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

                                                                     User: Samin Al Fattah
                                                                  Project: Samin Al Fattah

  ___  ____  ____  ____  ____ ©
 /__    /   ____/   /   ____/      17.0
___/   /   /___/   /   /___/       MP—Parallel Edition

 Statistics and Data Science       Copyright 1985-2021 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: Unlimited-user 64-core network perpetual
Serial number: 18461036
  Licensed to: TEAM BTCR
               TEAM BTCR

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 5,000; see help set maxvar.

 1 . cd "C:\Users\Administrator\Desktop\Stata exam"
   C:\Users\Administrator\Desktop\Stata exam

 2 . pwd
   C:\Users\Administrator\Desktop\Stata exam

 3 . use "Data set_BUPFS EXCEL MAESTROS ROUND 2"

 4 . help string

 5 . help f_sting

 6 . recast double case_id

 7 . recast float case_id

 8 . recast string case_id
   case_id:  string invalid
   r(109);

 9 . destring cluster, gen(cl)
   cluster: all characters numeric; cl generated as byte

10 . drop cluster

11 . rename cl cluster

12 . rename V001 gender
```

```
13 . rename V003 age

14 . rename age income

15 . rename v002 age

16 . rename v004 edu

17 . rename V005 occup

18 . rename v025 residence

19 . rename v225 mar_st

20 . label variable gender "Sex of the respondent"

21 . label variable age "Age of the respondent"

22 . label variable income "Monthly income"

23 . label variable edu "Highest education achieved"

24 . label variable occup "Current occupation"

25 . label variable residence "Place of residence"

26 . label variable mar_st "Ever married status"

27 . label define gen_label 1 "Female" 2 "Male"

28 . label value gender gen_label

29 . br

30 . ed

31 . label define occup_label 1 "Agricultue" 2 "Business" 3 "Privat
   > e job" 4 "Government job" 5 "Day laborer"

32 . label value occup occup_label

33 . ed

34 . label define res_label  0 "Urban" 2 "Rural"

35 . label value residence res_label

36 . ed

37 . label define res_label  1 "Urban" 2 "Rural"
   label res_label already defined
   r(110);
```

```
38 . label define res_labell  1 "Urban" 2 "Rural"

39 . label value residence res_labell

40 . ed

41 . label define mr_label 0 "No" 1 "Yes"

42 . label value mar_st mr_label

43 . ed

44 . summarize age income edu, detail
```

Age of the respondent

Percentiles		Smallest		
1%	**25.02747**	**25.00549**		
5%	**25.29695**	**25.02747**		
10%	**25.68148**	**25.09949**	Obs	**113**
25%	**27.18787**	**25.17426**	Sum of wgt.	**113**
50%	**29.53597**		Mean	**29.77941**
		Largest	Std. dev.	**3.057544**
75%	**32.53594**	**34.64049**		
90%	**34.27671**	**34.86572**	Variance	**9.348574**
95%	**34.43022**	**34.92676**	Skewness	**.1321498**
99%	**34.92676**	**34.95758**	Kurtosis	**1.75177**

Monthly income

Percentiles		Smallest		
1%	**10163**	**10066**		
5%	**10519**	**10163**		
10%	**11405**	**10252**	Obs	**113**
25%	**16131**	**10306**	Sum of wgt.	**113**
50%	**21768**		Mean	**21748.99**
		Largest	Std. dev.	**7255.816**
75%	**27074**	**34364**		
90%	**31853**	**34648**	Variance	**5.26e+07**
95%	**34007**	**34703**	Skewness	**.0294865**
99%	**34703**	**34947**	Kurtosis	**1.922893**

Highest education achieved

Percentiles		Smallest		
1%	**0**	**0**		
5%	**1**	**0**		
10%	**2**	**0**	Obs	**113**
25%	**6**	**0**	Sum of wgt.	**113**
50%	**9**		Mean	**8.548673**
		Largest	Std. dev.	**4.342463**
75%	**12**	**16**		
90%	**14**	**16**	Variance	**18.85698**
95%	**16**	**17**	Skewness	**-.023112**
99%	**17**	**17**	Kurtosis	**2.3383**

```
45 . tab age income edu
too many variables specified
r(103);

46 . tab age
```

Age of the respondent	Freq.	Percent	Cum.
25.01	1	0.88	0.88
25.03	1	0.88	1.77
25.10	1	0.88	2.65
25.17	1	0.88	3.54
25.24	1	0.88	4.42
25.30	1	0.88	5.31
25.39	1	0.88	6.19
25.55	1	0.88	7.08
25.55	1	0.88	7.96
25.56	1	0.88	8.85
25.62	1	0.88	9.73
25.68	1	0.88	10.62
25.80	1	0.88	11.50
25.87	1	0.88	12.39
25.90	1	0.88	13.27
26.03	1	0.88	14.16
26.29	1	0.88	15.04
26.35	1	0.88	15.93
26.38	1	0.88	16.81
26.40	1	0.88	17.70
26.49	1	0.88	18.58
26.66	1	0.88	19.47
26.68	1	0.88	20.35
26.70	1	0.88	21.24
26.71	1	0.88	22.12
26.82	1	0.88	23.01
26.91	1	0.88	23.89
26.99	1	0.88	24.78
27.19	1	0.88	25.66
27.20	1	0.88	26.55
27.21	1	0.88	27.43
27.22	1	0.88	28.32
27.26	1	0.88	29.20
27.47	1	0.88	30.09
27.49	1	0.88	30.97
27.53	1	0.88	31.86
27.71	1	0.88	32.74
27.75	1	0.88	33.63
27.79	1	0.88	34.51
27.86	1	0.88	35.40
27.99	1	0.88	36.28
28.25	1	0.88	37.17
28.33	1	0.88	38.05
28.69	1	0.88	38.94
28.77	1	0.88	39.82
28.81	1	0.88	40.71
28.91	1	0.88	41.59
28.92	1	0.88	42.48
28.93	1	0.88	43.36
29.15	1	0.88	44.25
29.16	1	0.88	45.13
29.24	1	0.88	46.02
29.36	1	0.88	46.90
29.43	1	0.88	47.79

29.44	**1**	**0.88**	**48.67**
29.46	**1**	**0.88**	**49.56**
29.54	**1**	**0.88**	**50.44**
29.59	**1**	**0.88**	**51.33**
29.61	**1**	**0.88**	**52.21**
29.74	**1**	**0.88**	**53.10**
29.96	**1**	**0.88**	**53.98**
29.98	**1**	**0.88**	**54.87**
30.18	**1**	**0.88**	**55.75**
30.26	**1**	**0.88**	**56.64**
30.36	**1**	**0.88**	**57.52**
30.45	**1**	**0.88**	**58.41**
30.49	**1**	**0.88**	**59.29**
30.61	**1**	**0.88**	**60.18**
30.62	**1**	**0.88**	**61.06**
30.74	**1**	**0.88**	**61.95**
30.76	**1**	**0.88**	**62.83**
30.80	**1**	**0.88**	**63.72**
31.10	**1**	**0.88**	**64.60**
31.18	**1**	**0.88**	**65.49**
31.18	**1**	**0.88**	**66.37**
31.18	**1**	**0.88**	**67.26**
31.23	**1**	**0.88**	**68.14**
31.36	**1**	**0.88**	**69.03**
31.51	**1**	**0.88**	**69.91**
31.54	**1**	**0.88**	**70.80**
31.74	**1**	**0.88**	**71.68**
32.16	**1**	**0.88**	**72.57**
32.34	**1**	**0.88**	**73.45**
32.52	**1**	**0.88**	**74.34**
32.54	**1**	**0.88**	**75.22**
32.56	**1**	**0.88**	**76.11**
32.69	**1**	**0.88**	**76.99**
32.72	**1**	**0.88**	**77.88**
32.88	**1**	**0.88**	**78.76**
33.36	**1**	**0.88**	**79.65**
33.49	**1**	**0.88**	**80.53**
33.56	**1**	**0.88**	**81.42**
33.69	**1**	**0.88**	**82.30**
33.70	**1**	**0.88**	**83.19**
33.73	**1**	**0.88**	**84.07**
33.78	**1**	**0.88**	**84.96**
33.83	**1**	**0.88**	**85.84**
33.84	**1**	**0.88**	**86.73**
33.93	**1**	**0.88**	**87.61**
33.98	**1**	**0.88**	**88.50**
34.09	**1**	**0.88**	**89.38**
34.28	**1**	**0.88**	**90.27**
34.31	**1**	**0.88**	**91.15**
34.34	**1**	**0.88**	**92.04**
34.35	**1**	**0.88**	**92.92**
34.36	**1**	**0.88**	**93.81**
34.40	**1**	**0.88**	**94.69**
34.43	**1**	**0.88**	**95.58**
34.44	**1**	**0.88**	**96.46**
34.64	**1**	**0.88**	**97.35**
34.87	**1**	**0.88**	**98.23**
34.93	**1**	**0.88**	**99.12**
34.96	**1**	**0.88**	**100.00**
Total	**113**	**100.00**	

```
47 . ed

48 . gen inc_cat = income

49 . ed

50 . replace inc_cat=1 if income < 15000
   (24 real changes made)

51 . replace inc_cat=2 if 15000 < income < 20000
   (113 real changes made)

52 . replace inc_cat=1 if income < 15000
   (24 real changes made)

53 . ed

54 . replace inc_cat=5 if income < 40000
   (113 real changes made)

55 . replace inc_cat=4 if income < 30000
   (97 real changes made)

56 . replace inc_cat=3 if income < 25000
   (69 real changes made)

57 . replace inc_cat=2 if income < 20000
   (48 real changes made)

58 . replave inc_cat=1 if income < 15000
   command replave is unrecognized
   r(199);

59 . replace inc_cat=1 if income < 15000
   (24 real changes made)

60 . ed

61 .  gen edu_cat = edu

62 . replace edu_cat="Secondary pass or above" if edu < 20
   type mismatch
   r(109);

63 . replace edu_cat=4 if edu < 20
   (106 real changes made)

64 .
65 . . replace edu_cat=3 if edu <= 7
   (45 real changes made)
```

```
66 . replace edu_cat=2 if edu <= 4
   (22 real changes made)

67 . replace edu_cat=1 if edu == 0
   (5 real changes made)

68 . label define educ_label 1 "Uneducated" 2 "Below primary" 3 "Pr
   > imary pass" 4 "Secondary pass or above"

69 . label value edu_cat educ_label

70 . ed

71 . gen age_cat = age

72 . replace age_cat=1 if age <50
   (113 real changes made)

73 . replace age_cat=2 if age < 50
   (113 real changes made)

74 . replace age_cat=1 if age <= 30
   (62 real changes made)

75 . ed

76 . label define agec_label 1 "30 years of below" 2 "30 years abov
   > e"

77 . label value age_cat agec_label

78 . ed

79 .
80 . . label define agec_label 1 "30 years of below" 2 "30 years abov
   label agec_label already defined
   r(110);

81 .
82 . label define agec_labell 1 "30 years or below" 2 "30 years above"

83 . label value age_cat agec_labell

84 . label variable inc_cat "Generated Income Category"

85 . label variable edu_cat "Generated Education Category"

86 . label variable age_cat "Generated Age Category"

87 . tab residence
```

Place of residence	Freq.	Percent	Cum.
Urban	**32**	**28.32**	**28.32**
Rural	**81**	**71.68**	**100.00**
Total	**113**	**100.00**	

```
88 . tab gender
```

Sex of the respondent	Freq.	Percent	Cum.
Female	62	54.87	54.87
Male	51	45.13	100.00
Total	113	100.00	

```
89 . tab occup
```

Current occupation	Freq.	Percent	Cum.
Agricultue	7	6.19	6.19
Business	44	38.94	45.13
Private job	34	30.09	75.22
Government job	20	17.70	92.92
Day laborer	8	7.08	100.00
Total	113	100.00	

```
90 . tab inc_cat
```

Generated Income Category	Freq.	Percent	Cum.
1	24	21.24	21.24
2	24	21.24	42.48
3	21	18.58	61.06
4	28	24.78	85.84
5	16	14.16	100.00
Total	113	100.00	

```
91 . tab edu_cat
```

Generated Education Category	Freq.	Percent	Cum.
Uneducated	5	4.42	4.42
Below primary	17	15.04	19.47
Primary pass	23	20.35	39.82
Secondary pass or above	68	60.18	100.00
Total	113	100.00	

```
92 . tab edu_cat gender, row col chi2
```

Key
frequency *row percentage* *column percentage*

Generated Education Category	Sex of the respondent Female	Male	Total
Uneducated	3	2	5
	60.00	40.00	100.00
	4.84	3.92	4.42
Below primary	8	9	17
	47.06	52.94	100.00
	12.90	17.65	15.04
Primary pass	9	14	23
	39.13	60.87	100.00
	14.52	27.45	20.35
Secondary pass or abo	42	26	68
	61.76	38.24	100.00
	67.74	50.98	60.18
Total	62	51	113
	54.87	45.13	100.00
	100.00	100.00	100.00

Pearson chi2(3) = 4.0783 Pr = 0.253

```
93 .  tab edu_cat gender, row col chi2
```

Key
frequency *row percentage* *column percentage*

Generated Education Category	Sex of the respondent Female	Male	Total
Uneducated	3	2	5
	60.00	40.00	100.00
	4.84	3.92	4.42
Below primary	8	9	17
	47.06	52.94	100.00
	12.90	17.65	15.04
Primary pass	9	14	23
	39.13	60.87	100.00
	14.52	27.45	20.35
Secondary pass or abo	42	26	68
	61.76	38.24	100.00
	67.74	50.98	60.18
Total	62	51	113
	54.87	45.13	100.00
	100.00	100.00	100.00

Pearson chi2(3) = 4.0783 Pr = 0.253

```
94 .
95 .  tab inc_cat residence , row col chi2
```

Key
frequency *row percentage* *column percentage*

Generated Income Category	Place of residence: Urban	Rural	Total
1	**7**	**17**	**24**
	29.17	**70.83**	**100.00**
	21.88	**20.99**	**21.24**
2	**7**	**17**	**24**
	29.17	**70.83**	**100.00**
	21.88	**20.99**	**21.24**
3	**8**	**13**	**21**
	38.10	**61.90**	**100.00**
	25.00	**16.05**	**18.58**
4	**8**	**20**	**28**
	28.57	**71.43**	**100.00**
	25.00	**24.69**	**24.78**
5	**2**	**14**	**16**
	12.50	**87.50**	**100.00**
	6.25	**17.28**	**14.16**
Total	**32**	**81**	**113**
	28.32	**71.68**	**100.00**
	100.00	**100.00**	**100.00**

Pearson chi2(**4**) = **2.9790** Pr = **0.561**

```
96 .  tab inc_cat edu_cat , row col chi2
```

Key
frequency *row percentage* *column percentage*

Generated Income Category	Generated Education Category: Uneducate	Below pri	Primary p	Total
1	**2**	**2**	**6**	**24**
	8.33	**8.33**	**25.00**	**100.00**
	40.00	**11.76**	**26.09**	**21.24**
2	**1**	**3**	**4**	**24**
	4.17	**12.50**	**16.67**	**100.00**
	20.00	**17.65**	**17.39**	**21.24**

3	**1**	**2**	**8**	**21**
	4.76	**9.52**	**38.10**	**100.00**
	20.00	**11.76**	**34.78**	**18.58**
4	**0**	**9**	**1**	**28**
	0.00	**32.14**	**3.57**	**100.00**
	0.00	**52.94**	**4.35**	**24.78**
5	**1**	**1**	**4**	**16**
	6.25	**6.25**	**25.00**	**100.00**
	20.00	**5.88**	**17.39**	**14.16**
Total	**5**	**17**	**23**	**113**
	4.42	**15.04**	**20.35**	**100.00**
	100.00	**100.00**	**100.00**	**100.00**

Generated Income Category	Generated Education Category Secondary	Total
1	**14**	**24**
	58.33	**100.00**
	20.59	**21.24**
2	**16**	**24**
	66.67	**100.00**
	23.53	**21.24**
3	**10**	**21**
	47.62	**100.00**
	14.71	**18.58**
4	**18**	**28**
	64.29	**100.00**
	26.47	**24.78**
5	**10**	**16**
	62.50	**100.00**
	14.71	**14.16**
Total	**68**	**113**
	60.18	**100.00**
	100.00	**100.00**

Pearson chi2(**12**) = **18.2393**   Pr = **0.109**

```
97 . graph pie gender

98 . ed
```

```
 99 . graph pie residence, over(gender)

100 . graph export "C:\Users\Administrator\Desktop\Stata exam\Graph.
    > png", as(png) name("Graph")
    file C:\Users\Administrator\Desktop\Stata exam\Graph.png saved
        as PNG format

101 . histogram income
    (bin=10, start=10066, width=2488.1)

102 . graph save "Graph" "C:\Users\Administrator\Desktop\Stata exam\
    > Histogram.gph"
    file C:\Users\Administrator\Desktop\Stata exam\Histogram.gph sav
    > ed

103 . reg income age edu residence occup

          Source │       SS           df       MS      Number of obs
    >    =       113
    ─────────────┼──────────────────────────────────────  F(4, 108)
    >    =      0.83
           Model │   175653305         4  43913326.3   Prob > F
    >    =    0.5095
        Residual │  5.7208e+09       108  52970324.7   R-squared
    >    =    0.0298
    ─────────────┼──────────────────────────────────────  Adj R-squared
    >    =   -0.0061
           Total │  5.8964e+09       112  52646860.5   Root MSE
    >    =    7278.1

    ─────────────┬──────────────────────────────────────────────────────────
    > ─────────────
          income │ Coefficient  Std. err.      t    P>|t|     [95% c
    > on
    >   f. interval]
    ─────────────┼──────────────────────────────────────────────────────────
    > ─────────────
             age │  -327.3089   232.9013    -1.41   0.163    -788.95
    > 97
    >       134.3419
             edu │  -16.01918   159.3633    -0.10   0.920     -331.9
    > 05
    >       299.8666
       residence │   2123.722   1572.621     1.35   0.180    -993.48
    > 45
    >       5240.929
           occup │  -329.6182   667.5158    -0.49   0.622     -1652.
    > 75
    >       993.5139
           _cons │   28911.64   7293.245     3.96   0.000     14455.
    > 17
    >       43368.12
    ─────────────┴──────────────────────────────────────────────────────────
    > ─────────────
```

```
104 . save "C:\Users\Administrator\Desktop\Stata exam\stata test.dta
    > "
    file C:\Users\Administrator\Desktop\Stata exam\stata test.dta
        saved

105 . export excel using "stata test", firstrow(variables)
    file stata test.xls saved

106 . graph bar (mean) occup

107 . graph save "Graph" "C:\Users\Administrator\Desktop\Stata exam\
    > bar plot.gph"
    file C:\Users\Administrator\Desktop\Stata exam\bar plot.gph save
    > d

108 . save "C:\Users\Administrator\Desktop\Stata exam\stata test.dta
    > ", replace
    file C:\Users\Administrator\Desktop\Stata exam\stata test.dta
        saved

109 . putexcel set "Results"

110 . putexcel set "C:\Users\Administrator\Desktop\Stata exam\Results.xlsx"

111 . save "C:\Users\Administrator\Desktop\Stata exam\stata test.dta", replace
    file C:\Users\Administrator\Desktop\Stata exam\stata test.dta saved

112 .
```