```
                                                                  ___  ____  ____  ____  ____(R)
                                                                 /__    /   ____/   /   ____/
                                                                ___/   /   /___/   /   /___/
                                                                  Statistics/Data analysis

                                                                    User: Jiayi
                                                                 Project: Project 6-1

  ___  ____  ____  ____  ____ ®
 /__    /   ____/   /   ____/      18.0
___/   /   /___/   /   /___/       MP—Parallel Edition

 Statistics and Data Science       Copyright 1985-2023 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: Single-user 2-core  perpetual
Serial number: 501806300203
  Licensed to: 501806300203


Notes:
      1. Unicode is supported; see help unicode_advice.
      2. More than 2 billion observations are allowed; see help obs_advice.
      3. Maximum number of variables is set to 30,000 but can be increased;
          see help set_maxvar.

Running D:\stata18\profile.do ...

1 . import delimited "C:\Users\Administrator\Desktop\marketing_and_sales_data_evaluation_lr.csv", clear
(encoding automatically selected: ISO-8859-1)
(4 vars, 4,572 obs)

2 . list in 1/5

     +-------------------------------------------------+
     | tv       radio   social_~a       sales          |
     |-------------------------------------------------|
  1. | 16   6.5662308   2.9079828   54.732757          |
  2. | 13   9.2377646   2.4095672   46.677897          |
  3. | 41   15.886446   2.9134102   150.17783          |
  4. | 83   30.020028    6.922304   298.24634          |
  5. | 15   8.4374077   1.4059982   56.594181          |
     +-------------------------------------------------+

3 . describe

Contains data
 Observations:         4,572
    Variables:             4
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
tv              byte    %8.0g                 TV
radio           double  %10.0g                Radio
social_media    double  %10.0g                Social_Media
sales           double  %10.0g                Sales
-------------------------------------------------------------------------------
Sorted by:
     Note: Dataset has changed since last saved.

4 . display "Rows: " _N " Columns: " _Nvars
Rows: 4572 Columns: _Nvars not found
r(111);
```

```
5 . browse

6 . codebook
```

---

tv

```
                  Type: Numeric (byte)

                 Range: [10,100]                      Units: 1
         Unique values: 91                         Missing .: 10/4,572

                  Mean: 54.0669
             Std. dev.: 26.1251

           Percentiles:     10%       25%       50%       75%       90%
                             19        32        53        77        91
```

---

radio

```
                  Type: Numeric (double)

                 Range: [.00068395,48.871161]         Units: 1.000e-09
         Unique values: 4,568                      Missing .: 4/4,572

                  Mean: 18.1604
             Std. dev.: 9.67696

           Percentiles:     10%       25%       50%       75%       90%
                         5.30937   10.5249   17.8595   25.6507   31.3573
```

---

social_media

```
                  Type: Numeric (double)

                 Range: [.0000313,13.981662]          Units: 1.000e-09
         Unique values: 4,566                      Missing .: 6/4,572

                  Mean: 3.32396
             Std. dev.: 2.21267

           Percentiles:     10%       25%       50%       75%       90%
                         .627606   1.52764   3.05557   4.80766   6.39581
```

---

sales

```
                  Type: Numeric (double)

                 Range: [31.199409,364.07975]         Units: 1.000e-08
         Unique values: 4,566                      Missing .: 6/4,572

                  Mean: 192.467
             Std. dev.: 93.1331

           Percentiles:     10%       25%       50%       75%       90%
                         66.3175   112.315   189.231   272.537   323.855
```

```
. summarize
```

Variable	Obs	Mean	Std. dev.	Min	Max
tv	4,562	54.06686	26.12505	10	100
radio	4,568	18.16036	9.676958	.0006839	48.87116
social_media	4,566	3.323956	2.21267	.0000313	13.98166
sales	4,566	192.4666	93.13309	31.19941	364.0798

```
. summarize tv, detail
```

TV

Percentiles		Smallest		
1%	11	10		
5%	14	10		
10%	19	10	Obs	4,562
25%	32	10	Sum of wgt.	4,562
50%	53		Mean	54.06686
		Largest	Std. dev.	26.12505
75%	77	100		
90%	91	100	Variance	682.5184
95%	96	100	Skewness	.0691637
99%	100	100	Kurtosis	1.812995

```
. summarize radio, detail
```

Radio

Percentiles		Smallest		
1%	.6758737	.0006839		
5%	2.797291	.0144865		
10%	5.309371	.0218826	Obs	4,568
25%	10.52493	.0262955	Sum of wgt.	4,568
50%	17.85951		Mean	18.16036
		Largest	Std. dev.	9.676958
75%	25.6507	44.8613		
90%	31.35729	45.08292	Variance	93.64352
95%	34.14759	47.11629	Skewness	.143103
99%	38.24273	48.87116	Kurtosis	2.180604

```
. summarize social_media, detail
```

Social_Media

Percentiles		Smallest		
1%	.0463245	.0000313		
5%	.2988564	.0009766		
10%	.6276056	.0013154	Obs	4,566
25%	1.527642	.0031768	Sum of wgt.	4,566
50%	3.055565		Mean	3.323956
		Largest	Std. dev.	2.21267
75%	4.807659	12.05475		
90%	6.395812	12.10802	Variance	4.89591
95%	7.445375	13.08396	Skewness	.6447013
99%	9.072549	13.98166	Kurtosis	3.043857

```
. summarize sales, detail
```

Sales

Percentiles		Smallest		
1%	37.08405	31.19941		
5%	50.64876	31.40226		
10%	66.31752	32.23703	Obs	4,566
25%	112.3154	32.56683	Sum of wgt.	4,566

50%	**189.2312**		Mean	**192.4666**
		Largest	Std. dev.	**93.13309**
75%	**272.5368**	**360.1097**		
90%	**323.8554**	**360.4001**	Variance	**8673.773**
95%	**342.6976**	**362.0421**	Skewness	**.0694783**
99%	**354.7367**	**364.0798**	Kurtosis	**1.812587**

```
12 . display Percentage of missing values in 'sales'
Percentage not found
r(111);

13 . display "Percentage of missing values in 'sales': " (r(N_missing) / r(N)) * 100
Percentage of missing values in 'sales': .

14 .
15 . drop if missing(variable)
variable not found
r(111);

16 . drop if missing (smallest)
missing not found
r(111);

17 . to display percentage of missing values in sales count if missing(sales)
command to is unrecognized
r(199);

18 . misstable summarize
```

				Obs<.		
Variable	Obs=.	Obs>.	Obs<.	Unique values	Min	Max
tv	**10**		**4,562**	91	10	100
radio	**4**		**4,568**	>500	.0006839	48.87116
social_media	**6**		**4,566**	>500	.0000313	13.98166
sales	**6**		**4,566**	>500	31.19941	364.0798

```
19 . list if missing(TV) | missing(Radio) | missing(Social_Media) | missing(Sales)
TV not found
r(111);

20 . list if missing(tv) | missing(radio) | missing(social_media) | missing(sales)
```

	tv	radio	social_~a	sales
14.	.	22.351667	3.0318153	276.16535
27.	.	34.111674	4.6241483	342.91337
47.	.	34.859637	7.7814167	318.96978
76.	.	6.4822934	.86684499	91.177216
100.	.	7.6358194	1.5541458	56.18673
120.	.	30.470485	6.806919	336.81869
142.	.	9.164464	1.0966814	65.259189
164.	.	38.118424	6.6766114	328.55518
183.	81	26.425422	.	288.64944
184.	.	1.2870605	.39617874	56.545293
185.	25	.41384939	.	92.357092
187.	73	25.340209	.	258.35897
190.	89	29.682384	.	320.2644
193.	41	13.142657	.	142.62686
196.	34	12.660398	.	117.11414
198.	98	.	6.3997296	345.95284
201.	89	.	4.535458	316.72562
204.	22	.	4.1325256	78.031498
207.	12	.	1.2300262	50.00921
209.	77	25.598379	5.5147866	.
211.	.	17.007075	5.1996348	208.81638
215.	100	36.466753	5.6359925	.

222.	**74**	**24.220634**	**.28589761**	**.**
227.	**19**	**8.7267827**	**.49783849**	**.**
232.	**22**	**6.8097517**	**.01145052**	**.**
237.	**27**	**1.3844155**	**2.3981292**	**.**

```
21 . drop if missing(tv) | missing(radio) | missing(social_media) | missing(sales)
   (26 observations deleted)

22 . misstable summarize
   (variables nonmissing or string)

23 . histogram sales, frequency normal title("Distribution of sales") xtitle("sales") ytitle("Frequency")
   (bin=36, start=31.199409, width=9.2466762)

24 . kdensity sales, normal title("Density of sales")

25 . graph box sales, title("Box Plot of sales")

26 . graph matrix tv radio social_media sales, half

27 . correlate tv radio social_media sales
   (obs=4,546)
```

	tv	radio	social~a	sales
tv	**1.0000**			
radio	**0.8692**	**1.0000**		
social_media	**0.5277**	**0.6063**	**1.0000**	
sales	**0.9995**	**0.8686**	**0.5274**	**1.0000**

```
28 . regress sales tv radio social_media
```

Source	SS	df	MS		
Model	**39286984.8**	**3**	**13095661.6**	Number of obs	= **4,546**
Residual	**39522.2825**	**4,542**	**8.7015153**	F(3, 4542)	> **99999.00**
Total	**39326507**	**4,545**	**8652.69682**	Prob > F	= **0.0000**
				R-squared	= **0.9990**
				Adj R-squared	= **0.9990**
				Root MSE	= **2.9498**

sales	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
tv	**3.56257**	**.0033893**	**1051.12**	**0.000**	**3.555925**	**3.569214**
radio	**-.0039704**	**.0097805**	**-0.41**	**0.685**	**-.023145**	**.0152042**
social_media	**.004964**	**.0248836**	**0.20**	**0.842**	**-.0438199**	**.0537479**
_cons	**-.1339631**	**.1028203**	**-1.30**	**0.193**	**-.3355409**	**.0676148**

```
29 . predict residuals, residuals

30 . histogram residuals, normal
   (bin=36, start=-11.615811, width=.63538665)

31 . scatter residuals sales

32 . correlate tv radio social_media sales
   (obs=4,546)
```

	tv	radio	social~a	sales
tv	**1.0000**			
radio	**0.8692**	**1.0000**		
social_media	**0.5277**	**0.6063**	**1.0000**	
sales	**0.9995**	**0.8686**	**0.5274**	**1.0000**

```
33 . regress sales VAR_NAME
variable VAR_NAME not found
r(111);

34 . regress sales tv

      Source |       SS           df       MS      Number of obs   =
>      4,546
-------------+----------------------------------   F(1, 4544)      >
>   99999.00
       Model |  39286983.3         1  39286983.3   Prob > F        =
>     0.0000
    Residual |  39523.7484     4,544  8.69800802   R-squared       =
>     0.9990
-------------+----------------------------------   Adj R-squared   =
>     0.9990
       Total |    39326507     4,545  8652.69682   Root MSE        =
>     2.9492

------------------------------------------------------------------------------
>------------
       sales | Coefficient  Std. err.      t    P>|t|     [95% conf.
>  interval]
------------------------------------------------------------------------------
>------------
          tv |   3.561514   .0016758  2125.27   0.000     3.558229
>   3.564799
       _cons |  -.1324925    .100605    -1.32   0.188    -.3297273
>   .0647422
------------------------------------------------------------------------------
>------------

35 . regress sales radio

      Source |       SS           df       MS      Number of obs   =
>      4,546
-------------+----------------------------------   F(1, 4544)      =
>   13967.55
       Model |  29673094.2         1  29673094.2   Prob > F        =
>     0.0000
    Residual |  9653412.83     4,544  2124.43064   R-squared       =
>     0.7545
-------------+----------------------------------   Adj R-squared   =
>     0.7545
       Total |    39326507     4,545  8652.69682   Root MSE        =
>     46.092

------------------------------------------------------------------------------
>------------
       sales | Coefficient  Std. err.      t    P>|t|     [95% conf.
>  interval]
------------------------------------------------------------------------------
>------------
       radio |   8.361628   .0707507   118.18   0.000     8.222922
>   8.500333
       _cons |    40.5868    1.45522    27.89   0.000     37.73386
>   43.43974
------------------------------------------------------------------------------
>------------

36 . regress sales social_media

      Source |       SS           df       MS      Number of obs   =
>      4,546
-------------+----------------------------------   F(1, 4544)      =
>    1751.37
       Model |  10940623.4         1  10940623.4   Prob > F        =
>     0.0000
    Residual |  28385883.6     4,544   6246.8934   R-squared       =
>     0.2782
-------------+----------------------------------   Adj R-squared   =
>     0.2780
       Total |    39326507     4,545  8652.69682   Root MSE        =
>     79.037
```

```
------------------------------------------------------------------------------
------------
       sales | Coefficient  Std. err.      t    P>|t|     [95% conf.
>  interval]
-------------+----------------------------------------------------------------
------------
social_media |   22.18786   .5301838    41.85   0.000     21.14844
>   23.22728
       _cons |   118.6726   2.116359    56.07   0.000     114.5235
>   122.8217
------------------------------------------------------------------------------
------------

37 . scatter sales tv || lfit sales tv, title("tv Advertising vs sales"
   > )

38 . regress sales tv

      Source |       SS           df       MS      Number of obs   =
>      4,546
-------------+----------------------------------   F(1, 4544)      >
>   99999.00
       Model |  39286983.3         1  39286983.3   Prob > F        =
>     0.0000
    Residual |  39523.7484     4,544  8.69800802   R-squared       =
>     0.9990
-------------+----------------------------------   Adj R-squared   =
>     0.9990
       Total |    39326507     4,545  8652.69682   Root MSE        =
>     2.9492

------------------------------------------------------------------------------
------------
       sales | Coefficient  Std. err.      t    P>|t|     [95% conf.
>  interval]
-------------+----------------------------------------------------------------
------------
          tv |   3.561514   .0016758  2125.27   0.000     3.558229
>   3.564799
       _cons |  -.1324925    .100605    -1.32   0.188    -.3297273
>   .0647422
------------------------------------------------------------------------------
------------

39 . ================================================================
   > =============
   == is not a valid command name
   r(199);

40 .
41 .                             OLS Regression Results
   >
   command OLS is unrecognized
   r(199);

42 .
43 . ================================================================
   > =============
   == is not a valid command name
   r(199);
```

```
44 .
45 . Number of observations: 4546
   command Number is unrecognized
   r(199);

46 .
47 . F(3, 4542): 1504986.33
   command F is unrecognized
   r(199);

48 .
49 . Prob > F: 0.0000
   command Prob not defined by Prob.ado
   r(199);

50 .
51 . R-squared: 0.9990
   command R is unrecognized
   r(199);

52 .
53 . Adjusted R-squared: 0.9990
   command Adjusted is unrecognized
   r(199);

54 .
55 . Root MSE: 2.9498
   command Root is unrecognized
   r(199);

56 .
57 . ------------------------------------------------------------------
   > -------------
   - is not a valid command name
   r(199);

58 .
59 .       Sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf
   > . Interval]
   command Sales is unrecognized
   r(199);

60 .
61 . -------------+----------------------------------------------------
   > -----------
   - is not a valid command name
   r(199);

62 .
63 . _cons       |     -0.1340      0.1028   -1.30   0.385     -0.3355
   >      0.0676
   command _cons is unrecognized
   r(199);

64 .
65 . TV          |      3.5626      0.0034  1051.12   0.000      3.5559
   >       3.5692
   command TV is unrecognized
   r(199);

66 .
67 . Radio       |     -0.0040      0.0098   -0.41   1.370     -0.0231
   >      0.0152
   command Radio is unrecognized
   r(199);
```

```
68 .
69 . Social_Media|      0.0050      0.0249    0.20   1.684     -0.0438
   >      0.0537
```

command Social_Media is unrecognized

```
r(199);
```

```
70 .
71 . ================================================================
   > =============
```

== is not a valid command name

```
r(199);
```

```
72 .
```