

به نام خدا

نرگس غلامی ۸۱۰۱۹۸۴۴۷

درس آمار و احتمال - پروژه دوم 😊

استاد راهنما: استاد بهرک



بخش اول:

۲- اگر ورودی تابعی مقادیر NA داشته باشد با `na.rm = TRUE` می‌توانیم آن را در محاسبه در نظر نگیریم.

یک راه حل دیگر `(na.omit)` است.

همچنین می‌توانیم داده‌های گمشده را با میانگین و یا میانه پر کنیم.

۳- ماتریس همبستگی را در پایین مشاهده می‌کنید.

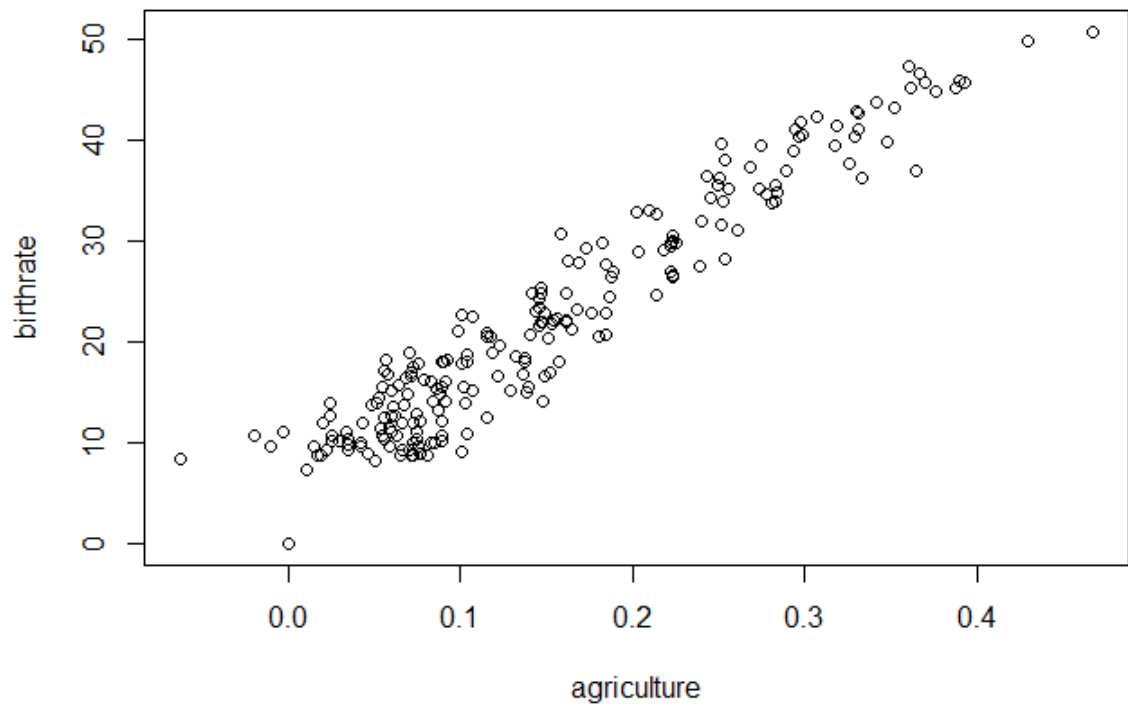
	population	area	density	coastline	net_migration	infant_mortality	GDP_	literacy	phones	arable	crops
population	1.000000e+00	0.469882964	-0.02840257	-0.0678291600	2.787508e-05	0.025073091	-0.03887470	0.024091831	-0.02667799	0.18904789	-0.057604567
area	4.698830e-01	1.000000000	-0.06706308	-0.0949327376	4.746921e-02	-0.004147634	0.07263979	0.080608144	0.05820727	-0.07653135	-0.139940075
density	-2.840257e-02	-0.067063078	1.000000000	0.2420910611	1.780186e-01	-0.141769912	0.19598421	0.019500873	0.28070731	-0.08299839	-0.030825640
coastline	-6.782916e-02	-0.094932738	0.24209106	1.0000000000	-1.340760e-01	-0.136383054	0.05125458	-0.009996931	0.15368234	-0.12372388	0.339369012
net_migration	2.787508e-05	0.047469208	0.17801862	-0.1340759759	1.000000e+00	-0.024834932	0.38153271	-0.020708094	0.23704279	-0.06760298	-0.349701635
infant_mortality	2.507309e-02	-0.004147634	-0.14176991	-0.1363830536	-2.483493e-02	1.000000000	-0.59021658	-0.361046481	-0.65166515	-0.10820285	-0.063625706
GDP_	-3.887470e-02	0.072639794	0.19598421	0.0512545764	3.815327e-01	-0.590216577	1.000000000	0.309446163	0.82794270	0.02076714	-0.212628022
literacy	2.409183e-02	0.080608144	0.01950087	-0.0099969309	-2.070809e-02	-0.361046481	0.30944616	1.000000000	0.23782318	0.16660871	0.002434332
phones	-2.667799e-02	0.058207267	0.28070731	0.1536823395	2.370428e-01	-0.651665153	0.82794270	0.237823182	1.000000000	0.05655300	-0.141818646
arable	1.890479e-01	-0.076531347	-0.08299839	-0.1237238762	-6.760298e-02	-0.108202846	0.02076714	0.166608708	0.05655300	1.000000000	0.107673262
crops	-5.760457e-02	-0.139940075	-0.03082564	0.3393690118	-3.497016e-01	-0.063625706	-0.21262802	0.002434332	-0.14181865	0.10767326	1.000000000
other	-8.991593e-02	0.139939259	0.06867917	-0.0802421473	1.510750e-01	0.130374291	0.09500348	0.068529091	0.00275561	-0.67577457	-0.480857364
climate	4.165680e-03	-0.165295374	-0.08820325	-0.0405721889	-7.777141e-02	-0.184508817	0.12447374	0.084445679	0.11607740	0.19961010	0.064256955
birthrate	-4.076392e-02	-0.059890923	-0.15304738	-0.0694048945	-6.288375e-02	0.841493249	-0.62219510	-0.416395505	-0.69328483	-0.18585694	0.122051168
deathrate	-2.243628e-02	0.046891693	-0.06387178	-0.1567005997	3.526494e-02	0.659608153	-0.17887769	-0.167256162	-0.24134550	0.04637654	-0.202013679
agriculture	-6.049711e-03	-0.013188643	-0.15501645	-0.1094899294	-6.927963e-02	0.811695576	-0.58735721	-0.351306042	-0.64331162	-0.13165853	0.094873580
industry	1.180841e-01	0.138573597	-0.17993875	-0.2603475549	-5.318217e-02	0.090041769	-0.06145650	0.166347561	-0.21085037	0.01889628	-0.067818352
service	-1.847283e-02	0.003447593	0.01794705	-0.0007417245	3.980470e-02	-0.351821967	0.34600821	0.203735977	0.32973864	0.15192553	0.078159091
other	population	0.004165680	-0.04076392	-0.022436283	-0.006049711	0.11808407	-0.0184728277				
area	0.13993926	-0.165295374	-0.05989092	0.046891693	-0.013188643	0.13857360	0.0034475926				
density	0.06867917	-0.088203251	-0.15304738	-0.063871783	-0.155016445	-0.17993875	0.0179470473				
coastline	-0.08024215	-0.040572189	-0.06940489	-0.156700600	-0.109489929	-0.26034755	-0.0007417245				
net_migration	0.15107503	-0.077771415	-0.06288375	0.035264940	-0.069279630	-0.05318217	0.0398046964				
infant_mortality	0.13037429	-0.184508817	0.84149325	0.659608153	0.811695576	0.09004177	-0.3518219672				
GDP_	0.09500348	0.124473739	-0.62219510	-0.178877687	-0.587357205	-0.06145650	0.3460082128				
literacy	0.06852909	0.084445679	-0.41639551	-0.167256162	-0.351306042	0.16634756	0.2037359767				
phones	0.00275561	0.116077400	-0.69328483	-0.241345503	-0.643311620	-0.21085037	0.3297386435				
arable	-0.67577457	0.199610101	-0.18585694	0.046376538	-0.131658527	0.01889628	0.1519255259				
crops	-0.48085736	0.064256955	0.12205117	-0.202013679	0.094873580	-0.06781835	0.0781590914				
other	1.00000000	-0.148743163	0.11036134	0.102602876	0.129759984	0.12638625	-0.0613806931				
climate	-0.14874316	1.000000000	-0.15767373	0.003326035	-0.142025176	0.02604803	0.2705061150				
birthrate	0.11036134	-0.157673726	0.100000000	0.412038702	0.905602830	0.014024861	-0.163419262				
deathrate	0.10260288	0.003326035	0.41203870	1.000000000	0.425374564	0.10531609	-0.1658607126				
agriculture	0.12975998	-0.142025176	0.90560283	0.425374564	1.000000000	0.11674867	-0.1655172659				
industry	0.12638625	0.026048029	0.01402861	0.105316094	0.116748672	1.000000000	-0.0525356209				
service	-0.06138069	0.270506115	-0.31634193	-0.165860713	-0.165517266	-0.05253562	1.0000000000				

در ماتریس زیر نیز اعداد تا دو رقم اعشار رند شده‌اند. همان گونه که می‌بینید ضریب همبستگی بین یک و منفی یک است

	population	area	density	coastline	net_migration	infant_mortality	GDP_	literacy	phones	arable	crops	other	climate	birthrate	deathrate	agriculture	industry	service
population	1.00	0.47	-0.03	-0.07	0.00	0.03	-0.04	0.02	-0.03	0.19	-0.06	-0.09	0.00	-0.04	-0.02	-0.01	0.12	-0.02
area	0.47	1.00	-0.07	-0.09	0.05	0.00	0.07	0.08	0.06	-0.08	-0.14	0.14	-0.17	-0.06	0.05	-0.01	0.14	0.00
density	-0.03	-0.07	1.00	0.24	0.18	-0.14	0.20	0.02	0.28	-0.08	-0.03	0.07	-0.09	-0.15	-0.06	-0.16	-0.18	0.02
coastline	-0.07	-0.09	0.24	1.00	-0.13	-0.14	0.05	-0.01	0.15	-0.12	0.34	-0.08	-0.04	-0.07	-0.16	-0.11	-0.26	0.00
net_migration	0.00	0.05	0.18	-0.13	1.00	-0.02	0.38	-0.02	0.24	-0.07	-0.35	0.15	-0.08	-0.06	0.04	-0.07	-0.05	0.04
infant_mortality	0.03	0.00	-0.14	-0.14	-0.02	1.00	-0.59	-0.36	-0.65	-0.11	-0.06	0.13	-0.18	0.84	0.66	0.81	0.09	-0.35
GDP_	-0.04	0.07	0.20	0.05	0.38	-0.59	1.00	0.31	0.83	0.02	-0.21	0.10	0.12	-0.62	-0.18	-0.59	-0.06	0.35
literacy	0.02	0.08	0.02	-0.01	-0.02	-0.36	0.31	1.00	0.24	0.17	0.00	0.07	0.08	-0.42	-0.17	-0.35	0.17	0.20
phones	-0.03	0.06	0.28	0.15	0.24	-0.65	0.83	0.24	1.00	0.06	-0.14	0.00	0.12	-0.69	-0.24	-0.64	-0.21	0.33
arable	0.19	-0.08	-0.08	-0.12	-0.07	-0.11	0.02	0.17	0.06	1.00	0.11	-0.68	0.20	-0.19	0.05	-0.13	0.02	0.15
crops	-0.06	-0.14	-0.03	0.34	-0.35	-0.06	-0.21	0.00	-0.14	0.11	1.00	-0.48	0.06	0.12	-0.20	0.09	-0.07	0.08
other	-0.09	0.14	0.07	-0.08	0.15	0.13	0.10	0.07	0.00	-0.68	-0.48	1.00	-0.15	0.11	0.10	0.13	0.13	-0.06
climate	0.00	-0.17	-0.09	-0.04	-0.08	-0.18	0.12	0.08	0.12	0.20	0.06	-0.15	1.00	-0.16	0.00	-0.14	0.03	0.27
birthrate	-0.04	-0.06	-0.15	-0.07	-0.06	0.84	-0.62	-0.42	-0.69	-0.19	0.12	0.11	-0.16	1.00	0.41	0.91	0.01	-0.32
deathrate	-0.02	0.05	-0.06	-0.16	0.04	0.66	-0.18	-0.17	-0.24	0.05	-0.20	0.10	0.00	0.41	1.00	0.43	0.11	-0.17
agriculture	-0.01	-0.01	-0.16	-0.11	-0.07	0.81	-0.59	-0.35	-0.64	-0.13	0.09	0.13	-0.14	0.91	0.43	1.00	0.12	-0.17
industry	0.12	0.14	-0.18	-0.26	-0.05	0.09	-0.06	0.17	-0.21	0.02	-0.07	0.13	0.03	0.01	0.11	0.12	1.00	-0.05
service	-0.02	0.00	0.02	0.00	0.04	-0.35	0.35	0.20	0.33	0.15	0.08	-0.06	0.27	-0.32	-0.17	-0.17	-0.05	1.00

همچنین دیده می‌شود که رابطه بین کشاورزی و زاد و ولد خطی است.

۴- شکل زیر scatter plot کشاورزی و زاد و ولد است.



۵- در قسمت پنج نیز با توجه به رابطه خطی زاد و ولد و کشاورزی می‌توانیم در مقادیر NA مقادیر زاد و ولد مطابق با آن خانه را در محل مورد نظر جایگذاری کرد.

بخش دوم:

- ۱- با استفاده از تابع `runif()` یک متغیر تصادفی با توزیع یکنواخت استاندارد ساخته شد.
 - ۲- با استفاده از `if` و `else` و تابع `runif()` متغیر تصادفی با توزیع برنولی با $p=0.6$ ساخته شد
 - ۳- با استفاده از `if` و `else` و تابع `runif()` و `for` یک متغیر تصادفی با توزیع دوجمله‌ای ساخته شد
- میانگین و واریانس متغیر تصادفی برنولی به ترتیب np و $np(1-p)$ است که در این مثال به صورت تئوری باید ۶ و ۲.۴ شود
- برنامه نیز به ما مقدارهای میانگین و واریانس متغیر تصادفی برنولی را ۵.۸۱ و ۲.۲۱۶۰۶۱ می‌دهد.

```
> mean(z)
[1] 5.81
> var(z)
[1] 2.216061
```

علت این تفاوت به علت کم بودن مقدار n است. هر چه n بیشتر شود مقدار میانگین و واریانس به مقدار تئوری نزدیکتر می‌شود.

بخش سوم:

۱-قسمت اول:

اگر $U \sim \text{Uniform}[0,1]$ و $F_n(x)$ تابع توزیع تجمعی باشد
ثابت می‌کنیم

$$F_n(x) = P(X \leq x) = P(T(U) \leq x)$$

$$= P(U \leq T^{-1}(x)) = T^{-1}(x)$$

$$F_n(x) = T^{-1}(u)$$

$$F^{-1}(x) = T(u) = u$$

۲- قسمت دوم:

$$F(x) = 1 - e^{-\lambda x} = u$$

$$F(F^{-1}(u)) = 1 - e^{-\lambda F^{-1}(u)} = u$$

$$\ln(1-u) = -\lambda F^{-1}(u) \quad \frac{-1}{\lambda} \ln(1-u) = x$$

$$U \sim \text{uniform}[0,1] \rightarrow v = 1-u \quad F_v(v) = P\{1-u \leq v\}$$

$$P\{u \geq 1-v\} = 1 - P\{u \leq 1-v\} = 1 - (1-v) = v$$

$$F(u) = F(v) = F(1-u) \Rightarrow F^{-1}(u) = F^{-1}(1-u)$$

$$X = \frac{-1}{\lambda} \ln(v)$$

همانگونه که مشاهده می‌شود اعداد به صورت توزیع exp توزیع شده‌اند.

