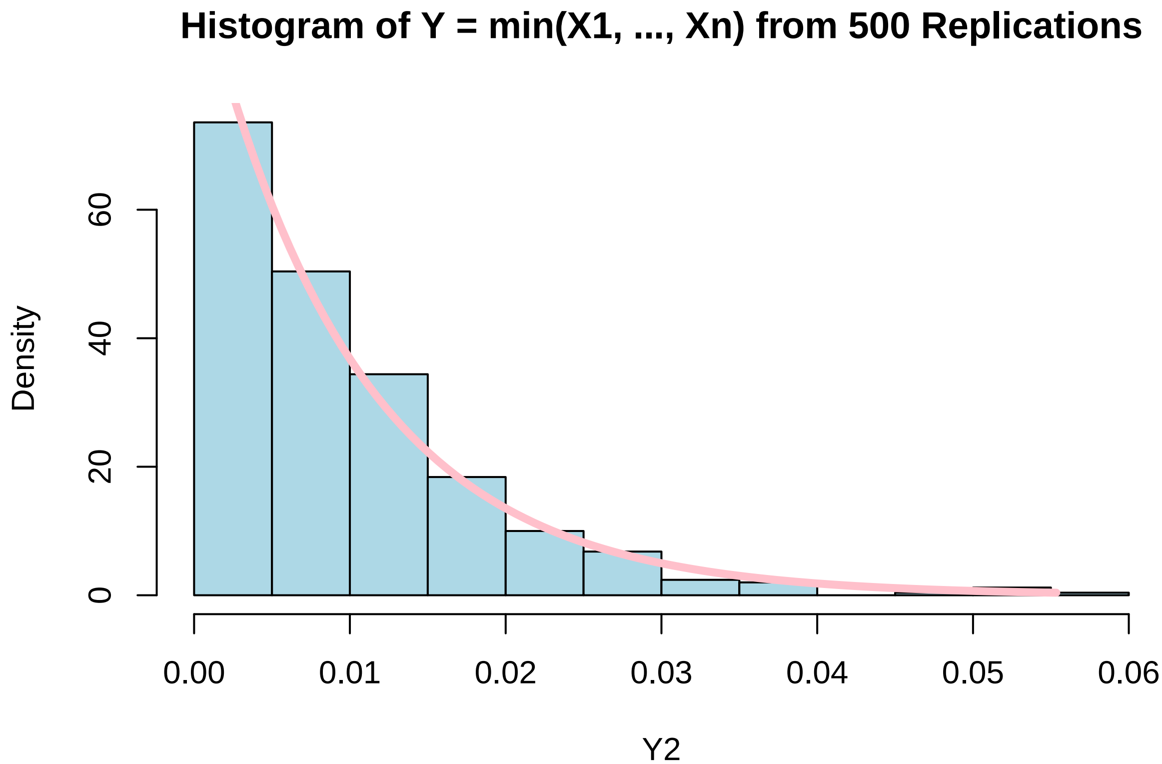
Let X1,...,Xn be independent and identically distributed (iid) random variables following Exp(lambda). Assume lambda = 1. a) Sample n = 100 random numbers from Exp(lambda). Visualize the data with a histogram and superimpose the theoretical density.

```
set.seed(1)
n =  100
lambda = 1
x_exp <- rexp(n, rate = lambda)

x_max <- max(x_exp)

xs <- seq(0,x_max, by=0.05)

pmf <- dexp(xs, lambda)
hist(x_exp, probability = TRUE, col = "lightblue",
     main = "Histogram of a Exponential Distribution")
lines(xs, pmf, col = "pink", lwd = 4)
```



b) Repeat the experiment 500 times. For each replication, record Y=min(X1, X2, ..., Xn). Plot a histogram of the 500 values of Y. What distribution do you expect for Y, and what is its parameter value? Superimpose the theoretical density on the histogram to check.

```
n <- 100
lambda <- 1
n_rep <- 500

Y1 <- replicate(n_rep, min(rexp(n, rate = lambda))) # more R style
# repeat n_rep times taking the min of rexp(n, rate = lambda)

#or, more or less equivalent:

Y2 <- numeric(n_rep) #prepare space for storing Y values
for (i in 1:n_rep){
  Y2[i] <- min(rexp(n=100,1))
}


xs <- seq(0, max(Y2), length=100)
den <- dexp(xs, rate = n * lambda)
hist(Y2, probability = TRUE, col = "lightblue",
     main = "Histogram of Y = min(X1, ..., Xn) from 500 Replications")

lines(xs, den, col = "pink", lwd = 4)
```



We expected the waiting times to be smaller, when there are 100 events for wich we are waiting, and we are waiting only for the first one. Actually, the rate gets 100 times higher, and waiting time on average 100 times shorter. We can see the density with rate lambda*100 corresponding to the histogram of min(X1, . . . , X100).