Axioms of probability

1. $0 \leq Pr(A) \leq 1$ for any event A
2. $Pr(S) = 1$ for any sample space S
3. If $A_1, \ldots$ are mutuxally exclusive, then:

$$Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr(A_i)$$

- e.g. At the simplest case, $Pr(A \cup B) = Pr(A) + Pr(B)$ if A and B are mutually exclusive.

Properties

- $Pr(\emptyset) = 0$
- $Pr(A') = 1 - Pr(A)$
- $Pr(A \cap B) + Pr(A \cap B') = Pr(A)$
- $Pr(A) + Pr(B) - Pr(A \cap B) = Pr(A \cup B)$
- If $A \subset B$, then $Pr(A) \leq Pr(B)$

Discrete probability distribution

A random variable X that has **finite or countable infinite** number of possible values **discrete**.

Each value of X is associated with a certain probability. $f(x)$ is the **probability function** of x

Properties

- $f(x_i) \geq 0$ for all x_i
- $\sum f(x_i) = 1$

Continuous probability distribution

For a **countinuous** random variable X, $f(x)$ represents the **probability density function**.

Note: $f(x)$ outputs **proability density**, not the actual probability

Properties

- $f(x) \geq 0$ for all $x \in R_X$
- $f(x) = 0$ for all $x \notin R_X$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(c \leq X \leq d) = \int_c^d f(x)dx$
- For any value $x_0 \in X$, $Pr(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$
  - **Corollary**: $Pr(X \leq x) = Pr(X < x)$ in the continuous case.
  - **Corollary**: $Pr(A) = 0$ does not necessary imply that $A = \emptyset$.

Cumulative distribution function

For a random variable X, $F(x) = Pr(X \leq x)$

Properties

- $F(x)$ is **non-decreasing**
- $0 \leq F(x) \leq 1$

Discrete c.d.f

$$F(x) = \sum_{k \leq x} Pr(X = k)$$

Graphically, $F(x)$ takes the shape of a step function.

For any number a, b, $a \leq b$,

$$Pr(a \leq X \leq b) = Pr(X \leq b) - Pr(X < a) = F(b) - F(a^-)$$

where a^- is the largest value of X that is strictly less than a .

If all possible values of $X = x$ are integers, then we get

$$Pr(a \leq X \leq b) = F(b) - F(a - 1)$$

Corollary: Setting $a = b$, we get

$$Pr(X = a) = F(a) - F(a - 1)$$

Piecewise Function Form

$$F_X(x) = \begin{cases} 0, & x < 1 \\ 0.1, & 1 \leq x < 3, \\ 0.4, & 3 \leq x < 7, \\ 0.9, & 7 \leq x < 10, \\ 1, & \text{otherwise} \end{cases}$$

Table Form

X	1	3	7	10
$P(X = x)$	0.1	0.3	0.5	0.1

Continuous c.d.f

$$F(x) = \int_{-\infty}^{x} f(k) dk$$

$$Pr(a \leq X \leq b) = Pr(a < X \leq b) = F(b) - F(a)$$

Corollary:

$$f(x) = \frac{d}{dx} F(x)$$

when the derivative exists

Expectation

Given a random variable X which has $R_X = \{x_1, x_2, \dots\}$, and a probability function $f(x)$.

Discrete Let μ_X and $E(X)$ denote the mean/expected value.

$$\mu_X = E(X) = \sum_i x_i f(x_i) = \sum_x x f(x)$$

Continuous

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Expectation of a function of X

Given another function $g(X)$ of a random variable X with $f_X(x)$.

Discrete

$$E(g(X)) = \sum_x g(x) f_x(x)$$

Continuous

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

Properties

- $E(aX + b) = aE(X) + b$

Variance

$$\sigma_x^2 = V(X) = E((X - \mu_X)^2)$$

Properties

- $V(X) \geq 0$
- $V(X) = E(X^2) - (E(X))^2$
- $V(aX + b) = a^2 V(x)$

Chebyshev's inequality

For any **positive number** k ,

$$Pr(|X - E(X)| \geq k\sigma_X) = Pr(X \leq \mu - k\sigma \cup X \geq \mu + k\sigma) \leq \frac{1}{k^2}$$

Corollary:

$$Pr(|X - E(X)| < k\sigma_X) = Pr(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Joint probability density function

Properties (Discrete)

- $f_{X,Y}(x_i, y_j) \geq 0$ for all $(x_i, y_j) \in R_{X,Y}$
- $\sum_i \sum_j f_{X,Y}(x_i, y_j) = 1$

Properties (Continuous)

- $f_{X,Y}(x_i, y_j) \geq 0$ for all$(x_i, y_j) \in R_{X,Y}$
- $\int\int_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)dxdy = 1$

Marginal probability distribution

Discrete

$$f_X(x) = \sum_y f_{X,Y}(x,y) \quad f_Y(y) = \sum_x f_{X,Y}(x,y)$$

Continuous

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

Conditional distribution

The **conditional distribution** of X given that $Y = y$ is defined as:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{if } f(Y)(y) > 0$$

$f_{X,Y}(x,y)$		x	
		2	4
y	1	0.1	0.15
	3	0.2	0.3
	5	0.1	0.15

x	2	4
$f_X(x)$	0.4	0.6

y	1	3	5
$f_Y(y)$	0.25	0.5	0.35

x	2	4
$f_{X\|Y=3}(x)$	0.4	0.6

Independence

Random variables X, Y are independent if and only if:

$$f_{X,Y}(x,y) = f_X(x)f_Y(x) \quad \textbf{for all } x, y$$

Expectation

Discrete

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) f_{X,Y}(x,y)$$

Continuous

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$$

Covariance

The **covariance** of (X, Y) is defined as

$$Cov(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Discrete

$$Cov(X,Y) = \sum_x \sum_y (x - E(X))(y - E(Y)) f_{X,Y}(x,y)$$

Continuous

$$Cov(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) f_{X,Y}(x,y) dx dy$$

Properties

- If X and Y are independent, then $Cov(X,Y) = 0$
  - However, the converse is not true
- $Cov(aX + b, cY + d) = ac(Cov(X,Y))$
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$

Correlation coefficient

The **correlation coefficient** of X, Y is defined as follows

$$Cor(X,Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Properties

- $-1 \leq \rho_{X,Y} \leq 1$
- It measures the degree of linear relationship between X and Y
- If X and Y are independent, then $\rho_{X,Y} = 0$
  - The converse is not true

Common Probability Distributions

Discrete uniform distribution

Discrete uniform distribution is simply a random variable which can assume values $x_1, x_2, \dots, x_n$ with equal probability. Formally, the probability function is defined as,

$$f_X(x) = \begin{cases} \frac{1}{n} & \text{for } x_1, x_2, \dots, x_n \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \sum x f_X(x) = \sum_{i=0}^{n} x_i \frac{1}{n} = \frac{1}{n} \sum_{i=0}^{n} x_i$$

$$V(X) = \sum (x-\mu)^2 f_X(x) = \frac{1}{n} \sum_{i=0}^{n} (x_i - \mu)^2$$

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{n} \left(\sum_{i=0}^{n} x_i^2 \right) - \mu^2$$

Bernoulli distribution

A Bernoulli experiment is a random experiment with only two outcomes. A classic example is the experiment of flipping a (possibly biased) coin. A simple way to encode the outcome is to set $x = 0$ or 1 .

The generic form of the **Bernoulli distribution** is defined as,

$$f_X(x) = p^x (1-p)^{1-x} \quad x \in \{0, 1\}$$

$$E(X) = p \quad V(X) = p(1-p) = pq$$

Binomial distribution

Let X be the random variable denoting the **number of successes**out of that we performed n **independent** Bernoulli trials with constant probability of success p

$$X \sim B(n, p)$$
$$Pr(X = x) = f_X(x) = C_x^n p^x (1-p)^{n-x}, \quad x \in \mathbb{Z}_{\geq 0}$$

$$E(X) = np \quad V(X) = np(1-p) = npq$$

Negative binomial distribution

Suppose that our experiment has all the properties of a binomial experiment. But instead, we repeat the trials **until a fixed number of successes has occurred**.

Let X denote the **number of trials needed to achieve k successes**

$$X \sim NB(k, p)$$

$$Pr(X = x) = C_{k-1}^{x-1} p^k q^{x-k}, \quad x >= k, x \in \mathbb{Z}_{\geq 0}$$

$$E(X) = \frac{k}{p} \quad V(X) = \frac{(1-p)k}{p^2}$$

Poisson distribution

Properties

Poisson experiments has the following properties:

- The number of successes occurring in an interval (**time OR space**) are **independent** of those occurring in another disjoint interval
- The probability of a single success occurring in an interval is **proportional** to the length of the interval.
- The probability of more than 1 success occurring in a short interval is **negligible**

Let X denote the **number of successes** in an interval of fixed-length

$$X \sim P(\lambda)$$

$$Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in \mathbb{Z}_{\geq 0}$$

where λ is the average number of successes occurring in a given interval.

$$E(X) = \lambda \quad V(X) = \lambda$$

Approximation of binomial distribution

For a binomial distribution $B(n, p)$, approximately follows the poisson distribution $P(np)$ as $n \to \infty, p \to 0$

Notes: if p is close to 1, we can swap the probability of p and q and also the number of successes with the number of failures, then we can use the approximation above.

Continuous uniform distribution

$$X \sim U(a, b)$$

$$f_X(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$
$$0 \text{ otherwise}$$

$$E(X) = \frac{a+b}{2} \quad V(X) = \frac{1}{12}(b-a)^2$$

Exponential distribution

$$X \sim Exp(\alpha)$$

$$f_X(x) = \alpha e^{-\alpha x}, \quad x > 0$$
$$0 \text{ otherwise}$$

$$E(X) = \mu = \frac{1}{\alpha} \quad V(X) = \frac{1}{\alpha^2} = \mu^2$$

Memory-less Property Notice that for any two positive integers a and b , we have

$$Pr(X > a + b | X > a) = Pr(X > b)$$

Consider the probability of a car breaking down, where supposed the car has already driven for 5 hours, and the probability of it being intact in the next 5 hours is the same as the probability of it being able to drive the 1st 5 hours brand new.

Normal distribution

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

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty$$

$$E(X) = \mu$$
$$V(X) = \sigma^2$$

Properties

- The graph is symmetrical about $x = \mu$.
- The maximum point is also at $x = \mu$, with value $\frac{1}{\sigma\sqrt{2\pi}}$
- The graph approaches a horizontal asymptote in both directions
- The area under the curve is 1
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Approximation of binomial distribution

As $n \to \infty$ and $p \to 1/2$, $X \sim B(n,p)$ approximately follows $N(np, np(1-p))$

The heuristic for a good approximation is $np > 5 \quad \text{and} \quad nq > 5$

Continuity correction

$$Pr(X = k) \approx Pr(k - 0.5 < X < k + 0.5)$$

$$Pr(a \leq X \leq b) \approx Pr(a - 0.5 < X < k + 0.5)$$

$$Pr(a < X \leq b) \approx Pr(a + 0.5 < X < k + 0.5)$$

$$Pr(a \leq X < b) \approx Pr(a - 0.5 < X < k - 0.5)$$

$$Pr(a < X < b) \approx Pr(a + 0.5 < X < k - 0.5)$$

$$Pr(X \leq k) \approx Pr(-0.5 < X < k + 0.5)$$

$$Pr(X > k) \approx Pr(k + 0.5 < X < n + 0.5)$$

Distribution of Sample Means

Given a population that has mean of μ and variance of σ^2 ; when random samples of size n are drawn with replacement, the sampling distribution of the sample mean $\bar{X}$ has the following properties,

$$\mu_{\bar{X}} = \mu_X \quad \sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n}$$

Law of large numbers

Given a sample of size n from a population with mean μ and **finite** variance σ^2

The **law of large number** states that for any $\epsilon \in \mathbb{R}$

$$Pr(|\bar{X} - \mu| > \epsilon) \to 0 \quad \text{as } n \to \infty$$

Corollary

$$Pr(|\bar{X} - \mu| < \epsilon) \to 1 \quad \text{as } n \to \infty$$

In other words, as the same size gets larger, it becomes more likely that the sample mean is close to the population mean.

Central limit theorem

Given a sample of size n from a population with mean μ and **finite** variance σ^2.

If n is **sufficiently large** ($n \geq 30$),

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \text{ approximately}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ approximately}$$

Sampling distribution from normal population

If all $X_i \sim N(\mu, \sigma^2)$ (i.e. all observations are drawn from the same **normal** distribution), then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ for any sample size n .

Sampling distribution of difference of two sample means

Suppose that we have two populations with means μ_1, μ_2 and σ_1^2, σ_2^2 respectively. If we take samples of size n_1, n_2 from each respective population, then

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \text{ approximately}$$

If $n_1, n_2 \geq 30$, then the normal approximation of $\bar{X}_1 - \bar{X}_2$ is rather good regardless of the shape of the two population distributions.

Chi-square distribution

The **chi-square** distribution with n degrees of freedom is denoted by $\chi^2(n)$

$$E(X) = n \quad V(X) = 2n$$

Properties

- For large n, $\chi^2(n) \sim N(n, 2n)$ approximately
- If $X_1, \ldots, X_k$ are independent chi-square random variables with $n_1, \ldots n_k$ degrees of freedom, then $X_1 + \cdots + X_k$ also has a chi-square distribution, with $n_1 + \ldots n_k$ degrees of freedom.

$$\sum X_i \sim \chi^2(\sum n_i)$$

- If $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$
  - By standardizing, if $X \sim N(\mu, \sigma^2)$, then $(\frac{X-\mu}{\sigma^2})^2 \sim \chi^2(1)$

Connection to sampling

- Given a sample $X_1, X_2, \ldots, X_n$ taken from a normal distribution $N(\mu, \sigma^2)$,

$$\sum (\frac{X_i - \mu}{\sigma^2})^2 \sim \chi^2(n)$$
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

t-distribution

Let $Z \sim N(0,1)$, and $U \sim \chi^2(n)$. If Z and U are independent, then

$$T = \frac{Z}{\sqrt{U/n}} \sim t(n)$$

$$E(T) = 0 \quad V(T) = \frac{n}{n-2}, \quad n > 2$$

Properties

- The graph of the t -distribution is symmetric about $t = 0$
- As $n \to \infty$, $\quad T \sim N(0,1)$ approximately

Connection to sampling

- If X is drawn from a **normal** population, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} =\sim t_{n-1}$$

Given a random sample $X_1, \ldots, X_n$, the **sample variance** is defined as

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

F-distribution

Let $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, then

$$F = \frac{U/n_1}{V/n_2} \sim F(n_1 - 1, n_2, -1)$$

Connection to sampling

Suppose that we have two random samples of sizes n_1, n_2 , both obtained from two **normal** populations with variance σ_1^2, σ_2^2 respectively.

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Theorems

If $F \sim F(n, m)$, then $1/F \sim F(m, n)$

Unbiased estimator

An **unbiased estimator** ($\hat{\Theta}$) of θ satisfy

$$E(\hat{\Theta}) = \theta$$

Confidence interval for mean

Known variance

1. Population variance is known
2. Population is normal or $n \geq 30$

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Finding sample size

$$e \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

For a given margin of error e, the smallest possible sample size is given by

$$n \geq \left(z_{\alpha/2}\frac{\sigma}{e}\right)^2$$

Unknown variance

1. Population variance is unknown
2. Population is normal/approximately normal

Small sample size ($n < 30$)

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

$$\bar{X} - t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

Large sample size ($n \geq 30$)

$$\bar{X} - z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

Confidence intervals for difference of two means

Suppose that we have two populations with means μ_1, μ_2 , variance σ_1^2, σ_2^2 . Then $\bar{X}_1 - \bar{X}_2$ is a point estimator of $\mu_1 - \mu_2$

Known variance

1. σ_1^2, σ_2^2 are known and not equal
2. Populations are normal or $n_1, n_2 \geq 30$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Unknown variance

1. σ_1^2, σ_2^2 are unknown
2. **Large sample:** $n_1, n_2 \geq 30$

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Unknown but equal variance

1. σ_1^2, σ_2^2 are unknown **but equal**
2. Populations are normal (for small sample case)

Small sample size $(n_1, n_2 < 30)$

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim t(n_1 + n_2 - 2)$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$\bar{X}_1 - \bar{X}_2 - t_{n_1+n_2-2;\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + t_{n_1+n_2-2;\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Large sample size $(n_1, n_2 \geq 30)$

$$\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Paired data

Let $\mu_d = \mu_1 - \mu_2$, and the point estimate of μ_d be $\bar{d} = \frac{1}{n}\sum d_i$

$$s_d^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2$$

$$T = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} \sim t_{n-1}$$

Small, normal sample For $n \leq 30$ and population is approximately normal,

$$\bar{d} - t_{n-1;\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right) < \mu_d < \bar{d} + t_{n-1;\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right)$$

Large sample For $n \geq 30$,

$$\bar{d} - z_{\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right) < \mu_d < \bar{d} + z_{\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right)$$

Confidence interval for variance

Normal population

Let $X_1, \ldots, X_n$ be a random sample from a (approximately) normal distribution.

Known mean Suppose that μ is known.

$$\frac{(\sum X_i - \mu)^2}{\chi^2_{n;1-\alpha/2}} < \sigma^2 < \frac{(\sum X_i - \mu)^2}{\chi^2_{n;\alpha/2}}$$

Unknown mean Suppose that μ is unknown. Then

$$\frac{(n-1)S^2}{\sigma^2} = \sum\frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$$

Note: The above is true for both small and large n

Confidence interval for ratio of variance

Suppose that a sample is drawn from each normal population, of unknown means.

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

$$\frac{S_1^2}{S_2^2}\left(\frac{1}{F_{n_1-1,n_2-1;a/2}}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}(F_{n_2-1,n_1-1;a/2})$$

Type of errors

They are 2 types of errors in hypothesis testing

	H_0 **is true**	H_0 **is false**
Reject H_0	**Type I error**	Correct
Do not reject H_0	Correct	**Type II error**

Type I errors

Occurs when H_0 is rejected with H_0 is true.

The **level of significance** is denoted as

$$\alpha = Pr(\text{Type I Error}) = Pr(\text{reject } H_0|H_0)$$

Type II errors

Occurs when H_0 is not rejected when H_0 is false.

The **power of the test** is denoted as $1 - \beta$, where

$$\beta = Pr(\text{ do not reject } H_0|H_1)$$

Hence, the power of the test corresponds to the probability of committing a type II error.β is not computable unless we have a specific alternative hypothesis.

Procedure for statistical experiment

1. Select a suitable test statistic for the parameter in question
2. Set a significance level α
3. Determine the decision rule that divides the set of all possible values of the test statistic into 2 regions
    - the **rejection region/critical region** and the **acceptance region**
4. Collect samples
5. Compute test statistic
6. If test statistics assumes a value in the rejection region, reject null hypothesis

The **critical value** is the value which separates the rejection and acceptance region.

Note that this is similar to a proof by contradiction, where we assume that H_0 is true, and try to obtain a contradiction using our observed sample statistic.

Hypothesis testing concerning mean

Known variance

$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$

Under $H_0 : \mu = \mu_0$,

$$\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$$

Critical value approach

By setting a significance level of α, we can find two critical value $\bar{x}_1, \bar{x}_2$, such that $\bar{x}_1 < \bar{X} < \bar{x}_2$ defines the **acceptance region**. The **critical region/rejection region** is $\bar{X} < \bar{x}_1$ and $\bar{X} > \bar{x}_2$. For a two tailed test, there will be 2 critical regions.

$$Z = \frac{\bar{X} - \mu_0}{\sigma\sqrt{n}} \sim N(0,1)$$

$$\bar{x}_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \bar{x}_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

By comparing the two inequalities, we will realize that $\bar{x}_1 < \bar{X} < \bar{x}_2$ is equivalent to $-z_{\alpha/2} < Z < z_{\alpha/2}$

Hence, we will reject H_0 if z (the observed value of Z), is $> z_{\alpha/2}$ or $< -z_{\alpha/2}$

Note: the two-sided test procedure is equivalent to finding a $(1-\alpha)100\%$ confidence interval for μ . H_0 will be accepted if μ_0 is in the confidence interval.

$p-$value approach

Instead of finding the an interval for the sample mean in order to support the hypothesis, we can instead compute the **probability of obtaining a test statistic that is more extreme that what we have observed in the sample, assuming H_0 is true**.

This is also called the **observed level of significance**.

1. Convert the sample statistic (*e.g.* $\bar{X}$) to a test statistic (*e.g.* $\bar{Z}$)
2. Obtain the p -value
3. If p-value $< \alpha$, then reject H_0

Note: Compare p-value against α instead of $\alpha/2$, since the process of determining a test statistic that is more extreme has incorporated the two-tailed characteristic.

H_1	Critical (Rejection) region
$\mu > \mu_0$	$t > z_\alpha$
$\mu < \mu_0$	$t < z_\alpha$
$\mu \neq \mu_0$	$t < z_{(1-\alpha/2)}$ or $t > z_{(\alpha/2)}$

Unknown variance

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

H_1	Critical (Rejection)region
$\mu > \mu_0$	$t > t_{(n-1;\alpha)}$
$\mu < \mu_0$	$t < t_{(n-1;1-\alpha)}$
$\mu \neq \mu_0$	$t < t_{(n-1;1-\alpha/2)}$ or $t > t_{(n-1;\alpha/2)}$

Hypothesis testing concerning difference of two mean

Known variance

Large n , unknown variance

Unknown but equal variance

Paired data

Hypothesis testing concerning variance

One variance

$H_0 : \sigma^2 = \sigma_0^2$

If the underlying distribution is normal,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Under assumption of H_0 , the test statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

H_1	Critical (Rejection) region
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{(n-1;\alpha)}$
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{(n-1;1-\alpha)}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{(n-1;1-\alpha/2)}$ or $\chi^2 > \chi^2_{(n-1;\alpha/2)}$

Ratio of variance

$H_0 : \sigma_1^2 = \sigma_2^2$.

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

Under assumption of H_0, the test statistic

$$F = \frac{S_1^2}{S_2^2}$$

H_1	Critical (Rejection) region
$\sigma_1^2 > \sigma_2^2$	$F > F_{(n_1-1,n_2-1;\alpha)}$
$\sigma_1^2 < \sigma_2^2$	$F < F_{(n_1-1,n_2-1;1-\alpha)}$
$\sigma_1^2 \neq \sigma_2^2$	$F < F_{(n_1-1,n_2-1;1-\alpha/2)}$ or $F > F_{(n_1-1,n_2-1;\alpha/2)}$