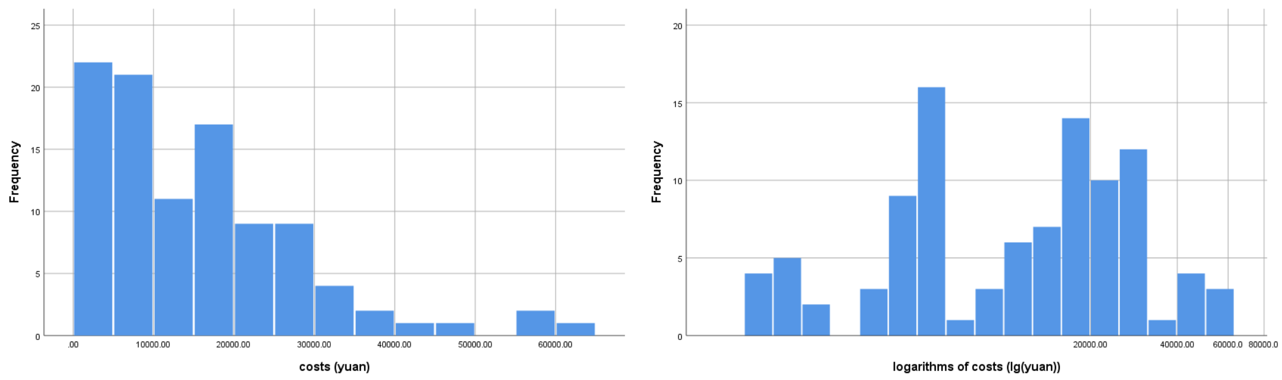
Medical Statistics
Homework 1

Class 3 莫润冰 20980131

1. Exercise 1



(a) (b)

Figure 1: Histogram of total treatment and rehabilitation costs in thousands

The histogram is higher around the center and shorter on two sides but not symmetric. The tail on the positive side is much longer than the negative side, and it is called positive skew. This represents that the majority of people's expenses in treatment and rehabilitation are below average.

1.1 Measurement of Average

As the histogram of cost is positive skew, and the logrithms is close to symmetric, the geometric mean may better represent the average level of total treatment and rehabilitation costs than the arithmetic mean.

$$arithmetic\ mean : \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 15550.30 \tag{1.1}$$

$$geometric\ mean : \bar{x} = \sqrt[n]{x_1 x_2 \cdots x_n} = 10445.64 \tag{1.2}$$

*total number of patients n = 100; x_i represents the cost of i^{th} patient

Since the histogram is taller around the center and shorter on two sides, but not pocessses symmetric feature. Median can also present the average level well.

*Github repo: https://github.com/MoRunbing/Medical_Statistics

†E-mail: morb@mail2.sysu.edu.cn

$$median : M_d = 13092.50 \tag{1.3}$$

Median is smaller than the arithmetic mean, which further proves the positive skew of the histogram.

1.2 Measurement of Variation

Range:

$$maximum\ value = 60790.00 \tag{1.4}$$

$$minimum\ value = 1435.00 \tag{1.5}$$

$$range = maximum\ value - minimum\ value = 59355.00 \tag{1.6}$$

$Q_3 - Q_1$:

Q_3 is 75 percentile of patients' costs, which indicates the value with a rank most closely to 75% of the patients, while Q_1 is 25 percentile. $Q_3 - Q_1$ discribes sample variance better than range since it excludes those extreme values.

$$Q_3 = 21176.25 \tag{1.7}$$

$$Q_1 = 5082.5 \tag{1.8}$$

$$Q_3 - Q_1 = 16093.75 \tag{1.9}$$

Varience and Standard Deviation:

Varience and standard deviation show more individual information than those values calculated above.

$$variance : S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = 172369611.53 \tag{1.10}$$

$$standard\ deviation : S = \sqrt{S^2} = 13128.96 \tag{1.11}$$

Coefficient of Variation:

Coefficient of variation is a dimentionless value suitable for comparison between different datasets.

$$coefficient\ of\ variation : CV = \frac{S}{\bar{x}} = 0.84 \tag{1.12}$$