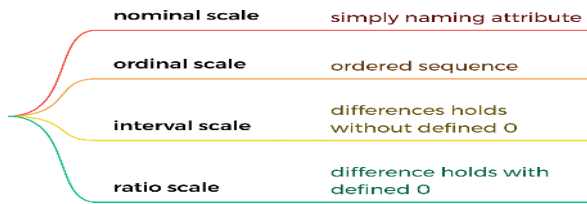
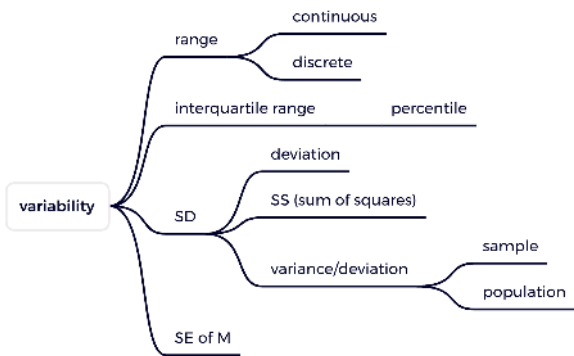


✓ Scale of measurement



✓ Variability



> z-score: $Z = \frac{x-\mu}{\sigma}$

✓ Bayes theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$

✓ Distributions

Binomial $X \sim B(n, p)$	$P(x) = C_n^x p^x (1-p)^{n-x}$	$E(X) = np,$ $Var(X) = np(1-p)$
Normal distribution $X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	$E(X) = \mu$ $Var(X) = \sigma^2$
Exponential distribution	$f(x) = \lambda e^{-\lambda x},$ for $x > 0, \lambda > 0$	$E(X) = \lambda^{-1}$ $Var(X) = \lambda^{-2}$
Poisson distribution	$P(k \text{ events}) = \frac{\lambda^k e^{-\lambda}}{k!}$	$E(X) = \lambda$ $Var(X) = \lambda$
t distribution	$X_i \sim N(\mu, \sigma^2), \text{ then } \frac{\sqrt{n}(X_n - \mu)}{s} \sim t_{n-1}$	

Uniform distribution

✓ CLT: $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), E(X) = \mu, Var(X) = \sigma^2$

✓ LLN (law of large number): the larger the sample size n, the closer \bar{X}_n be to the population mean.

✓ Hypothesis testing

	H_0 True	H_0 False
H_1 True	No error $1 - \alpha$	Type II error β
H_1 False	Type I error α	No error $1 - \beta$

> α and β : inversely correlated, reduced with greater n

✓ Power

Null Hypothesis	Power	Target Sample Size
$H_0: \mu < \mu_0$	$\phi\left(z_\alpha + \frac{ \mu_0 - \mu_1 }{\sigma/\sqrt{n}}\right)$	$n = \frac{(z_{1-\beta} + z_{1-\alpha})^2 \sigma^2}{(\mu_0 - \mu_1)^2}$
$H_0: \mu > \mu_0$	$\approx \phi\left(z_\alpha + \frac{ \mu_0 - \mu_1 }{\sigma/\sqrt{n}}\right)$	$n = \frac{(z_{1-\beta} + z_{1-\alpha/2})^2 \sigma^2}{(\mu_0 - \mu_1)^2}$

> factors

Significance level, α	-
Effect size (Alternative mean and null mean)	+
Measuring error (SD)	-
Sample size	+

✓ Effect size: depends on size of effect & size of sample

$$\text{Cohen's } d = \frac{M - \mu}{\sigma} \left(\sigma \xrightarrow{t \text{ test}} s \right)$$

> Measuring effect size with t statistic

$$\text{percentage of variance accounted for} = r^2 = \frac{t^2}{t^2 + df}$$

✓ Z-test: $z = \frac{M - \mu}{SEM}$

✓ Paired t test

$$D_i = X_{2i} - X_{1i} \sim N(\Delta, \sigma_D^2)$$

$$t = \frac{M_D - \mu_D}{S_{MD}}$$

✧ Plots

- histogram
 - width extends to the real limits
 - bars adjacent to each other
- polygon
 - height corresponds to frequency
 - additional two points of zero
- bar graph
 - gaps left between
 - widely used for nominal, ordinal scales and frequency, absolute levels
- stem and leaf display
 - leaf consisting of final digits
- regular vs grouped frequency distribution table

X	f	p = f/n	% = p(100)
5	1	1/10 = 0.10	10%
4	2	2/10 = 0.20	20%
3	3	3/10 = 0.30	30%
2	3	3/10 = 0.30	30%
1	1	1/10 = 0.10	10%

X	f
50-54	3
55-59	4
60-64	5
65-69	4
70-74	3
75-79	1
80-84	1
85-89	1
90-94	1

✧ Concepts:

Parameter, statistic, real limits (URL, LRL), apparent limits, **range**. PMF, CDF, PDF

Percentile (from lower), distribution of sample means.

连续函数刻画离散问题时, 注意精确上下限!