1 Measurement

Scales of Measurement

1. nominal: 没有 order 的 categories
2. ordinal: 有 order
3. interval: 数值按照等长区间分类
4. ratio: 单点的数值

数据的分类

1. Categorical / Qualitative: Nominal / Ordinal
2. Numerical / Quantitative: Discrete / Continuous

Basic Quantities

`quantile(arr, 0.25)`: Q_1
$Q_{1,2,3}$: 25%, 50%, 75% percentile
$\text{IQR} = Q_3 - Q_1$

Skewness: 看尾巴在哪边

1. Left-Skewed: Negative Skewness
2. Right-Skewed: Positive Skewness

Why trimmed mean?

1. Sometimes may have a lower SE when data is not normal
2. Balance between median and mean, protect against outliers

画图

1. Stem and leaf plot: 左边是数字第一位，右边是后面的，中间用 | 隔开（`stem(x)`）
2. Histogram: `hist(x)`

transformation

log 把中心往右，exp 把中心往左

Log-normal distribution: $\log X \sim \mathcal{N}(\mu, \sigma^2)$
$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)}{2\sigma^2}}$
$\mu = e^{\mu + \frac{\sigma^2}{2}}$, $\sigma^2 = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$
Coefficient of Variation (CV): $\frac{\sigma}{\mu}$
Geomean = $\sqrt[n]{\prod X_i}$

Check Normal

Imposing a Normal PDF on the Histogram

```
hist(x)
xpt <- seq(from, to, by=by)
n_den <- dnorm(xpt, mean(return), sd(return))
ypt <- n_den * length(x) * 10
lines(xpt, ypt, col="blue")
```