F19 STA 100 A01 Discussion 05

Yishan Huang

2019/10/29

Discussion Time: Tuesday 8:00 – 8:50 am, Haring Hall 1204.

Notes: https://github.com/Hahahuo-13316/sta100-a01-fall19

Office hour: Tuesday 12:00 – 1:00 pm, Mathematical Sciences Building 1117.

Email: yishuang@ucdavis.edu

Announcement: Today we will talk about the answer to quiz 1. If you have any question on your paper, please tell me before we talking about the answer.

Binomial Distribution

For binomial random variable $X \sim \text{Binomial}(n, p)$, we know that $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)}$. We call $p(k) = \mathbb{P}(X = k)$ the probability mass function (pmf) of X, and $F(k) = \mathbb{P}(X \leq k) = p(0) + p(1) + \cdots + p(k)$ be the cumulative distribution function (cdf).

```
dbinom(x = 0:20, size = 20, prob = 0.4)
```

```
##  [1] 3.656158e-05 4.874878e-04 3.087423e-03 1.234969e-02 3.499079e-02
##  [6] 7.464702e-02 1.244117e-01 1.658823e-01 1.797058e-01 1.597385e-01
## [11] 1.171416e-01 7.099488e-02 3.549744e-02 1.456305e-02 4.854351e-03
## [16] 1.294494e-03 2.696862e-04 4.230371e-05 4.700412e-06 3.298535e-07
## [21] 1.099512e-08
```

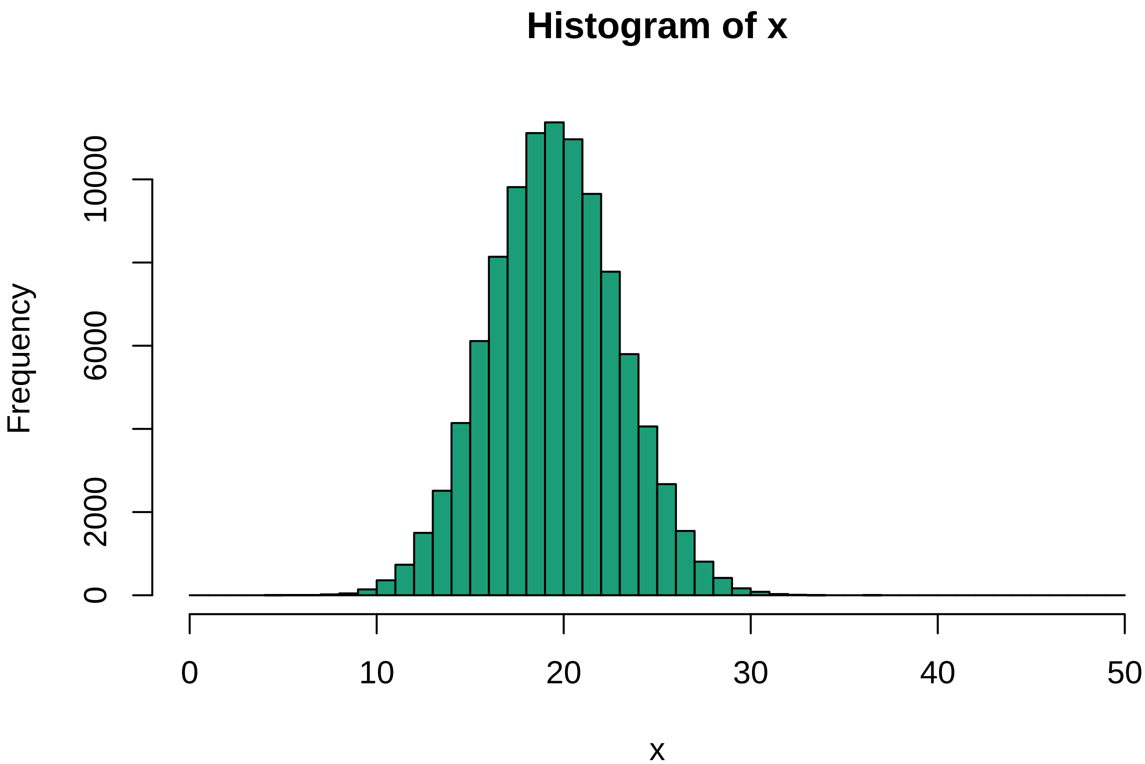
```
pbinom(q = 0:20, size = 20, prob = 0.4)
```

```
##  [1] 3.656158e-05 5.240494e-04 3.611472e-03 1.596116e-02 5.095195e-02
##  [6] 1.255990e-01 2.500107e-01 4.158929e-01 5.955987e-01 7.553372e-01
## [11] 8.724788e-01 9.434736e-01 9.789711e-01 9.935341e-01 9.983885e-01
## [16] 9.996830e-01 9.999527e-01 9.999950e-01 9.999997e-01 1.000000e+00
## [21] 1.000000e+00
```

```
qbinom(p = 0.01, size = 20, prob = 0.4)
```

```
## [1] 3
```

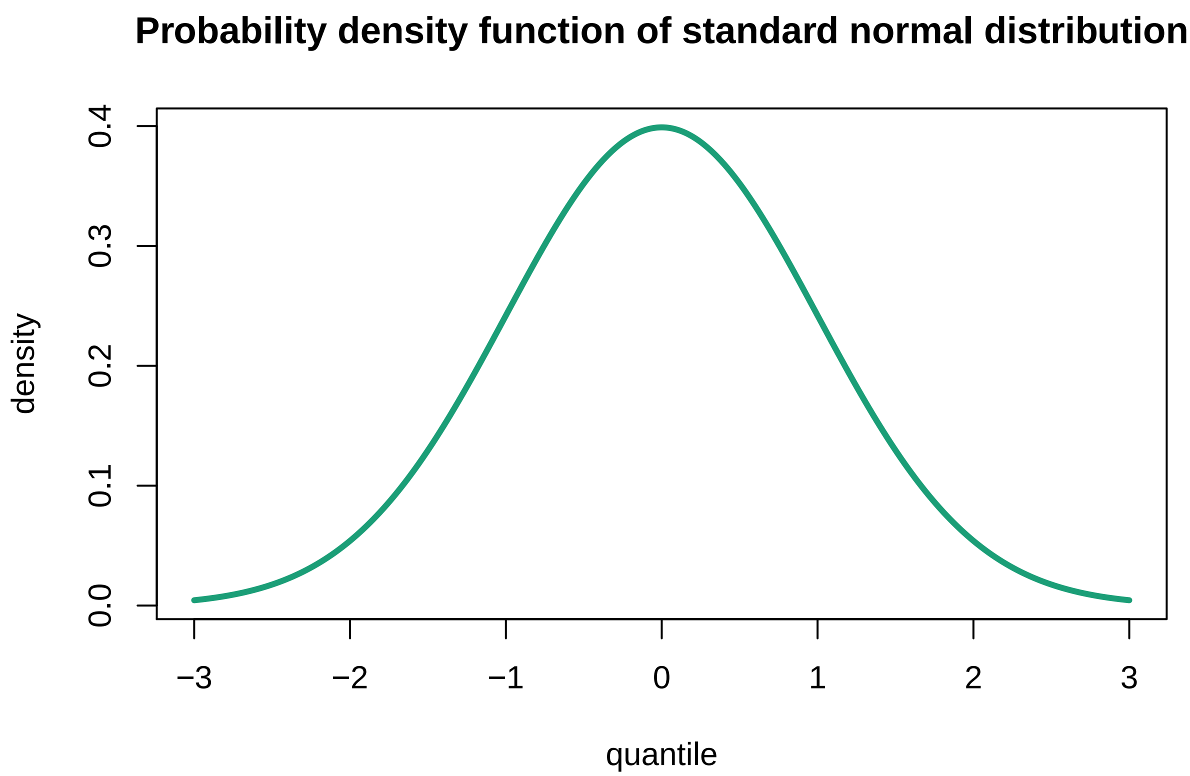
```
library(RColorBrewer)
pal <- brewer.pal(8, "Dark2")
x <- rbinom(100000, size = 50, prob = 0.4)
hist(x, breaks = c(0:50), col = pal[1])
```



Notice that, the quantile function $q(p)$ is the smallest x such that $F(x) \geq p$.

Normal Distribution

```
x1 <- (-300:300)/100
plot(x1, dnorm(x1), col = pal[1], type = "l", lwd = 3, xlab = "quantile", ylab = "density",
     main = "Probability density function of standard normal distribution")
```




```
2 * pnorm(c(1, 2, 3)) - 1
```

```
## [1] 0.6826895 0.9544997 0.9973002
```

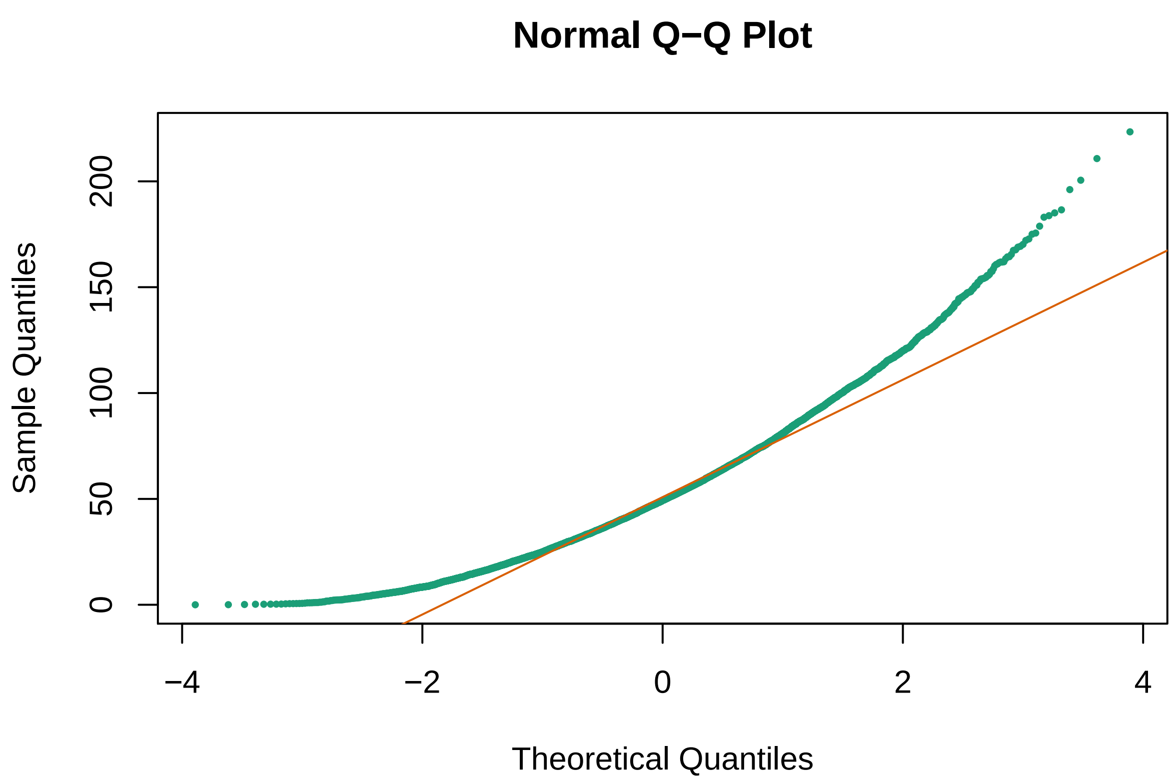
Example: the symmetric interval that covers 90% area of the density of $X = 3Z + 2$ is?

```
qnorm(p = c(0.05, 0.95), mean = 2, sd = 3)
```

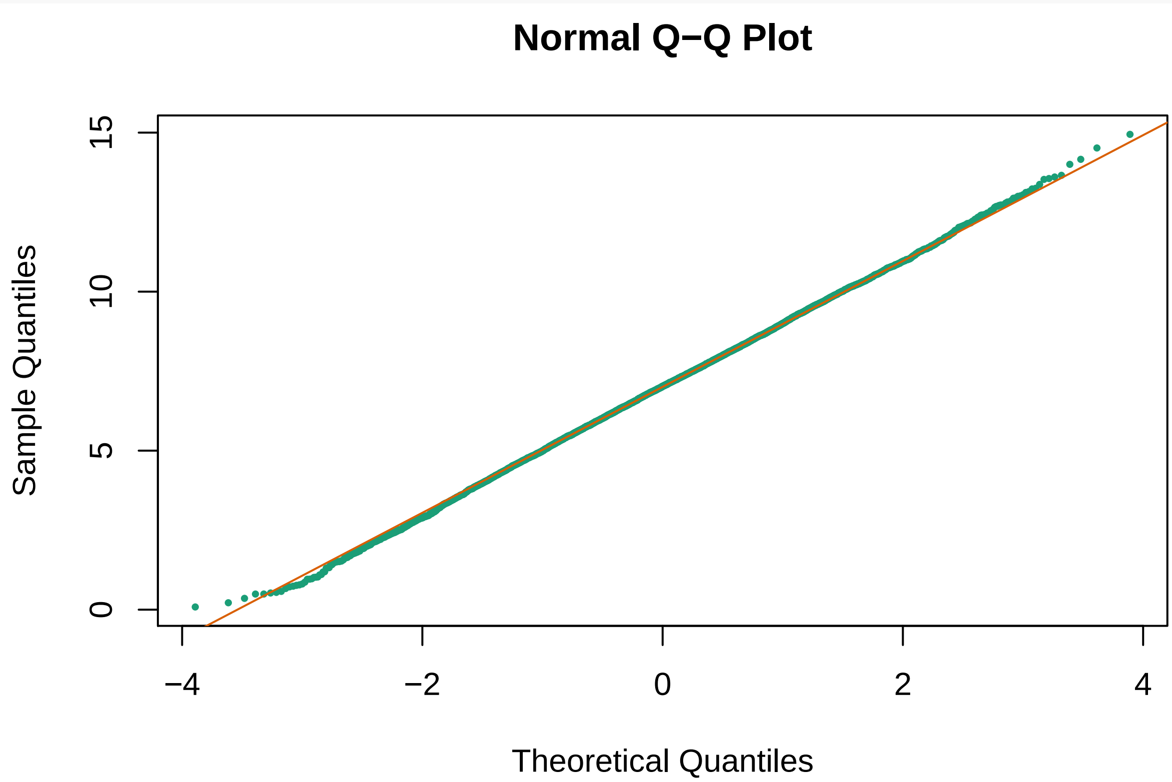
```
## [1] -2.934561  6.934561
```

Q-Q plots

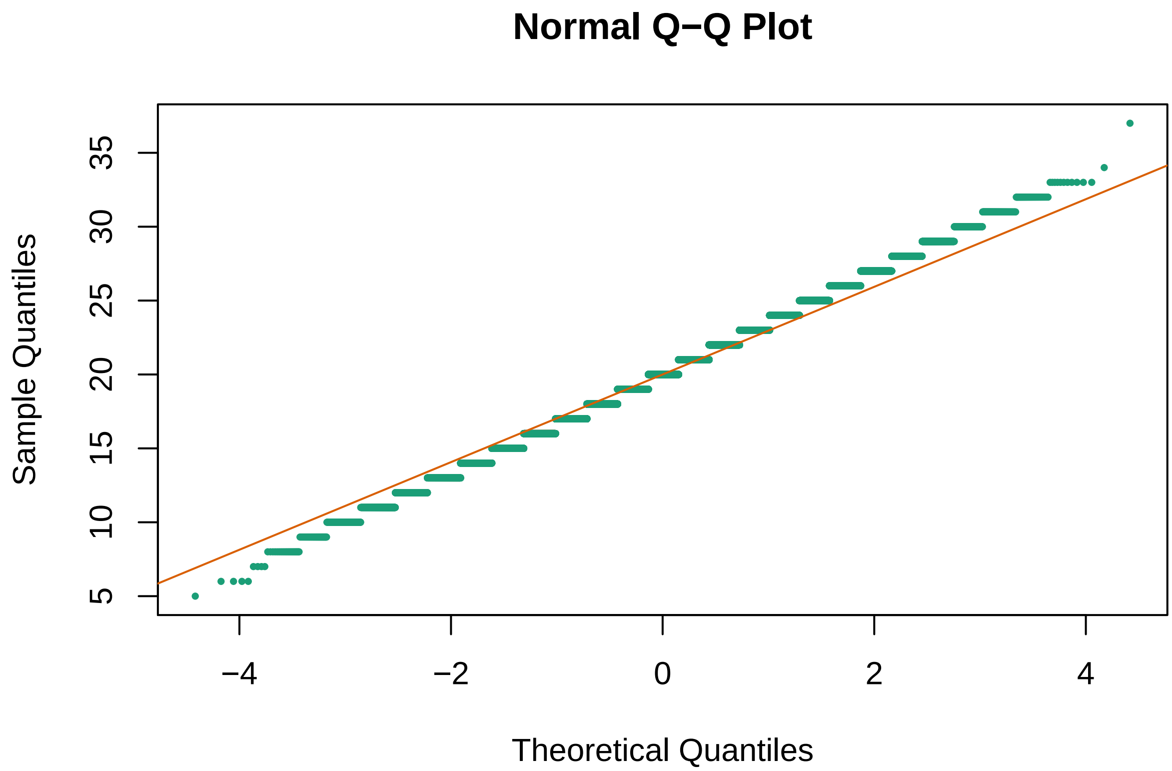
```
x2 <- rnorm(10000, mean = 7, sd = 2)
x0 <- x2^2
qqnorm(x0, pch = 16, cex = 0.5, col = pal[1])
qqline(x0, col = pal[2])
```




```
x3 <- sqrt(x0)
qqnorm(x3, pch = 16, cex = 0.5, col = pal[1])
qqline(x3, col = pal[2])
```



```r
qqnorm(x, pch = 16, cex = 0.5, col = pal[1])
qqline(x, col = pal[2])
```



Problems

(4.3.9) The serum cholesterol levels of 12- to 14-year-olds follow a normal distribution with mean 155 mg/dl and standard deviation 27 mg/dl. What percentage of 12 to 14-year-olds have serum cholesterol values (a) 164 or more? (b) 137 or less? (c) 186 or less? (d) 100 or more? (e) between 159 and 186? (f) between 100 and 132? (g) between 132 and 159?

```r
1 - pnorm(164, mean = 155, sd = 27)
```

```
## [1] 0.3694413
```

```r
pnorm(137, mean = 155, sd = 27)
```

```
## [1] 0.2524925
```

```r
pnorm(186, mean = 155, sd = 27)
```

```
## [1] 0.8745463
```

```r
1 - pnorm(164, mean = 155, sd = 27)
```

```
## [1] 0.3694413
```

```r
pnorm(186, mean = 155, sd = 27) - pnorm(159, mean = 155, sd = 27)
```

```
## [1] 0.3156592
```

```r
pnorm(132, mean = 155, sd = 27) - pnorm(100, mean = 155, sd = 27)
```

```
## [1] 0.176325
```

```r
pnorm(159, mean = 155, sd = 27) - pnorm(132, mean = 155, sd = 27)
```

```
## [1] 0.3617389
```