Mid-term for APSTA-GE 2001

Yewei Sun

N13994686

1 (a)

Stem-and-leaf plot for days\_skipped

0f | 44444444445

0s | 666677

0. | 88899

1\* | 001111

1t | 222233

1f | 55

1s | 67

1. | 8

2\* | 0

1(b) Mean=9.1

1(c) Median = 8.5

1(d) Mode = 4, the frequency is 10s

1(e) Q1 = 4.5

1(f) Q3 = 12

1(g) IQR = 7.5. It means that the middle 50% of values of days\_skipped have a spread of 7.5 days. And based on the box plot, we can see that the distribution is positively screwed.

1(h)

图表

描述已自动生成

1(i) Based on the output of “summarize” command, the skewness = 0.5573, and we can also conclude from the box plot that the distribution of this dataset is **positively skewed**.

1(j) Generate the new variable and summarize it by using:

gen days\_skipped\_double = 2 \* days\_skipped

summarize days\_skipped\_double, detail //1(j)

New Mean: 18.2

New Median: 17

New Mode: 8

1(k) New Standard Deviation = Old SD \* 2 = 9

1(l) The skewness will remain the same. Run the command:

summarize days\_skipped\_double, detail

we can gain the skewness = 0.5573

1(m) Because the transformation from 20 to 40 days is linear transformation, so the Skewness new = Skewness old.

1(n) I won’t. Because John skipped lunch 9 days = Mean(boy)+0.8929\*Std. Dev., and Mary skipped lunch 9 days = Mean(girl)-0.6395\*Std. Dev., so they do not skip lunch to the same extent.

2(a) B

2(b) Based on the graph and the given stem-and-leaf plot, the median for Distribution B equals to Q2, which is around 41.5

2(c) 44%

2(d)

图表, 条形图, 直方图

描述已自动生成

(bin=4, start=47, width=4.75)

3 (a) Because males are coded 1 and females are coded 0, if participants include half male and half female, the mean of nominal variable will equals to 0.5.

3(b)

3(c) When “male”=1, the will equals to 6.45. So, a male person will skip lunch 6.45 days.

3(d) When “male”=0, the will equals to 11.75. So, a female person will skip lunch 11.75 days.

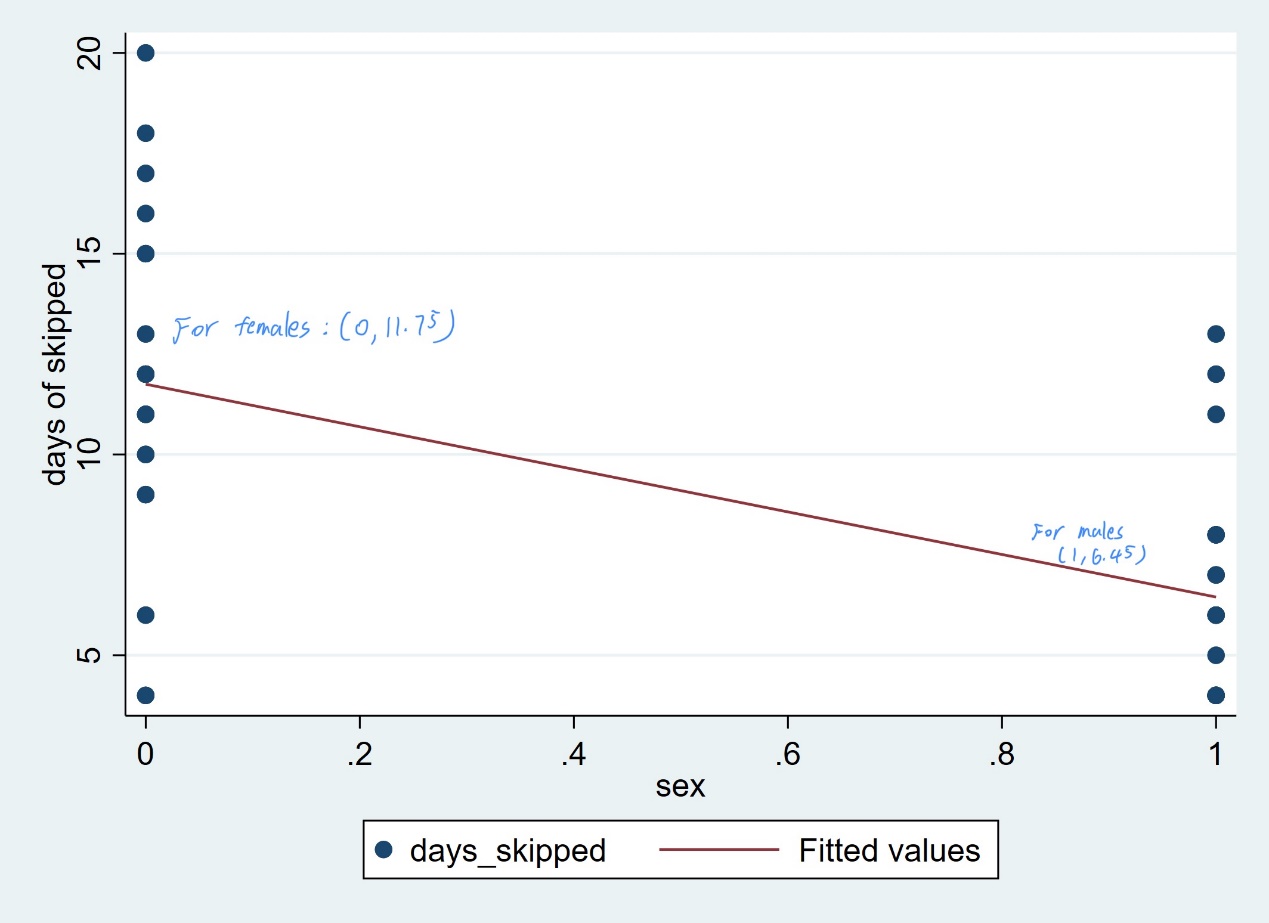
3(e) The predicted values of male are lower than those of female, indicating that male skip fewer meals than males. The value of difference equals to 5.3

3(f) Based on the =0.3576, we can know that 35.76% of the is contributed by variable “male” times its b-weight. In a scene of single dichotomous predictor, when the independent variable=1, the predicted dependent variable equals to the value of b-weight + constant, and when the independent variable=0, the predicted dependent variable equals to the value of constant. So the value of b-weight in a regression equation with a single dichotomous predictor as well as the constant contribute to the most part of the reliability of the predicted dependent variable.

3(g) Because the variable “male” is negatively associated with the predicted value.

3(h) Based on the =0.3576, we can know that 35.76% of the is contributed by variable “male” times its b-weight.

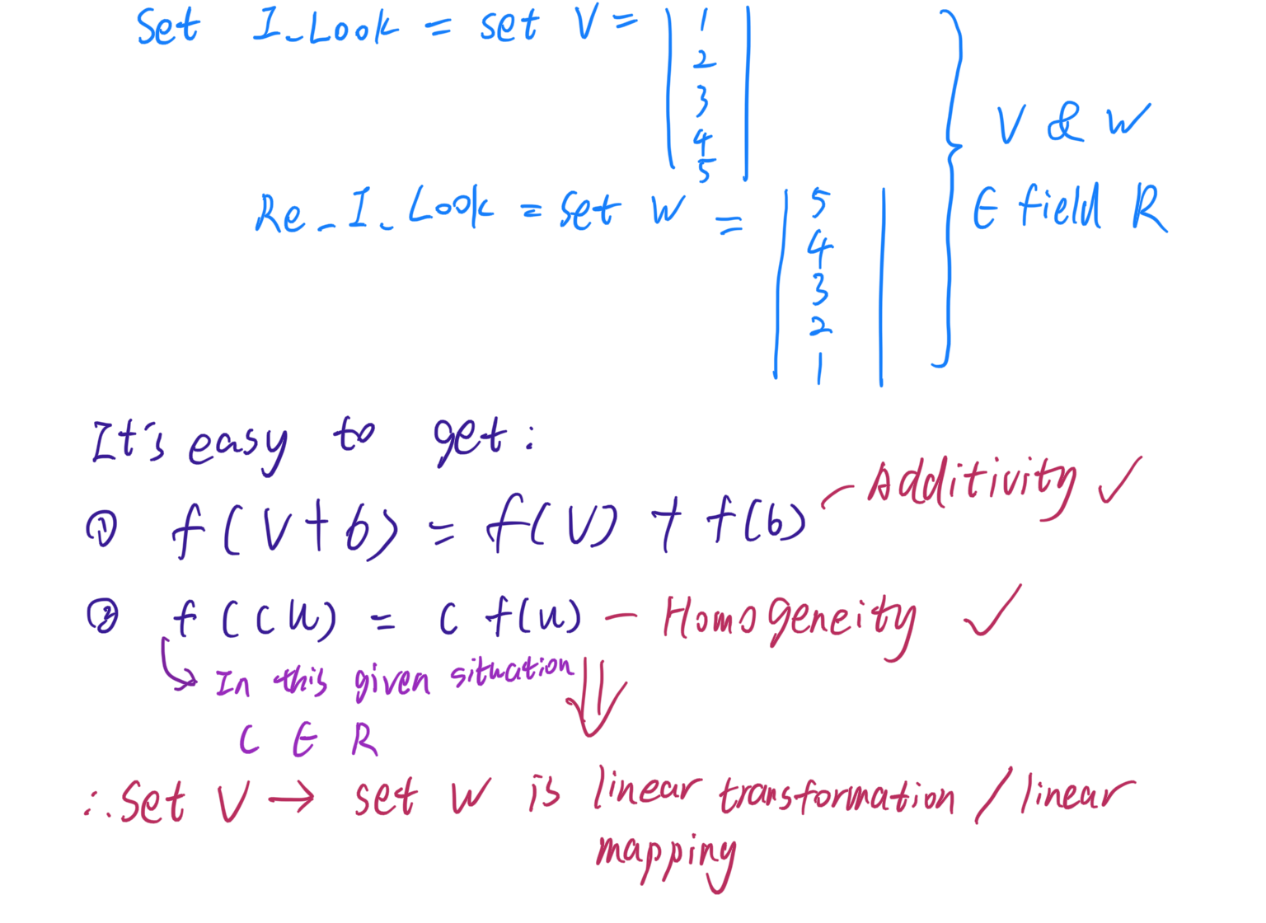
4



5(a) Re\_I\_Look = 6 - I\_Look

5(b) It is an example of a linear transformation. According to p119 of the textbook, a linear transformation are defined by rules that include only a combination of multiplication, division, addition, and subtraction to set up the one-to-one correspondence between numeric systems. This transformation is subtraction, so it is a linear transformation.

More precise and mathematical proof:



6(a) ordinal

6(b) nominal

6(c) ratio

6(d) interval

7 It will be less dramatic. In common sense, income follows the distribution of Pareto distribution, the equation is

形状

中度可信度描述已自动生成[[1]](#footnote-1)

Under this prerequisite, the median will be , . This equation is extremely positively skewed. It is easily to get that the median will locate much closer to the lowest fifth than the highest fifth. The highest fifth have a very low frequency in distribution and they are commonly extremely values. So, they will rise the mean of income to a unnecessarily high value, and under the situation of Pareto distribution, even the frequency of mean is relatively low. So, it is wiser to report the median instead of the mean.

1. Source code for this equation (I do not know why word cannot render it):

   \bar{F(x)}=\left\{\begin{matrix}

   \frac{x\_{m}}{x}^{a} & x\geq x\_{min}\\

   1 & x< x\_{min}

   \end{matrix}\right.\\

   x\_{min} = the \ minimal \ wage [↑](#footnote-ref-1)