Statistical distributions vignette

Contents

1 Purpose **1**

2 Reference **2**

3 Libraries **2**

4 Normal distribution **2**

4.1 Generate the data 2

4.2 prob (z < a certain point) 3

4.3 Get point where prob = 0.95 3

4.4 Now for a diff mean 3

5 t distribution **3**

5.1 Generate data 3

5.2 prob (t < a certain point) 4

5.3 Get point where prob = 0.95 4

6 Q-Q plot (Normal) **5**

6.1 Default 5

6.2 Add a line 5

7 Simulations: Normal, Uniform, Exponential, t distributions **6**

8 Normality tests **7**

8.1 Q-Q plot 7

8.2 Shapiro test 8

1 Purpose

This vignette aims to introduce you statistical distributions using R.

2 Reference

Dr. Bharatendra https://www.youtube.com/watch?v=rsfV57N7Uns&list=PL34t5iLfZddtUUABMikey6NtL05hPAp42&index=7

3 Libraries

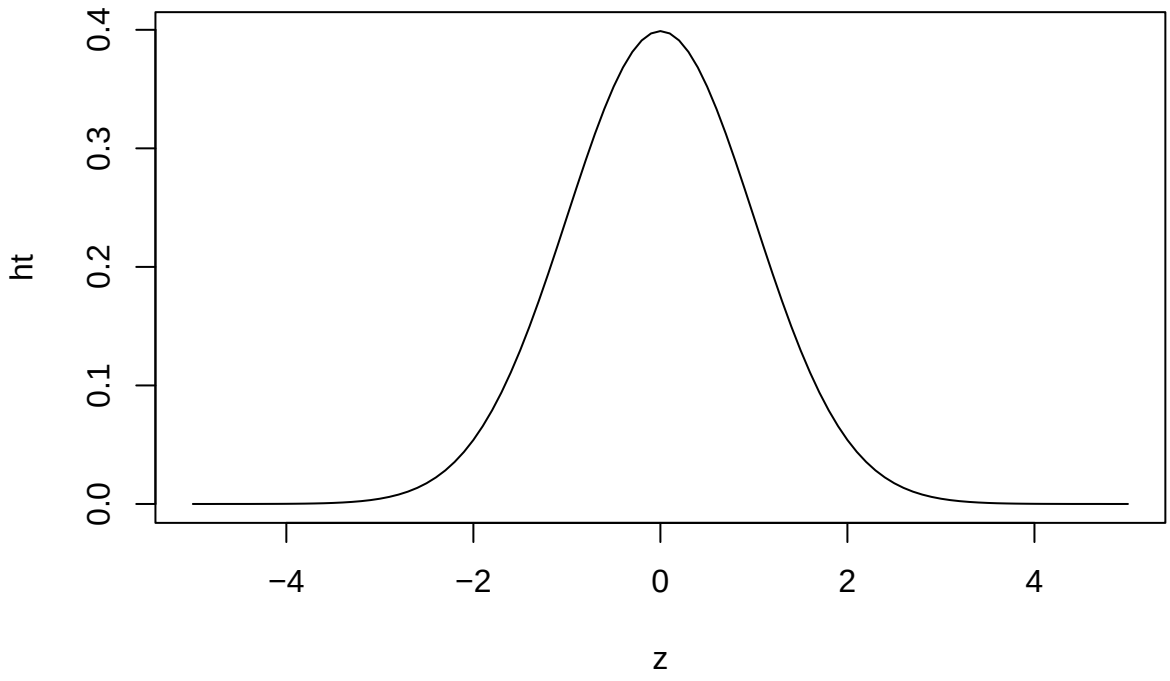
4 Normal distribution

4.1 Generate the data

```
z <- pretty(c(-5, 5), 100)

ht <- dnorm(z)

# plot type line
plot(z, ht, type='l')
```



4.2 prob (z < a certain point)

```
# P(z < -2)
pnorm(-2)
```

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

4.3 Get point where prob = 0.95

At what point will I have 95%

```
# P(z < ?) = 0.95
qnorm(.95)
```

```
## [1] 1.644854
```

4.4 Now for a diff mean

```
# Where mean is 0.2 and sigma is 0.5
qnorm(.95, 2, 0.5)
```

```
## [1] 2.822427
```

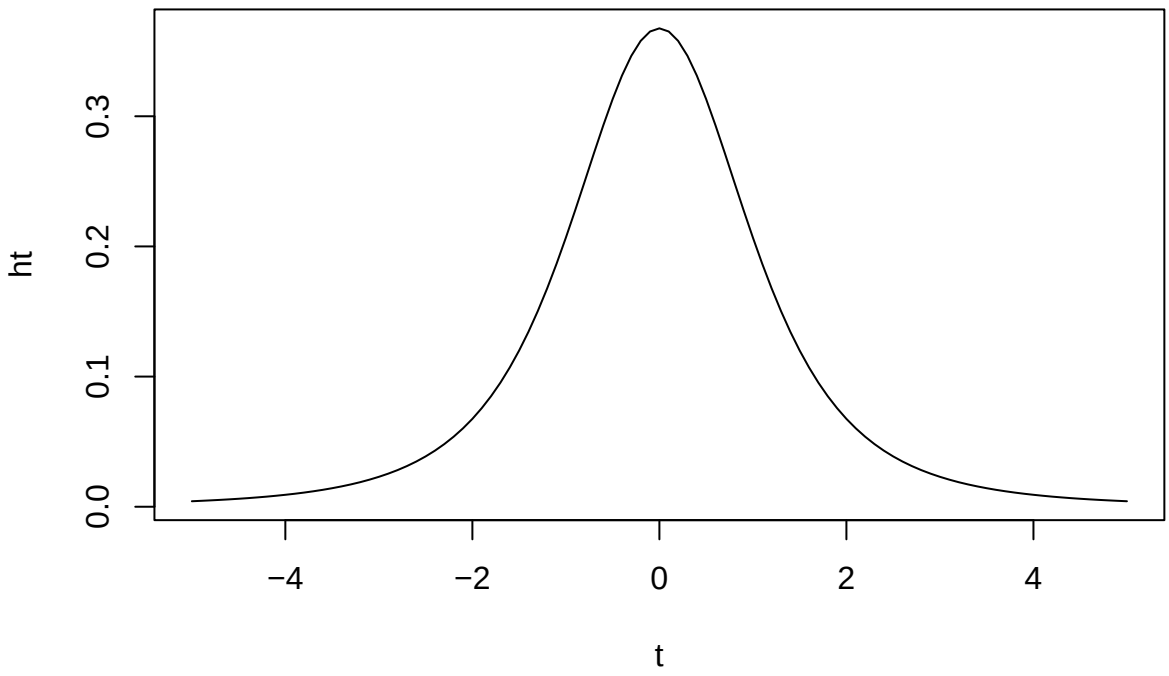
5 t distribution

5.1 Generate data

```
t <- pretty(c(-5, 5), 100)

ht <- dt(t, 3)

# plot type line
plot(t, ht, type='l')
```



5.2 prob (t < a certain point)

With 3 degrees of freedom

```
# P(z < -2)
pt(-2, 3)
```

```
## [1] 0.06966298
```

5.3 Get point where prob = 0.95

At what point will I have 95%, with 3 degrees of freedom

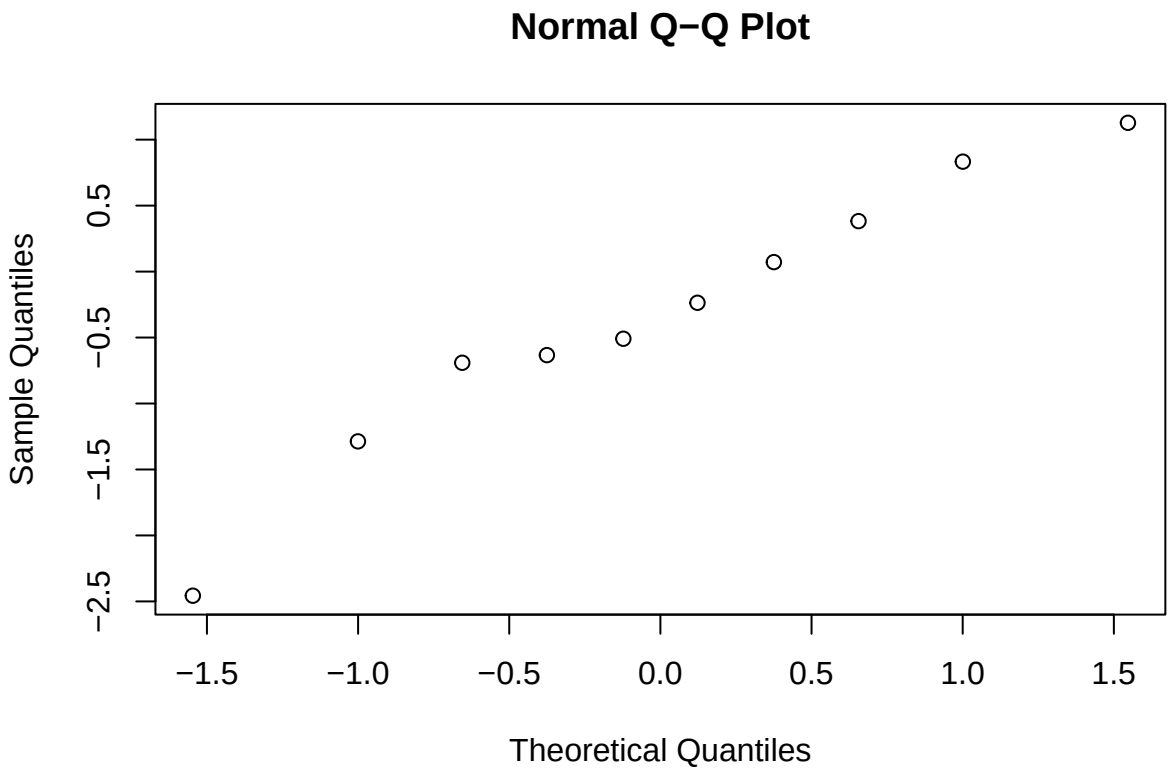
```
# P(z < ?) = 0.95
qt(.95, 3)
```

```
## [1] 2.353363
```

6 Q-Q plot (Normal)

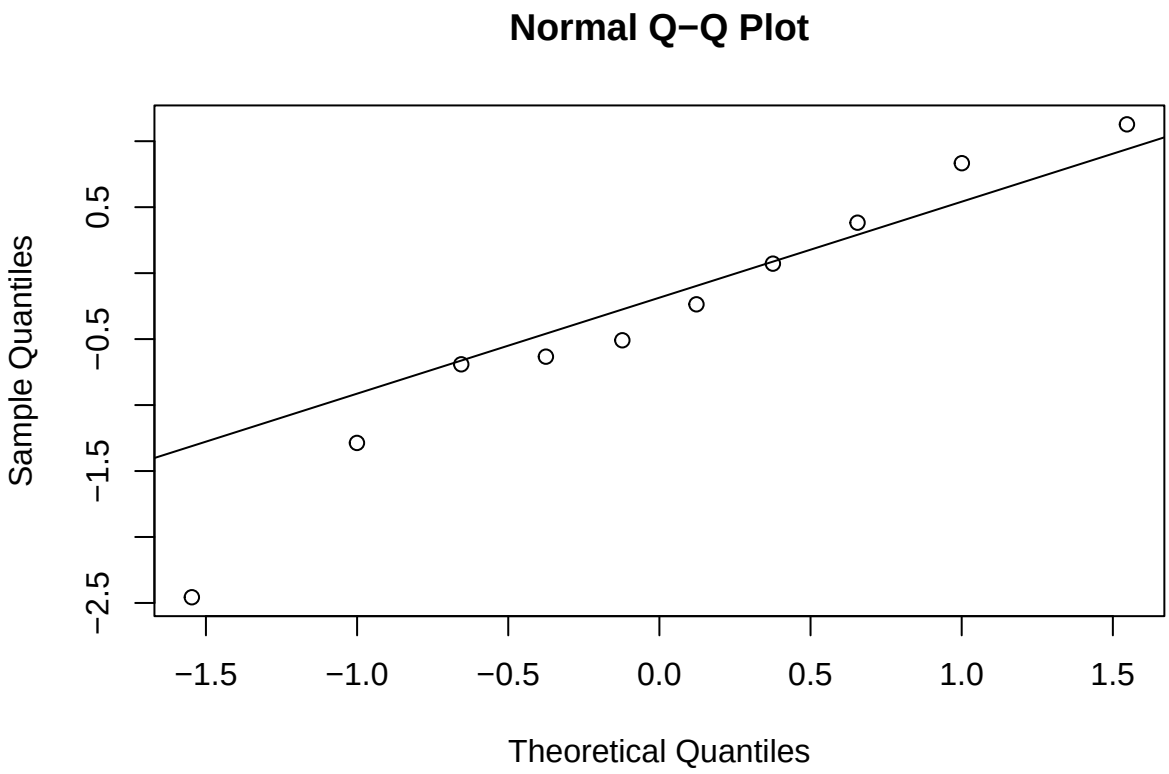
6.1 Default

```
x <- rnorm(10)
qqnorm(x)
```



6.2 Add a line

```
qqnorm(x)
qqline(x)
```

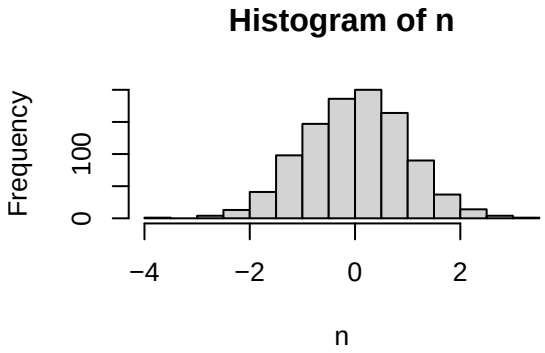


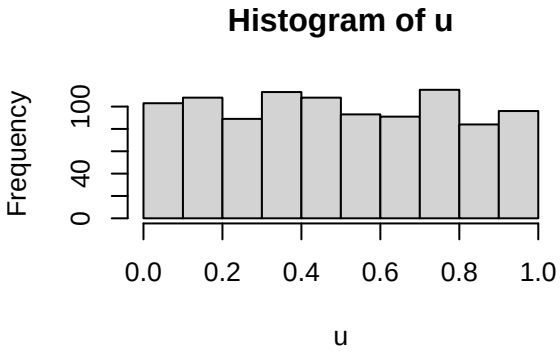
7 Simulations: Normal, Uniform, Exponential, t distributions

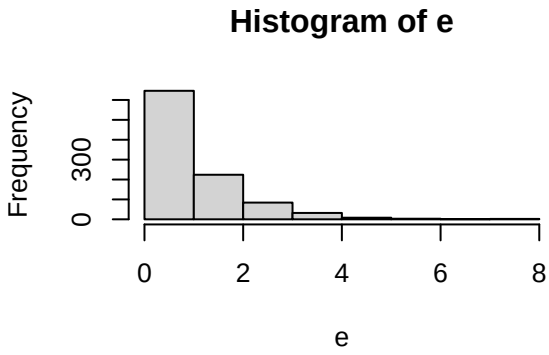
Reference: Dr. Bharatendra https://www.youtube.com/watch?v=XyBfmm1pk8g&list=PL34t5iLfZddtUUABMikey6NtL05hl
index=5

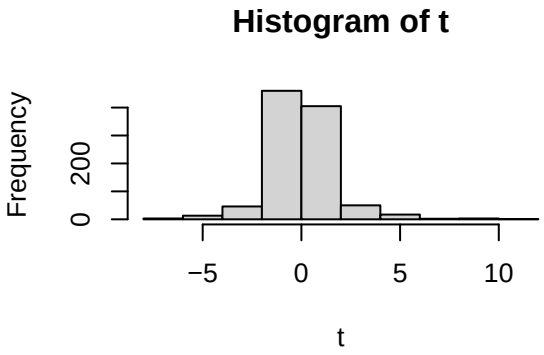
```
n <- rnorm(1000)
u <- runif(1000)
e <- rexp(1000)
t <- rt(1000, 3)

# Plot 2 x 2 plots
par(mfrow=c(2, 2))
hist(n)
hist(u)
hist(e)
hist(t)
```





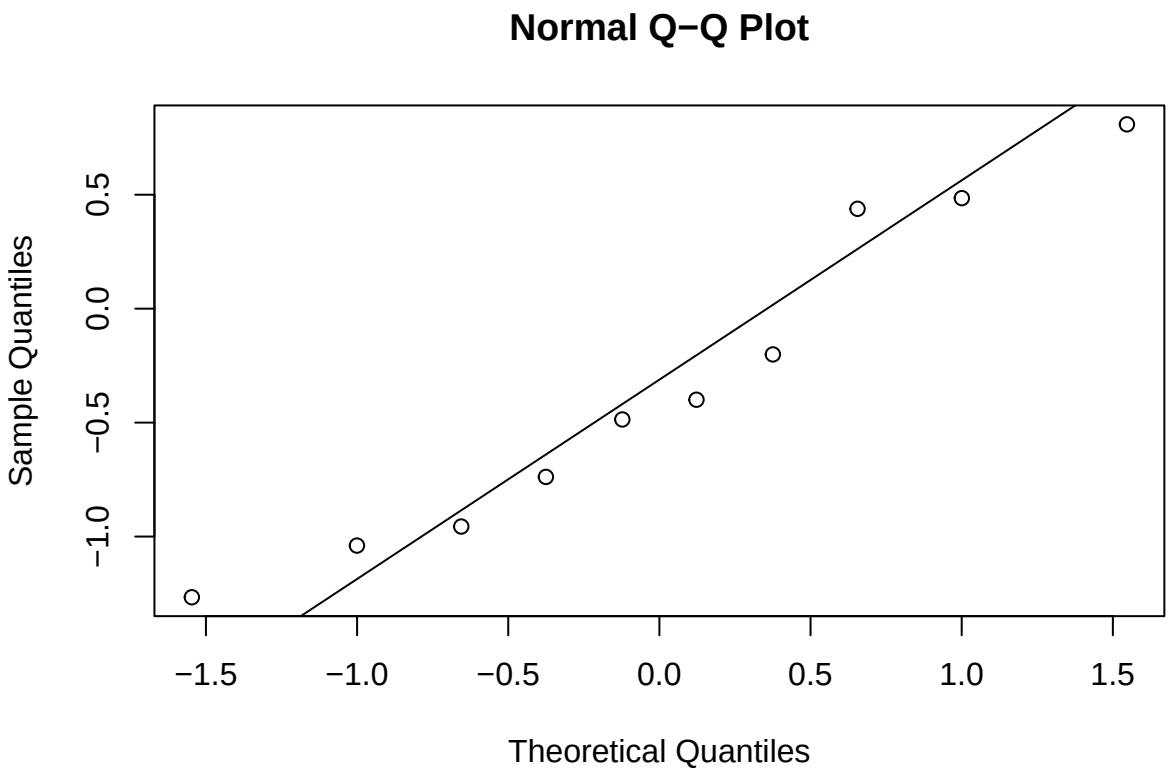




8 Normality tests

8.1 Q-Q plot

```
par(mfrow=c(1, 1))
x <- rnorm(10)
qqnorm(x)
qqline(x)
```



The result is subjective. This is a small dataset. So how do we know if this is Normal?

8.2 Shapiro test

```
# Do a Null Hypothesis test with Shapiro tests.
# Look at the p-value. If p-value > 0.05 then accept the NULL hypothesis.
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.93596, p-value = 0.509
```

Now we can look at the p-value.

If p-value > 0.05, then the distribution is NOT significantly different from a Normal distribution. That is the null hypothesis is possible.