

Expectation and Variance in R

WEEK 5

Wasinee Jantorn
Sartja Duangchaiyoosook

Exercise

ข้อมูล smoke.csv

	age	sex	bmi	children	smoker	region	charges
1	19	female	27.900	0	yes	southwest	16884.924
2	18	male	33.770	1	no	southeast	1725.552
3	28	male	33.000	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.471
5	32	male	28.880	0	no	northwest	3866.855
6	31	female	25.740	0	no	southeast	3756.622
7	46	female	33.440	1	no	southeast	8240.590
8	37	female	27.740	3	no	northwest	7281.506
9	37	male	29.830	2	no	northeast	6406.411
10	60	female	25.840	0	no	northwest	28923.137

จงหาค่าคาดหมายและค่าความแปรปรวนของตัวแปร **age** และ **charges** จากข้อมูลข้างต้น

R code

```
## read csv data ##
```

```
smokerData = read.csv(file='C:/smoker.csv',sep=',',header=T)
```

```
## create table data ##
```

```
mydata = table(smokerData$age,smokerData$charges)
```

```
## create 2 variables ##
```

```
age = smokerData$age
```

```
charges = smokerData$charges
```

	age	sex	bmi	children	smoker	region	charges
1	19	female	27.900	0	yes	southwest	16884.924
2	18	male	33.770	1	no	southeast	1725.552
3	28	male	33.000	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.471
5	32	male	28.880	0	no	northwest	3866.855
6	31	female	25.740	0	no	southeast	3756.622
7	46	female	33.440	1	no	southeast	8240.590
8	37	female	27.740	3	no	northwest	7281.506
9	37	male	29.830	2	no	northeast	6406.411
10	60	female	25.840	0	no	northwest	28923.137

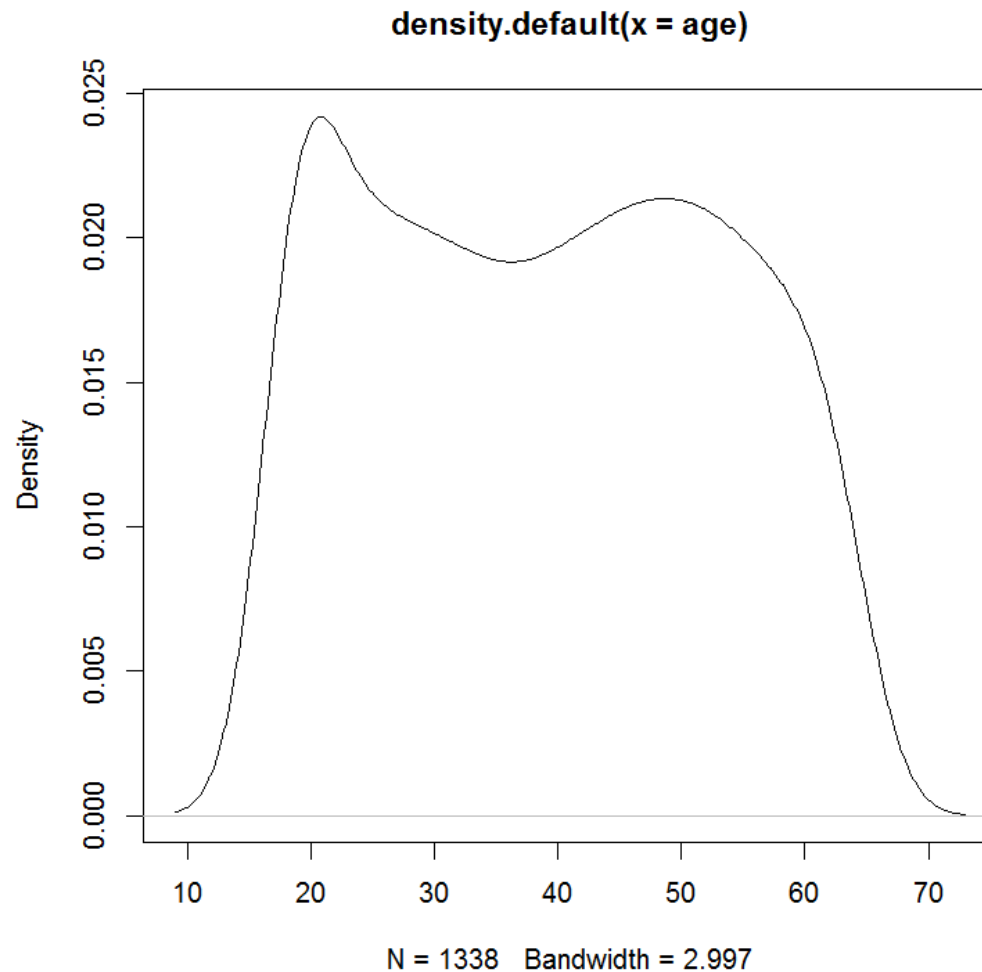


	age	charges
1	19	16884.924
2	18	1725.552
3	28	4449.462
4	33	21984.471
5	32	3866.855
6	31	3756.622
7	46	8240.590
8	37	7281.506
9	37	6406.411
10	60	28923.137

R code

```
## see the distribution ##  
plot(density(age))
```

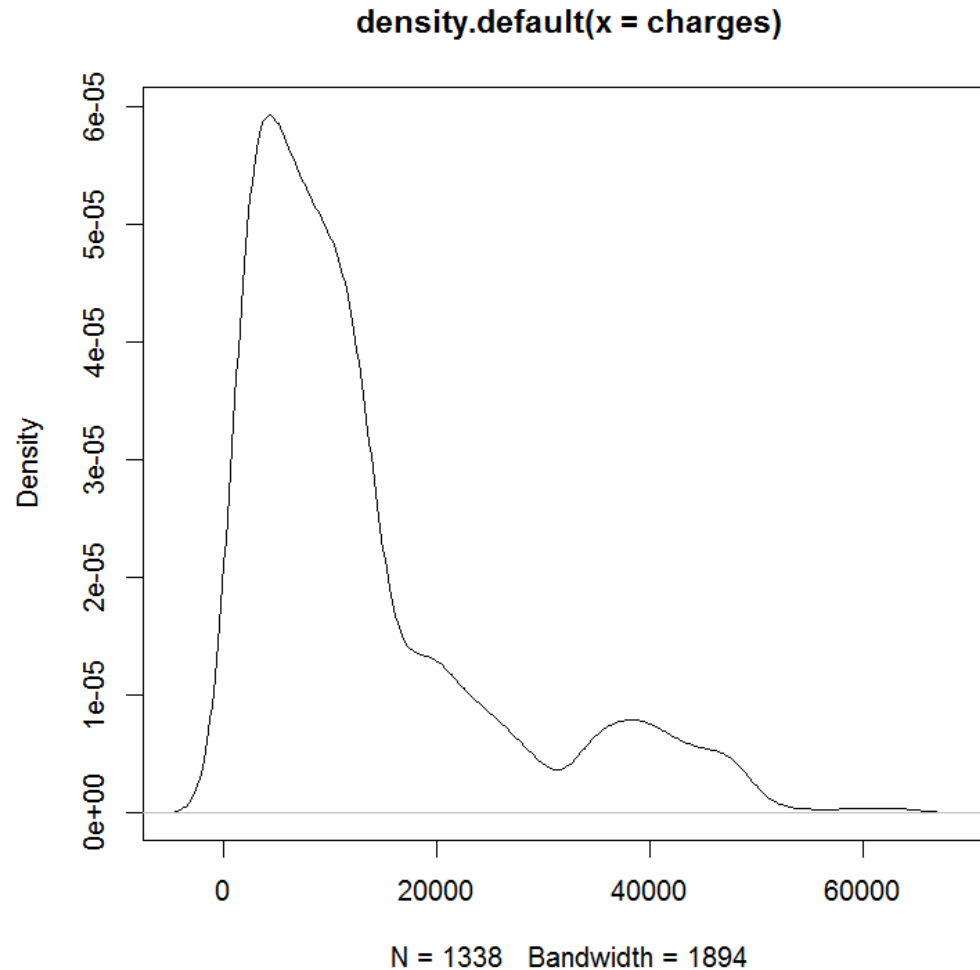
Distribution ?



R code

```
## see the distribution ##  
plot(density(charges))
```

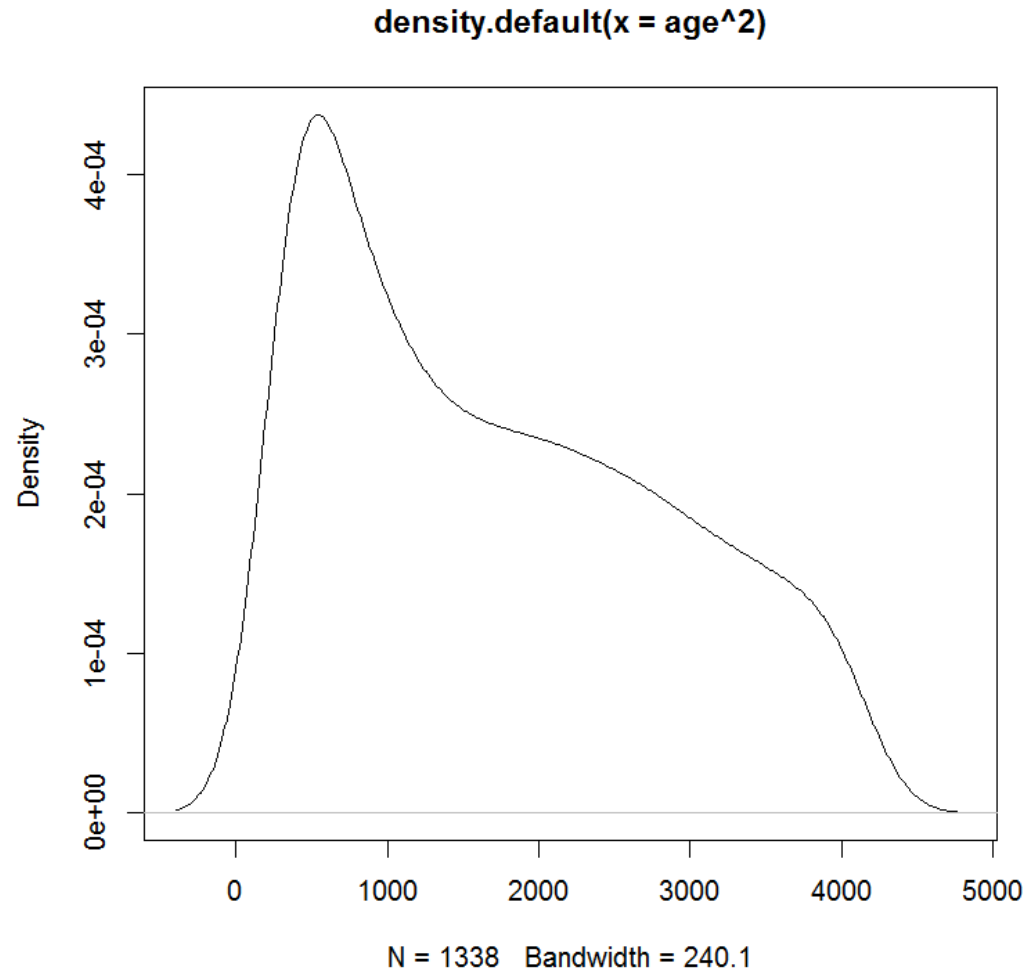
Distribution ?



R code

```
## see the distribution ##  
plot(density(age^2))
```

Distribution ?



R code

```
## normal density function ##
```

```
mean(age)
```

```
sd(age)
```

```
> mean(age)
```

```
[1] 39.20703
```

```
> sd(age)
```

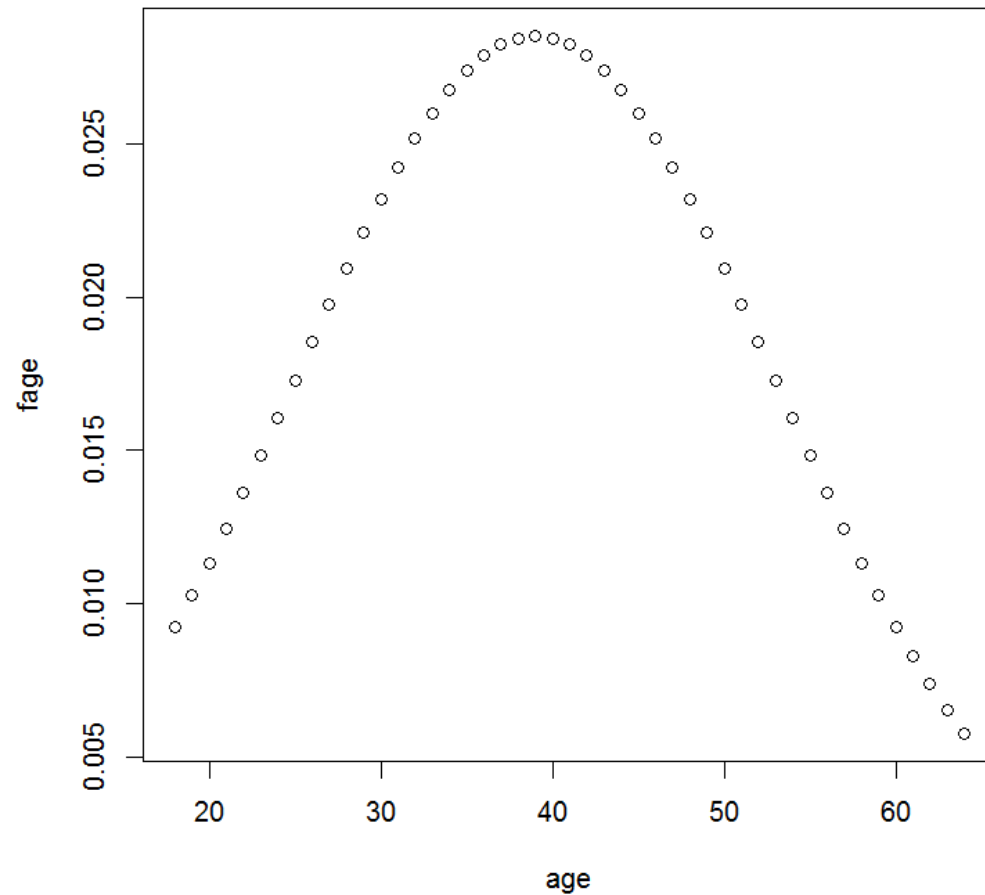
```
[1] 14.04996
```

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

```
fage = (1/sqrt(2*pi*14^2))*exp(-(age-39)^2/(2*14^2))
```

R code

```
## normal density function ##  
plot(age,fage)
```



R code

```
## log-normal density function ##
```

```
lcharges = log(charges)
```

```
mean(lcharges)
```

```
sd(lcharges)
```

```
> mean(lcharges)
```

```
[1] 9.098659
```

```
> sd(lcharges)
```

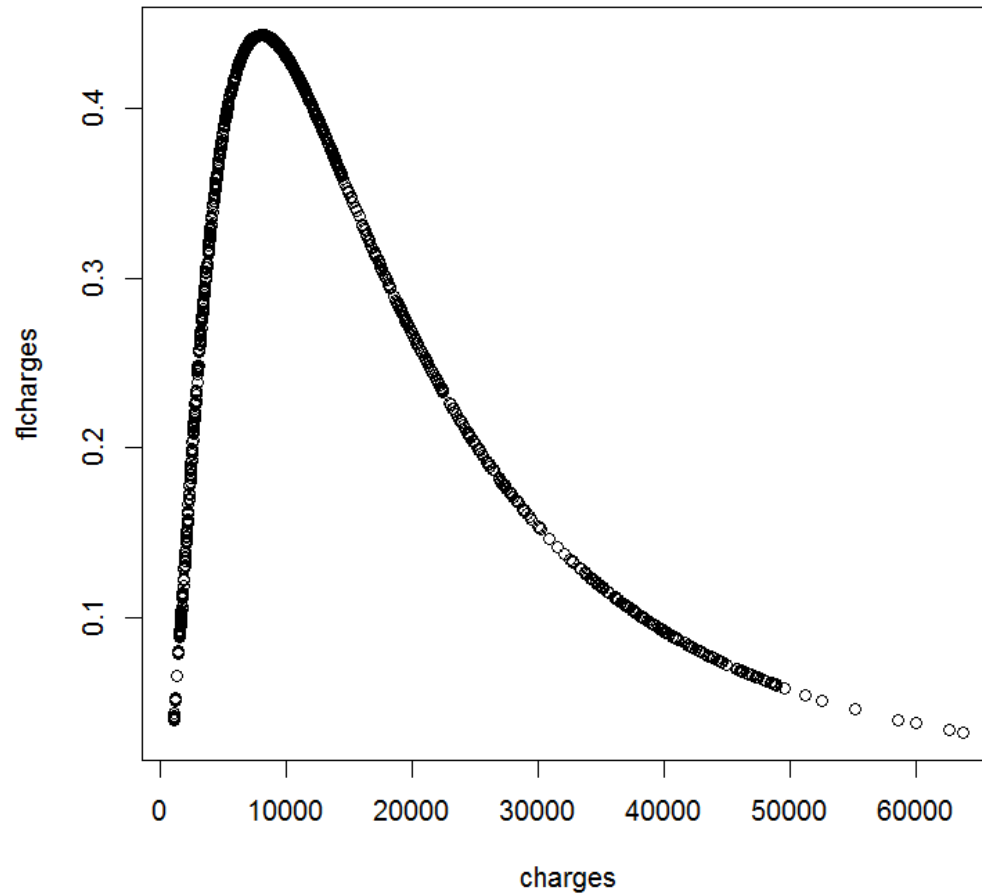
```
[1] 0.9195271
```

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right), \quad x > 0$$

```
flcharges = (1/sqrt(2*pi*0.9^2))*exp(-(lcharges-9)^2/(2*0.9^2))
```

R code

```
## log-normal density function ##  
plot(charges,flcharges)
```



R code

```
## chi-square density function ##
```

```
age2 = age^2
```

$$f(x) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad x > 0$$

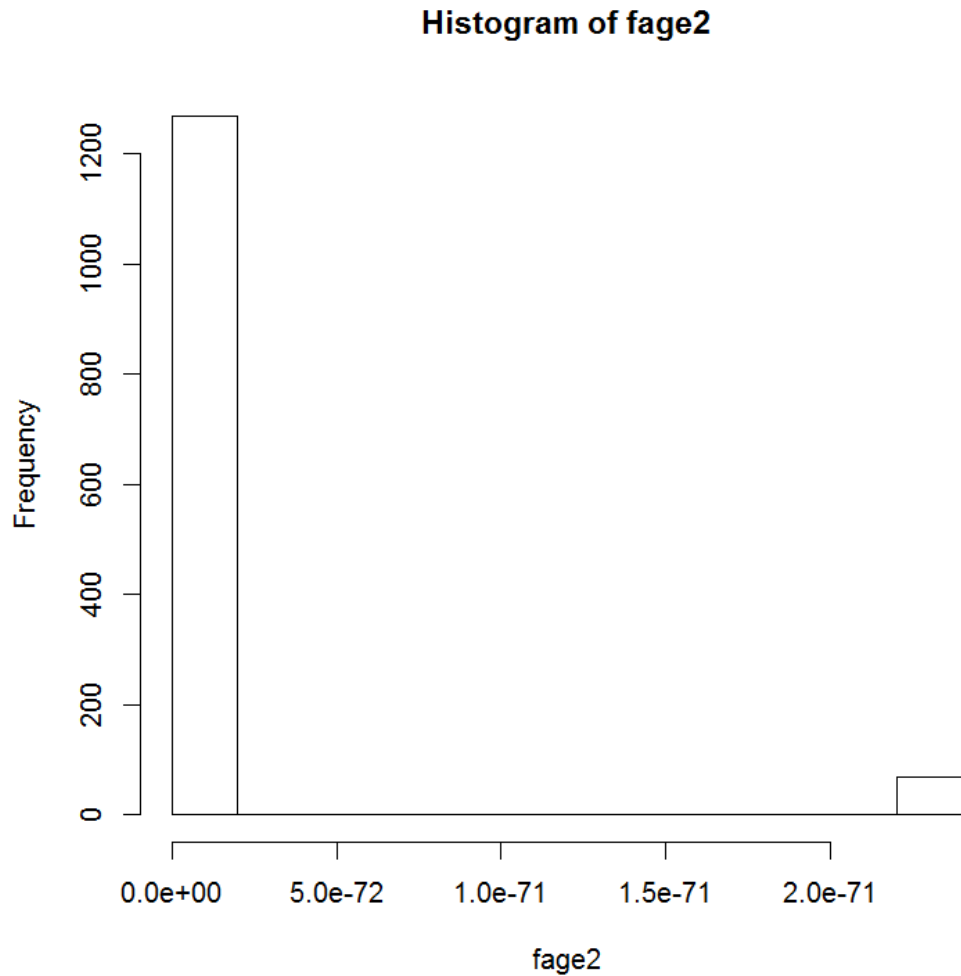
```
#degree of freedom
```

```
n=2
```

```
fage2 = (age2^((n/2)-1)*exp(-age2/2))/(2^(n/2)*gamma(n/2))
```

R code

```
## chi-square density function ##  
hist(fage2)
```



R code

กำหนดให้ X เป็นตัวแปรสุ่ม และ r เป็นฟังก์ชันค่าจริง ถ้า X มีการแจกแจงต่อเนื่อง (continuous distribution) และค่าคาดหวังมีค่าจำกัด แล้ว

$$E[r(X)] = \int_{-\infty}^{\infty} r(x)f(x) dx \quad (7)$$

ในทำนองเดียวกัน ถ้า X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) และค่าคาดหวังมีค่าจำกัด แล้ว

$$E[r(X)] = \sum r(x)f(x) \quad (8)$$

calculate expected value

```
expected_age = sum(age*fage)
```

```
expected_charges=sum(charges*flcharges)
```

R code

กำหนดให้ X เป็นตัวแปรสุ่มที่ค่าคาดหวังมีค่าจำกัดเท่ากับ $\mu = E[X]$ ความแปรปรวน (variance) ของ X นิยามได้เป็น

$$\text{Var}[X] = E[(X - \mu)^2] \quad (22)$$

ในกรณีที่ความแปรปรวนหาค่าได้ เราจะเรียกรากที่สอง (square root) ของความแปรปรวน $\sqrt{\text{Var}[X]}$ ว่า ค่าเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งมักแทนด้วยสัญลักษณ์ σ_X

calculate variance

```
Exage2 = sum(age^2*fage)
```

```
Exagepower2 = sum(age*fage)^2
```

```
variance_age = Exage2 - Exagepower2
```

```
Excharges2 = sum(charges^2*flcharges)
```

```
Exchargespower2 = sum(charges*flcharges)^2
```

```
variance_charges = Excharges2 - Exchargespower2
```

R code

```
## display value ##
```

```
expected_age
```

```
variance_age
```

```
expected_charges
```

```
variance_charges
```

```
> expected_age
```

```
[1] 978.5069
```

```
> variance_age
```

```
[1] -916029.1
```

```
> expected_charges
```

```
[1] 4330554
```

```
> variance_charges
```

```
[1] -1.86827e+13
```

END