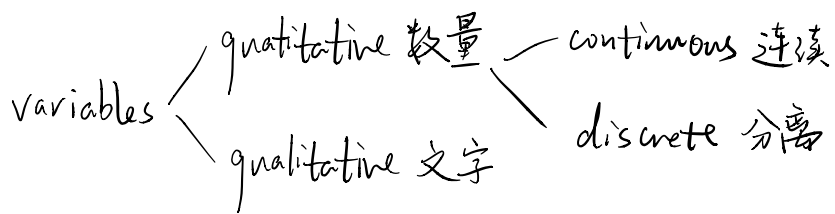


一. 模型法

二. 数据收集



三. 数据评价

• Quartiles

$$Q_1 (\text{lower } \sim) : \div 4$$

$$Q_3 (\text{upper } \sim) : \div 4 \times 3$$

• Variance

$$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

$$\text{s.d} = \sqrt{\sigma^2}$$

→ 注意, 要求 σ^2 还是 s.d

$\sigma^2 \rightarrow$ Variance

s.d \rightarrow standard deviation

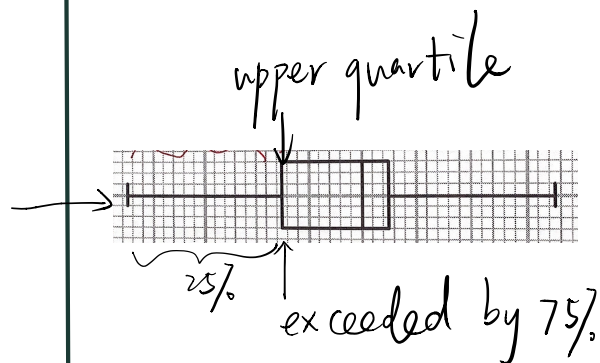
四. 数据表示

• stem & leaf diagram 茎叶图

• box plot 箱线图

• histogram 直方图

$$\text{Area} = k \times \text{frequency}$$






→ 计算要求公式

$$A = nk$$

用直方图的原因: (x轴的量) is continuous

基本特征: the area of bar is proportional to frequency

	symmetrical	posi	neg
mode, median, mean	= =	< <	> >
$\frac{3(\bar{x} - \text{median})}{s.d}$	0	+	-
$Q_2 - Q_1$ vs $Q_3 - Q_2$	=	<	>
图缘			

五. 概率

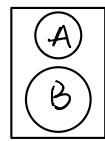
• $P(A|B) = \frac{P(A \cap B)}{P(B)}$
 "A given B"

given B, find the possibility that it was also A

分清是否有 replacement
 有时题中不会明确指出

2019.10 (2)

- $\begin{cases} \text{mutually exclusive} & P(A \cap B) = 0 \\ \text{independent} & P(A|B) = P(A) \\ & P(A \cap B) = P(A) \times P(B) \end{cases}$



每个事件概率相同 in 分布
 discrete uniform distribution

$\begin{cases} P \\ \uparrow \\ \text{probability distribution 概率分布} \\ \downarrow \\ \text{cumulative distribution 累积概率} \\ \downarrow F \end{cases}$

六. 关联性

• $S_{xx} = \sum x^2 - \frac{\sum x^2}{n}$

correlation

• $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

↑ product moment correlation

$\begin{cases} \text{close to } 1 & \text{strong} \\ \text{close to } 0 & \text{weak} \\ \approx 0 & \text{no} \end{cases}$

v. Regression

• regression line: $y = a + bx$
 $\bar{y} - b\bar{x} \rightarrow \frac{S_{xy}}{S_{xx}}$

- a reason to support fitting a regression line $|r|$ is close to 1

• coding 影响 (coding: $y = \frac{x-a}{b}$)

1. mean $\pm x$ 都变
2. σ . s.d \pm 不变 x 变
3. r 不变
4. $y = a + bx$ 代入 coding 计算
5. $\text{Var}(ax+b) = a^2 \text{Var}(x)$

λ. Discrete random variables

$$E(x) = \sum x P(x)$$

$$\text{Var}(x) = E(x^2) - [E(x)]^2$$

$$E(ax+b) = aE(x) + b$$

$$\text{Var}(ax+b) = a^2 \text{Var}(x)$$

μ. Normal distribution

$$X \sim N(\mu, \sigma^2)$$

\downarrow
 $\frac{\mu}{x}$

$$Z = \frac{x - \mu}{\sigma}$$

注意要求 in line 是
coding 前 还是后 2019.10 ①

* 论述题

1. 用直方图的原因:

(x轴的量) is continuous

2. 直方图的基本特征:

the area of bar is proportional to frequency

3. how to improve histogram to describe data more accurately
use shorter intervals

3. reason for there's a linear relation
points are close to a straight line

4. a reason to support fitting a regression line
 $|r|$ is close to 1. strong correlation

5. give an interpretation of the gradient of the regression line
x 每增加 1, y 增加 ...
* 一定要带单位

$$- y = a + bx$$

give an interpretation of the value of b the regression line
 b 每增加 1, y 增加 ...

give an interpretation of the value of a the regression line
当 $x=0$ 时, $y=a$

6. $y = at + bx$

- state response variable 自变量 (x)
x is depend on y
- state explanatory variable 因变量 (y)
x is set and y varies

7. comment on the relationship between x & y

$\begin{cases} x \uparrow & y \uparrow \\ x \uparrow & y \downarrow \end{cases}$ positive correlation
negative correlation

8. comment on the reliability of the estimate value x

$\begin{cases} \text{reliable} & x \text{ is within the range} \\ \text{unreliable} & x \text{ is beyond the range} \end{cases}$

9. 改变量对 s.d 的影响 \rightarrow 看是否接近 mean.

接近 \bar{x} , s.d \downarrow . value replaced is more corrected to mean
data is more concentrated

10. 替换/移除数据对 r 的影响 \rightarrow 看是否接近 mean.

接近 \bar{x} : replacement / removing result a better linear fit
So $|r|$ is closer to 1.