Financial Econometrics

Basic Probability and Statistics

Tim C.C. Hung 洪志清

February 14^{th}, 2022

- **Outcomes**: the mutually exclusive potential results of a random process.
- **Sample Space**: the set of all possible outcomes.
- **Event**: a subset of the sample space.

- **Probability Distribution**:
  the proportion of the time that the outcome occurs in the long run.

- **Probability Space**:
  a triple (Ω, F, P) consists of the sample space Ω, event F, and a probability measure $P \to [0, 1]$

- **Venn Diagram**: unions, intersections, and complements.

Venn Diagram

National Taiwan University 國立臺灣大學

- Operations: $\cap$:interaction; $\cup$:Union; A^C:Event A's complement

- **Principles**:
  - $Pr(A) + Pr(A^C) = Pr(\Omega) = 1$
  - $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
    – *Principle of Inclusion and Exclusion (排容原理)*
  - $Pr((A \cap B)^C) = Pr(A^C \cup B^C)$
  - $Pr((A \cup B)^C) = Pr(A^C \cap B^C)$
    – *De Morgan's Laws*
  - $Pr(A \cup (B \cap C)) = Pr((A \cup B) \cap (A \cup C))$
  - $Pr(A \cap (B \cup C)) = Pr((A \cap B) \cup (A \cap C))$
    – *Distributive Property of Sets*

Probability Distribution - Example

National Taiwan University

- **Random Experiment**: *NBA's Most Valuable Player*.
- **Sample Outcome**: *Nikola Jokic* (2021).
- **Sample Space**:
  $\Omega =$\{*LeBron James, Stephen Curry, Kevin Durant, Giannis Antetokounmpo, Nikola Jokic, ...*\}
- **Events**:
  Event A: An European player wins.
  ($A =$\{*Giannis Antetokounmpo, Nikola Jokic, ...*\})
  Event B: A point guard wins.
  ($B =$\{*Stephen Curry, Derrick Rose, Steph Nash, ...*\})

Random Variables

National Taiwan University

- **Random Variable**: a real-valued *function* from some sample space Ω to some measurable space.
- The value depends on the particular outcome we happen to observe.
- Random variables can be discrete or continuous.
- Random variables can be scalar (univariate) or vectors (multivariate).
- Conventions: Let capital letters (X) denote the random variable and small letters (x) as a particular realization.

Random Variables - Discrete Example

National Taiwan University

- Possible outcomes for one coin toss: $\Omega = \{heads, tails\}$.
- Random variable X takes the value $x = \begin{cases} 1, if\ head \\ 0, if\ tail \end{cases}$
- X has a probability mass function $f(x) = \begin{cases} \frac{1}{2}, if\ x = 1 \\ \frac{1}{2}, if\ x = 0 \end{cases}$

Random Variables - Continuous

- Probability Function: $f(x) = P[X = x]$
- For continuous X, the measure of any given x is 0.
- Probability Density Function (PDF): describes positive probabilities to intervals in the range of x.
  - $f(x) \geq 0$
  - $P[X \in [a, b]] = P[a \leq X \leq b] = \int_a^b f(x)dx$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
- Cumulative Distribution Function (CDF): $F(x) = P[X \leq x]$

Expected Value, Mean, and Variance

- The expected value of a random variable X is a *suitably weighted average* over the range of x, or the *center* of the distribution.
  - Usually denoted as $E[X]$, μ_X, or μ.
  - Also called the **expectation of** X, or the **mean of** X
- Definition: $E[X] = \begin{cases} \sum x * f(x) & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} x * f(x)dx & if\ X\ is\ continuous \end{cases}$
- We can also take expectations of functions.

Expected Value

- Discrete example: X is the result of one toss of a fair die: $E[X] = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = \frac{7}{2}$
- Continuous example: If the pdf of Y is given by $f(y) = 1$ for $y \in [1, 2]$ and $f(y) = 0$ otherwise, $E[Y] = \int_1^2 (1 * y) dy = \frac{3}{2}$
- Bernoulli random variable B: $E[B] = 1 * p + 0 * (1 - p) = p$
- **Linearity of Expectation**: $E[X] + E[Y] = E[X + Y]$

Variance

- The variance of X measures the dispersion of X.
  It is usually denoted by $Var[X]$, or σ_X^2, or σ_X^2.

- Definition: $Var[X] = E[(X - E(X))^2]$

$$= \begin{cases} \sum (x - \mu_X)^2 f(x) & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx & if\ X\ is\ continuous \end{cases}$$

- We can also show that:

$$\begin{aligned} Var[X] &= E[(X - E(X))^2] \\ &= E[X^2 - 2(XE(X)) + (E(X))^2] \\ &= E[X^2] - E[X] * 2E[X] + (E[X])^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

Variance

- Standard Deviation: Square Root of Variance.

- Discrete example: X is the result of one toss of a fair die:
  $Var[X] = \frac{1}{6} * [(1-\frac{7}{2})^2 + (2-\frac{7}{2})^2 + (3-\frac{7}{2})^2] + (4-\frac{7}{2})^2 + (5-\frac{7}{2})^2 + (6-\frac{7}{2})^2]$
  or alternatively, $Var[X] = E[X^2] - (E[X])^2 = \frac{91}{6} - (\frac{7}{2})^2$
  $Var[X] = \frac{35}{12}$

- Bernoulli random variable B:
  $Var[B] = (1-p)^2 * p + (0-p)^2 * (1-p) = p(1-p)$
  $\sigma_B = \sqrt{p(1-p)}$

Variance

National Taiwan University

- Variance Properties: for some constants a, b, c,
  - $Var[X + c] = Var[X]$
  - $Var[cX] = c^2 Var[X]$
  - $Var[aX + b] = a^2 Var[X]$
  - $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abCov(X, Y)$
- $Cov(X, Y)$: covariance of X and Y, defined as: $E[(X - \mu_X)(Y - \mu_Y)]$
  - $Cov(X, Y) = E[XY] - E[X]E[Y]$
  - $Cov(X, X) = Var[X]$
  - $Cov(X, Y) = Cov(Y, X)$
  - $Cov(aX, bY) = abCov(X, Y)$
- For some $Y = a + bX$, $\mu_Y = a + b\mu_X$, $\sigma_Y^2 = b^2 \sigma_X^2$
- Can you proof them?

Joint and Marginal Distribution

- Discrete Version: the joint probability distribution of two discrete random variables, say X and Y, is the probability that the random variables simultaneously take on certain values, say x and y.

- The joint probability distribution can be written as the function: $Pr(X = x, Y = y)$.

- The marginal probability distribution of a random variable Y is just another name for its probability distribution (without reference to X).
$$Pr(Y = y) = \sum_{i=1}^{I} Pr(X = x_i, Y = y)$$

Joint and Marginal Distribution

National Taiwan University

- Continuous Version:
  Suppose the joint probability function for X and Y is denoted as $f(x, y)$

- $Pr(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y)dydx.$

- $f(x, y) \geq 0.$

- $\int_x \int_y f(x, y)dydx = 1.$

- The marginal probability function for X is:
  $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$

Conditional Distribution

- **Bayes' theorem**: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$

- Alternatively, $Pr(A \cap B) = Pr(A|B) * Pr(B) = Pr(B|A) * Pr(A)$, or $Pr(Y = y \cap X = x) = Pr(Y = y|X = x) * Pr(X = x)$

- If X and Y are independent, $Pr(Y = y) = Pr(Y = y|X = x)$ for any given x, vice versa.

- Thus, if X and Y are independent, $Pr(Y = y \cap X = x) = Pr(Y = y) * Pr(X = x)$

Conditional Expectation

- Conditional expectation of Y given $X = x$ is:
  $E[Y|X = x] = \sum_{i=1}^{I} y_i \; Pr(Y = y_i|X = x)$
- The mean of Y is the weighted average of the conditional expectation of Y given X, weighted by the probability distribution of X:
  $E(Y) = \sum_{j=1}^{J} E[Y|X = x_j] * Pr(X = x_j)$
- This is called, the **Law of Iterated Expectation**:
  $E[Y] = E_X[E(Y|X)]$

Conditional Variance

National Taiwan University

- The variance of Y conditional on X is the variance of the conditional distribution of Y given X.
- $Var(Y|X = x) = \sum_{i}^{I}[y_i - E(Y|X = x)]^2 * Pr(Y = y_i|X = x)$

Covariance and Correlation

- Now, back to covariance!

- Covariance is a measure of the extent to which two random variables move together.

- In the discrete version:

$$
\begin{aligned}
Cov(X,Y) &\equiv \sigma_{XY} \\
&= E[(X-\mu_X)(Y-\mu_Y)] \\
&= \sum_{i}^{I}\sum_{j}^{J}(x_i-\mu_X)(y_j-\mu_Y) * Pr(X=x_i, Y=y_j)
\end{aligned}
$$

Covariance and Correlation

- Correlation, or $Corr(X,Y)$ is defined as:
  $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$
- Random variables X and Y are considered *uncorrelated* if $Corr(X,Y) = 0$
- Also, $-1 \leq Corr(X,Y) \leq 1$

Independence and Correlation

- Independence implies non-correlation!
- Suppose $E[Y|X] = \mu_Y$ for any x

$$\begin{aligned} Cov(X,Y) &= E[XY] - \mu_X \mu_Y \\ &= E[X * E(Y|X)] - \mu_X \mu_Y \\ &= E[X] * E[Y|X] - \mu_X \mu_Y = 0 \end{aligned}$$

- However, the reverse is not necessarily true.
- *Correlation does not imply causality!*

Example

- Suppose the following is a survey of our students with/without intern experience.

	Junior (X=0)	Senior (X=1)	Total
No Intern (Y=0)	10 (25%)	2 (5%)	12 (30%)
Intern (Y=1)	10 (25%)	18 (45%)	28 (70%)
Total	20 (50%)	20 (50%)	40 (100%)

- Can you see the joint, marginal, and conditional probability of X and Y?

- Can you calculate the expected value and variance of X and Y?

- Can you calculate the covariance of X and Y? Are they independent?

Normal Distribution

- The probability density function of a normal distributed random variable (the normal PDF) is:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y}} exp[-\frac{1}{2}(\frac{y-\mu_Y}{\sigma_Y})^2]$$

- A **Standard Normal Distribution** has a mean of 0 and a standard deviation of 1, usually denoted by Z.

- Its cumulative distribution function is denoted by Φ.
  For some constant c, $Pr(Z \leq c) = \Phi(c)$.

Chi-Squared Distribution

National Taiwan University

- The Chi-squared distribution is the distribution of the sum of m squared independent standard normal random variables.
- The distribution depends on m, which is called the *degrees of freedom* of the Chi-squared distribution.
- A Chi-squared distribution with m degrees of freedom is denoted as χ_m^2.

Student t Distribution

- The Student t distribution with m degrees of freedom is defined to be the distribution of the ratio of a standard normal random variable, divided by the square root of an independently distributed Chi-squared random variable with m degrees of freedom divided by m.

- Say, there's a standard normally distributed Z and a random variable W with a Chi-squared distribution and degrees of freedom of m,

$$\frac{Z}{\sqrt{\frac{W}{m}}} \sim t_m$$

- For some m that is large enough, $t_\infty \to Z$.

F Distribution

National Taiwan University

- $F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n}$, where χ_m^2 and χ_n^2 are independent.
- When $n \to \infty$, $\chi_n^2/n \to 1$.
- $F_{m,\infty}$ is the distribution of a random variable with a Chi-squared distribution with m degrees of freedom, divided by m.
- Equivalently, the $F_{m,\infty}$ distribution is the distribution of the average of m squared standard normal random variables.

Random Sampling

National Taiwan University

- Simple random sampling is the simplest sampling scheme in which n objects are selected at random from a population and each member of the population is equally likely to be included in the sample.
- Since the members of the population included in the sample are selected at random, the values of the observations $Y_1, Y_2, ..., Y_n$ are themselves random.
- Because $Y_1, ..., Y_n$ are randomly drawn from the same population, the marginal distribution of Y_i is the same for each $i = 1, 2, ..., n$. $Y_1, ..., Y_n$ are said to be **identically distributed**.
- When $Y_1, ..., Y_n$ are drawn from the same distribution and are independently distributed, they are said to be independently and identically distributed, or **i.i.d**.

Random Sampling

- Denote the sample average $\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + ... + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$
- $\bar{Y}$ also has a distribution (called sampling distribution).
  - $E[\bar{Y}] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \frac{1}{n} * n\mu_Y = \mu_Y$
  - $Var(\bar{Y}) = Var(\frac{1}{n}\sum_{i=1}^{n} Y_i)$
    $= \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} Cov(Y_i, Y_j)$
    $= \frac{\sigma_Y^2}{n}$

Convergence in Probability

- The **Law of Large Numbers**:
  The property that $\bar{Y}$ is near μ_Y with increasing probability as n increases is called convergence in probability, or consistency.

- More formally, consider a sequence $\{S_n\}$. $\{S_n\}$ is said to converge in probability to a limit, μ, if the probability that S_n is within $\pm\delta$ of μ tends to one as $n \rightarrow \infty$ as long as the constant δ is positive.

- $S_n \xrightarrow{p} \mu$ if and only if $Pr[|S_n - \mu| \geq \delta] \rightarrow 0$ as $n \rightarrow \infty$ for all $\delta > 0$.

- If $S_n \xrightarrow{p} \mu$, then S_n is said to be a **consistent estimator** of μ.

Convergence in Distribution

- The **Central Limit Theorem**:
  The property that the distribution of $\bar{Y}$ is well approximated by a normal distribution when n is large enough.

- Since the mean of $\bar{Y}$ is μ_Y and its variance is $\frac{\sigma^2_Y}{n}$, when n is large the distribution of $\bar{Y}$ is approximately $N(\mu_Y, \sigma^2_{\bar{Y}})$

- Accordingly, $\frac{\bar{Y}-\mu_Y}{\sigma_{\bar{Y}}}$ is well approximated by the standard normal distribution $N(0, 1)$.

- Let $\{F_n\}$ be a sequence of cumulative distribution functions corresponding to a sequence of random variables $\{S_n\}$. Then, the sequence of random variables S_n is said to converge in distribution to S (denoted as $S_n \xrightarrow{d} S$) if the distribution functions $\{F_n\}$ converge to F.

Convergence in Distribution

National Taiwan University

- $S_n \xrightarrow{d} S$ if and only if $\lim_{n\to\infty} F_n(t) = F(t)$, where the limit holds at all t at which the limiting distribution F is continuous.

- The distribution F is called the **asymptotic distribution** of S_n.

- If $Y_1, ..., Y_n$ are i.i.d. and $0 < \sigma_Y^2 < \infty$, then

$$\sqrt{n}(\bar{Y} - \mu_Y) \xrightarrow{d} N(0, \sigma_Y^2)$$

- In other words, the asymptotic distribution of

$$\sqrt{n}\frac{\bar{Y} - \mu_Y}{\sigma_Y} = \frac{\bar{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} = \frac{\bar{Y} - \mu_Y}{\sigma_{\bar{Y}}} \sim N(0, 1)$$

Estimator

National Taiwan University

- An **estimator** is a rule for calculating an estimate of a given quantity based on observed data.
- Example: Estimating Population Mean (μ_Y):
  One natural way is to simply compute the sample average ($\bar{Y}$)!
- $\bar{Y}$ is not the only possible estimator. For example, the first observation Y_1 can be another estimator of μ_Y.
- What makes one estimator **better** than another? What are desirable **characteristics** of the sampling distribution of an estimator?

Estimator

- In general, we want an estimator that gets as close as possible to the unknown true value, at least in some average sense.
- In other words, we want the sampling distribution of an estimator to be as tightly centered around the unknown value as possible.
- This leads to three specific desirable characteristics of an estimator: unbiasedness, consistency, and efficiency!

Estimator

- Let $\hat{\mu_Y}$ denote some estimator of μ_Y.
- **Unbiasedness**: $E[\hat{\mu_Y}] = \mu_Y$
- **Consistency**: $\hat{\mu_Y} \xrightarrow{p} \mu_Y$
- **Efficiency**: Let $\tilde{\mu_Y}$ be another estimator of μ_Y. Suppose that both $\tilde{\mu_Y}$ and $\hat{\mu_Y}$ are unbiased estimators. If $Var(\hat{\mu_Y}) < Var(\tilde{\mu_Y})$, then $\hat{\mu_Y}$ is said to be more efficient than $\tilde{\mu_Y}$.

Estimator

- Let's compare $\bar{Y}$ and Y_1 under these criterion.
- We can see that $E[\bar{Y}] = \mu_Y$ and $\bar{Y} \xrightarrow{p} \mu_Y$
- We can also see that $E[Y_1] = \mu_Y$
- However, $Var(\bar{Y}) = \frac{\sigma_Y^2}{n} < \sigma_Y^2 = Var(Y_1)$ for $n > 1$.
- Therefore, for $n \geq 2$, $\bar{Y}$ is more efficient than Y_1.

Estimator

- Another example,
  $\tilde{Y} = \frac{1}{n}(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + ... + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n)$ for some even number n.

- $E[\tilde{Y}] = \frac{1}{2n} * \sum_{odd\ i} Y_i + \frac{3}{2n} * \sum_{even\ i} Y_i = \frac{1}{2n} * \frac{n}{2} * \mu_Y + \frac{3}{2n} * \frac{n}{2} * \mu_Y = \mu_Y$

- $Var(\tilde{Y}) = (\frac{1}{2n})^2 * \frac{n}{2} * \sigma_Y^2 + (\frac{3}{2n})^2 * \frac{n}{2} * \sigma_Y^2 = \frac{5}{4n}\sigma_Y^2 > \frac{1}{n}\sigma_Y^2 = Var(\bar{Y})$

- In fact, $\bar{Y}$ is the most efficient estimator of μ_Y among all unbiased estimators that are weighted averages of $Y_1, ..., Y_n$.

- $\bar{Y}$ is also called the **Best Linear Unbiased Estimator (BLUE)** for μ_Y.

- $\bar{Y}$ is also the least square estimator (the solution of minimizing $\sum_1^n (Y_i - M)^2$) for μ_Y. We will discuss about it shortly.

- **Hypothesis Testing**:
  A method of statistical inference used to determine a possible conclusion from two different, and likely conflicting, hypotheses.

- Null Hypothesis: usually denoted as H_0

- Alternative Hypothesis: usually denoted as H_1 or H_A.
  Could be more than one.

- E.g. Hypothesis testing for population mean.
  H_0: $E[Y] = \mu_{Y,0}$
  H_1: $E[Y] \neq \mu_{Y,0}$ (Two-sided test)

- p-Value: the probability of drawing a statistic at least as adverse to the null as the value actually computed with your data, *assuming that the null hypothesis is true.*

- If the p-value is small (e.g., less than or equal to a pre-specified level, say, 5%), then we say the null hypothesis is *unlikely* to be true, and we are in favor of the alternative hypothesis. That is, **we reject the null hypothesis.**

- Please be aware that even if the null hypothesis is not rejected, this does not mean that the null hypothesis is true. It is accepted tentatively with the recognition that it might be rejected later based on additional data.

Calculating p-Value

- Calculating the p-value based on $\bar{Y}$ (Two-sided Test):

$$p - value = Pr(\quad |\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}| \quad | \quad H_0 \ is \ true \quad)$$

where $\bar{Y}^{act}$ is the actual $\bar{Y}$ we observe (from the data).

- To compute the p-value, we need to know the sampling distribution of $\bar{Y}$ under the null hypothesis.

- If n is large, $\bar{Y}$ is well approximated by a normal distribution.

National Taiwan University

$$
\begin{aligned}
p-value &= Pr(\quad |\bar{Y}-\mu_{Y,0}| > |\bar{Y}^{act}-\mu_{Y,0}| \quad | \quad H_0 \ is \ true \quad) \\
&= Pr(\quad |\frac{\bar{Y}-\mu_{Y,0}}{\sigma_Y}| > |\frac{\bar{Y}^{act}-\mu_{Y,0}}{\sigma_Y}| \quad | \quad H_0 \ is \ true \quad) \\
&= Pr(\quad |\frac{\bar{Y}-\mu_{Y,0}}{\sigma_Y/\sqrt{n}}| > |\frac{\bar{Y}^{act}-\mu_{Y,0}}{\sigma_Y/\sqrt{n}}| \quad | \quad H_0 \ is \ true \quad) \\
&= Pr(\quad |\frac{\bar{Y}-\mu_{Y,0}}{\sigma_{\bar{Y}}}| > |\frac{\bar{Y}^{act}-\mu_{Y,0}}{\sigma_{\bar{Y}}}| \quad | \quad H_0 \ is \ true \quad) \\
&= Pr(\ |t| > |t^{act}| \ | \ H_0 \ is \ true \ )
\end{aligned}
$$

$t = \frac{\bar{Y}-\mu_{Y,0}}{\sigma_Y/\sqrt{n}}$ is the t-statistic or t-ratio.

- Question: Do we know σ_Y or $\sigma_{\bar{Y}}$?

Sample Variance

National Taiwan University

- In practice, $\sigma_{\bar{Y}}$ is unknown and needs to be estimated.

- Estimator of the variance of Y:

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

- Why $n-1$?

- If we use n instead of $n-1$, say we have $S_Y^{*2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

  See next page.

Sample Variance

$$
\begin{aligned}
E[S_Y^{*2}] &= E[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2)] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}Y_i^2 - 2\frac{1}{n}\sum_{i=1}^{n}Y_i\bar{Y} + \frac{1}{n}\sum_{i=1}^{n}\bar{Y}^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}Y_i^2 - 2\bar{Y}^2 + \bar{Y}^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}Y_i^2 - \bar{Y}^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}E[Y_i^2] - E[\bar{Y}^2] \\
&= \frac{1}{n} * n * (\mu_Y^2 + \sigma_Y^2) - (\mu_Y^2 + \frac{\sigma_Y^2}{n}) \\
&= \frac{n-1}{n}\sigma_Y^2
\end{aligned}
$$

National Taiwan University

- Try again. We can show that $E[S_Y^2] \equiv E[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2] = \sigma_Y^2$
- Furthermore, with some moderate assumption:
  If $Y_1, ..., Y_n$ are i.i.d. and $E[Y^4] < \infty$, then
  $S_Y^2 \xrightarrow{p} \sigma_Y^2$
- **Standard Deviation** of $\bar{Y}$: $\sigma_Y = \frac{\sigma_Y}{\sqrt{n}} \rightarrow$ Not Observable!
- **Standard Error** of $\bar{Y}$: $\hat{\sigma_Y} = \frac{S_Y}{\sqrt{n}}$

Type I Type II Error

- **Type I error**:
  the null hypothesis (無罪) is rejected when in fact it is true (誤判).

- **Type II error**:
  the null hypothesis (無罪) is not rejected when in fact it is false (縱放).

- The pre-specified probability of type I error is the **significance level**.

- With a pre-specified significance level (e.g. 5%):
  - Reject if $|t| > 1.96$, or equivalently, reject if < 0.05

- The probability that the test incorrectly rejects the null when it is true is the **size** of the test.

- The probability that the test correctly rejects the null when the alternative is true is the **power** of the test.

Confidence Interval

- Because of the random sampling error, it is impossible to learn the exact value of the population mean of Y using only information in a sample.
- It is possible to use data from a random sample to construct a set of values that contains the true population mean μ_Y with a certain pre-specified probability.
- $\{\mu_Y \ \mid \ |\frac{\bar{Y}-\mu_Y}{S_Y/\sqrt{n}}| \leq 1.96 \ \}$
  $\rightarrow \{\mu_Y \in (\bar{Y} - 1.96 * \frac{S_Y}{\sqrt{n}}, \bar{Y} + 1.96 * \frac{S_Y}{\sqrt{n}})\}$
- The probability that this interval contains the true value of the population mean is 95%.
- So, where is the randomness here? The confidence interval!
  It will differ from one sample to the next; the population parameter, μ_Y, is not random!

Sample Covariance and Correlation

- For two random variables X and Y.

- The population covariance and correlation can be estimated by the sample covariance and sample correlation.

- The sample covariance is:

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

- The sample correlation is:

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}, |r_{XY}| \leq 1$$