uniform distribution: (minimum, maximum)

probability density function (PDF) $\begin{cases} \frac{1}{max-min+1} & min \le x \le max \\ 0 & (otherwise) \end{cases}$

p.m.f: $P(X=k) = \frac{1}{b-a+1}$

mean: $\mu = \frac{a+b}{2}$

cumulative distribution function (CDF) $P(X \le x) = \begin{cases} \frac{x - min}{max - min} & min \le x \le max \\ 1 & x > max \end{cases}$

poisson distribution: (mean $= \lambda$)

pdf: $P(X = k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!} & (k \geq 0) \\ 0 & (otherwise) \end{cases}$

cdf: $P(X \le k) = \begin{cases} Q(\lfloor k \rfloor + 1, \lambda) & k \geq 0 \\ 0 & (otherwise) \end{cases}$

($\lfloor k \rfloor$ is a floor function, $Q(a,x)$ is the regularized incomplete grammar function)

binomial distribution: (number of trials "n" (positive integer), probability of success "p" (0 or 1))

pdf: $P(X = x) = \begin{cases} p^x \binom{n}{x} (1-p)^{n-x} & 0 \le x \le n \\ 0 & (otherwise) \end{cases}$

$\binom{n}{m}$ is the binomial coefficient

cdf: $P(X = x) = \begin{cases} I_{1-p}(n - \lfloor x \rfloor, \lfloor x \rfloor + 1) & 0 \le x < n \\ 1 & x \geq n \end{cases}$

$I_z(a,b)$ is the regularized incomplete beta function

exponential distribution: (rate "λ" (positive))

$P(T=t)$

pdf: $\begin{cases} \lambda e^{-t\lambda} & t \geq 0 \\ 0 & otherwise \end{cases}$

$P(T \le t)$

cdf: $\begin{cases} 1 - e^{-t\lambda} & t \geq 0 \\ 0 & otherwise \end{cases}$

查找0.2行和0.5列

Z distribution: (mean : $\mu = 0$) (standard deviation : $\sigma = 1$)

pdf: $\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$

Z score: $Z = \frac{(x - \mu)}{\sigma} = \frac{(6-5)}{4} = 0.25$

然后去 normal table 查相应数字

cdf: $P(X \le x) = \int_{-\infty}^{x} \frac{e^{-(\frac{x^2}{2})}}{\sqrt{2\pi}} = \frac{1}{2} erfc\left(-\frac{x}{\sqrt{2}}\right)$

t distribution: (v (positive))

pdf: $\frac{\left(\frac{v}{v+x^2}\right)^{(v+1)/2}}{\sqrt{v} \quad B(v/2, 1/2)}$

$B(a,b)$ is the beta function

cdf: $P(X \le x) = \begin{cases} \frac{1}{2} I_{\frac{v}{x^2+v}}\left(\frac{v}{2}, \frac{1}{2}\right) & x \le 0 \\ \frac{1}{2}\left(I_{\frac{x^2}{x^2+v}}\left(\frac{1}{2}, \frac{v}{2}\right) + 1\right) & \text{(otherwise)} \end{cases}$

confidence intervals: (with t distribution?)

$x \pm \frac{t_{(1-c/2)} S}{\sqrt{n}}$ t table 查询

n: sample size
x: sample mean
s: sample standard deviation
c: confidence level

confidence intervals for proportion: (proportion problems are never t-test problems — alway use z

$\hat{p} \pm \sqrt{\frac{(1-\hat{p})\hat{p}}{n}} \quad Z_{(1-c/2)}$ z table 查询

$\hat{p}$: sample proportion (hypothesised population proportion)
n: sample size
c: confidence level

$\hat{p} \pm (CV) \cdot SE(\hat{p})$

point estimate of mean; critical value; standard error of the sampling mean

95% CI for proportion

$\bar{p} \pm 1.96 * SE(\bar{p}) = \bar{p} \pm 1.96 * \sqrt{\frac{\bar{p} * (1-\bar{p})}{n}}$

Confidence Intervals for z distribution:

example: (95% CI) $\mu \pm 1.96 \boxed{\frac{\sigma}{\sqrt{n}}}$ CV, SE

μ: mean of the sample
σ: standard deviation (sample)
size: n

Confidence Intervals for t distribution:

$\begin{cases} \alpha\text{-level} = 1 - \text{desired confidence level} \\ \text{df (degree of freedom)} = \text{sample size} - 1 \ (n-1) \end{cases}$

$\bar{x} \pm CV \frac{SD}{\sqrt{n}}$

Z distribution:
90% CI → look up 95% in the table, get 1.645
95% CI → look up 97.5% in the table, get 1.96
99% CI → look up 99.5% in the table, get 2.575

Z-test:
normal proportion
population standard deviation
t-test:
normal proportion
larger sample size

Test statistic: $\frac{\bar{x} - \mu_0}{\frac{sd}{\sqrt{n}}}$

μ_0 = H_0 value for population mean
$\bar{x}$ = sample mean
n = sample size
sd = population sd

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|A')P(A')}$$

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|B')P(B')}$$

Event A: chip comes from Z $P(A) = 0.33$
Event A′: chip comes from X $P(A') = 0.67$
Event B: chip is defective $P(B|A) = 0.20$
Event B′: chip is not defective $P(B|A') = 0.10$

$$P(A|B) = \frac{0.20 \times 0.33}{0.20 \times 0.33 + 0.10 \times 0.67}$$

Standard Error

$$SE = \sqrt{\frac{(\text{Phat})(1-\text{Phat})}{n}}$$

estimate of SE of mean

$$SE = \frac{SD}{\sqrt{n}}$$

SD → sample standard deviation
$\sqrt{n}$ → sample size

(Finite Population Correction)

$$SE = \frac{SD}{\sqrt{n}} \sqrt{\frac{(N-n)}{(N-1)}}$$

N: population size
n: sample size

Critical Values depend on two decisions

1. What confidence level you choose
   - 90% CI
   - 95% CI
   - 99% CI
2. What statistical distribution you use
   - Z distribution
   - t-distribution

Type of data
- means → $n \geq 30$?
  - Yes → population SD known
    - Yes → Z distribution
    - NO → t distribution
  - NO → population normally distributed?
    - Yes → population SD known
    - NO → not in course
- proportions → $np > 10$; $n1 > 10$?
  - Yes → Z distribution
  - NO → not in course

- Types of test

  One sample tests (location tests)
  - One sample Z-test
  - One sample t-test

- Two sample tests
  - Independent
    - Two-sample independent Z-test
      - Equal variance
      - Unequal variance
    - Two-sample independent t-test
      - Equal variance
      - Unequal variance
    - Chi square test of independence
  - Paired
    - Two sample paired Z-test
    - Two sample paired t-test

Test Statistic:

$$\frac{\bar{x} - \mu_0}{\frac{sd}{\sqrt{n}}}$$

μ_0 = Ho value for population mean
$\bar{x}$ = sample mean
n = sample size
sd: Population sd if known. Sample sd if population sd unknown

$$\frac{\hat{\mu}_1 - \hat{\mu}_2 - D}{SE(\hat{\mu}_1 - \hat{\mu}_2)}$$ independent two sample test (D is often just 0, null hypothesis is no difference)

$$CI = (\bar{x}_A - \bar{x}_B) \pm z^* \, SE(\bar{x}_A - \bar{x}_B)$$

$$CI = (\bar{x}_A - \bar{x}_B) \pm t^* \, SE(\bar{x}_A - \bar{x}_B)$$

- 回归模型 (the regression model) 是线性的
- 正确 specified 回归模型
- 残差平均值为 0
- 残差应为正态分布
- 残差的均方差 (等方差) (equal variance)
- X变量和残差不相关

$$SE(\bar{\pi}_1 - \bar{\pi}_2) = \sqrt{\frac{\bar{\pi}_1 \times (1-\bar{\pi}_1)}{n_1} + \frac{\bar{\pi}_2 \times (1-\bar{\pi}_2)}{n_2}}$$

Chi square:

- Expected values: $\frac{\text{(row total)} \times \text{(column total)}}{\text{observation total}}$
- Test statistic: $\sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expect})^2}{\text{Expected}}$
- Degree of freedom = (rows −1) × (columns −1)

Type 1 Error: the probability of rejecting the Null Hypothesis when the Null Hypothesis is in fact true

Type 2 Error: the probability of failing to reject the null hypothesis when the Null Hypothesis is in fact false

	H_0 is true	H_0 is false
Fail to reject H_0	Confidence Level $(1-\alpha)$	Type 2 error (β)
reject H_0	Type 1 error (α)	power $(1-\beta)$