



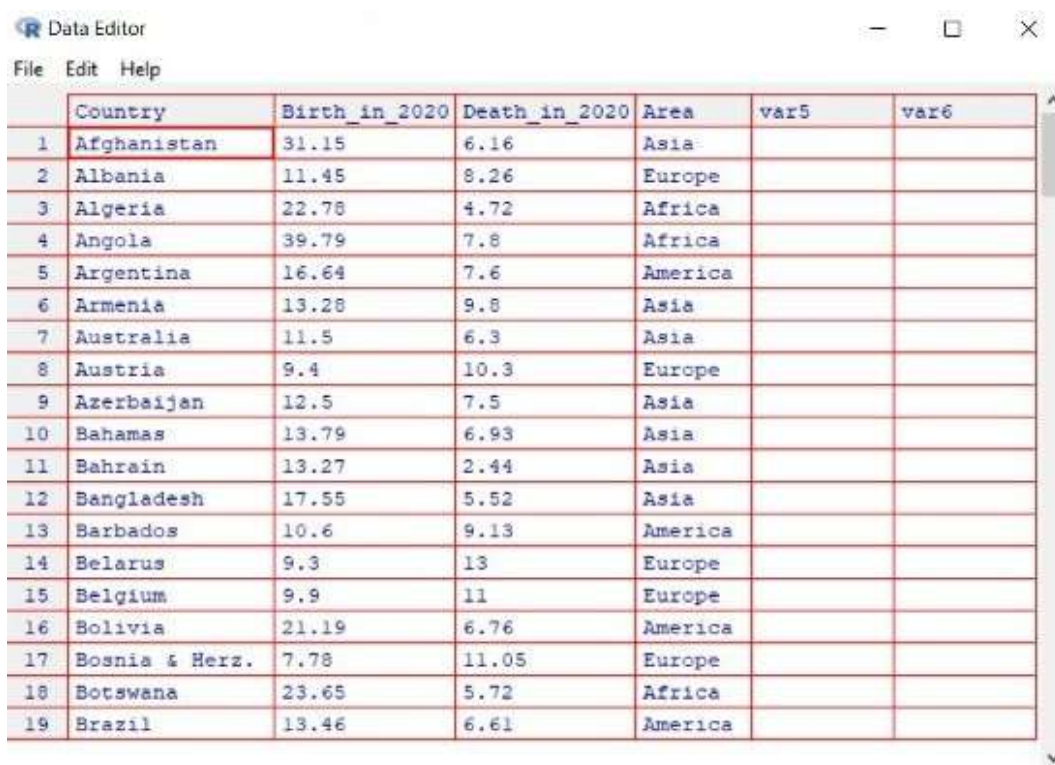
# احتمال و آمار مهندسی

سال ۱۴۰۱

استاد درس: استاد کلکین نما

اعضای گروه: آرش آذرپور (شماره دانشجویی: ۴۰۰۰۳۵۹۳) - علی طاهری (شماره دانشجویی: ۴۰۰۱۲۳۲۳)

ما در این پروژه قصد داریم روی داده های آماری مربوط به میزان زاد و ولد و مرگ و میر را در سطح جهان در سال ۲۰۲۰ در کشورها و قاره های مختلف بررسی کنیم. برای این منظور ۱۵۱ کشور مختلف در جهان در نظر گرفته شده و برای هر کدام دو متغیر پیوسته، **میزان زاد و ولد** و **میزان مرگ و میر**، برای متغیرگسسته هم از تفاوت بین قاره ها استفاده می کنیم.



	Country	Birth_in_2020	Death_in_2020	Area	var5	var6
1	Afghanistan	31.15	6.16	Asia		
2	Albania	11.45	8.26	Europe		
3	Algeria	22.78	4.72	Africa		
4	Angola	39.79	7.8	Africa		
5	Argentina	16.64	7.6	America		
6	Armenia	13.28	9.8	Asia		
7	Australia	11.5	6.3	Asia		
8	Austria	9.4	10.3	Europe		
9	Azerbaijan	12.5	7.5	Asia		
10	Bahamas	13.79	6.93	Asia		
11	Bahrain	13.27	2.44	Asia		
12	Bangladesh	17.55	5.52	Asia		
13	Barbados	10.6	9.13	America		
14	Belarus	9.3	13	Europe		
15	Belgium	9.9	11	Europe		
16	Bolivia	21.19	6.76	America		
17	Bosnia & Herz.	7.78	11.05	Europe		
18	Botswana	23.65	5.72	Africa		
19	Brazil	13.46	6.61	America		

تصویر 1 قسمتی از دیتا های استفاده شده در پروژه

برای وارد کردن دیتاهای مورد نظر در برنامه به ترتیب از دستورهای `read.csv()` و `head()` استفاده می کنیم.

```
#Importing libraries:
library(epiDisplay)
library(modeest)

#Importing data from the dataset:
DataFrame <- read.csv("E:\\R\\DataBase.csv")

head(DataFrame)

fix(DataFrame)
```

تصویر 2 کد مربوطه برای وارد کردن دیتا ها

## بخش اول (آمار توصیفی) :

### ۱- داده های کیفی:

برای بدست آوردن جدول فراوانی برحسب فراوانی، درصد فراوانی و فراوانی تجمعی می توان از دستور `tab1()` که در پکیج `epiDisplay` قرار دارد، استفاده کرد.

	Frequency	Percent	Cum. percent
Asia	54	36.0	36.0
Europe	38	25.3	61.3
Africa	34	22.7	84.0
America	24	16.0	100.0
Total	150	100.0	100.0

```
tab1(DataFrame$Area, sort.group = "decreasing" , cum.percent = T)
```

تصویر 3 کد مربوطه برای بدست آوردن جدول فراوانی

تصویر 4 جدول فراوانی

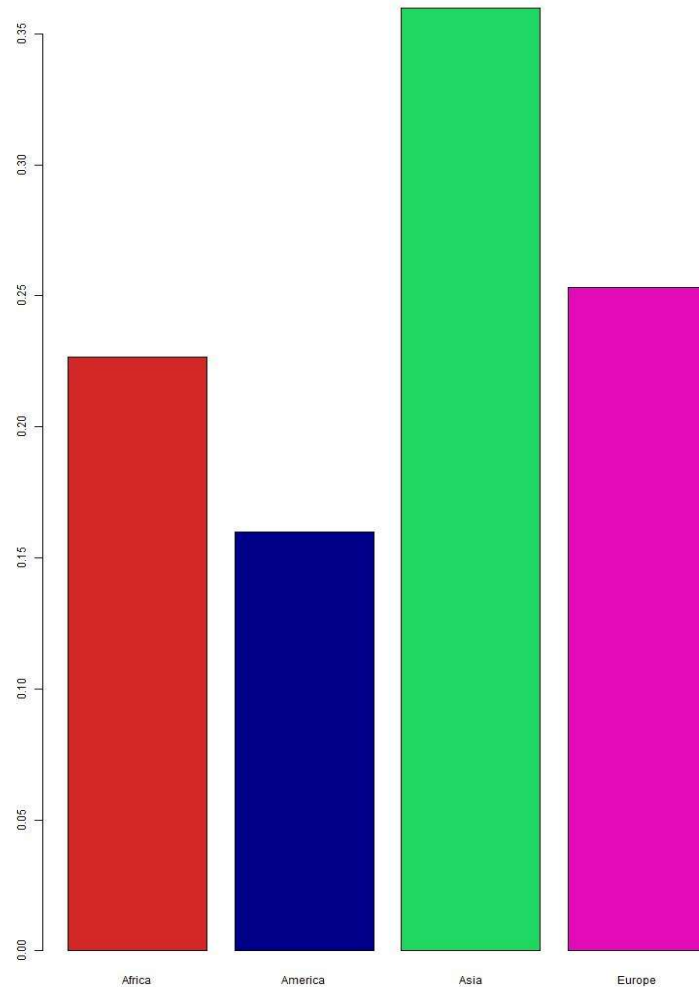
برای بدست آوردن نمودار میله ای برحسب درصد فراوانی نیز از دستور `barplot()` که در کتابخانه `epiDisplay` قرار دارد استفاده می کنیم.

```
Area<- table(DataFrame$Area)

names(Area)<-c("Africa","America","Asia","Europe")

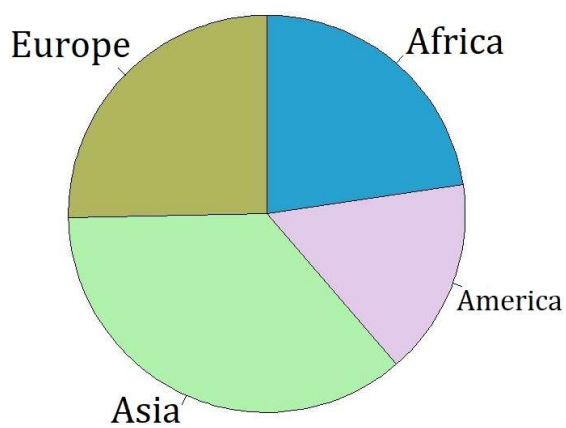
barplot(Area / length(DataFrame$Area),col = c("#d12828", "blue4" , "#1fd762" , "#e30cb8"))
```

تصویر 5 کد مربوطه برای بدست آوردن نمودار میله ای



تصویر 6 نمودار میله ای برحسب درصد فراوانی

برای رسم نمودار دایره ای هم از دستور `pie()` استفاده می کنیم.



تصویر 8 نمودار دایره ای

```
#Pie graph:|
Area<-table(DataFrame$Area)

pie(Area,clockwise = T ,
col = c("#28a1d1", "#8e2eb141" ,
"#25d71f5c" , "#848b00a2"))
```

تصویر 7 کد مربوطه برای رسم نمودار دایره ای

## ۲- داده های پیوسته:

برای بدست آوردن میانگین، میانه و مد به ترتیب از دستورات `mean()` و `median()` و `mfv()` استفاده می کنیم. (لازم به ذکر است برای استفاده از `mfv()` باید از کتابخانه `modeest` استفاده کنیم.)

```
Avg of Birth in 2020 : 18.25147
Avg of Death in 2020 : 7.809133
Median of birth : 16.62
Median of death : 7.05
Mode of Birth : 8.9 9.2 9.4 9.9 10.9
> Mode of Death : 4.71 5.14 5.48 5.98 6.26 6.29 6.3 6.35 7.05 7.3 7.5 10.3 10.4 11.9 12.6
```

تصویر 9 مقادیر میانگین، میانه و مد برای نرخ مرگ و میر و زاد و ولد

```
#Average, Median and Mode:

Avg_of_Birth <- mean(DataFrame$Birth_in_2020)
Avg_of_death <- mean(DataFrame$Death_in_2020)
Median_of_birth <- median(DataFrame$Birth_in_2020)
Median_of_death <- median(DataFrame$Death_in_2020)
Mode_of_birth <- mfv(DataFrame$Birth_in_2020)
Mode_of_death <- mfv(DataFrame$Death_in_2020)

cat("\033[1;33m ",
    "Avg of Birth in 2020 : ",
    Avg_of_Birth, "\n Avg of Death in 2020 : ",
    Avg_of_death, '\n', "Median of birth : ",
    Median_of_birth, "\n Median of death : ",
    Median_of_death, '\n', "Mode of Birth : ",
    Mode_of_birth, "\n Mode of Death : ",
    Mode_of_death,
    "\033[0;0m")
```

تصویر 10 کد مربوطه برای میانگین، میانه و مد

Variable Value	Mean	Median	Mode
<b>Birth Rate</b>	<b>18.25147</b>	<b>16.62</b>	<b>8.9, 9.2, 9.4, 9.9, 10.9</b>
<b>Death Rate</b>	<b>7.809133</b>	<b>7.05</b>	<b>4.71, 5.14, 5.48, 5.98, 6.26, 6.29, 6.3, 6.35, 7.05, 10.3, 10.4, 11.9, 12.6</b>

جدول 1 مقادیر میانگین، میانه و مد برای دیتا های استفاده شده

برای بدست آوردن ماکسیمم، مینیمم، دامنه، انحراف معیار و واریانس به ترتیب از دستورات `max()` و `min()` و `range()` و `sd()` و `var()` استفاده می کنیم.

```
# Standard deviation and Variance:
Standard_dev_Birth <- sd(DataFrame$Birth_in_2020)

Standard_dev_Death <- sd(DataFrame$Death_in_2020)

Variance_Birth <- var(DataFrame$Birth_in_2020)

Variance_Death <- var(DataFrame$Death_in_2020)

cat("\033[1;35m " ,
    "\n Variance of Birth : ",Variance_Birth ,
    "\n Variance of Death : ",Variance_Death ,
    "\nStandard deviation of Birth : ",Standard_dev_Birth,
    "\n Standard deviation Death : " , Standard_dev_Death,
    "\033[0;0m")
```

تصویر 15 کد مربوطه برای بدست آوردن واریانس و انحراف معیار

```
# Domain:

Domain_of_Birth <- range(DataFrame$Birth_in_2020)

Domain_of_Death <- range(DataFrame$Death_in_2020)

cat("\033[1;31m " ,
    "\n Birth in range " ,
    Domain_of_Birth[1],
    "\t to \t",Domain_of_Birth[2],
    "\n","Death in range " ,
    Domain_of_Death[1],
    "\t to \t",Domain_of_Death[2],
    "\n","\033[0;0m")
```

تصویر 13 کد مربوطه برای بدست آوردن دامنه

```
Maximum of Birth : 41.41
Maximum of Death : 18
Minimum of Birth : 5.3
> Minimum of Death : 1.29

Variance of Birth : 86.14908
Variance of Death : 9.994702
Standard deviation of Birth : 9.281653
> tandard deviation Death : 3.16144

Birth in range 5.3 to 41.41
Death in range 1.29 to 18
```

تصویر 17 خروجی مربوط به ماکسیمم، مینیموم، انحراف معیار، واریانس و دامنه

```
#Min and Max:

Minimum_of_Birth <- min(DataFrame$Birth_in_2020)

Minimum_of_Death <- min(DataFrame$Death_in_2020)

Maximum_of_Birth <- max(DataFrame$Birth_in_2020)

Maximum_of_Death <- max(DataFrame$Death_in_2020)

cat("\033[1;36m " ,
    "\n Maximum of Birth : " , Maximum_of_Birth,
    "\n Maximum of Death : " , Maximum_of_Death,
    "\n Minimum of Birth : " , Minimum_of_Birth,
    "\n Minimum of Death : ",Minimum_of_Death ,
    "\033[0;0m")
```

تصویر 11 کد مربوطه برای بدست آوردن ماکسیمم و مینیمم

Variable Value	Min	Max	Standard deviation	Variance	Range
Birth Rate	5.3	41.41	9.281653	86.14908	5.3 to 41.41
Death Rate	1.29	18	3.16144	9.994702	1.29 to 18

جدول 2 مقادیر مربوط به ماکسیمم، مینیموم، انحراف معیار، واریانس و دامنه

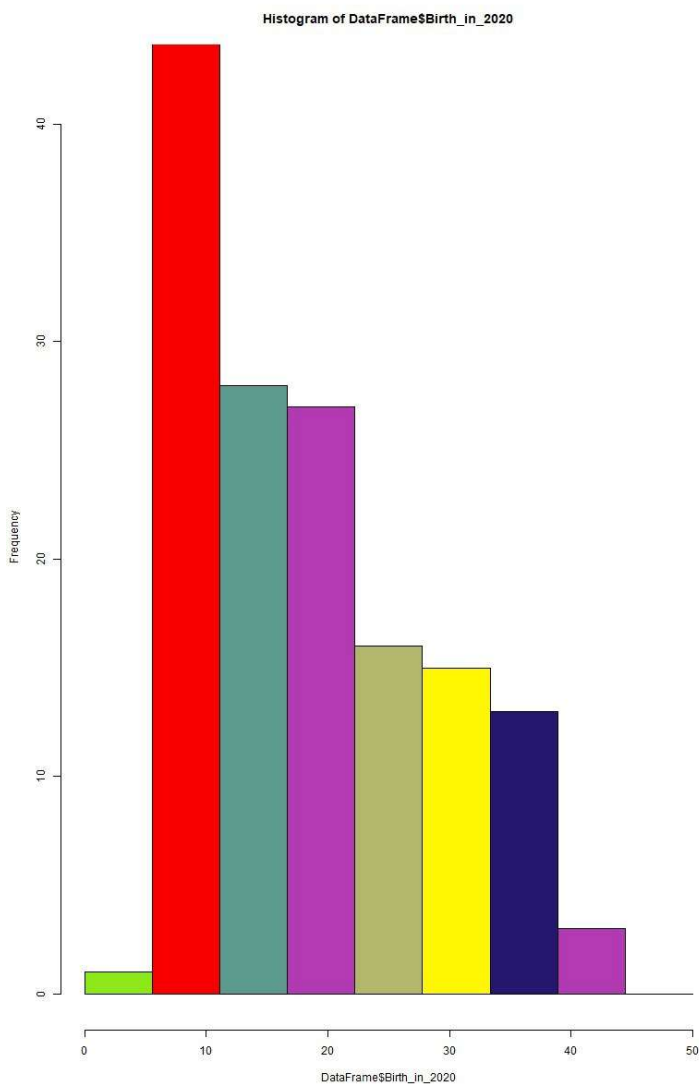


برای رسم نمودار مستطیلی از دستور `hist()` استفاده می کنیم.

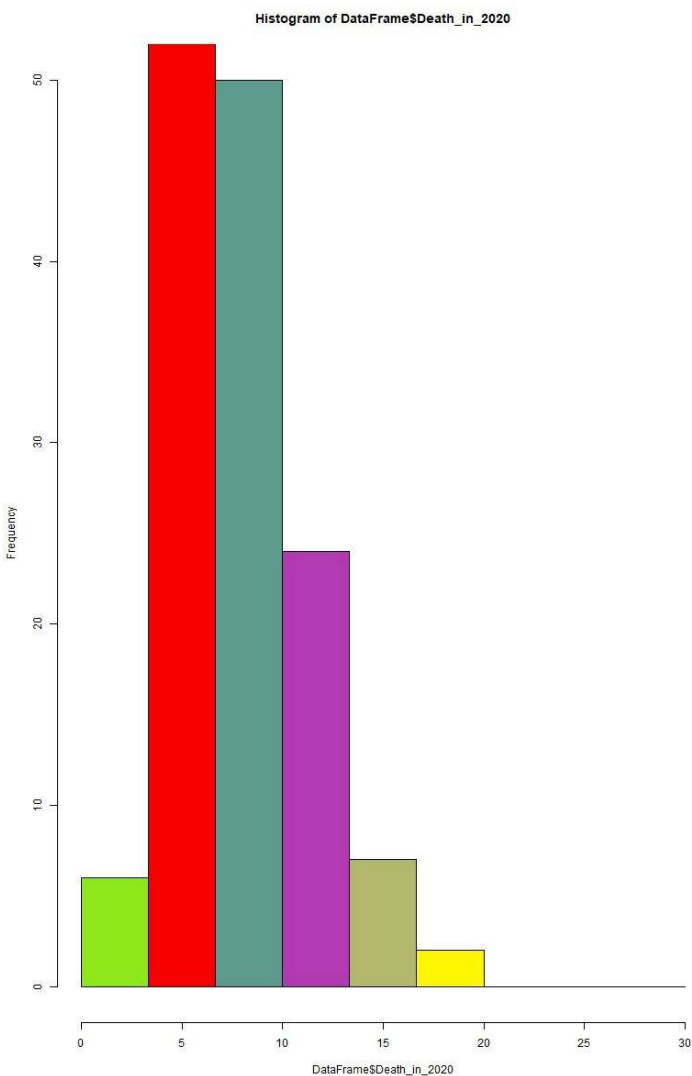
```
# Histograms:
hist(DataFrame$Birth_in_2020,breaks = seq(0,50,l = 10),
     xlim = c(0,50),ylim = c(0,42), col = c("#8ee71b", "#fa0000", "#156d5bb0", "#b23ab2",
     "#8189109c","#fff700", "#24156dfc", "#b23ab2", "#00e5ff", "#11ff00"))

hist(DataFrame$Death_in_2020,breaks = seq(0,30,l = 10),
     xlim = c(0,30),ylim = c(0,50), col = c("#8ee71b", "#fa0000", "#156d5bb0", "#b23ab2",
     "#8189109c","#fff700", "#24156dfc", "#b23ab2", "#00e5ff", "#11ff00"))
```

تصویر 20 کد مربوطه برای رسم نمودار مستطیلی

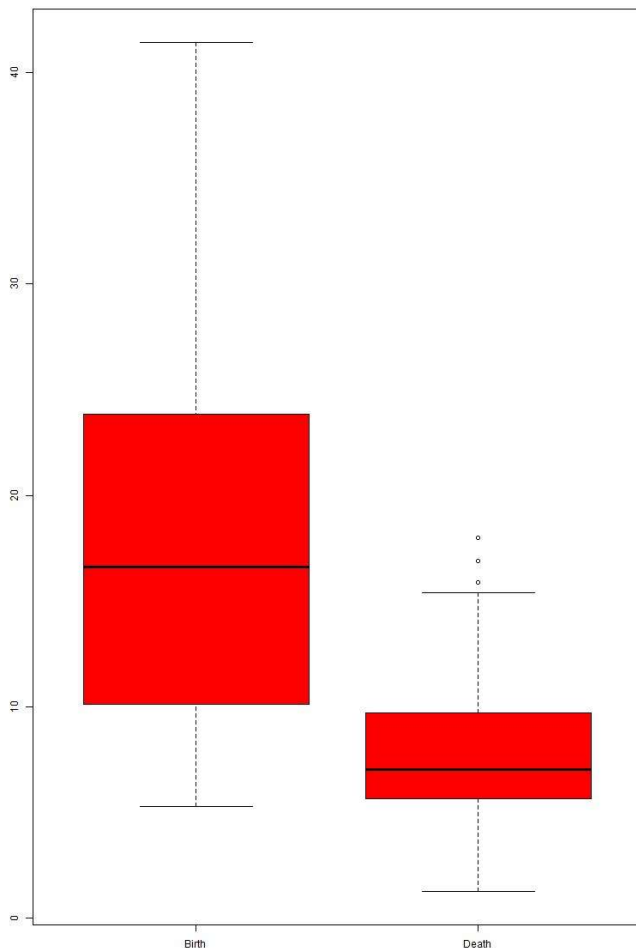


تصویر 22 نمودار مستطیلی برای متغیر Birth rate



تصویر 21 نمودار مستطیلی برای متغیر Death rate

برای رسم نمودار جعبه ای نیز از دستور `boxplot()` استفاده می کنیم.



*#Box Plot:*

```
boxplot(DataFrame$Birth_in_2020,  
        DataFrame$Death_in_2020,  
        names = c("Birth","Death"),  
        col = c("#ff0000"))
```

تصویر 23 کد مربوط به رسم نمودار جعبه ای

تصویر 24 نمودار جعبه ای برای متغیرهای (سمت راست: Death) و (سمت چپ: Birth)

برای بدست آوردن کواریانس و ضریب همبستگی به ترتیب از دستورهای `cov()` و `cor()` استفاده می کنیم.

```
# Covariance and Correlation coefficient:  
Covariance <- cov(DataFrame$Birth_in_2020,DataFrame$Death_in_2020)  
cat(" Covariance is : ",Covariance)  
CorrelationCoefficient <- cor(DataFrame$Birth_in_2020,DataFrame$Death_in_2020)  
cat("\n CorrelationCoefficient is : ",CorrelationCoefficient)
```

تصویر 25 کد مربوطه به کواریانس و ضریب همبستگی

```
> Covariance is : -9.096461
> CorrelationCoefficient is : -0.3100004
```

تصویر 26 مقادیر کواریانس و ضریب همبستگی

Covariance	Correlation Coefficient
-9.096461	-0.3100004

## بخش دوم :

۱- فاصله اطمینان ۹۵ درصد برای میانگین برای متغیر Birth rate :

فرمول فاصله اطمینان  $100(1-\alpha)\%$  برای میانگین بصورت کلی بصورت رو به رو می باشد :

$$\mu \in (\bar{x} - t_{1-\alpha/2}(n-1) s/\sqrt{n}, \bar{x} + t_{1-\alpha/2}(n-1) s/\sqrt{n})$$

در این روش ما با استفاده از دستورهایی چون `mean()`، `sd()` و... دقیقاً فرمول را بازسازی می کنیم.

```
Min_Birth <- mean(DataFrame$Birth_in_2020)
Counter <- length(DataFrame$Birth_in_2020)
Standard_dev_of_Birth <- sd(DataFrame$Birth_in_2020)
standardError <- Standard_dev_of_Birth/sqrt(Counter)

Alpha <- 0.05
degrees.freedom <- Counter - 1
distribution_t <- qt(p=1-(Alpha/2), df=degrees.freedom, lower.tail=TRUE)

ERROR <- distribution_t * standardError
cat("\tμ ∈ (", Min_Birth - ERROR,
    ",", Min_Birth + ERROR, ")")
```

```
> μ ∈ ( 16.75396 , 19.74898 )
```

تصویر 28 مقدار بازه اطمینان ۹۵ درصد برای میانگین

تصویر 27 کد مربوطه به بازه اطمینان ۹۵ درصد برای میانگین

پس از اجرا مشاهده می کنیم:  $\mu \in (16.75396, 19.74898)$



## ۲- فاصله اطمینان ۹۵ درصد برای واریانس برای متغیر Birth rate :

فرمول فاصله اطمینان  $100(1-\alpha)\%$  برای واریانس بصورت کلی بصورت زیر می باشد :

$$\sigma^2 \in \frac{(n-1)s^2}{x_{1-\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)s^2}{x_{\frac{\alpha}{2}}^2(n-1)}$$

مانند `var()`، `length()` و ... بازسازی می کنیم.

```
Alpha <- 0.05
Counter <- length(DataFrame$Birth_in_2020)
degrees.freedom <- Counter - 1
Variance <- var(DataFrame$Birth_in_2020)
Lower_value <- qchisq(1-(Alpha/2), degrees.freedom ,lower.tail = TRUE)
Upper_value <- qchisq(Alpha/2, degrees.freedom ,lower.tail = TRUE)
cat("\to^2 ∈ (",
    (degrees.freedom*Variance)/Lower_value,
    ", ", (degrees.freedom*Variance)/Upper_value,
    ")\n")
```

```
> 0^2 ∈ ( 69.50255 , 109.6194 )
```

تصویر 30 فاصله اطمینان ۹۵ درصد برای واریانس

تصویر 29 کد مربوطه به محاسبه فاصله اطمینان ۹۵ درصد برای واریانس

## ۳- آزمون فرض میانگین جامعه برای متغیر Death rate :

ما در این قسمت می خواهیم آزمون  $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$  را در سطح معنی داری ۰,۰۵ انجام دهیم. برای این کار از فرمول اصلی آزمون فرض که در کتاب هم در جدولی در فصل هشتم به آن اشاره شده است استفاده خواهیم کرد و با استفاده از دستور هایی مانند `length()`، `mean()` و ... عین فرمول را بازسازی می کنیم.

$\mu = \mu_0$ یا $\mu \leq \mu_0$	$\mu > \mu_0$	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ ، $\sigma$ نامعلوم	$T > t_{1-\alpha}(n-1)$
-----------------------------------	---------------	---	-------------------------

تصویر 31 فرمول مربوط به آزمون فرض برای میانگین جامعه

```
# H0 : u=u0
# H1 : u>u0

Alpha = 0.5

Average <- mean(DataFrame$Death_in_2020)

mu0 <- Average+10

Number <- length(DataFrame$Death_in_2020)

SD <- sd(DataFrame$Death_in_2020)

SE <- SD / sqrt(Number)

t_stat <- (Average - mu0) / SE

degree_of_freedom <- Number - 1

t_score <- qt(p=1-(Alpha), df=degree_of_freedom, lower.tail=TRUE)

if(t_stat >= t_score)
  cat("\033[1;32m ", "t_stat >= t_score : reject H0", "\033[0m")

if(t_stat < t_score)
  cat("\033[1;32m ", "t_stat < t_score : confirm H0", "\033[0m")
```

```
> t_stat < t_score : confirm H0
```

تصویر 33 خروجی آزمون فرض

تصویر 32 کد مربوط به آزمون فرض برای میانگین جامعه

در کد مربوطه  $t\_stat$  همان مقدار  $T$  و  $t\_score$  همان مقدار توزیع  $t$  با درجه آزادی  $n-1$  در فرمول صفحه قبل است. با توجه به خروجی چون  $T$  کوچکتر از  $t$  است پس فرض  $H_0$  در سطح معنی داری  $0,05$  تایید می شود.

### نکاتی درمورد پروژه:

- ۱- داده های مورد استفاده از سایت [https://www.theglobaleconomy.com/rankings/Death\\_rate/](https://www.theglobaleconomy.com/rankings/Death_rate/) و سایت [https://www.theglobaleconomy.com/rankings/birth\\_rate/](https://www.theglobaleconomy.com/rankings/birth_rate/) می باشند. درمورد مقادیر birth rate و death rate برابر با تعداد مرگ و میر یا تعداد زاد و ولد به ازای هر ۱۰۰۰ نفر است.