ST2334 Probability and Statistics

Final Examination Helpsheet

AY2024/25 Semester 1 · Prepared by Tian Xiao *@snoidetx*

1 Probability

Terminology	Definition	Example
Statistical experiment	Procedure that produces data or observations	Rolling a dice
Sample space S	Set of all possible outcomes of a statistical experiment	$\{1,2,3,4,5,6\}$
Sample point	An outcome in the sample space	1
Event	A subset of the sample space	An odd number facing up

- The sample space itself is an event and called a *sure event*.
- An event that contains no element is called a *null event* $\emptyset$.

Event: union $A \cup B$, intersection $A \cap B$, complement A'.

- Mutually exclusive/disjoint: $A \cap B = \emptyset$.
  - ▷ $P(A \cap B) = 0$ does not mean mutually exclusive (e.g., continuous).
- Contained: $A \subset B$.
- Equivalent: $A \subset B$ and $B \subset A \Leftrightarrow A = B$.

$A \cap A' = \emptyset$	$A \cap \emptyset = \emptyset$	$A \cup A' = S$	$(A')' = A$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup B = A \cup (B \cap A')$	$A = (A \cap B) \cup (A \cap B')$

- De Morgan's law: $(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n'$; $(A_1 \cap A_2 \cap \cdots \cap A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n'$.

Counting: Multiplication principle + Addition principle.

- Permutation: $P_r^n = \frac{n!}{r!(n-r)!}$.
- Combination: $C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$.

Probability: How likely event A occurs, $P(A)$.

- Relative frequency: $f_A = \frac{n_A}{n} \to P(A)$ as $n \to \infty$.
  - ▷ $0 \le f_A \le 1$;
  - ▷ $f_A = 1$ if A occurs in every repetition;
  - ▷ If A and B are mutually exclusive, then $f_{A \cup B} = f_A + f_B$.
- Axioms of probability:
  - ① For any event A, $0 \le P(A) \le 1$;
  - ② $P(S) = 1$;
  - ③ $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$.
  - ▷ $P(\emptyset) = 0$;
  - ▷ If $A_1, A_2, \cdots, A_n$ are mutually exclusive events, then $P_{A_1 \cup A_2 \cup \cdots A_n} = P(A_1) + P(A_2) + \cdots + P(A_n)$.
  - ▷ For any event A, $P(A') = 1 - P(A)$.
  - ▷ For any events A, B, $P(A) = P(A \cap B) + P(A \cap B')$.
  - ▷ For any events A, B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  - ▷ If $A \subset B$, then $P(A) \le P(B)$.
- Conditional probability: $P(B|A) = \frac{P(A \cap B)}{P(A)}$.
  - ▷ $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ if $P(A), P(B) \neq 0$.
  - ▷ $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.
- Independence: A and B are *independent* if and only if $P(A \cap B) = P(A)P(B)$. We denote this by $A \perp B$.
  - ▷ If $P(A) \neq 0$, $A \perp B$ if and only if $P(B|A) = P(B)$.
- Law of total probability: Suppose $A_1, A_2, \cdots, A_n$ is a partition of S. Then, $P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$.
- Bayes' theorem: Suppose $A_1, A_2, \cdots, A_n$ is a partition of S. Then, $P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$.

2 Random Variables

Random variable: A random variable $X : S \to \mathbb{R}$ assigns a real number to every $s \in S$.

- Range space: $R_X = \{x \mid x = X(s), s \in S\}$. Either finite or countable.
- Discrete r.v.: $R_X = \{x_1, x_2, x_3, \cdots\}$.
  - ▷ Probability (mass) function: $f(x) = \begin{cases} P(X = x) & \text{for } x \in R_X; \\ 0 & \text{for } x \notin R_X. \end{cases}$
  - ▷ Probability distribution: Collection of pairs $(x_i, f(x_i))$.
  - ▷ Properties of p.m.f.:
    - ① $f(x_i) \ge 0$ for all $x_i \in R_X$;
    - ② $f(x_i) = 0$ for all $x_i \notin R_X$;
    - ③ $\sum_{x_i \in R_X} f(x_i) = 1$.
    - ④ For any $B \subset \mathbb{R}$, $P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i)$.
- Continuous r.v.:
  - ▷ Probability density function:
    - ① $f(x) \ge 0$ for all $x \in R_X$; and $f(x) = 0$ for all $x \notin R_X$;
    - ② $\int_{R_X} f(x)\,\mathrm{d}x = 1$;
    - ③ $P(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}x$.
- Cumulative distribution function: $F(x) = P(X \le x)$.
  - ▷ $P(a \le X \le b) = P(X \le b) - P(X < a) = F(b) - F(a-)$.
  - ▷ Continuous: $P(a \le X \le b) = P(a < X < b) = F(b) - F(a)$.
  - ▷ Right continuous: $F(a) = \lim_{x \to a^+} F(x)$.
  - ▷ Convergence to 0 and 1 in the limits: $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to +\infty} F(x) = 1$.

Expectation:

- Discrete: $\mathbb{E}[X] = \mu_X = \sum_{x_i \in R_X} x_i f(x_i)$.
- Continuous: $\mathbb{E}[X] = \mu_X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_{x \in R_X} x f(x)\,\mathrm{d}x$.
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- Discrete: $\mathbb{E}[g(X)] = \sum_{x \in R_X} g(x)f(x)$; continuous: $\mathbb{E}[g(X)] = \int_{R_X} g(x)f(x)\,\mathrm{d}x$.

Variance: $\sigma_X^2 = \mathrm{V}[X] = \mathbb{E}[(X - \mu_X)^2]$.

- Discrete: $\mathrm{V}[X] = \sum_{x \in R_X} (x - \mu_X)^2 f(x)$;
- Continuous: $\mathrm{V}[X] = \mathrm{d}_{R_X}(x - \mu_X)^2 f(x)\,\mathrm{d}x$.
- $V(aX + b) = a^2\mathrm{V}[X]$.
- $\mathrm{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- Standard deviation: $\sigma_X = \sqrt{\mathrm{V}[X]}$.

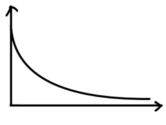
3 Special Probability Distributions

Discrete distributions:

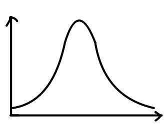
- Discrete uniform distribution:
  - ▷ P.m.f.: $f_X(x) = \begin{cases} 1/k & , x = x_1, x_2, \cdots, x_k; \\ 0 & \text{otherwise}. \end{cases}$
  - ▷ Mean: $\frac{1}{k}\sum_{i=1}^k x_i$; var: $\frac{1}{k}\sum_{i=1}^k x_i^2 - \mu_X^2$.
- Bernoulli distribution: $X \sim \mathrm{Bern}(p); p \in [0, 1]$.
  - ▷ P.m.f.: $f_X(x) = \begin{cases} p & , x = 1; \\ 1 - p & , x = 0 \end{cases} = p^x(1-p)^{(1-x)}$.
  - ▷ Mean: p; var: $p(1-p) = pq$.
- Binomial: No. of successes in n Bernoulli trials; $X \sim \mathrm{Bin}(n, p)$.
  - ▷ P.m.f.: $f_X(x) = \binom{n}{x}p^x(1-p)^{1-x}$ for $x = 0, 1, \cdots, n$.
  - ▷ Mean: np; var: $np(1-p)$.
- Negative binomial: No. of i.i.d. Bernoulli trials until k successes.
  - ▷ P.m.f.: $f_X(x) = \binom{x-1}{k-1}p^k(1-p)^{x-k}$ for $x = k, k+1, \cdots$.
  - ▷ Mean: k/p; var: $(1-p)k/p^2$.
- Geometric: No. of Bernoulli trials until first success.
  - ▷ P.m.f.: $f_X(x) = p(1-p)^{x-1}$.
  - ▷ Mean: $1/p$; var: $(1-p)/p^2$.
- Poisson: No. of occurrences in fixed time/region; $X \sim \mathrm{Poisson}(\lambda)$.
  - ▷ Mean: λ; var: λ.
  - ▷ Poisson process: Continuous-time process with rate α: $\mathrm{Poisson}(\alpha T)$.
  - ▷ Poisson approximation to binomial: Let $X \sim \mathrm{Bin}(n, p)$. Suppose $n \to \infty, p \to 0$ s.t. np constant, then
    $$\lim_{n \to \infty; p \to 0} \Pr[X = x] = \frac{e^{-np}(np)^x}{x!}.$$
    - ∗ Good when $n \ge 20, p \le 0.05$ or $n \ge 100, np \le 10$.

Continuous distributions:

- Continuous uniform: $X \sim U(a, b)$
  - ▷ p.d.f.: $f_X(x) = \begin{cases} \frac{1}{b-a} & , a \le x \le b \\ 0 & , \text{otherwise}. \end{cases}$
  - ▷ c.d.f.: $F_X(x) = \begin{cases} 0 & , x < a \\ \frac{x-a}{b-a} & , a \le x \le b \\ 1 & , x > b. \end{cases}$
  - ▷ Mean: $\frac{a+b}{2}$; variance: $\frac{(b-a)^2}{12}$.
- Exponential: $X \sim \mathrm{Exp}(\lambda)$.
  - ▷ p.d.f.: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & , x \ge 0 \\ 0 & , x < 0. \end{cases}$
  - ▷ c.d.f.: $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & , x \ge 0 \\ 0 & , x < 0. \end{cases}$
  - ▷ Mean: $\frac{1}{\lambda}$; variance: $\frac{1}{\lambda^2}$.
  - ▷ Alternative form: Parameter $\mu = \frac{1}{\lambda}$. $f_X(x) = \begin{cases} \frac{1}{\mu}e^{-\frac{x}{\mu}} & , x \ge 0. \\ 0 & , x < 0. \end{cases}$
  - ▷ Theorem: $\Pr[X > s + t \mid X > s] = \Pr[X > t]$. ("memoryless").



- Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - ▷ p.d.f.: $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < +\infty$.
  - ▷ Mean: μ; variance: σ^2.
  - ▷ Let $Z = \frac{X - \mu}{\sigma}$, then $Z \sim \mathcal{N}(0, 1)$.
    $$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$
    $$\Pr[Z \ge 0] = \Pr[Z \le 0] = \Phi(0) = 0.5.$$



  - ▷ $\Phi(z) = \Pr[Z \le z] = \Pr[Z \ge -z] = 1 - \Phi(-z)$.
  - ▷ If $Z \sim \mathcal{N}(0, 1)$, then $-Z \sim \mathcal{N}(0, 1)$.
  - ▷ If $Z \sim \mathcal{N}(0, 1)$, then $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$.
  - ▷ α-upper quantile: $\Pr[Z \ge z_\alpha] = \alpha$.
    - ∗ $z_{0.05} = 1.645$; $z_{0.01} = 2.326$.
  - ▷ Normal approximation to binomial:
    Let $X \sim \mathrm{Bin}(n, p)$ s.t. $\mathbb{E}[X] = np$, $\mathbb{V}[X] = np(1-p)$.
    As $n \to \infty$, $Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}} = \frac{X - np}{\sqrt{np(1-p)}} \to \sim \mathcal{N}(0, 1)$.
  - ▷ Continuity correction:
    - $\Pr[X = k] \approx \Pr[k - \frac{1}{2} < X < k + \frac{1}{2}]$
    - $\Pr[a \le X \le b] \approx \Pr[a - \frac{1}{2} < X < b + \frac{1}{2}]$
    - $\Pr[a < X \le b] \approx \Pr[a + \frac{1}{2} < X < b + \frac{1}{2}]$
    - $\Pr[a \le X < b] \approx \Pr[a - \frac{1}{2} < X < b - \frac{1}{2}]$
    - $\Pr[a < X < b] \approx \Pr[a + \frac{1}{2} < X < b - \frac{1}{2}]$
    - $\Pr[X \le c] = \Pr[0 \le X \le c] \approx \Pr[-\frac{1}{2} \le X < c + \frac{1}{2}]$
    - $\Pr[X > c] = \Pr[c < X \le n] \approx \Pr[c + \frac{1}{2} < X < n + \frac{1}{2}]$

4 Sampling and Sampling Distributions

Population: The totality of all possible outcomes or observations.

- Population can be finite or infinite.

Sample: Any subset of a population.

- Simple random sample (SRS): Every subset of n observations of the population has the same probability of being selected.
- SRS from infinite population: $(X_1, X_2, \cdots, X_n)$ independent r.v.'s.
  - ▷ Joint p.f.: $f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$.

Statistic: A function of $(X_1, \cdots, X_n)$.

- Sample mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$.
  - ▷ Realization: $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$.
- Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$.
  - ▷ Realization: $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$.

Sampling distribution: The probability distribution of a statistic.

Distribution of $\overline{X}$: $\mu_{\overline{X}} = \mathbb{E}[\overline{X}] = \mu_X$; $\sigma^2_{\overline{X}} = \mathbb{V}[\overline{X}] = \frac{\sigma^2_X}{n}$.

- Standard error: $\sigma_{\overline{X}}$, spread of sampling distribution.
- Law of large numbers: $\Pr[|\overline{X} - \mu| > \epsilon] \to 0$ as $n \to \infty$.
- Central limit theorem: $\overline{X} \to \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ as $n \to \infty$. Equivalently, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to Z \sim \mathcal{N}(0,1)$.
  - ▷ Population is symmetric with no outlier: $15 \sim 20$;
  - ▷ Population is moderately skewed such as exponential or χ^2: $30 \sim 50$;
  - ▷ Population is extremely skewed: $\gg 1000$.

χ^2-distribution: A r.v. with same distribution as i.i.d. $Z_1^2 + \cdots + Z_n^2$ is called a χ^2 r.v. with n degrees of freedom, denoted as $\chi^2(n)$.

- If $Y \sim \chi^2(n)$, then $\mathbb{E}[Y] = n$ and $\mathbb{V}[Y] = 2n$.
- For large n, $\chi^2(n)$ is approximately $\mathcal{N}(n, 2n)$.
- If Y_1 and Y_2 are independent χ^2 r.v.'s with m and n degress of freedom respectively, then $Y_1 + Y_2$ is a χ^2 r.v. with $m+n$ degrees of freedom.
- All χ^2 p.d.f.'s have a long right tail.
- Define $\chi^2(n;\alpha)$ s.t. for $Y \sim \chi^2(n)$, $\Pr[Y > \chi^2(n;\alpha)] = \alpha$.

Distribution of $(n-1)S^2/\sigma^2$: $\chi^2(n-1)$.

t-distribution: Suppose $Z \sim \mathcal{N}(0,1)$ and $U \sim \chi^2(n)$. If Z and U are independent, then $T = \frac{Z}{\sqrt{U/n}}$ follows the (Student's) t-distribution with n degrees of freedom, denoted as $t(n)$.

- If $T \sim t(n)$, then $\mathbb{E}[T] = 0$ and $\mathbb{V}[T] = \frac{n}{n-2}$ for $n > 2$.
- $t(n) \to \mathcal{N}(0,1)$ as $n \to \infty$. When $n \geq 30$, replace it by $\mathcal{N}(0,1)$.
- Its graph is symmetric.
- Define $t_{n;\alpha}$ s.t. for $T \sim t(n)$, $\Pr[T > t_{n;\alpha}] = \alpha$.
- If $X_1, \cdot, X_n$ are i.i.d. normal r.v. $\sim \mathcal{N}(\mu, \sigma^2)$, then $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ follows a t-distribution with $n-1$ degrees of freedom.

F-distribution: Suppose $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are independent. Then the distribution of $F = \frac{U/m}{V/n}$ is called a F-distribution with (m,n) degrees of freedom, denoted as $F(m,n)$.

- If $X \sim F(m,n)$, then $\mathbb{E}[X] = \frac{n}{n-2}$ for $n > 2$ and $\mathbb{V}[X] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ for $n > 4$.
- if $F \sim F(m,n)$. then $1/F \sim F(n,m)$.
- Define $F(m,n;\alpha)$ s.t. for $F \sim F(m,n)$, $\Pr[F > F(m,n;\alpha)] = \alpha$.
- $F(m,n;1-\alpha) = 1/F(m,n;\alpha)$.

5 Estimation of Population Parameters

Point estimation: Estimate population parameter as a single number.

- Estimator: An *estimator* is a rule, usually expressed as a formula, that tells us how to calculate an *estimate* based on information in the sample.
- Unbiased estimator: If $\mathbb{E}[\hat{\Theta}] = \theta$.
  - ▷ S^2 is an unbiased estimator of σ^2.
- Maximum error of estimate: $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.
- Minimum sample size: $n \geq \left(\frac{z_{\alpha/2}\cdot\sigma}{E_0}\right)^2$.

	Population	σ	n	**Statistic**	E	**n given E_0 & α**
I	normal	known	any	$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$	$\left(\frac{z_{\alpha/2}\cdot\sigma}{E_0}\right)^2$
II	any	known	large	$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$	$\left(\frac{z_{\alpha/2}\cdot\sigma}{E_0}\right)^2$
III	normal	unknown	small	$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$	$t_{n-1;\alpha/2}\cdot\frac{s}{\sqrt{n}}$	$\left(\frac{t_{n-1;\alpha/2}\cdot s}{E_0}\right)^2$
IV	any	unknown	large	$Z = \frac{\overline{X}-\mu}{s/\sqrt{n}}$	$z_{\alpha/2}\cdot\frac{s}{\sqrt{n}}$	$\left(\frac{z_{\alpha/2}\cdot s}{E_0}\right)^2$

Interval estimation: A rule for calculating from the sample an interval (a,b) which you are fairly certain population parameter lies in.

- Confidence interval (CI): If $\Pr[a < \mu < b] = 1-\alpha$, then (a,b) is called the $(1-\alpha)$ *confidence interval*. $(1-\alpha)$ is *confidence level*.
- CI for the mean:

	Population	σ	n	**Confidence interval**
I	normal	known	any	$\overline{x} \pm z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$
II	any	known	large	$\overline{x} \pm z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$
III	normal	unknown	small	$\overline{x} \pm t_{n-1;\alpha/2}\cdot\frac{s}{\sqrt{n}}$
IV	any	unknown	large	$\overline{x} \pm z_{\alpha/2}\cdot\frac{s}{\sqrt{n}}$

- Interpretation: If we take a sample and compute a different CI many times, about $(1-\alpha)$ of them contain μ. "Confidence" refers to a confidence in the method used.

Comparing two populations: Make inference on $\mu_1 - \mu_2$.

- Pooled estimator: $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.
- We can roughly assume equal variance if $1/2 \leq S_1/S_2 \leq 2$.

Sample	**Confidence interval**
• Independent samples of sizes n_1 and n_2; • Pop. vars are known and unequal: $\sigma_1^2 \neq \sigma_2^2$; • Both pop. normal/both samples large (≥ 30).	$(\overline{x}-\overline{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
• Independent samples of sizes n_1 and n_2; • Pop. vars are unknown and unequal: $\sigma_1^2 \neq \sigma_2^2$; • Both samples large ($\geq 30$).	$(\overline{x}-\overline{y}) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
• Independent samples of sizes n_1 and n_2; • Pop. vars are unknown and equal: $\sigma_1^2 = \sigma_2^2 = \sigma^2$; • Both samples small ($< 30$).	$(\overline{x}-\overline{y}) \pm t_{n_1+n_2-2;\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
• Independent samples of sizes n_1 and n_2; • Pop. vars are unknown and equal: $\sigma_1^2 = \sigma_2^2 = \sigma^2$; • Both samples large ($\geq 30$).	$(\overline{x}-\overline{y}) \pm z_{\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
• Independent pairs $(X_1,Y_1),\cdots,(X_n,Y_n)$; • X_i and Y_i are dependent; • Define $D_i = X_i - Y_i, \mu_D = \mu_1 - \mu_2$; • Treat $D_1,\cdots,D_n$ as from a pop. with mean μ_D and variance σ_D^2.	n small and pop. normal: $\overline{d} \pm t_{n-1;\alpha/2}\cdot\frac{S_D}{\sqrt{n}}$ n large: $\overline{d} \pm z_{\alpha/2}\cdot\frac{S_D}{\sqrt{n}}$

6 Hypothesis Tests

① Null Hypothesis vs. Alternative Hypothesis:
- We either reject or fail to reject the null hypothesis.
- Two-sided test: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.
- One-sided test: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$.

② Level of Significance:

	Do not reject H_0	Reject H_0
H_0 is true	correct decision	Type I error
H_0 is false	Type II error	correct decision

- Level of significance: $\alpha = \Pr[\text{Type I}] = \Pr[\text{Reject } H_0 | H_0 \text{ true}]$.
  - ▷ α is usually set to be 0.05 or 0.01.
- Power of test: $1-\beta = 1-\Pr[\text{Type II}] = \Pr[\text{Reject } H_0 | H_0 \text{ false}]$.

③ Test Statistic, Distribution and Reject Region:
- Test statistic quantifies how unlikely it is to observe the sample assuming H_0 is true.
- Based on α, a decision rule divides possible values of test statistic into one *rejection/critical region* and one *acceptance region*.

④ Compute the Observed Test Statistic Value:

⑤ Conclusion:
- If the computed test statistic is within our reection region, then our sample is too improbable assuming H_0 is true, hence we reject H_0;
- Otherwise, we fail to reject H_0.

Testing mean:

Case	**Test statistic**	**Rejection region**
• Known σ^2; • Pop. normal/n large; • $H_0 : \mu = \mu_0$.	$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$	$H_1 : \mu \neq \mu_0 \Rightarrow$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ $H_1 : \mu < \mu_0 \Rightarrow z < -z_\alpha$ $H_1 : \mu > \mu_0 \Rightarrow z > z_\alpha$
• Unknown σ^2; • Pop. normal; • $H_0 : \mu = \mu_0$.	$T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$	$H_1 : \mu \neq \mu_0 \Rightarrow$ $t < -t_{n-1;\alpha/2}$ or $t > t_{n-1;\alpha/2}$ $H_1 : \mu < \mu_0 \Rightarrow t < -t_{n-1;\alpha}$ $H_1 : \mu > \mu_0 \Rightarrow t > t_{n-1;\alpha}$ $n \geq 30 \Rightarrow$ use Z

- Alternatively, we can use the p-value approach:
  - ▷ Two-sided: p-value $= \Pr[|Z| > |z|] = 2\Pr[Z > |z|]$.
  - ▷ $H_1 : \mu < \mu_0$: p-value $= \Pr[Z < -|z|]$.
  - ▷ $H_1 : \mu > \mu_0$: p-value $= \Pr[Z > |z|]$.
  - ▷ If p-value $< \alpha$, reject H_0; else, do not reject H_0.
- For two-sided test, if CI contains μ_0, then H_0 will not be rejected at level α.

Testing comparing mean:

Case	**Test statistic**	**Rejection region**
• Known σ_1^2, σ_2^2; • Pop. normal/n large; • $H_0 : \mu_1 - \mu_2 = \delta_0$.	$Z = \frac{(\overline{X}-\overline{Y})-\delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1)$	
• Unknown $\sigma_1^2 = \sigma_2^2$; • Pop. normal/n small; • $H_0 : \mu_1 - \mu_2 = \delta_0$.	$Z = \frac{(\overline{X}-\overline{Y})-\delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $\sim t_{n_1+n_2-2}$	H_1 / Rejection / p-value: $\mu_1-\mu_2 \neq \delta_0$ / $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$ / $2\Pr[Z < \|z\|]$ $\mu_1-\mu_2 > \delta_0$ / $z > z_\alpha$ / $\Pr[Z > \|z\|]$ $\mu_1-\mu_2 < \delta_0$ / $z < -z_\alpha$ / $\Pr[Z < -\|z\|]$
• Paired data; • $H_0 : \mu_D = \mu_{D_0}$.	$T = \frac{\overline{D}-\mu_{D_0}}{S_D/\sqrt{n}} \sim t_{n-1}$ if n small & pop. normal $T \sim \mathcal{N}(0,1)$ if n large	