ا**حتمال و آمار مهندسی**



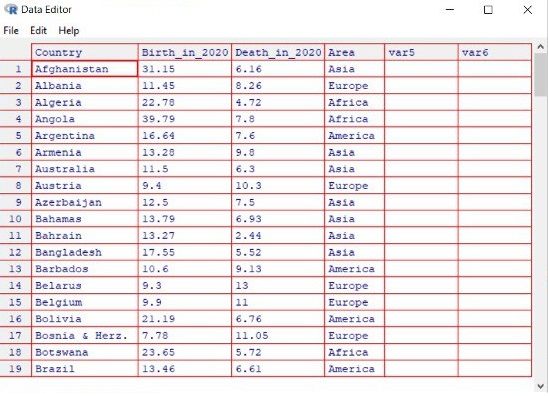


**سال 1401**

استاد درس: استاد کلکین نما

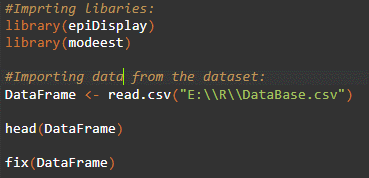
اعضای گروه: آرش آذرپور (شماره دانشجویی: 40003593) - علی طاهری (شماره دانشجویی: 40012323)

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 2020 در کشورها و قاره های مختلف بررسی کنیم. برای این منظور 151 کشور مختلف در جهان در نظر گرفته شده و برای هر کدام دو متغییر پیوسته، **میزان زاد و ولد** و **میزان مرگ و میر،** برای متغیرگسسته هم از تفاوت بین قاره ها استفاده می کنیم.



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.



تصویر 2 کد مربوطه برای وارد کردن دیتا ها

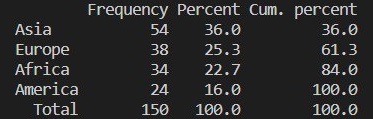
**بخش اول (آمار توصیفی) :**

**1- 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.



تصویر 3 کد مربوطه برای بدست آوردن جدول فراوانی

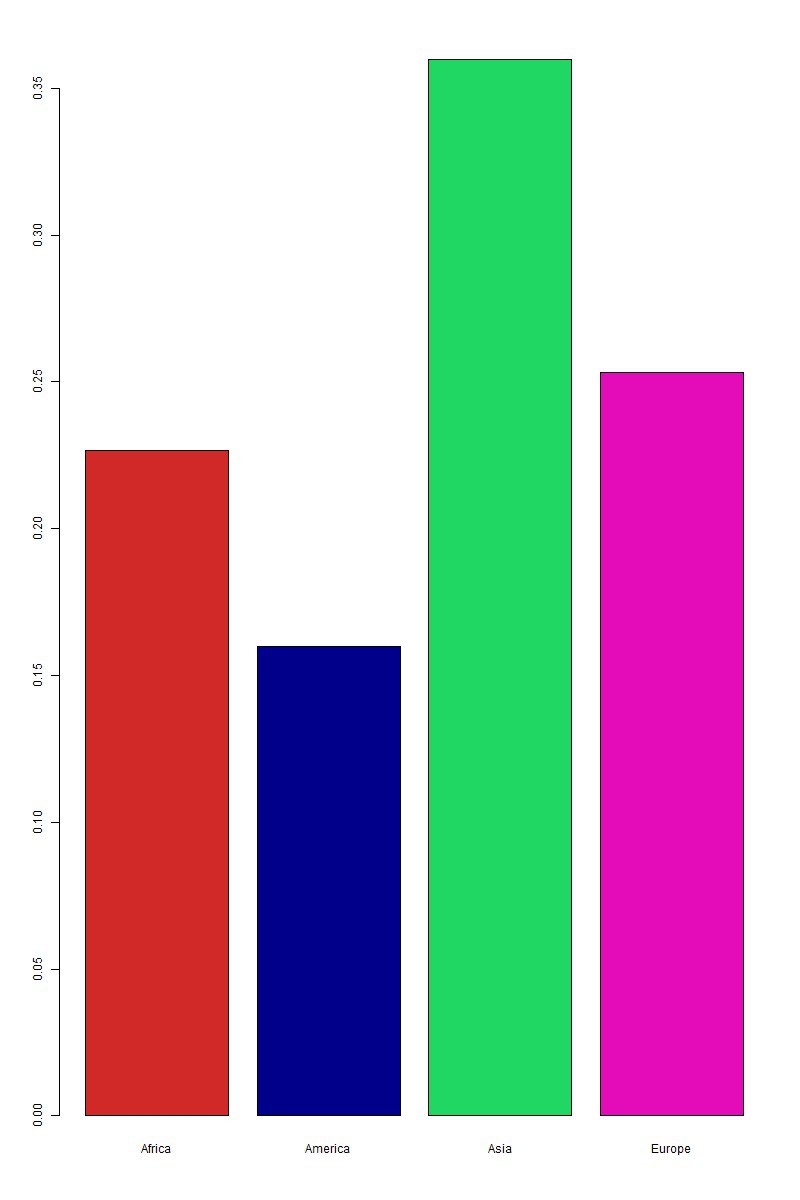


تصویر 4 جدول فراوانی

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

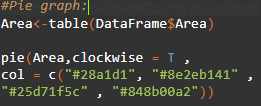


تصویر 5 کد مربوطه برای بدست آوردن نمودار میله ای

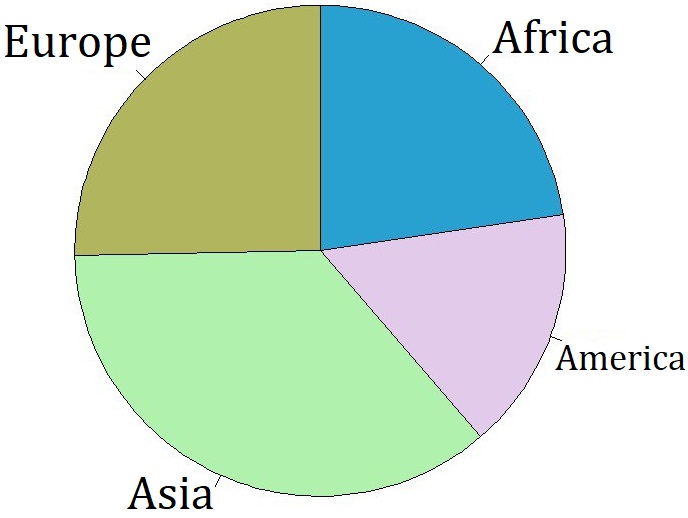


تصویر 6 نمودار میله ای برحسب درصد فراوانی

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



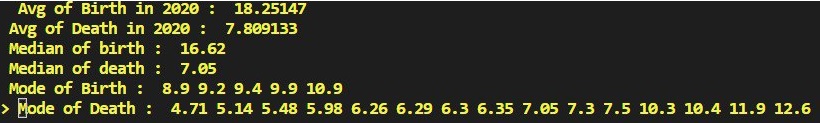
تصویر 7 کد مربوطه برای رسم نمودار دایره ای



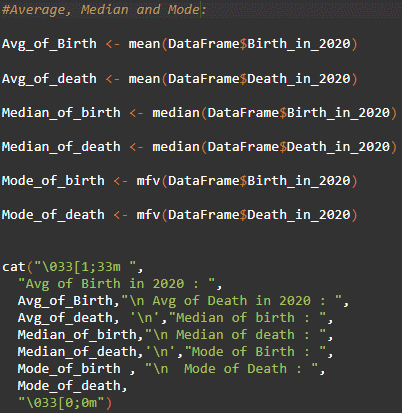
تصویر 8 نمودار دایره ای

2- داده های پیوسته:

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



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



تصویر 10کد مربوطه برای میانگین، میانه و مد

**Median**

**Mean**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

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

**Mode**

**Variable**

**Value**

**Birth Rate**

**8.9, 9.2, 9.4, 9.9, 10.9**

**16.62**

**….**

**18.25147**

**….**

**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**

**Death Rate**

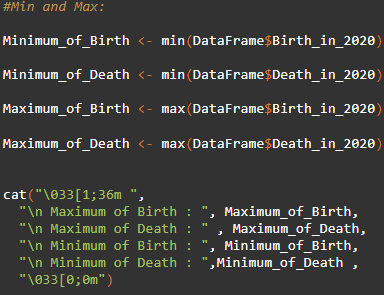
**7.05**

**….**

**7.809133**

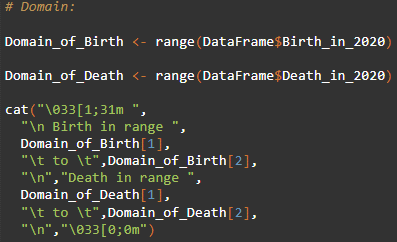
**….**

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.



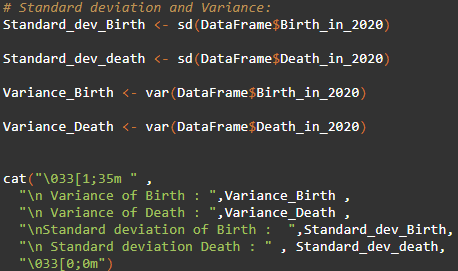
تصویر 11 کد مربوطه برای بدست آوردن ماکسیمم و مینیمم

تصویر 123 کد مربوطه برای بدست آوردن ماکسیمم و مینیمم



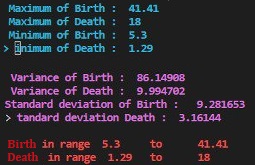
تصویر 13 کد مربوطه برای بدست آوردن دامنه

تصویر 141کد مربوطه برای بدست آوردن دامنه



تصویر 15کد مربوطه برای بدست آوردن واریانس و انحراف معیار

تصویر 162 کد مربوطه برای بدست آوردن واریانس و انحراف معیار



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

تصویر 184 خروجی مربوط به ماکسیمم، مینیموم، انحراف معیار، واریانس و دامنه

**Variable**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

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

تصویر 19مقادیر مربوط به ماکسیمم، مینیموم، انحراف معیار، واریانس و دامنه

**Range**

**Variance**

**Standard**

**deviationd**

**deviation**

**Max**

**Min**

**Value**

**to**

**5.3**

**86.14908**

**9.281653**

**41.41**

**5.3**

**Birth Rate**

**1.29**

**41.41**

**to**

**9.994702**

**3.16144**

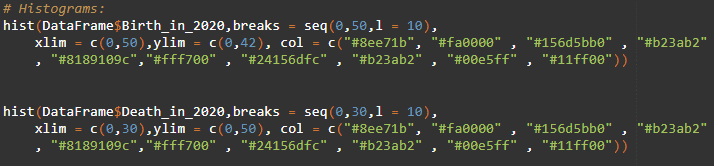
**18**

**1.29**

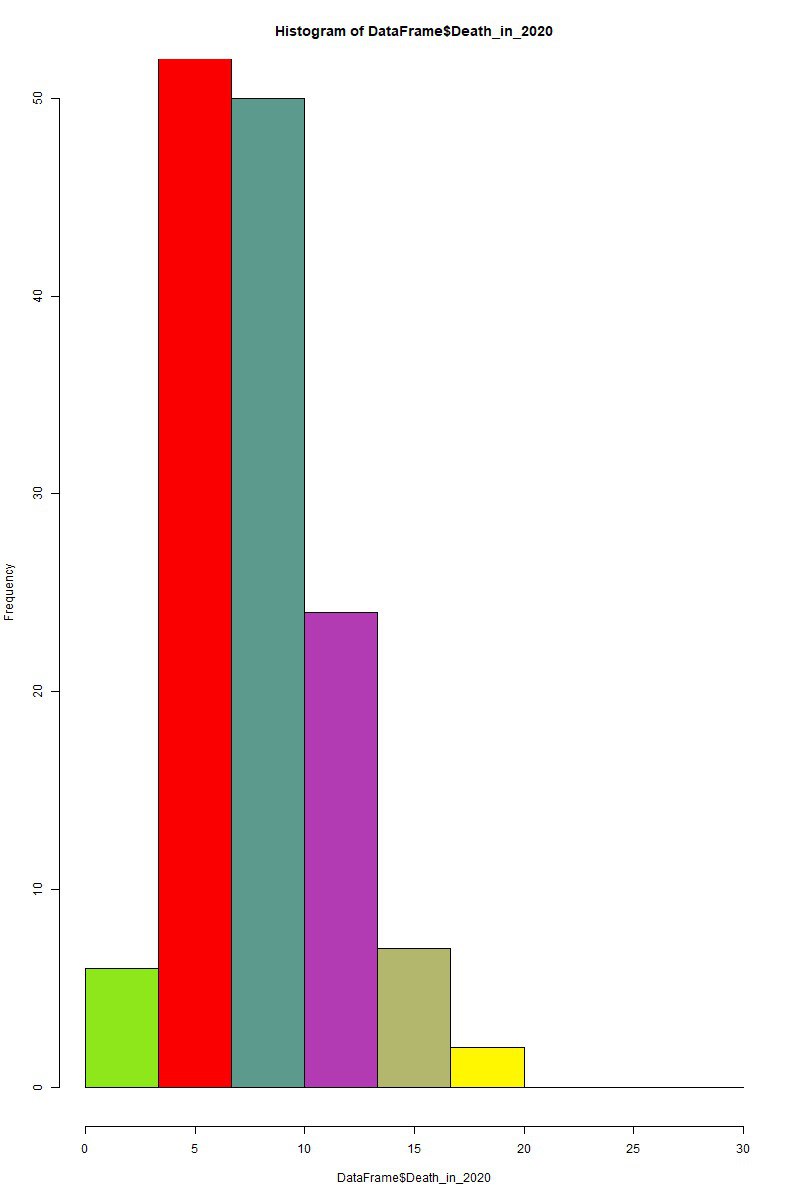
**Death Rate**

**18**

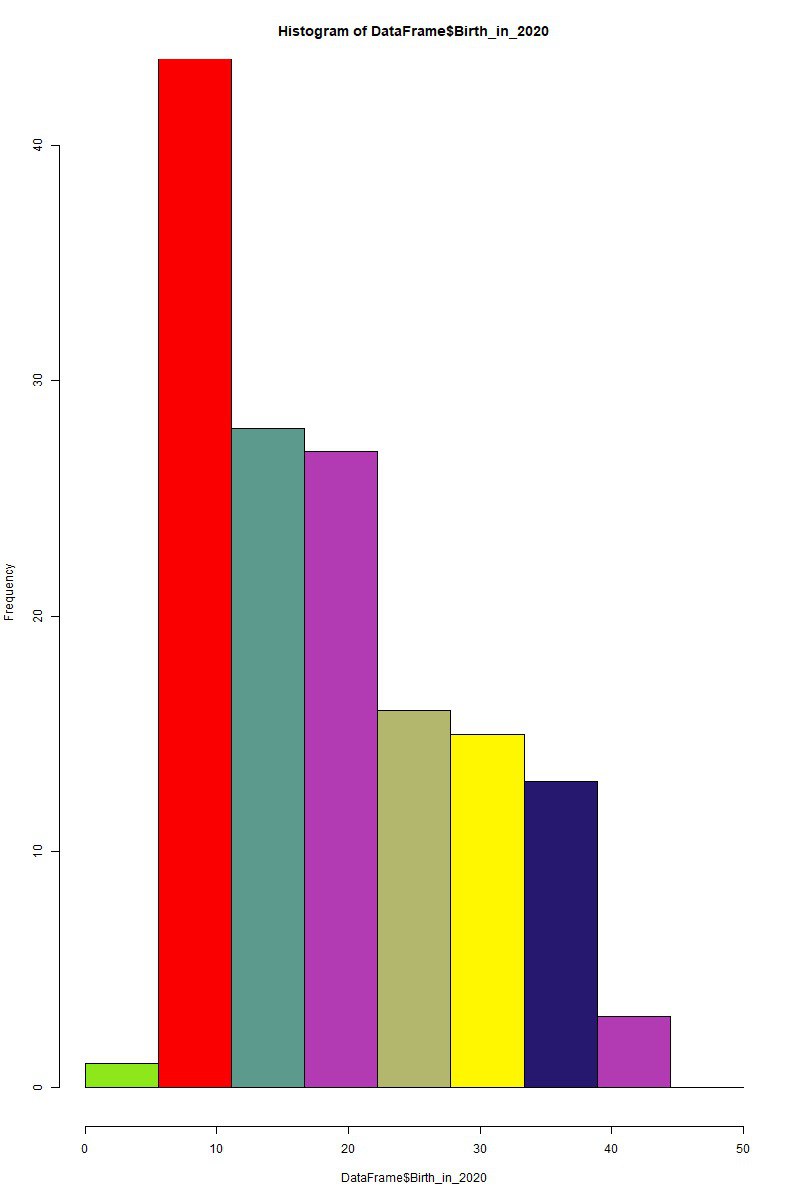
برای رسم نمودار مستطیلی از دستور hist() استفاده می کنیم.



تصویر 20 کد مربوطه برای رسم نمودار مستطیلی

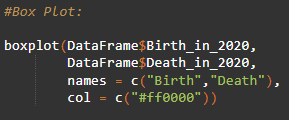


تصویر 21 نمودار مستطیلی برای متغیر Death rate

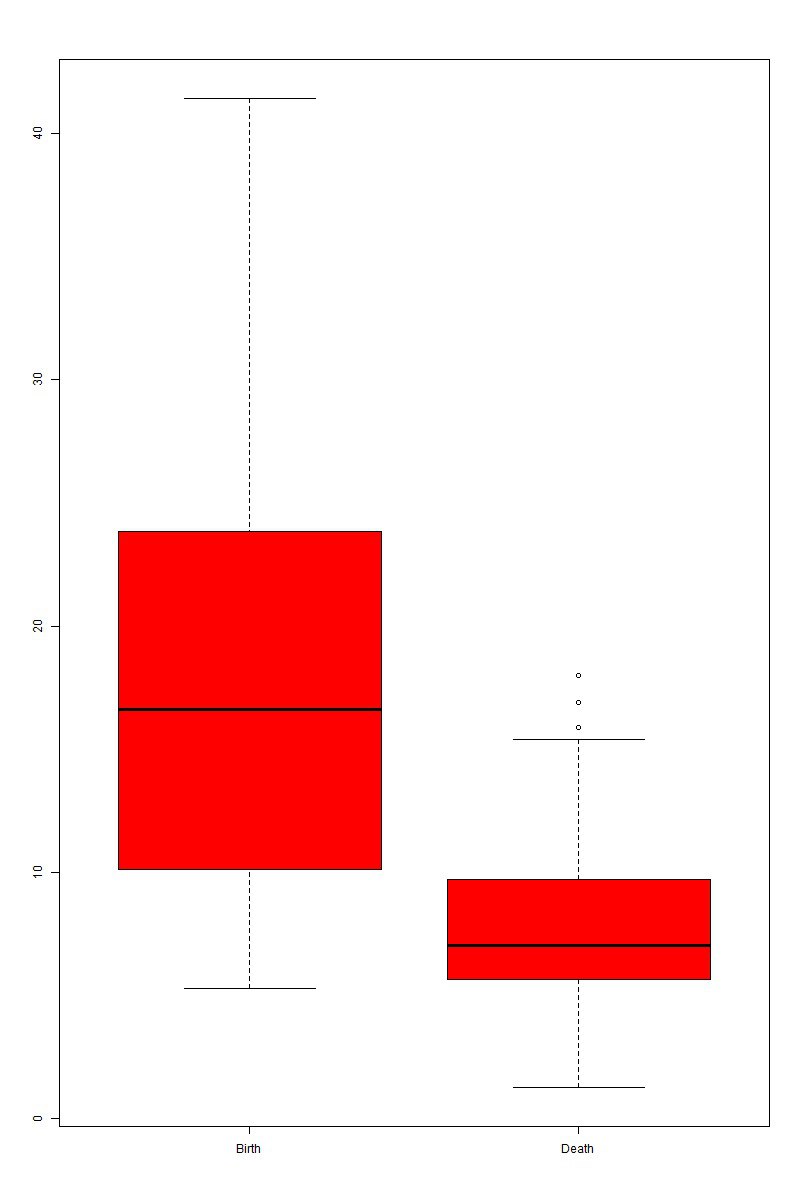


تصویر 22نمودار مستطیلی برای متغیر Birth rate

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

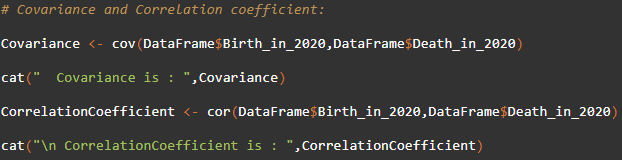


تصویر 23کد مربوط به رسم نمودار جعبه ای



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

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



تصویر 25 کد مربوطه به کواریانس و ضریب همبستگی



تصویر 26مقادیر کواریانس و ضریب همبستگی

**Covariance**

**Correlation Coefficient**

|  |  |
| --- | --- |
|  |  |
|  |  |

**-0.3100004**

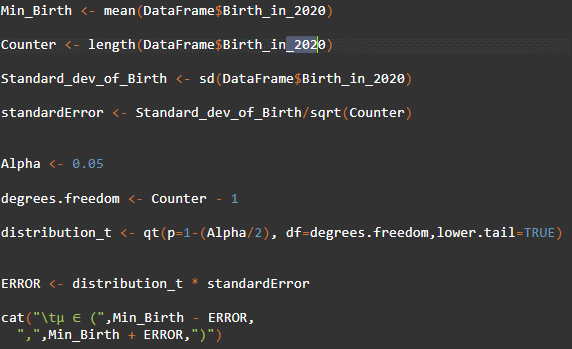
**-9.096461**

**بخش دوم :**

1- فاصله اطمینان 95درصد برای میانگین برای متغیر Birth rate :

فرمول فاصله اطمینان %100(1-𝛼) برای میانگین بصورت کلی بصورت رو به رو می باشد : 𝝁∈(𝒙̅−𝒕𝟏−𝜶𝟐(𝒏−𝟏)𝒔√𝒏 ,𝒙̅+𝒕𝟏−𝜶𝟐(𝒏−𝟏)𝒔√𝒏)

دراین روش ما با استفاده از دستورهایی چون mean()، sd() و... دقیقا فرمول را بازسازی می کنیم.



تصویر 27 کد مربوطه به بازه اطمینان 95 درصد برای میانگین



تصویر 28 مقدار بازه اطمینان 95درصدبرای میانگین

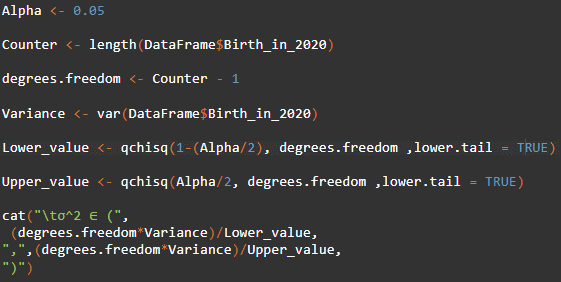
پس از اجرا مشاهده می کنیم:

2- فاصله اطمینان 95درصد برای واریانس برای متغیر Birth rate :

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

برای این قسمت مانند قسمت قبل فرمول کلی را با دستور هایی

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



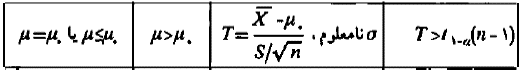
تصویر 29کد مربوطه به محاسبه فاصله اطمینان 95درصد برای واریانس



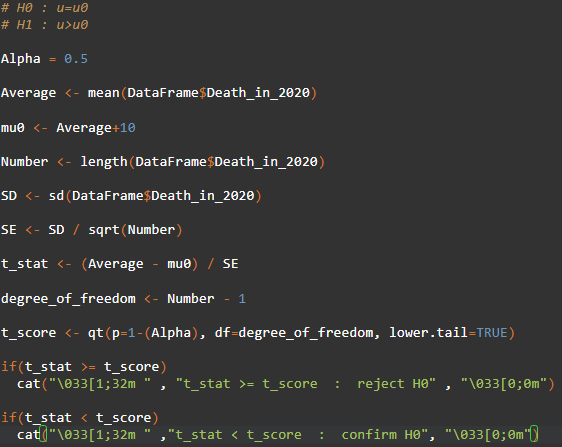
تصویر 30فاصله اطمینان 95درصد برای واریانس

3- آزمون فرض میانگین جامعه برای متغیر Death rate :

ما در این قسمت می خواهیم آزمون را در سطح معنی داری 0.05 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 mean()، length() و ... عین فرمول را بازسازی می کنیم.



تصویر 31فرمول مربوط به آزمون فرض برای میانگین جامعه



تصویر 32 کد مربوط به آزمون فرض برای میانگین جامعه



تصویر 33 خروجی آزمون فرض

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 H0 در سطح معنی داری 0.05 تایید می شود.

**نکاتی درمورد پروژه:**

**1-** داده های مورد استفاده از سایت و سایت

<https://www.theglobaleconomy.com/rankings/Death_rate/>

می باشند. درمورد مقادیر death rate و birth rate برابر با تعداد مرگ و میر یا تعداد زاد و ولد به ازای هر 1000 نفر است.

<https://www.theglobaleconomy.com/rankings/birth_rate/>