ST233
AY22/23 Sem 2
github.com/SeekSaveServe

Chapter 1: Basic Concepts of Probability

Definition 1 (Experiment, Sample space, Event)
A statistical experiment is any procedure that produces data or observations.

The sample space, denoted by S, is the set of all possible outcomes of a statistical experiment. The sample space depends on the problem of interest!

A sample point is an outcome (element) in the sample space.

An event is a subset of the sample space.

Union: $A \cup B = x : x \in A \vee x \in B$
Union of n events: $\cup_{i=1}^{n} = x : x \in A_1 \vee x \in A_2 ...$
Union: $A \cap B = x : x \in A \wedge x \in B$
Interection of n events: $\cap_{i=1}^{n} = x : x \in A_1 \wedge x \in A_2 ...$
Complement: $A' = x : x \in S \wedge x \notin A$
Mutually Exclusive: $A \cap B = \emptyset$

More Event Operations
(a) $A \cap A' = \emptyset$ (b) $A \cap \emptyset = \emptyset$
(c) $A \cup A' = S$ (d) $(A')' = A$
(e) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
(f) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
(g) $A \cup B = A \cup (B \cap A')$
(h) $A = (A \cap B) \cup (A \cap B')$

De Morgan's Law
For any n events $A_1, A_2, \ldots, A_n$,
(i) $(A_1 \cup A_2 \cup \ldots \cup A_n)' = A_1' \cap A_2' \cap \ldots \cap A_n'$.
A special case: $(A \cup B)' = A' \cap B'$.
(j) $(A_1 \cap A_2 \cap \ldots \cap A_n)' = A_1' \cup A_2' \cup \ldots \cup A_n'$.
A special case: $(A \cap B)' = A' \cup B'$.

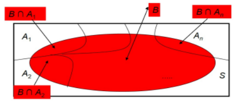
Multiplication	Addition
Sequential experiments $n_1 * n_2 * ... * n_r$ outcomes	Independent experiments $n_1 + n_2 + ... + n_r$ outcomess

Permutation: $P_r^n = \frac{n!}{(n-r)!} = n(n-1)...(n-(r-1))$
Selection and arrangement of r objects out of n. Order is considered.
Combination: $C_r^n = \frac{n!}{r!(n-r)!} = \frac{n(n-1)...(n-(r-1))}{r!}$
Selection of r objects out of n. Order is not considered.
$C_r^n = C_{n-r}^n$

Axioms and Properties of Probability
1 $0 \leq P(A) \leq 1$
2 P(S)=1
3 $P(A \cup B) = P(A) + P(B)$ if A and B are **mutex** (not to be confused with independence)
4 P($\emptyset$)=0
5 P($A_1 \cup A_2 \cup ... A_n$)=$\sum_{i=1}^{n} A_i$
6 P(A')=$1 - P(A)$
7 $P(A \cup B) = P(A) + P(B) - P(AB)$
8 P(A)=$P(A \cap B) + P(A \cap B')$
9 $A \subset B \rightarrow P(A) \leq P(B)$
10 $P(A_1) = P(A_2) = ... = P(A_k) \rightarrow for B \subset S, P(B) = \frac{|B|}{|S|}$

Conditional Probability (B occurs given A):
$P(B|A) = \frac{P(A \cap B)}{P(A)}$
Multiplication rules
$P(A \cap B) = P(A)P(B|A)$ if $P(A)! = 0$
$P(A \cap B) = P(B)P(A|B)$ if $P(B)! = 0$
Inverse Probability Formula: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
Indepedence $A \perp B iff P(A \cap B) = P(A)P(B)$ This implies $P(A|B) = P(A)$ and $P(B|A) = P(B)$
Partition $A_1, A_2 ... A_r$ are mutually exclusive and $\sum_{i=1}^{n} A_i = S$

Theorem 11 (Law of Total Probability)
Suppose $A_1, A_2, \ldots, A_n$ is a partition of S. Then for any event B, we have
$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



$P(B) = P(A)P(B|A) + P(A')P(B|A')$
Bayes' theorem
$P(A_k|B) = \frac{P(B|A)P(A_k)}{P(A)P(B|A) + P(A')P(B|A')}$
Bayes' Theorem(n=2) $P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{n} P(A_i)P(B|A_k)}$
Denom is P(B)
$P(A|B) = 1 - P(A'|B)$

L2:Random Variables

Random variable X is a function from S to R
Uppercase letters denote random variables
Lowercase letters denote observed values of random variables

Definition 3 (Probability mass function)
For a discrete random variable X, define
$$f(x) = \begin{cases} P(X = x), & \text{for } x \in R_X; \\ 0, & \text{for } x \notin R_X. \end{cases}$$
Then $f(x)$ is known as the probability function (pf), or probability mass function (pmf) of X.

The collection of pairs $(x_i, f(x_i)), i = 1, 2, 3, \ldots$, is called the probability distribution of X.

Properties of the Probability Mass Function
The probability mass function $f(x)$ of a discrete random variable X **must** satisfy:
(1) $f(x_i) \geq 0$ for all $x_i \in R_X$;
(2) $f(x) = 0$ for all $x \notin R_X$;
(3) $\sum_{i=1}^{\infty} f(x_i) = 1$, or $\sum_{x_i \in R_X} f(x_i) = 1$.
For any set $B \subset \mathbb{R}$, we have
$$P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i).$$

Definition 4 (Probability Density Function)
The probability density function of a continuous random variable X, denoted by $f(x)$, is a function that satisfies:
(1) $f(x) \geq 0$ for all $x \in R_X$; and $f(x) = 0$ for $x \notin R_X$;
(2) $\int_{R_X} f(x)\,dx = 1$;
(3) For any a and b such that $a \leq b$,
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

note: $P(a < X < b) = P(a \leq X \leq B)$
To check that a function is a PDF, check conditions 1 and 2
Cumulative Distribution Function: $F(x) = P(X \leq x)$ and
$P(a \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a-)$
Deriving CDF from probability distribution

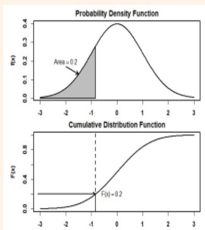
x	0	1	2
$f(x)$	1/4	1/2	1/4

We have
$$F(0) = f(0) = 1/4, F(1) = f(0) + f(1) = 3/4, F(2) = f(0) + f(1) + f(2) = 1.$$

Deriving probability distribution from cdf
a
$$F(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}.$$
- f(1)=P(X=0)=F(0)-F(0-)=$\frac{1}{4}$-0=$\frac{1}{4}$
- f(2)=P(X=1)=F(1)-F(1-)=$\frac{3}{4}$-$\frac{1}{4}$=$\frac{1}{2}$
- f(3)=P(X=2)=F(2)-F(2-)=1-$\frac{3}{4}$=$\frac{1}{4}$

CDF - Continuous Random Variable
- F(X) assumes different values in R_x when F(X)!=0
- F(X) is always non-decreasing (for discrete varaible too) $x_1 < x_2 \rightarrow F(x_1) < F(x_2)$
- Probability function and CDF is one-to-one and uniquely determined
- $0 < F(x) < 1$
- (discrete), $0 < f(x) < 1$
- (continuous), $f(x) \geq 0$ but $f(x)$ is not necessarily ≤ 1. While the PDF itself cannot exceed 1, there are points within its range that are greater than 1

CDF: Continuous Random Variable
If X is a continuous random variable,
$$F(x) = \int_{-\infty}^{x} f(t)dt,$$
and
$$f(x) = \frac{dF(x)}{dx}.$$
Further
$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a).$$



Find CDF of a continuous random variable
$$f(x) = \begin{cases} \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}, & \text{for } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x \frac{t}{\theta^2} e^{-t^2/(2\theta^2)}\,dt = \left[-e^{-t^2/(2\theta^2)}\right]_{t=0}^{x} = 1 - e^{-x^2/(2\theta^2)}.$$

Discrete: Expectation (mean) and Variance
- $E(X) = \sum_{x \in R_x} x_i f(x_i) = \mu_x$

Random: Expectation (mean) and Variance
- $E(X) = \mu_x = \int_{-\infty}^{\infty} x f(x)dx = \infty_{x \in R_x} x f(x)dx$
- The mean of X is not necessarily a possible value

Properties of Expectation
1 $E(aX + b) = aE(x) + b$
2 $E(X + Y) = E(X) + E(Y)$
3 (discrete) $E|g(x)| = \sum_{x \in R_x} g(x)f(x)$
3 (continuous) $E|g(x)| = \int_{x \in R_x} g(x)f(x)dx$
- Using properties 1, 2 we have $E(a_1X_1 + a_2X_2 + ... + a_kX + k) = a_1E(X_1) + ... + a_kE(X_k)$

Variance
- $Var(X) = \sigma_x^2 = E(X - \mu_x)^2 = E(X^2) - E(X)^2$
- (discrete) $Var(X) = \sum_{x \in R_x} (x - \mu_x)^2 f(x)$
- (continuous) $Var(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)dx$
- V(X) le 0 for any X, V(X)=0 when X is a constant

Joint Probability Distribution

Definition 5 (Discrete joint probability function)
*Let (X, Y) be a 2-dimensional **discrete** random variable. Its joint probability (mass) function is defined by*
$$f_{X,Y}(x, y) = P(X = x, Y = y),$$
for $(x, y) \in R_{X,Y}$.

Properties of the Discrete Joint Probability Function
The joint probability mass function has the following properties:
(1) $f_{X,Y}(x, y) \geq 0$ for any $(x, y) \in R_{X,Y}$.
(2) $f_{X,Y}(x, y) = 0$ for any $(x, y) \notin R_{X,Y}$.
(3) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = 1$.
Equivalently, $\sum\sum_{(x,y) \in R_{X,Y}} f(x, y) = 1$.
(4) Let A be any subset of $R_{X,Y}$, then
$$P((X, Y) \in A) = \sum\sum_{(x,y) \in A} f_{X,Y}(x, y).$$

Definition 6 (Continuous joint probability function)
*Let (X,Y) be a 2-dimensional **continuous** random variable. Its **joint probability (density) function** is a function $f_{X,Y}(x,y)$ such that*

$$P((X,Y) \in D) = \iint_{(x,y)\in D} f_{X,Y}(x,y)\,dy\,dx,$$

for any $D \subset \mathbb{R}^2$. More specifically,

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx.$$

Properties of the continuous joint probability function
The joint probability density function has the following properties:

(1) $f_{X,Y}(x,y) \ge 0$, for any $(x,y) \in R_{X,Y}$.
(2) $f_{X,Y}(x,y) = 0$, for any $(x,y) \notin R_{X,Y}$.
(3) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$.

Equivalently, $\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)\,dx\,dy = 1$.

Marginal Probability Distribution

Definition 7 (Marginal probability distribution)
*Let (X,Y) be a two-dimensional random variable with joint probability function $f_{X,Y}(x,y)$. We define the **marginal distribution** of X as follows.*

If Y is a discrete random variable, then for any x, hold x to be constant

$$f_X(x) = \sum_y f_{X,Y}(x,y).$$

If Y is a continuous random variable, then for any x,

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

x	y = 0	1	2	3	Row Total
0	0	3/84	6/84	1/84	10/84
1	4/84	24/84	12/84	0	40/84
2	12/84	18/84	0	0	30/84
3	4/84	0	0	0	4/84
Column Total	20/84	45/84	18/84	1/84	1

- Each cell is $P(X = x, Y = y)$
- Column total represents $f_Y(y)$
- Row total represents $f_X(x)$

Conditional Distribution

$f_{Y,X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, $f_{X,Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

- Only defined for $f_X(x) > 0$ and $f_Y(y) > 0$
- the summation of the PDF is 1
- $P(Y \le y|X = x) = \int_{-\infty}^{y} f_{Y|X}(y|x)dy$
- $E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$
- Replace bound with actual bounds

Normal Distribution

- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
- Shape is identical if two curves share the same σ
- Curve flattens as σ increases
- $\Phi(x)$ is the CDF of the normal distribution
- $\Phi(x) = P(x < 2)$, $\Phi(-x) = 1 - \Phi(x)$
- $P(|x| < 2) = 2\Phi(x) - 1 = p(-2 < x < 2)$
- $P(x_1 < x < x_2) = \Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})$
- $z = \frac{x-\mu}{\sigma}$
- $\phi(z) = f_Z(z)$
- $\Phi(z) = \int_{-\infty}^{z}(\phi(t))dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}(e^{-\frac{t^2}{2}})dt$

Normal apprximation to Binomail

As $n \to \infty$, $Z = \frac{X-np}{\sqrt{np(1-p)}}$

Continuity Correction - approximating normal using binomial

Continuity Correction
We apply continuity correction when approximating the binomial using the normal.

$$P(X = k) \approx P(k - 1/2 < X < k + 1/2)$$

$$P(a \le X \le b) \approx P(a - 1/2 < X < b + 1/2)$$
$$P(a < X \le b) \approx P(a + 1/2 < X < b + 1/2)$$
$$P(a \le X < b) \approx P(a - 1/2 < X < b - 1/2)$$
$$P(a < X < b) \approx P(a + 1/2 < X < b - 1/2)$$

$$P(X \le c) = P(0 \le X \le c) \approx P(-1/2 < X < c + 1/2)$$

$$P(X > c) = P(c < X \le n) \approx P(c + 1/2 < X < n + 1/2)$$

L-Example 4.19
A system is made up of 100 components, and each of which has a reliability equal to 0.90. These components function independently of one another, and the entire system functions only when at least 80 components function. What is the probability that the system functioning?

Solution:
Let X be the number of components functioning. Then $X \sim \text{Bin}(100, 0.9)$.
Thus $E(X) = (100)(0.9) = 90$ and $V(X) = (100)(0.9)(0.1) = 9$.
The system is functioning if $80 \le X \le 100$, with probability

$$P(80 \le X \le 100) \approx P\left(\frac{79.5-90}{3} < \frac{X-90}{3} < \frac{100.5-90}{3}\right)$$
$$= P(-3.5 < Z < 3.5) = \Phi(3.5) - \Phi(-3.5) = 0.9995.$$

Properties of Independent Random Variables

Definition 9 (Independent random variables)
*Random variables X and Y are **independent** if and only if for any x and y,*

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

*Random variables $X_1, X_2, \ldots, X_n$ are **independent** if and only if for any $x_1, x_2, \ldots, x_n$,*

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

- $R_{x,y}$ must be a product space (not a square)
- Any zeroes in the table indicate that $R_{x,y}$ is not product space
- If f(x,y) can be easily factorised to $c * g(x) * g(y)$, it is a product space
- **Cov(x,y)=0, E(XY)=E(X)E(Y)** if X and Y are independent

Properties of independent random variables
Suppose X, Y are independent random variables.

(1) If A and B are arbitrary subsets of $\mathbb{R}$, the events $X \in A$ and $Y \in B$ are independent events in S. Thus

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B).$$

In particular, for any real numbers x, y,

$$P(X \le x; Y \le y) = P(X \le x)P(Y \le y).$$

(2) For arbitrary functions $g_1(\cdot)$ and $g_2(\cdot)$, $g_1(X)$ and $g_2(Y)$ are independent. For example,
- X^2 and Y are independent.
- $\sin(X)$ and $\cos(Y)$ are independent.
- e^X and $\log(Y)$ are independent.

(3) Independence is connected with conditional distribution.
- If $f_X(x) > 0$, then $f_{Y|X}(y|x) = f_Y(y)$.
- If $f_Y(y) > 0$, then $f_{X|Y}(x|y) = f_X(x)$.

Uniform Distribution

Mean and Variance of the Uniform
We derive $E(X)$ and $V(X)$ for the continuous uniform distribution.
The mean is given as

$$E(X) = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

Recall that $V(X) = E(X^2) - [E(X)]^2$. So we compute

$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a} \cdot \left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} = \frac{a^2+ab+b^2}{3}.$$

Thus

$$V(X) = E(X^2) - [E(X)]^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(a^2-2ab+b^2)}{12} = \frac{(b-a)^2}{12}.$$

Cumulative Distribution Function of the Uniform
We derive the cumulative distribution function of a continuous uniform distribution.

Take note that $F_X(x) = 0$ when $x < a$, and $F_X(x) = 1$ when $x > b$.

When $a \le x \le b$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{a} 0\,dt + \int_a^x \frac{1}{b-a}\,dt = \frac{1}{b-a}[t]_a^x = \frac{x-a}{b-a}.$$

Exponential Distribution

- **Theorem 15** $\lambda > 0$ $P(X > (s+t)|X > s) = P(X > t)$ for positive values of s and t
- This is because exponential distribution is memoryless

Mean and Variance of the Exponential
We derive $E(X)$ and $V(X)$ for the exponential distribution.
The mean is given as

$$E(X) = \int_0^{\infty} x\lambda e^{-\lambda x}\,dx = \left[-xe^{-\lambda x}\right]_0^{\infty} - \int_0^{\infty}(-e^{-\lambda x})\,dx = \int_0^{\infty} e^{-\lambda x}\,dx = \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}.$$

Recall that $V(X) = E(X^2) - [E(X)]^2$. So we compute

$$E(X^2) = \int_0^{\infty} x^2\lambda e^{-\lambda x}\,dx = \left[-x^2e^{-\lambda x}\right]_0^{\infty} - \int_0^{\infty}\left(-2xe^{-\lambda x}\right)dx = \frac{2}{\lambda}\int_0^{\infty} x\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda} \cdot \frac{1}{\lambda} = \frac{2}{\lambda^2}.$$

Thus

$$V(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

We derive the cumulative distribution function of the exponential distribution with parameter λ.

For $x \ge 0$,

$$F_X(x) = P(X \le x) = \int_0^x \lambda e^{-\lambda t}\,dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}.$$

For $x < 0$, $F_X(x) = 0$.

This is a handy result to remember:

$$P(X > x) = e^{-\lambda x}, \quad \text{for } x > 0.$$

Alternative Form of the Exponential
The probability density function of the exponential distribution can be written in the following alternative form

$$f_X(x) = \begin{cases} \frac{1}{\mu}e^{-x/\mu}, & x \ge 0; \\ 0, & x < 0. \end{cases}$$

The parameters μ and λ have the relationship $\mu = 1/\lambda$.

We will then have

$$E(X) = \mu, \quad V(X) = \mu^2, \quad \text{and} \quad F_X(x) = 1 - e^{-x/\mu}, \quad \text{for } x \ge 0.$$

Central Limit Theorem

- Under CLT, sample mean approaches population mean as n increases
- Law of large numbers: $P(|\bar{x} - \mu_x| > \epsilon) \to 0$ as n increases
- Symmetric no outliers: 15-20
- Skewed: 30-50, heavily skewed: > 1000
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\to$ large n, higher accuracy

Margin of Error

Different Cases

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

$n \ge (\frac{z_\alpha * \sigma}{E})^2$

Confidence Interval

Confidence Intervals for the Mean
The table below gives the $(1-\alpha)$ confidence interval (formulas) for the population mean.

Case	Population	σ	n	Confidence Interval
I	Normal	known	any	$\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$
II	any	known	large	$\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$
III	Normal	unknown	small	$\bar{x} \pm t_{n-1;\alpha/2} \cdot s/\sqrt{n}$
IV	any	unknown	large	$\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$

Note that n is considered large when $n \ge 30$.

Alternative pooled variance:

$s_p^2 = \frac{\sum_{n_1}^{i=1}(x_i - \bar{x})^2 + \sum_{n_2}^{i=1}(y_i - \bar{y})^2}{n_1 + n_2 - 2}$

Sample Variance: $S^2 = \frac{\Sigma(X_i - \bar{x})^2}{n-1}$

P-value is the probability of obtaining a test statistic at least as extreme than the observed sample value, given H_0 is true. P-value of one sided test is half of that of a two sided.

Equal Variance Assumption: $\frac{1}{2} < \frac{s_1}{s_2} < 2$

Assumptions

- Normal population (esp for small samples)
- Random and independent sampling

	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

Definition 1 (Type I vs Type II error)
*The rejection of H_0 when H_0 is true is called a **Type I error**.*

*Not rejecting H_0 when H_0 is false is called a **Type II error**.*

Definition 2 (Significance level vs Power)
*The probability of making a Type I error is called the **level of significance**, denoted by α. That is,*

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true}).$$

Let

$$\beta = P(\text{Type II error}) = P(\text{Do not reject } H_0 \mid H_0 \text{ is false}).$$

*We define $1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ is false})$ to be the **power of the test**.*