

統計學與實習上
第一次作業

1.請說明雲端連結中

(https://docs.google.com/spreadsheets/d/1CGisbrbo9dpA10WmvWVvEGJxIhK7NqfxkvTrUwH_1_s/edit#gid=0) 你所提供的統計資訊為樣本或母體資料。(0.3 points)

2.請舉一個你關心或有興趣的議題 (可能是生活經驗或某個社會現象): (1)說明如何利用長條圖和/或次數多邊圖(折線圖)來呈現或描述這個議題;(2)簡略畫出圖示但不用特別設計數字的變化,只需標示變數名稱即可。(1 points)

3.(Text, p.14) A poll solicits a large number of college undergraduates for information on the following variables: the name of their cell phone provider (AT&T, Verizon, and so on), the numbers of minutes used last month (200, 400, for example), and their satisfaction with the service (Terrible, Adequate, Excellent, and so forth). What is the level of measurement for each of these three variables? (0.6 points)

The cell phone provider is nominal level data.

The minutes used are ratio level.

Satisfaction is ordinal level.

4. (Text, p.31) Wachesaw Manufacturing Inc. produced the following number of units in the last 16 days.

27	27	27	28	27	25	25	28
26	28	26	28	31	30	26	26

The information is to be organized into a frequency distribution. (0.6 points)

a. How many classes would you recommend?

$2^4=16$ suggests 5 classes

b. What class interval would you suggest?

$$i \geq \frac{31-25}{5} = 1.2 \quad \text{Use interval of 1.5}$$

c. What lower limit would you recommend for the first class?

24

d. Organize the information into a frequency distribution and determine the relative frequency distribution.

	f	Relative frequency
24 up to 25.5	2	0.125
25.5 up to 27	4	0.250
27 up to 28.5	8	0.500
28.5 up to 30	0	0.000
30 up to 31.5	2	0.125

Total 16 1.000

e. Comment on the shape of the distribution.

The number of units produced in the past 16 days range between 24 and 31 units. The largest concentration is in the 27 up to 28.5 class (8).

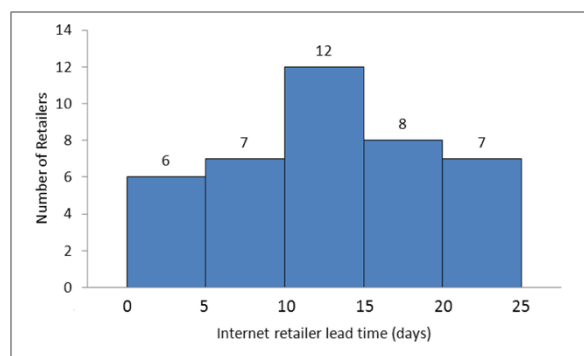
5. (Text, p.38,42) A large Internet retailer is studying the lead time (elapsed time between when an order is placed and when it is filled) for a sample of recent orders. The lead times are reported in days. **(1 points)**

Lead Time (days)	Frequency
0 up to 5	6
5 up to 10	7
10 up to 15	12
15 up to 20	8
20 up to 25	7
Total	40

a. What is the midpoint of the first class?

2.5 days

b. Draw a histogram.



c. Interpret the lead times using the histogram.

Based on the chart, the shortest lead time is 0 days, the longest 25 days. The concentration of lead times is 10-15 days.

d. Convert the frequency distribution to cumulative frequency and cumulative relative frequency distributions.

Lead Time	Cum. Freq	Cumulative Relative Frequency
Less than 5	6	.15 or 15%
Less than 10	13	.325 or 32.5%
Less than 15	25	.625 or 62.5%
Less than 20	33	.825 or 82.5%
Less than 25	40	1.00 or 100%

e. About 60% of the orders were filled in less than how many days?

14

6. 某學校針對某課程同學進行一項基本資料調查，記錄了學生的性別(0：生理女性, 1：生理男性)、身高(cm)、體重(kg)、血型(A, B, O, AB)。請使用資料 class.csv 完成以下問題。

- 請將此檔案讀入 R 語言中且命名為 class，並印出檔案的摘要(summary)。(注意：此檔案非 R 語言內建資料集) (0.3 points)
- 請使用適當的統計圖將某課程同學血型的資料進行視覺化。(注意：統計圖須包含適當的主標題與座標軸標題，並不應使圖形超出圖框) (0.5 points)
- 請繪製學生身高的累計次數多邊圖，其中身高應以每 5 公分為分群，自 150 公分至 190 公分進行區間之計算。 (0.7 points)

1. # 1.0.3 points

2. # loading data and print the summary of it

3. `class <- read.csv('~/.class.csv', header = T, sep = ',')`

4. `summary(class)`

```
> summary(class)
      sex      high      weight      blood
Min.   :0.00  Min.   :152.0  Min.   :40.00  A :24
1st Qu.:0.00  1st Qu.:164.0  1st Qu.:48.75  AB:10
Median :0.00  Median :172.0  Median :56.00  B :21
Mean    :0.45  Mean    :171.3  Mean    :58.15  O :45
3rd Qu.:1.00  3rd Qu.:178.2  3rd Qu.:65.00
Max.    :1.00  Max.    :190.0  Max.    :91.00
```

1. # 2.0.5 points

2. # Use pie chart or bar plot to visualize the blood type

3. `pie(table(class$blood),`

4. `main = 'Pie chart of blood type')`

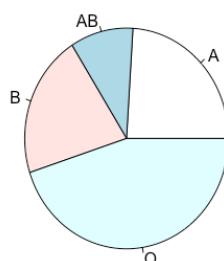
5.

6. `barplot(table(class$blood),`

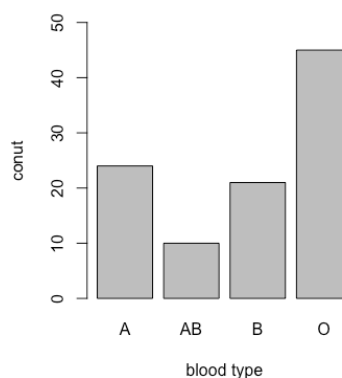
7. `main = 'Bar plot of blood type',`

8. `xlab = 'blood type', ylab = 'count', ylim = c(0,50))`

Pie chart of blood type



Bar plot of blood type



1. # 3.0.7 points

```

2. # build the cutpoint -> cutting numeric data -> build the table ->
3. # make the cumulative table and add 0 before it -> scatter plot -> line -> x axis
4. cutpoint <- seq(from=150,to=190,by=5)
5. class$high.cut <- cut(class$high,cutpoint)
6. table.high.cut <- table(class$high.cut)
7. cumu.high <- c(0,cumsum(table.high.cut))
8.
9. plot(cutpoint,cumu.high,
10.      main = 'cumulative frequency polygon of height ',
11.      xlab = 'height (cm)',ylab = 'cumulative frequency',
12.      ylim = c(0,100),
13.      pch = 16,
14.      xaxt= 'n')
15. lines(cutpoint, cumu.high)
16. axis(1,cutpoint)

```

