Covariance matrix of X:

```
                    income          age         sex
capital_gain  \
income            0.182820     1.364955   -0.043450
7.052093e+02
age               1.364955   186.055686   -0.570097
7.824578e+03
sex              -0.043450    -0.570097    0.221370
-1.684532e+02
capital_gain    705.209251  7824.578224 -168.453232
5.454086e+07
hours_per_week    1.212614    11.579774   -1.332141
7.149812e+03

                hours_per_week
income                1.212614
age                  11.579774
sex                  -1.332141
capital_gain       7149.812440
hours_per_week      152.454313
```

Correlation matrix of X:

```
                    income          age         sex
capital_gain  \
income            0.240811     10.656070    0.036211
9.647254e+02
age              10.656070   1674.644868   12.192892
4.940332e+04
sex               0.036211     12.192892    0.330798
1.880380e+02
capital_gain    964.725362  49403.322148  188.038036
5.570223e+07
```

```
hours_per_week   10.950638   1571.771326   12.044758
5.072853e+04

                hours_per_week
income               10.950638
age                1571.771326
sex                  12.044758
capital_gain      50728.528312
hours_per_week     1787.692368
```

Mean vector of X:

```
income               0.240817
age                 38.582832
sex                  0.330805
capital_gain      1077.681940
hours_per_week      40.438698
dtype: float64
```

Correlation Coefficient matrix of X:

```
                  income       age       sex
capital_gain  hours_per_week
income          1.000000  0.234056 -0.215965
0.223330        0.229707
age             0.234056  1.000000 -0.088759
0.077678        0.069020
sex            -0.215965 -0.088759  1.000000
-0.048476       -0.229199
capital_gain    0.223330  0.077678 -0.048476
1.000000        0.078410
```

```
hours_per_week  0.229707  0.069020 -0.229199
0.078410        1.000000
```

Error in estimation (Biased Vector):

income: 0.646168

age: 0.162906

sex: 0.694953

capital_gain: 0.92988

hours_per_week: 0.170216

Q. Are the histograms similar?

Ans. The histograms don't seem similar for all features. However, the features *age* and *hours_per_week* have the most overlap.

(Biased Vector)
Minimum relative error in column:

age

Corresponding relative error:

0.16290624401293272

Error in estimation (de-Biased Vector):

income: 0.841516

age: 0.942956

sex: 0.917897

capital_gain: 0.948651

hours_per_week: 0.912547

(de-Biased Vector)
Minimum relative error in column:

income

Corresponding relative error:

0.8415157333936801

• Feel free to comment on any difference on the relative error (when using X_{new} rather than X).
Ans: The error has increased with X_{new} when compared to **X.**

Feature with most predictability:

1. Biased vector:

Feature with most predictability:
`age`

2. de-Biased vector:

Feature with most predictability:
`income`
