Homework 1

2.5 (a)

graph bar (mean)alcbinge, over(gender)blabel(total)ytitle(“percent who ever binged on alcohol”) //2.5(a)

\*\* or maybe we can simply use "sum alcbinge if gender == 0" to gain the result?

We can generate an output of graph bar “percent who ever binged on alcohol”(Figure 1)



Figure 1

So, there are approximately 29.075% of males who once binged on alcohol.

2.5(b) According to the Figure 1, we can see the percentage of male and female binged on alcohol is approximately 29.075% and 15.38%, the former is bigger. So the answer is **male.**

2.5(c)

graph bar (mean)gender, over(alcbinge)blabel(total)ytitle(“female percent who ever binged on alcohol”) //2.5(c)

According to the output graph STATA gives, we gain a graph(see Figure 2) describing the percentage of females who ever binged or never binged to alcohol.



Figure 2

2.12(a)

stem achmat08 //2.12(a)

Stem-and-leaf plot for achmat08 (Eighth Grade Math Achievement)

achmat08 rounded to nearest multiple of .1

plot in units of .1

36\* | 6

37\* | 122

38\* | 478

39\* | 0357

40\* | 0225555799

41\* | 34689

42\* | 0335577

43\* | 000133445689

44\* | 00012244467788

45\* | 334467778899

46\* | 3445555667788889

47\* | 0022344456677888

48\* | 00000133666889

49\* | 01123444466889

50\* | 001112356667888899

51\* | 0000111233344455779

52\* | 111112223334455668889

53\* | 000022456677889

54\* | 00011223334445555778889999

55\* | 00000003345678899

56\* | 112233467778888999

57\* | 0000112233344444778

58\* | 00113334677888899999

59\* | 034446679

60\* | 001223334556777888

61\* | 000112333345689

62\* | 000111123334555889

63\* | 2223677777779

64\* | 01122566788899

65\* | 223455577889

66\* | 0001122333344667799

67\* | 122333456677889

68\* | 0022235666689

69\* | 00122446779

70\* | 0114899

71\* | 389

72\* | 146

73\* | 3335788

74\* | 02233335

75\* | 00004

76\* |

77\* | 222222

According to the stem-and-leaf output STATA gives, the lowest 8th grade math achievement score is **36.6**

2.12(b) According to the stem-and-leaf output STATA gives, just **1** student.

2.12(c) According to the stem-and-leaf output STATA gives, there’s no outliers.

2.12(d)

According to the stem-and-leaf output STATA gives, it’s almost symmetric.

kdensity achmat08, ytitle("kdensity graph of 8th maths score")

graph box achmat08, ytitle("box graph of 8th maths score") //2.12(d)

The above commands are used to generate figures which make the distribution of *achmat08* data more visible. Those output figures(Figure 3 & Figure 4) are attached below. According to these figures, we can see that the shape of distribution of *achmat08* data is almost symmetric.



Figure 3



Figure 4

2.12(e) According to the stem-and-leaf output STATA gives, 55 and 63.7 occur 7 times.

2.12(f) Stem-and-leaf plot provides us some quick but blurred information about the shape and distribution of the given data, but when we want to get some detailed information, or want the given data more visible, the frequency distribution table is a better choice.

2.23(a)

graph box late12 //2.23(a)

The value of the 50th percentile of *late12* is 1 according to the box graph (Figure 5) generated by STATA. According to this, we can conclude that half of the students in this dataset are late for school no more than 1-2 times since the 1 in *late12* represents “1-2 Times” late for schools in twelfth grade.



Figure 5

2.23(b) According to Figure 5, we can see that the IQR = 2-0 =2.

2.23(c) According to Figure 5, we can see there’s a long whisker on the right, which means these data are skewed positively.

2.23(d) The minimal number in *late12* data is equal to the value of the 25th percentile of the *late12* data, which means at least 25% of the participant in this dataset do not experience lateness for school in 12th grade.