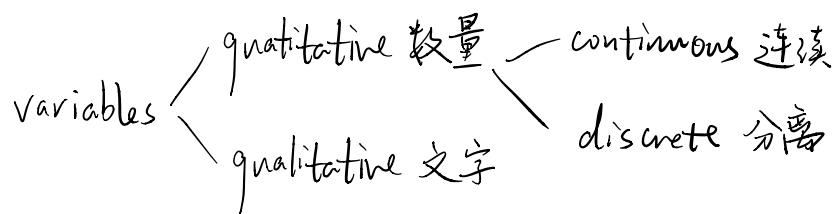


一. 模型法

二. 数据收集



三. 数据评价

- Quartiles

$$Q_1(\text{lower } \sim) := \frac{1}{4}$$

$$Q_3(\text{upper } \sim) := \frac{1}{4} \times 3$$

- Variance

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

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

→ 注意要求 σ^2 选定 s.d
 $\sigma^2 \rightarrow$ Variance
 $s.d \rightarrow$ standard deviation

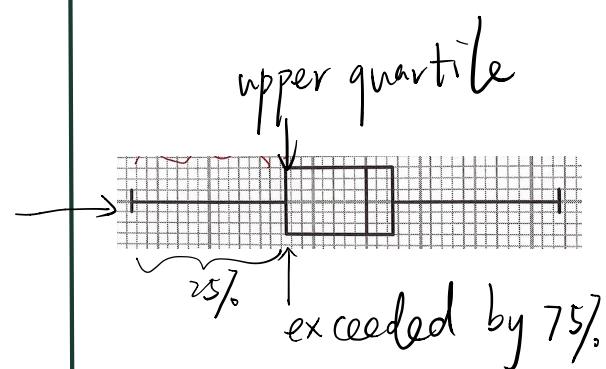
四. 数据表示

- stem & leaf diagram 茎叶图

- box plot 箱线图

- histogram 直方图

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



→ 计算量与公式
 $A = nk$

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

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

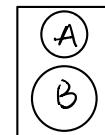
五. 概率

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

"A given B"

given B, find the possibility that it was also A

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



分清楚是否有 replacement
有时题中不会明确指出

2019.10 (2)

每个事件概率相同分布:
discrete uniform distribution

P
probability distribution 概率分布
 F
cumulative distribution 累积概率

六. 相关性

- $S_{xx} = \sum x^2 - \frac{\sum x^2}{n}$ correlation
- $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

close to 1	strong
close to 0	weak
≈ 0	no

product moment correlation

7. Regression

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

注意原 in line 及
coding 前 还是后 2019.10.①

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

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

1. mean $\pm x \div$ 都变

2. s. d. $+/-$ 不变 $x \div$ 变

3. r 不变

4. $y = a + b\bar{x}$ 代入 coding 计算

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

八. Discrete random variables

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

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

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

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

九. Normal distribution

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

\downarrow
 \bar{x}

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

* 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...

⇒ 一定要带二单位

$$- y = a + bx$$

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. $y = atbx$

- 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

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

8. comment on the reliability of the estimate value \bar{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 of $s.d$ in 影响 \rightarrow 离不接近 mean.

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

10. 裁換 / 移除 in 数据对 r 的影响 \rightarrow 离不接近 mean.

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