101學年　統計學(I)　授課老師：蔡碧紋

Statistical Inference：Tests of Hypotheses

Scientific questions

1. Whether the true average lifetime of a certain brand of tire is at least 22,000 kilometers.
2. whether fertilizer A produces a higher yield of soybeans than fertilizer B.
3. Pharmaceutical company: decide on the basis of samples whether at least 90% of all patients given a new medication will recover from a certain disease.
4. 廠商宣稱每杯citi coffee 重量至少 300 g.

Statistical Hypotheses testing

1. if the lifetime of the tire has pdf $f(x) = \lambda e^{-\lambda x}, x > 0$, then the expected lifetime, $\frac{1}{\lambda}$, is at least greater than 22000.
2. decide whether $\mu_A > \mu_B$, where μ_A, and μ_B are the means of the two populations
3. whether p, the parameter of a binomial distribution is greater or equal to 0.9.
4. whether $\mu > 300$.

In each case, it is assumed that the stated distribution correctly describes the experimental conditions, and the hypothesis concerns the parameter(s) of the distribution.

Hypotheses Testing

假設 (hypothesis) 就是我們對於母體參數的宣稱(claim, statement)

The Scientist formulates a statement concerning the value of a parameter.

A **test** of a statistical hypothesis is a procedure for deciding whether to “accept” or “reject” the hypothesis.

假設檢定 (hypothesis testing) 的目的就是要對這些宣稱提供統計上的檢驗, 以統計的檢定方法來推論假設的"真偽".

Terms you should know about hypotheses testing

1. Null and alternative hypotheses (H_0 vs H_1)
2. Test statistic, $T(\mathbf{X})$, and $T(\mathbf{x})$ (the distribution of $T(\mathbf{X})$)
3. Significance level of the test α
4. Rejection region (critical region) (RR) and acceptance region
5. Type I and Type II error probabilities.
6. p-value
7. power

Null and Alternative Hypotheses

Assume that the form of the distribution for the population is known , $X \sim F(x; \theta)$ where $\theta \in \Omega$, where Ω is the set of all possible values of θ can take, and is called the parameter space.

The statistical hypothesis is a statement about the value of the parameter(s) of the distribution, such as

$$\text{“}\theta \in \omega\text{”}$$

ω be a subset of Ω.
This is a statistical hypothesis and is denoted by H (H_0), called Null Hypothesis

Null and Alternative Hypotheses

On the other hand, the statement "$\theta \in \bar{\omega}$" (where $\bar{\omega}$ is the complement of ω w.r.t Ω) is called the alternative to H_0 and is denoted by H_a(or H_1).

we write

$$H_0 : \theta \in \omega \quad \text{and} \quad H_a : \theta \in \bar{\omega} \text{ (or } \theta \notin \omega)$$

H_0, H_1

1. In some case, we want to know the mean of something is as what people stated (or represents the *status quo*) we put it in the null hypothesis H_0
   主計處調查國民平均月所得為20000元
   H_0 : ________________ vs H_1 : ________________
   合歡山一月平均降雪量為20 公分。
   H_0 : ________________ vs H_1 : ________________
2. Often hypothesis arise in the form that we want to know if a new product, technique, teaching method, etc., is better than the existing one. In this context, H_0 is a statement that nullifies the theory and is sometimes called a null hypothesis. In this case, 我們把想要檢定的假設定為 H_1，H_0 則為其相反之假設。

H_0, H_1

1. if the lifetime of the tire has pdf $f(x) = \lambda e^{-\lambda x}, x > 0$, then the expected lifetime, $\frac{1}{\lambda}$, is at least greater than 22000.

   H_0 : ________________ vs H_1 : ________________
2. decide whether $\mu_A > \mu_B$, where μ_A, and μ_B are the means of the two populations

   H_0 : ________________ vs H_1 : ________________
3. whether p, the parameter of a binomial distribution is greater or equal to 0.9.

   H_0 : ________________ vs H_1 : ________________
4. 廠商宣稱每杯citi coffee 重量至少 300 g.

   H_0 : ________________ vs H_1 : ________________

Hypotheses Testing

- 假設檢定係指在尚未蒐集樣本資料、進行推論之前，就事先對母體的某種特徵性質作一合理的假設敘述，再利用隨機抽出的樣本及抽樣分配，配合機率原理，以判斷此項假設是否為真。

以統計方法進行決策的過程中，會提出兩個假設：

H_0: null hypothesis (虛無假設)。

H_1: alternative or research hypothesis(對立假設、研究假設)。

可能的結論：

1. 有足夠的統計證據可推論 H_1 為真 (reject H_0 and accept H_1)。
2. 沒有足夠的統計證據可推論 H_1 為真 (do not reject (Fail to reject, retain) H_0. The data doesn't provide

Hypotheses Testing

- 假設檢定的主要精神在於尋找證據來拒絕H_0而接受H_1， 我們無法證明H_0為絕對正確，只有不能拒斥它。

證據的角色: 假設 H_0 為真的情況下, 嘗試在其間找出矛盾,然後進行推論。

假設 H_0為真, 收集到此資料的可能性，如果是異常稀少事件（顯著的異常），則判定H_0的假設是錯誤，所以拒絕H_0.

因此假設檢定又稱為『顯著性檢定』（significant test）

simple and composite hypothesis

$$H_0 : \mu = 166 \text{ vs } H_1 : \mu > 166,$$
$$H_1 : \mu < 166 \text{ or}$$
$$H_1 : \mu \neq 166$$

If ω contains only **one point**, i.e., if $\omega = \{\theta : \theta = \theta_0\}$ then H_0 is called a **simple** hypothesis which completely specifies the null distribution. We write it as $H_0 : \theta = \theta_0$.

Otherwise, if it does not completely specify the distribution. It is called **composite** hypothesis.

Test Statistics

Test Statistics: $T(\mathbf{X})$,
a function of a set of i.i.d. random variables $X_1, \ldots, X_n$
which follow some distribution $F(x, \theta)$.
Such as $\overline{X}$ or S^2 etc.

Or a pivotal quantity for the test statistic

Significance level of the test

We defined an event with samll probability.

When H_0 is true, the probability of an extreme event such as that $\{\overline{X} > c\}$ is very small. If $\{\overline{X} > c\}$, we will reject the null hypothesis.

This small probability is called the significance level of the test, denoted by α.

$$\mathbf{P}(\overline{X} > c | H_0 : \mu = 160) = \alpha$$

Thus the hypothesis testing is also called the test of significance.

Rejection Region

If $\overline{X} > c$, then we will reject H_0. T

$\{\overline{X} > c\}$ is called the **rejection region, RR** (or **critical region**) denoted by R.

Note that critical value is defined before (ex-ante) we collect the data.

We will make the decision by comparing $\overline{x}$ with c

Decision rules

we use data (random sample) to test if the data provides significant evidence to reject the null hypothesis.

$$\text{If } \overline{X} > c \text{ reject } H_0$$

A test of hypotheses is a rule, or decision, based on a sample from a given distribution to show whether the data support our hypothesis.

Decision

Data:

$$\text{If } \overline{x} > c \text{ reject } H_0$$

Normally, the conclusion is either

1. Reject H_0 and conclude that H_1 is true, or
2. Do not reject H_0

Remark:
Note that, we don't say we"accept" H_0, because it implies stronger action than is really warranted. We can't find enough evidence to reject H_0 but it does not mean that H_0 is absolutely true.

Example: 7.2-1

Test if X, the breaking strength of a steel bar. $X \sim \mathcal{N}(50, 36)$ or $X \sim \mathcal{N}(55, 36)$. We want to know if new method increases the mean of the strength to 55.

Draw sample and test if they are from normally distributed populatioin with mean 55.

Example: 7.2-1

Test if X, the breaking strength of a steel bar. $X \sim \mathcal{N}(50, 36)$ or $X \sim \mathcal{N}(55, 36)$. We want to know if new method increases the mean of the strength to 55.

Draw sample and test if they are from normally distributed populatioin with mean 55.

1. two hypotheses: $H_0 : \mu = 50$ vs $H_a : \mu = 55$
2. test statistic: $\overline{X}$
3. Decision rule: If $\overline{X} > 53$, we will reject H_0.
   Rejection Region $R = \{\overline{x} > 53\}$

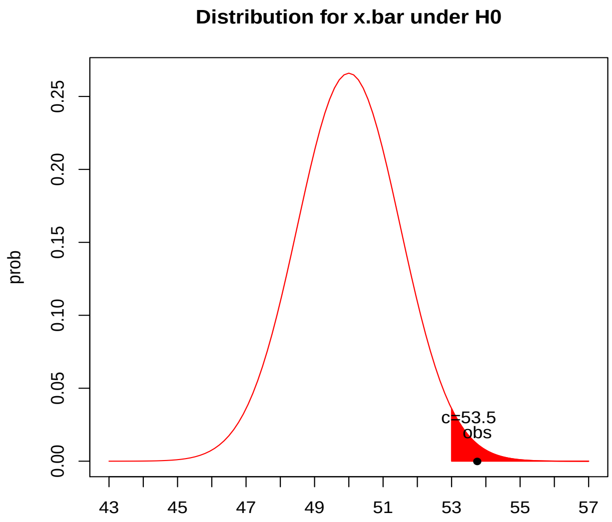
Example: 7.2-1

Test if X, the breaking strength of a steel bar.
$X \sim \mathcal{N}(50, 36)$ or $X \sim \mathcal{N}(55, 36)$. We want to know if new method increases the mean of the strength to 55.

Draw sample and test if they are from normally distributed populatioin with mean 55.

1. two hypotheses: $H_0 : \mu = 50$ vs $H_a : \mu = 55$
2. test statistic: $\overline{X}$
3. Decision rule: If $\overline{X} > 53$, we will reject H_0.
   Rejection Region $R = \{\overline{x} > 53\}$

Data: 16 random samples are drawn.
We observed $\overline{x} = 53.75$, then we will reject H_0.


Distribution for x.bar under H0
prob
0.00
0.05
0.10
0.15
0.20
0.25
43
45
47
49
51
53
55
57
c=53.5
obs

Example: The significance level of the test

The probability of a rare event when H_0 is true.

The significance level of the test is

$$\mathbf{P}\big(\overline{X} > 53 | H_0 : X_i \sim \mathcal{N}(50, 36)\big)$$

$$\mathbf{P}(\frac{\overline{X} - 50}{6/\sqrt{16}} > \frac{53 - 50}{6/\sqrt{16}}) = \mathbf{P}(Z > 12/3) = 0.023$$

The significance level of the test given $R = \{\overline{x} > 53\}$ is $\alpha =$0.02.

Type I and Type II error probabilities

Probability of making a wrong decision:

There are two types of errors that can occurs.

		Our decision	
		Reject H_0	Accept H_0
Actual situation	H_0 is true	Type I error	good
	H_0 is false	good	Type II error

Probabilities associated with the two incorrect decisions are denoted by type I and type II error probabilities.

Type I and Type II error probabilities

1. Reject H_0 when it is true.
   $\alpha = \mathbf{P}(\text{ Type I error }) = \mathbf{P}(\text{ reject a true } H_0) =$
   $\mathbf{P}(\text{ reject } H_0 | H_0 \text{ is true})$

2. Fail to reject H_0 when it is false (Fail to accept H_1 when H_1 is true)
   $\beta = \mathbf{P}(\text{ Type II error }) = \mathbf{P}(\text{ Retain a false } H_0) =$
   $\mathbf{P}(\text{ retain } H_0 | H_1 \text{ is true})$

Test statistics I

Often we work with the distribution of $T(\mathbf{X})$, pivotal quantity for the test statistic, such as a standard normal, t, χ^2, or F.

Pivotal quantity: a function of (data) observations and unknown parameters whose distribution does not depend on the unknown parameters

Test statistics II

1. If $X_1, \cdots X_n \sim \mathcal{N}(\mu, \sigma^2)$ with μ is the unknown parameter and σ^2 is some known constant, we have $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/\sqrt{n})$,
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$
2. If $X_1, \cdots X_n \sim \mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are unknown parameters, we have
$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Test statistics III

3. $X_1, \cdots, X_n \sim$ Bernoulli$p$ and $Y = X_1 + \cdots + X_n$ we have
$$\frac{(Y/n) - p}{\sqrt{\frac{p(1-p)}{n}}} \to \mathcal{N}(0,1)$$

4. $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are unknown parameters and let $S^2 = \sum(X_i - \overline{X})^2/(n-1)$.
$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Test statistics IV

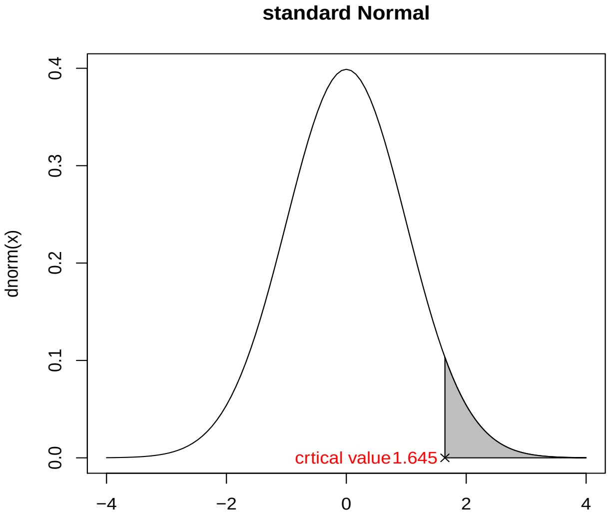
Based on the distribution of the test statistic we define the **Critical value** t^*, such as

$z_\alpha, t_\alpha, \chi^2_\alpha$, or F_α and

construct the decision rule (rejection region) for a given the significance level of the test α.

Critical value for α=0.05

EX: $\mathbf{P}(Z \geq 1.645) = 0.05, \quad z_{0.05} = 1.645$



The procedure

1. Specify the null and alternative hypotheses.
2. Specify the significant level of the test α. (control the Type I error)
3. Define a test statistic $T(\mathbf{X})$ and its distribution under H_0
4. Decision rule: obtain the rejection region $R = \{\mathbf{x} : T(\mathbf{x}) \in R(\theta_0)\}$.
5. Obtain the data and calculate the value of the test statistic $T(\mathbf{x})$ (T_{obs})
6. Conclusion: If the test statistic $T(\mathbf{x}) \in R(\theta_0)$ reject H_0 and conclude that there is strong evidence to reject the null hypothesis at the significant level α

Example (n=16, $\overline{x} = 53.75$)

1. Hypotheses: $H_0 : \mu = 50$ vs $H_a : \mu = 55$
2. Given $\alpha = 0.05$ (Significance level of the test)
3. Test statistic:
$$Z = \frac{\overline{X} - 50}{6/\sqrt{16}} \sim^{H_0} N(0,1)$$
4. Decision rule: if $Z_{\text{obs}} > Z_{0.05} = 1.645$ (critical value)
   Reject H_0
5. Data $\overline{x} = 53.75$ we have
$$z_{\text{obs}} = \frac{53.75 - 50}{6/4} = 2.5$$
6. Becuse $z_{\text{obs}} = 2.5 > 1.645$
   Reject $H_0 : \mu = 50$ at $\alpha = 0.05$.

Probability value (p-value) of the test

When H_0 is true, the probability that the test statistic is equal to or exceeds the actually observed value toward the direction of the alternative hypothesis.

Tail-end probability under H_0 toward H_1.

$$\mathbf{P}(\overline{X} \geq 53.75|H_0) = \mathbf{P}(Z \geq \frac{3.75}{6/4}) = \phi(2.5) = 0.006$$

$\overline{X} \sim^{H_0} \mathcal{N}(50, 36/16)$

The p-value of the test is 0.006.

Small *p*-value provides evidence to reject the null hypothesis H_0 given the data.

Decision rule

If the p-value of a thest if as small or smaller than the significance level of a test, α, we say the the data are statistically significant at an α significant level.

The p-values : if p-value $< \alpha$ reject H_0 at significance level α.
(We don't need to find different t^* for different significance level α.

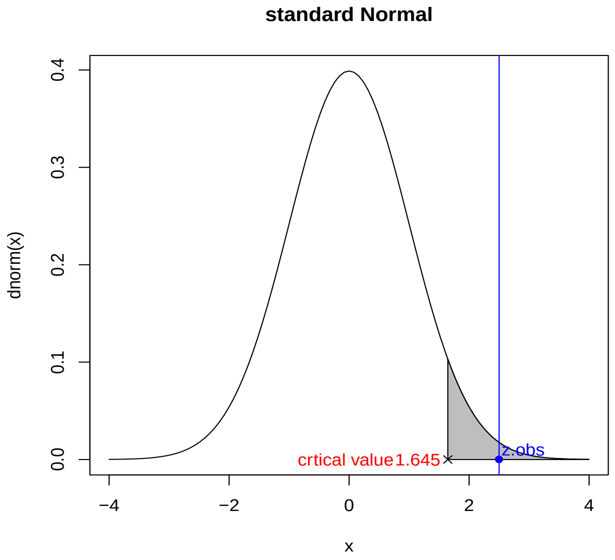
Decision rule

If the p-value of a thest if as small or smaller than the significance level of a test, α, we say the the data are statistically significant at an α significant level.

The p-values : if p-value < α reject H_0 at significance level α.
(We don't need to find different t^* for different significance level α.

Recall: Decision rule by critical value

$$\text{if } T(\mathbf{x}) \in R(\theta_0) \text{ reject } H_0 \text{ at significance level } \alpha.$$


standard Normal
0.4
0.3
0.2
0.1
0.0
dnorm(x)
crtical value1.645
z.obs
−4
−2
0
2
4
x

Power

Power=**P**(Accept H_1 when H_1 is true)

Power=$1-\beta$, $1-\beta$ is defined as the **power** of the test.

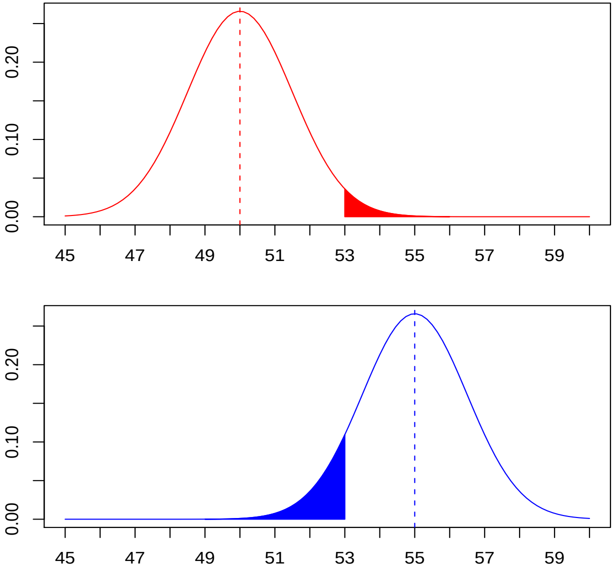
α, β and Power

Example (Ex 7.2-1)

Let X be the breaking strength of a steel bar. $H_0 : \mu = 50$ vs $H_a : \mu = 55$ Given $C = \{(x_1, \cdots, x_n) : \overline{x} \geq 53\}$ or $C = \{\overline{x} : \overline{x} \geq 53\}$ Data: n=16, what are the type I and type II error probabilities?

$\overline{X} \sim N(50, 36/16)$ under $H_0 : \mu = 50$

$\overline{X} \sim N(55, 36/16)$ under $H_1 : \mu = 55$


0.00
0.10
0.20
45
47
49
51
53
55
57
59
0.00
0.10
0.20
45
47
49
51
53
55
57
59

α, β and Power

Type I error rate = 0.0228

Type II error rate = 0.0912

The significance level of the test $\alpha =$0.0228.

The power of the test is $1 - \beta$=0.9088.

The relationship between α and β