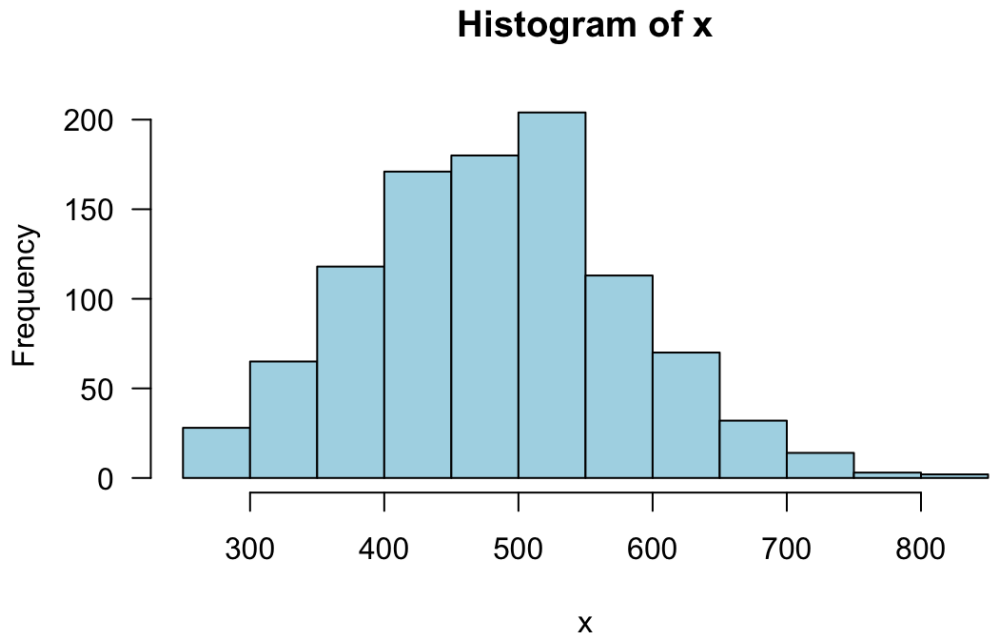
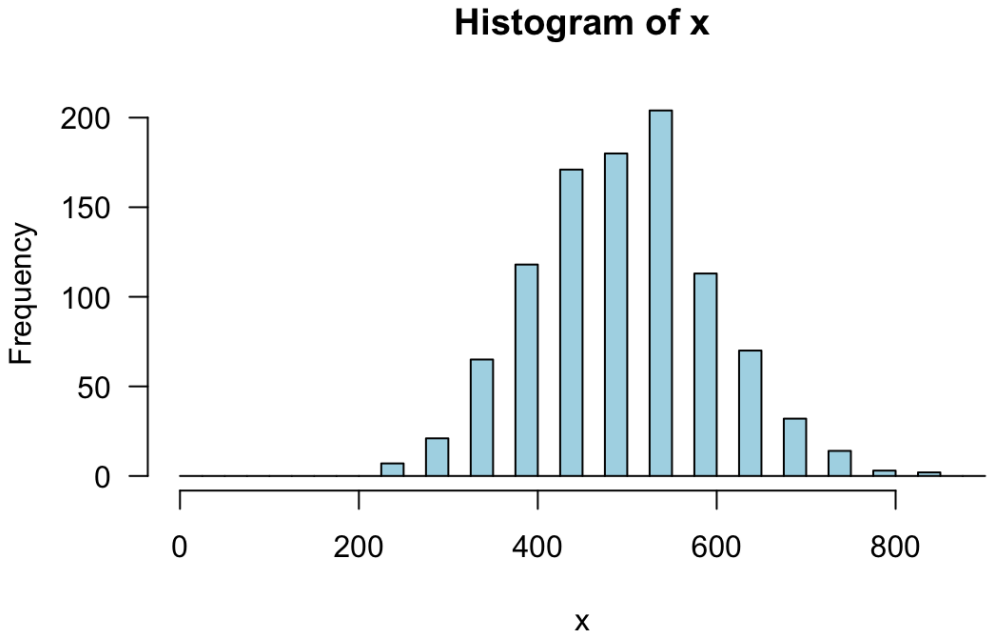
Continuous Variables, pt. 3

Joyce Robbins

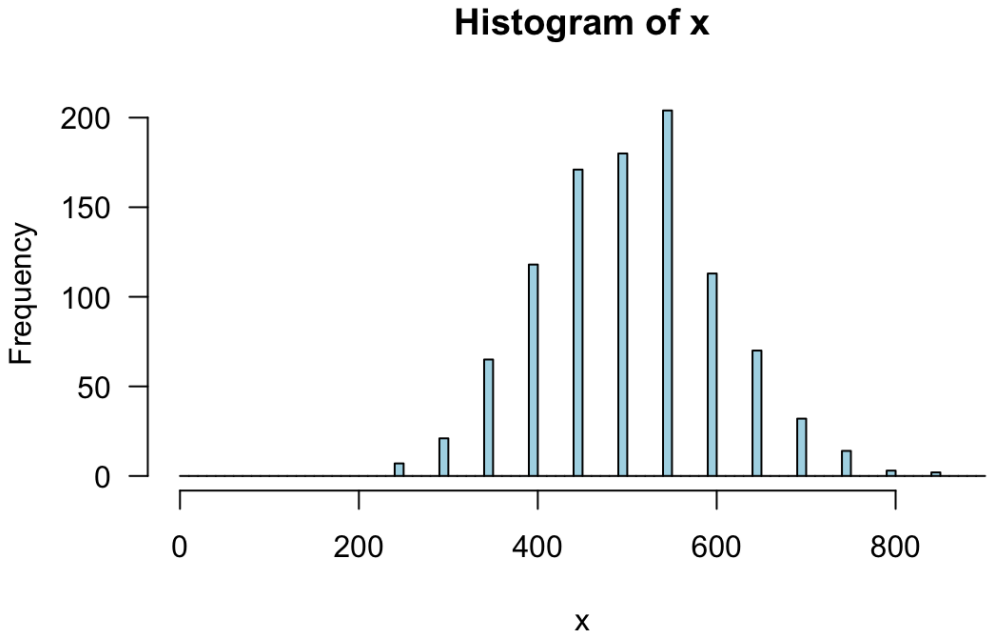
Rounding pattern



Change binwidth to 25



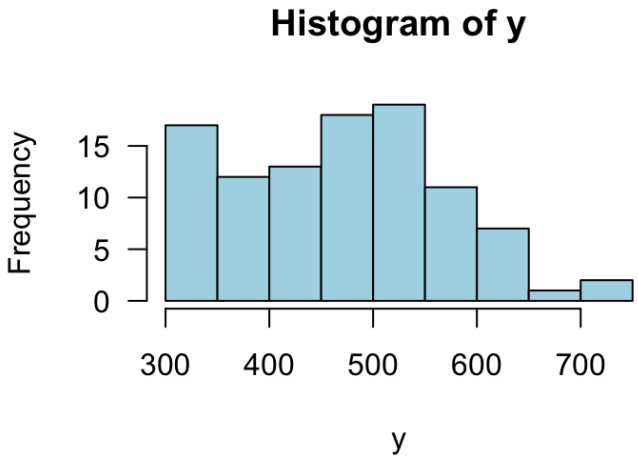
Change binwidth to 10



Stem and leaf

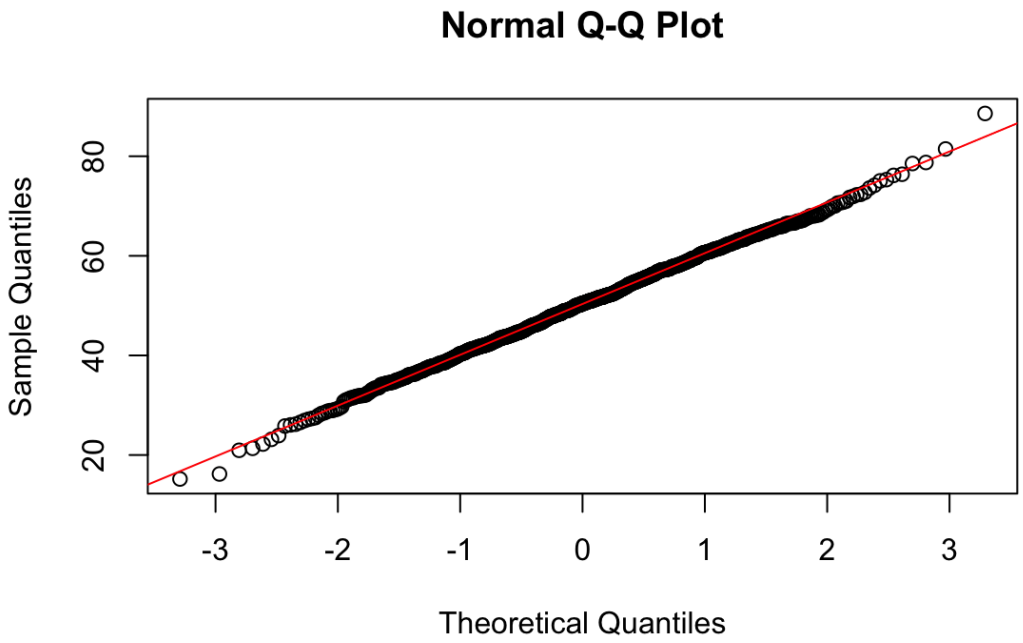
```
stem(y)
```

```
##
##   The decimal point is 2 digit(s) to the right of
the |
##
##    3 | 000
##    3 | 55555555555555
##    4 | 000000000000
##    4 | 5555555555555
##    5 | 0000000000000000000
##    5 | 5555555555555555555
##    6 | 00000000000
##    6 | 5555555
##    7 | 0
##    7 | 55
```



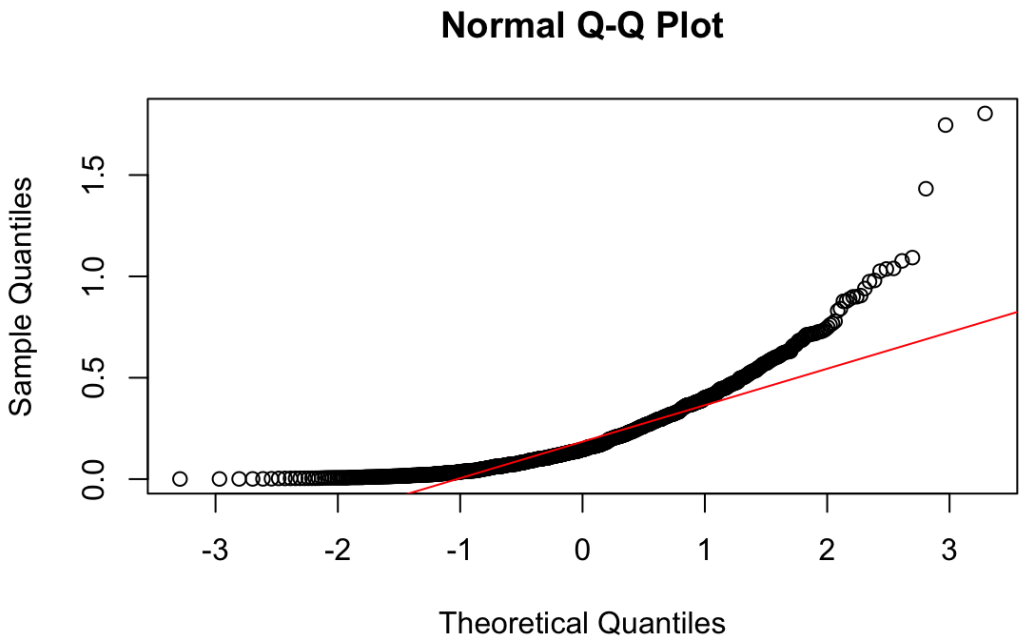
Q-Q plot (quantile-quantile)

normal



Q-Q plot (quantile-quantile)

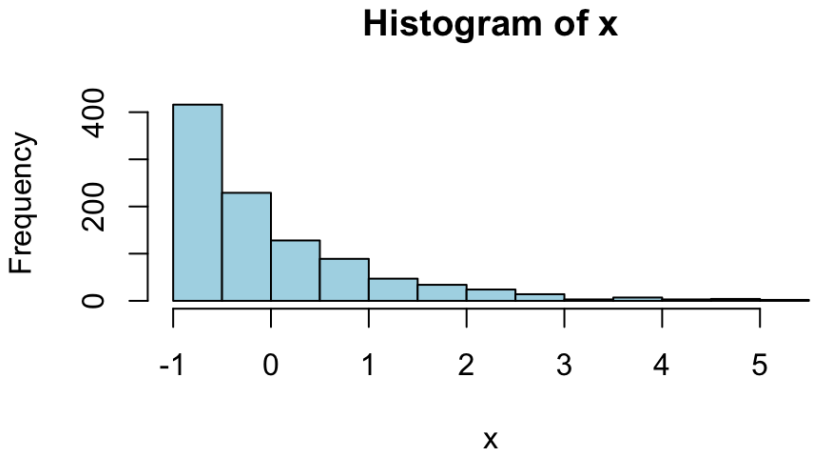
not normal



Exponential distribution

$$f(x) = \lambda e^{-\lambda x} - 1$$

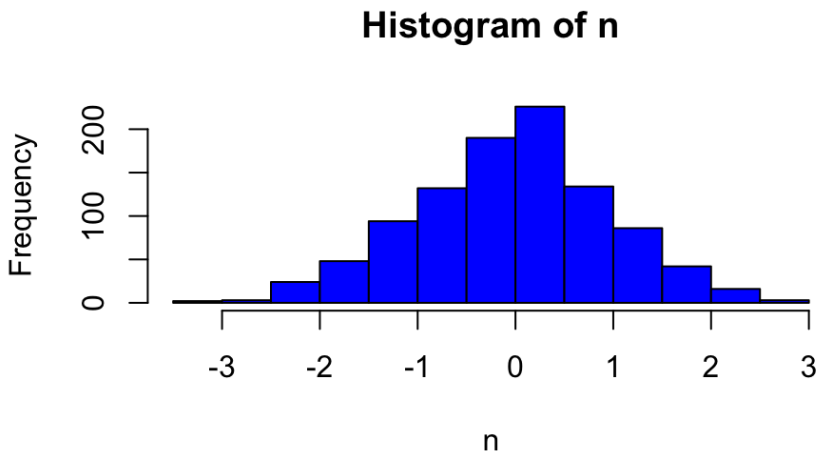
```r
set.seed(5702)
x <- rexp(1000, rate = 1) - 1    # rate = lambda
hist(x, col = "lightblue")
```



Normal distribution

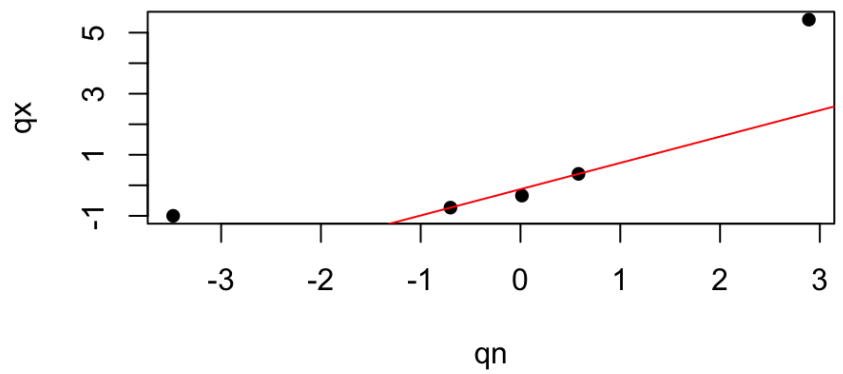
```r
n <- rnorm(1000)

hist(n, col = "blue")
```



DIY QQ plot

```
qx <- quantile(x)
qn <- quantile(n)
plot(qn, qx, pch = 16)
mod <- lm(c(qx[2], qx[4])~c(qn[2], qn[4]))
abline(mod, col = "red")
```




```
qx
```

```
##      0%     25%     50%     75%    100%
## -0.999 -0.731 -0.337   0.375   5.426
```

```
qn
```

```
##       0%      25%      50%      75%     100%
## -3.4811 -0.7018   0.0145   0.5814   2.8903
```

DIY QQ plot

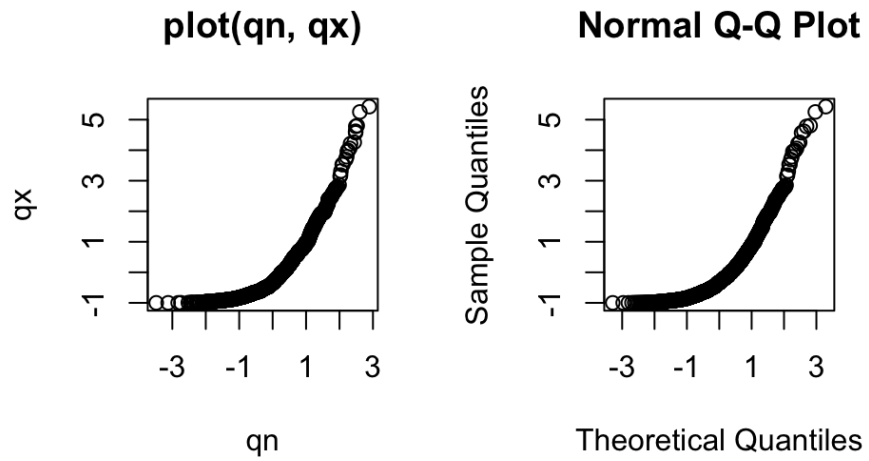
```
qx <- quantile(x, probs = seq(0, 1, .001))
qn <- quantile(n, probs = seq(0, 1, .001))
tail(qn)
```

```
##  99.5%  99.6%  99.7%  99.8%  99.9% 100.0%
##   2.48   2.48   2.49   2.53   2.61   2.89
```

```
tail(qx)
```

```
##  99.5%  99.6%  99.7%  99.8%  99.9% 100.0%
##   4.57   4.62   4.78   4.80   5.25   5.43
```

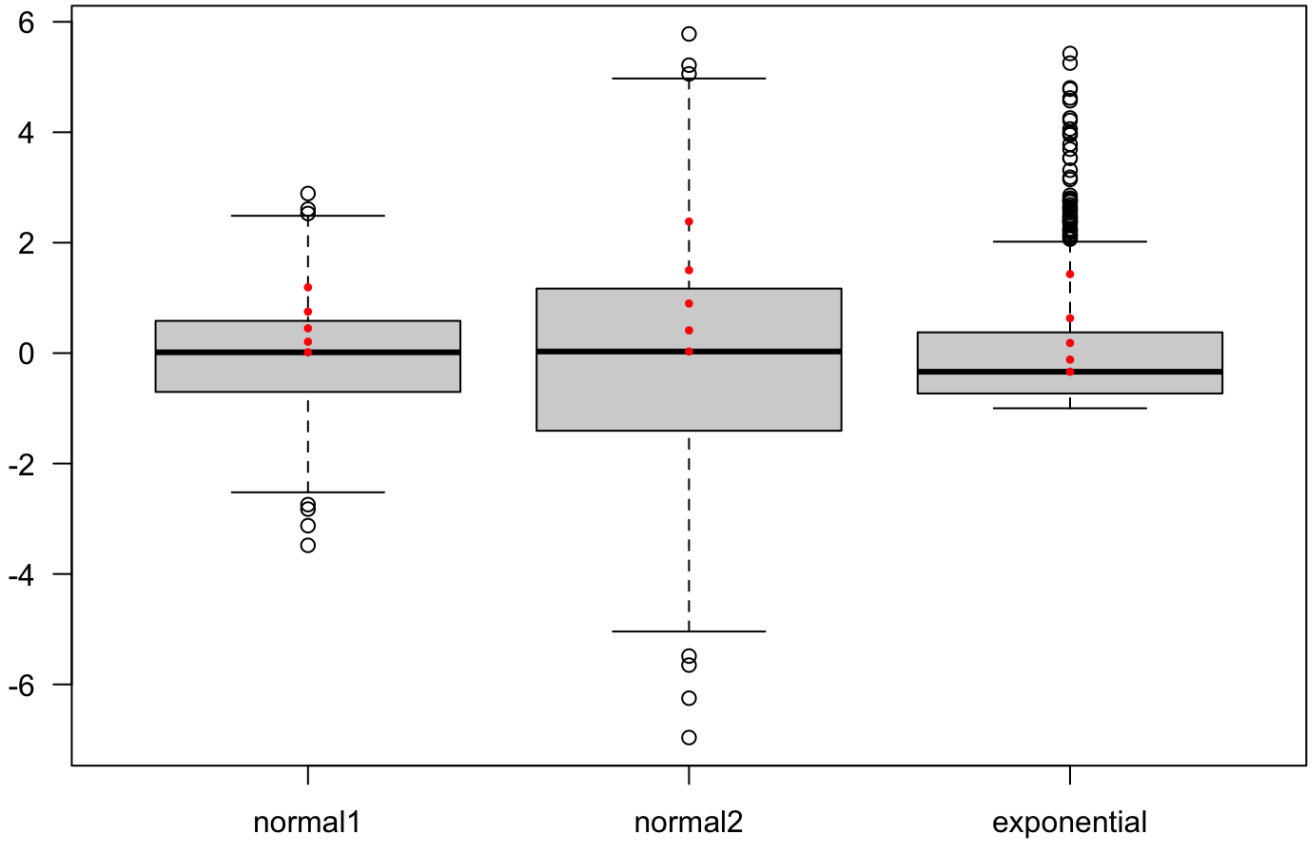
```
oldpar <- par(mfrow = c(1, 2))
plot(qn, qx, main = "plot(qn, qx)")
qqnorm(x)
```




```
par(oldpar)
```

Boxplot comparison

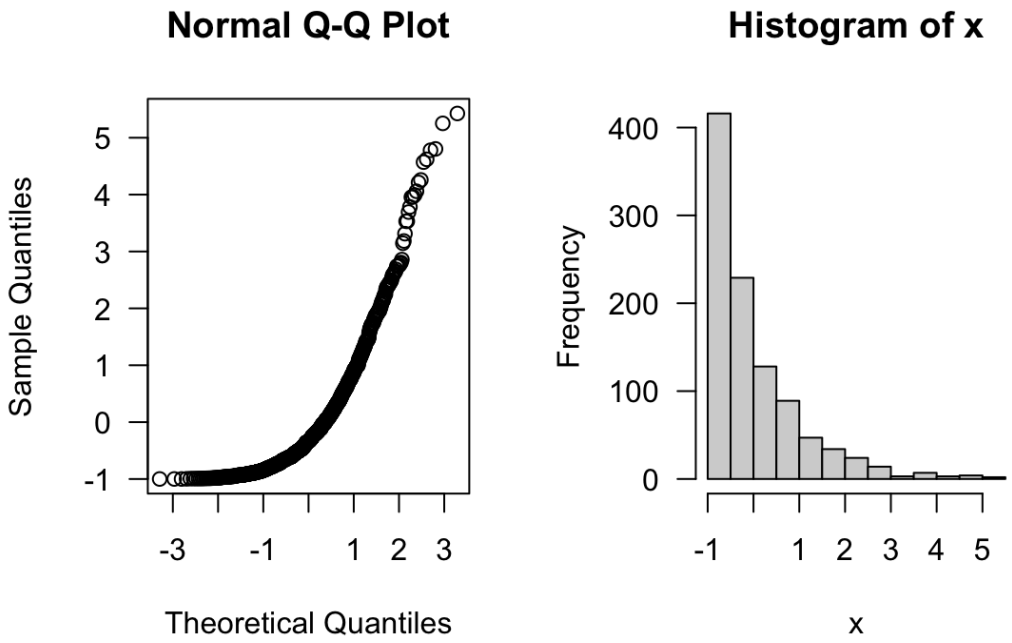
50th, 60th, 70th, 80th, 90th percentiles shown in red



Right skew

Exponential distribution

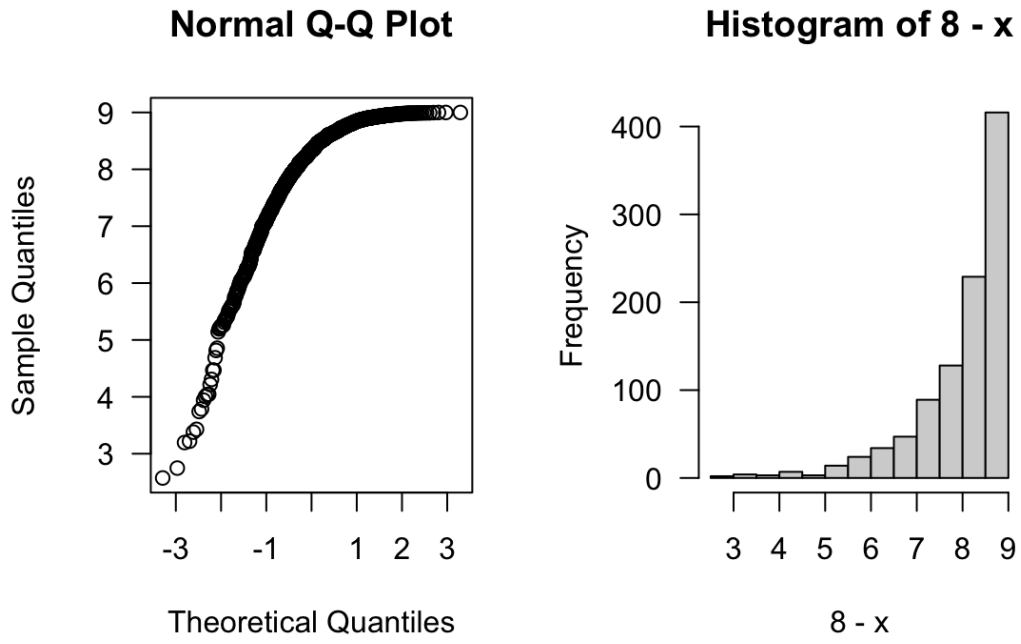
$$f(x) = \lambda e^{-\lambda x}$$



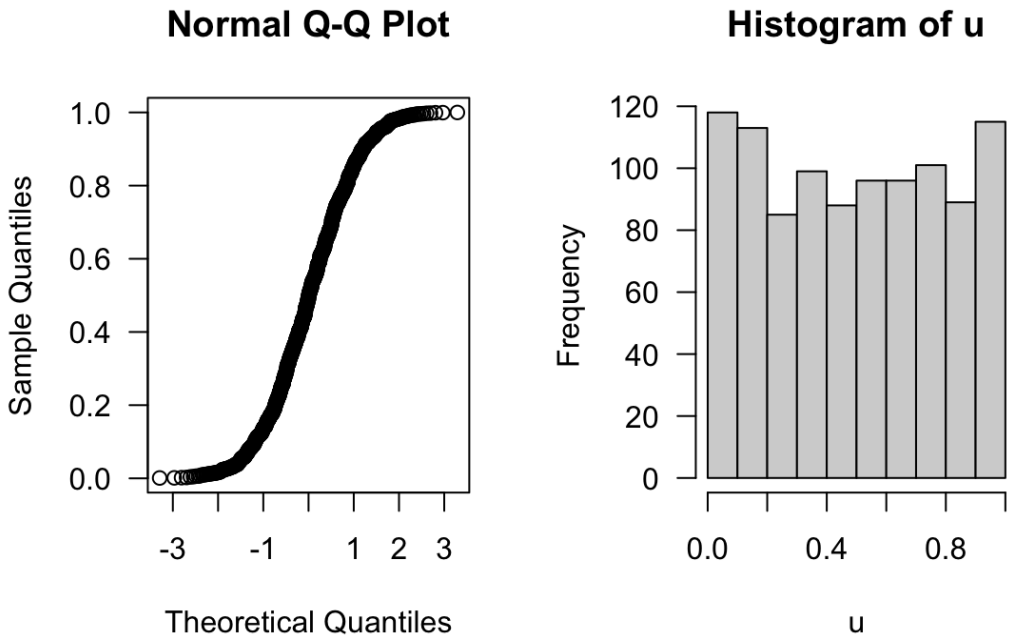
Left skew

Exponential distribution

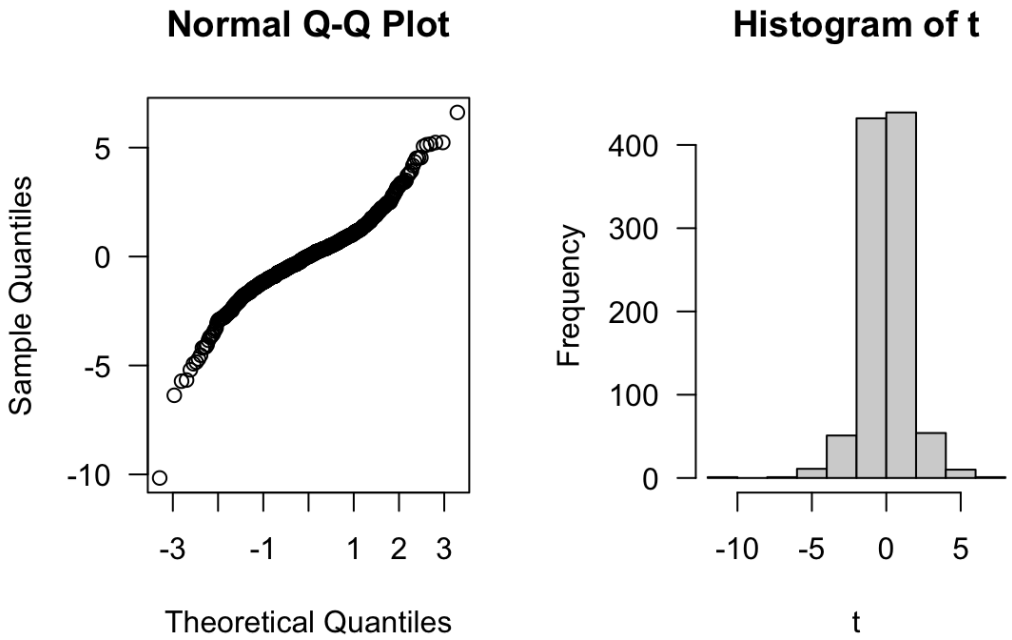
$$f(x) = 8 - \lambda e^{-\lambda x}$$



Uniform distribution

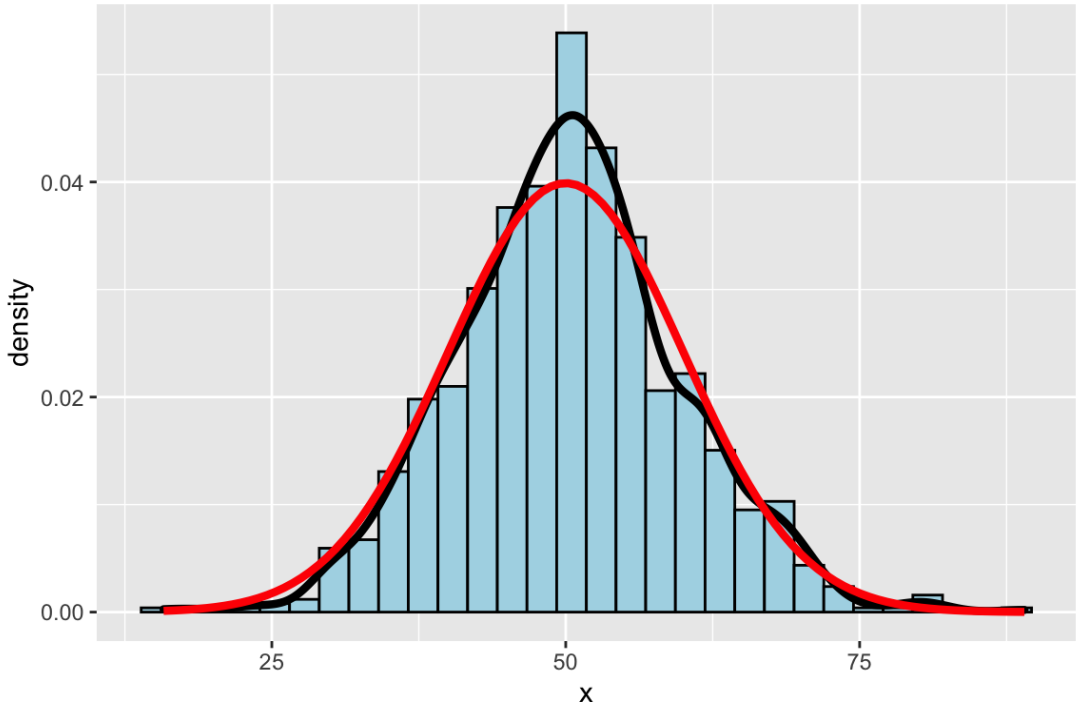


t-distribution (fat tails)



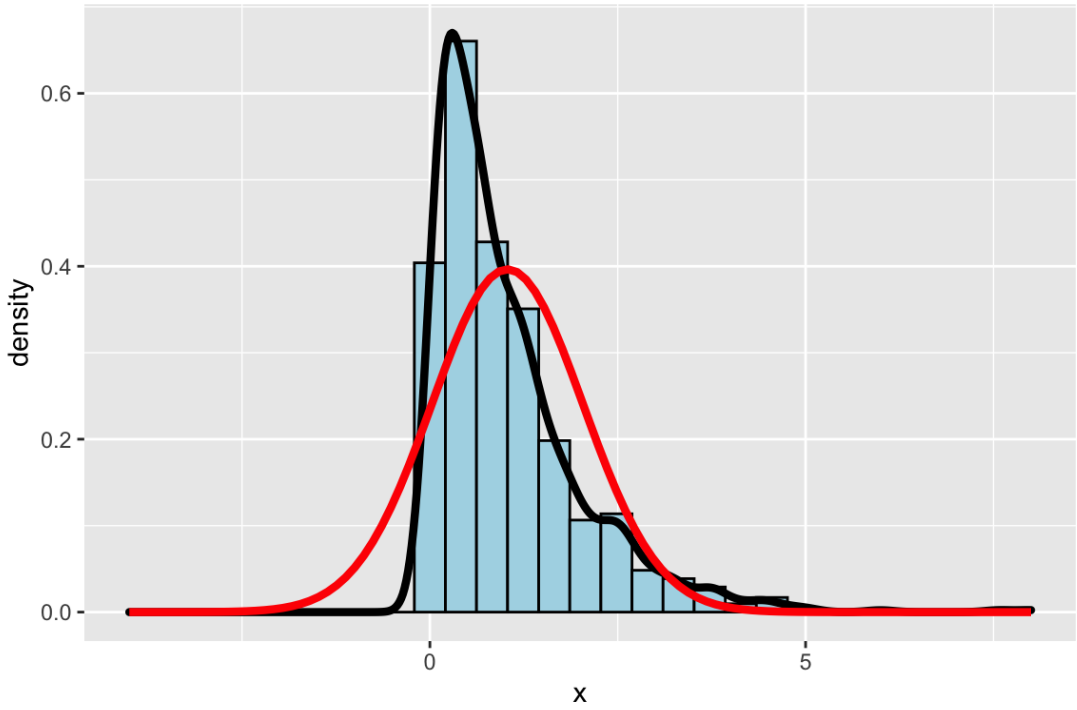
Density Curve + Normal Curve

Normal



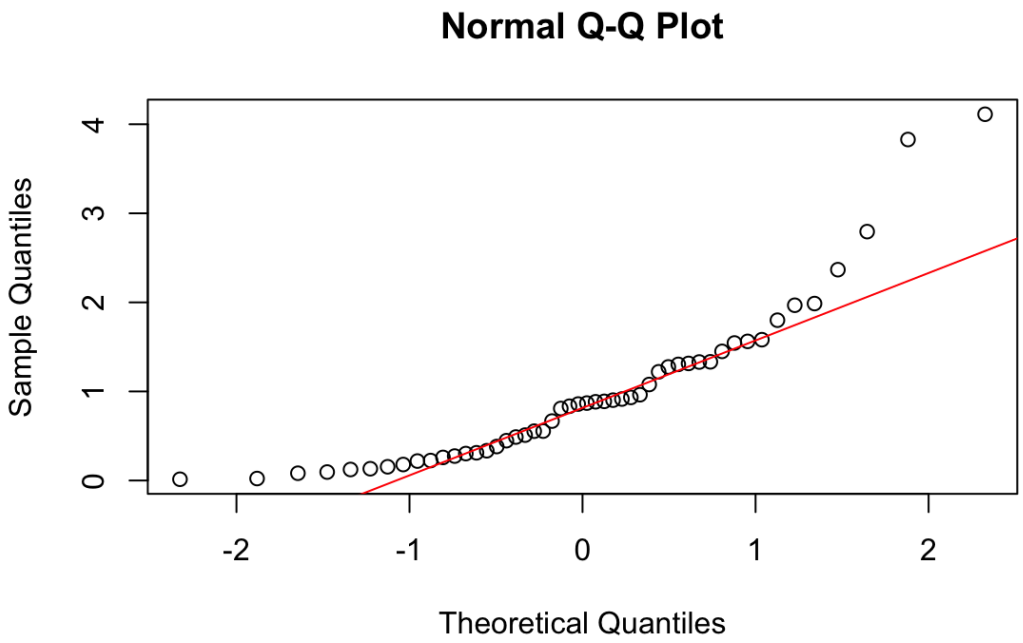
Density Curve + Normal Curve

Not normal



Shapiro Wilk test

```r
x <- rexp(50)
qqnorm(x)
qqline(x, col = "red")
```



Null hypothesis: data is normally distributed

Alternative hypothesis: data is not normally distributed

Can we reject the null hypothesis?

```r
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.8, p-value = 0.000009
```

W is the test statistic

p-value depends on W and n