Lecture 11: 12/1/2021

由信賴區間到假設檢定

Distribution theory of $\bar{X}$: Overview

Case 1: σ is known

Fact 1a: $X_i \sim^{iid} N(\mu,\sigma^2)$,

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

Fact 1b: $X_i \sim^{iid}$ non-normal with $E(X_i)=\mu$ and $Var(X_i)=\sigma^2,\ n\to\infty$,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \underset{\substack{n\to\infty \\ \text{approximately}}}{\sim} N(0,1) \quad \text{(by the \textit{Central Limit Theorem})}$$

Case 2: σ is unknown and estimated by

$$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-X)^2}{n-1}} \quad (\text{分母一定為 } n-1)$$

Fact 2a: $X_i \sim^{iid} N(\mu,\sigma^2)$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \underset{\substack{\forall n \\ \text{exactly}}}{\sim} T_{(n-1)} \quad \text{(T distribution with } n-1 \text{ degree of freedom)}$$

Fact 2b: $X_i \sim^{iid}$ non-normal with $E(X_i)=\mu$ and $Var(X_i)=\sigma^2,\ n\to\infty$

$$S\to\sigma \quad \& \quad \frac{\bar{X}-\mu}{S/\sqrt{n}} \underset{\substack{n\to\infty \\ \text{approximately}}}{\sim} T_{\infty}=Z$$

Case 3: σ is unknown and n is small

Fact 3a: If the data $(X_1,\ldots,X_n)$ are *roughly symmetric*, then

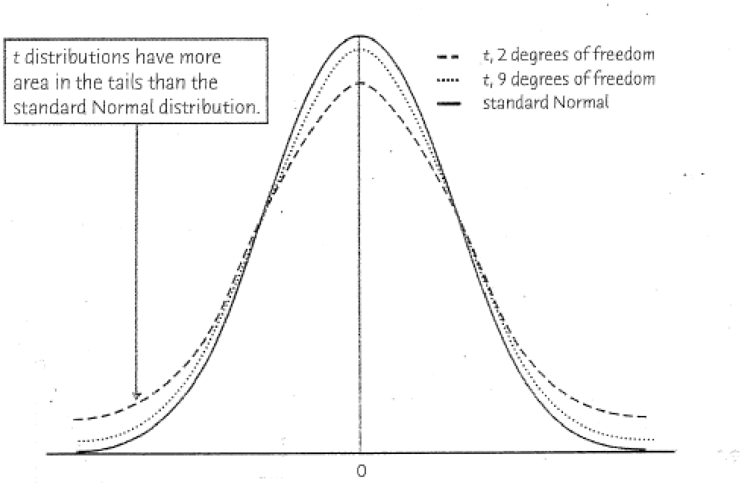
the distribution of $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ may be approximated by $T_{(n-1)}$ （近似不會太離譜）

Fact 3b: If the distribution the sample is skew, we can NOT use T intervals.

(通常題目要求畫出分佈圖的原因就是檢驗是否適用 T 分配)

(Student's) T Distribution

The density curves of the t distributions are similar in shape to the standard Normal curve. They are symmetric about 0, single-peaked, and bell-shaped.



Confidence Interval ($(1-\alpha)\cdot 100\%$) for μ

→ Need to know how to find the table values

Normal population: $X_i \sim^{iid} N(\mu, \sigma^2)$

- σ^2 is known

  $[\bar{X} - z_{\alpha/2}\cdot\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\cdot\sigma/\sqrt{n}]$

- σ^2 is unknown

  $[\bar{X} - t_{\alpha/2,n-1}\cdot S/\sqrt{n}, \bar{X} + t_{\alpha/2,n-1}\cdot S/\sqrt{n}]$

Non-Normal population: $X_i \sim^{iid} E(X_i) = \mu, Var(X_i) = \sigma^2$

- n is large, the above two intervals provide reasonable approximations.
- n is small, the formula is not suitable.

Confidence Intervals

Note that when $X_i \sim^{iid} N(\mu, \sigma^2)$,

Formula: $(1-\alpha)\cdot 100\%$ confidence interval for μ with unknown σ^2

$$[\bar{X} - t_{\alpha/2,n-1}\cdot S/\sqrt{n}, \bar{X} + t_{\alpha/2,n-1}\cdot S/\sqrt{n}],$$

EXAMPLE 17.2 Healing of skin wounds

Let's look again at the biological study we saw in Example 14.3. We follow the four-step process for a confidence interval, outlined on page 361.

STATE: Biologists studying the healing of skin wounds measured the rate at which new cells closed a razor cut made in the skin of an anesthetized newt. Here are data from 18 newts, measured in micrometers (millionths of a meter) per hour (μm/h):[4]

29	27	34	40	22	28	14	35	26
35	12	30	23	18	11	22	23	33

This is one of several sets of measurements made under different conditions. We want to estimate the mean rate for comparison with rates under other conditions.

1	1 2 4
1	8
2	2 2 3 3
2	6 7 8 9
3	0 3 4
3	5 5
4	0

FIGURE 17.2 Stemp of the healing rates in Example 17.2.

To apply the T-interval, your need to check whether the normality assumption is plausible.

FORMULATE: We will estimate the mean rate μ for all newts of this species by giving a 95% confidence interval.

SOLVE: We must first check the conditions for inference.

- As in Chapter 14 (page 361), we are willing to regard these newts as an SRS from their species.
- The stemplot in Figure 17.2 does not suggest any strong departures from Normality.

We can proceed to calculation. For these data,

$$\bar{x} = 25.67 \text{ and } s = 8.324$$

The degrees of freedom are $n - 1 = 17$. From Table C we find that for 95% confidence, $t^* = 2.110$. The confidence interval is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 25.67 \pm 2.110 \frac{8.324}{\sqrt{18}}$$
$$= 25.67 \pm 4.14$$
$$= 21.53 \text{ to } 29.81 \text{ micrometers per hour}$$

Exercise:

Rats eating oat bran. Exercise 17.2 gave the summary data (mean 89.01, standard error 5.36 mg/dl) for the cholesterol levels of 6 rats fed a diet enriched in fiber from oat bran. Cholesterol levels are usually approximately Normal, and we can regard these 6 rats as an SRS of the population of lab rats fed a diet enriched in oat fiber. Give the 95% confidence interval for the mean cholesterol level μ in this population.

$n = 6$, $\bar{x}_{obs} = 89.01$, $s = 5.36$, $\alpha = 0.05$, $t_{0.025,5} = 2.571$

$$[\bar{x}_{obs} - t_{\alpha/2,n-1} \cdot S/\sqrt{n}, \bar{x}_{obs} + t_{\alpha/2,n-1} \cdot S/\sqrt{n}] \Rightarrow 89.01 \pm 2.571 \cdot 5.36/\sqrt{6}]$$

Compare $\bar{X} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ & $\bar{X} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}$

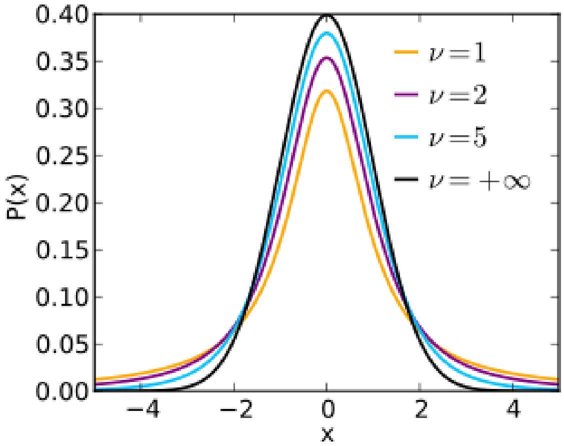
If it happens that the values of σ and S are close, T-intervals will be <u>wider</u> than Z-intervals since $t_{\alpha/2,n-1} > z_{\alpha/2}$

- This is the price paid by the extra estimation of σ.

  信賴區間寬度的統計意義：區間寬度變長代表估計的不確定性增加，這是從“確知σ”到“估計σ”所付出的<u>代價</u>

Note: T_v has heavier tails than N(0,1)

(T 尾巴機率較大，所以切點要更往外挪，才會切出一樣的機率)



Z Table and T Table → 查表 https://youtu.be/Y5knZOokAEY

https://read01.com/zh-tw/PManD37.html#.YaTxkdBBxdh

我們今天要說的，是另一個享譽全球的「不務正業」的公司，便是大名鼎鼎的愛爾蘭啤酒廠健力士（Guinness Brewery），金氏世界紀錄的冠名者。此外，我們還將介紹健力士酒廠的一位傳奇釀酒師，一位使用筆名「學生」的忍隱多年的掃地僧級的統計學家威廉-希利-戈塞（William Sealy Gosset）。

Framework of Hypothesis Testing

H_0: Null hypothesis (虛無假說)

– usually based on previous knowledge about μ

H_a: Alternative hypothesis (對立假說)

— usually based on the researcher's belief about μ

假設檢定的兩種錯誤 (Two types of errors in testing)

		Truth	
		H_0 **is true**	H_0 **is false**
Decision	**Accept** H_0	OK	Type II error β
	Reject H_0	Type I error α	OK

$\alpha = \Pr(\text{reject } H_0 \mid H_0 \text{ is true})$ → type I error rate =型一錯誤機率

$\beta = \Pr(\text{accept } H_0 \mid H_0 \text{ is false})$ → type I error rate =型二錯誤機率

Decision Rule:

- fix $\alpha = \Pr(\text{reject } H_0 \mid H_0 \text{ is true})$ = level of significance (顯著水準)

Convention: fix $\alpha = 0.05$

常見的現象: 雖然新方法看來是有效, 但是證據尚不具有統計顯著性.

Teens with at least one close friend can better cope with stress ...
The Conversation AU - 14 小時前
Boys had significantly higher resilience scores than girls – 27.6 out of ... did not have a good friend (this is a statistically significant difference).

Diabetes Diet: Avocados May Prevent Or Delay Diabetes
NDTV Food - 3 小時前
Interestingly, the team also observed nominal weight loss in human subjects, although Spagnuolo said the result was not statistically significant ...

http://www.labmed.org.tw/knowledge_1.ASP?mno=99

年度/日期	2018/06/07	回上頁
作　者	張斌 副教授 辛辛那提兒童醫學中心 (美國)，兒童醫學科-生物統計暨流行病學組 / 蔡雅雯 醫檢師 永康奇美醫院病理中心	
標　題	從p值談起	
	統計學中不少艱澀複雜的概念公式讓多數人覺得枯燥、退怯，唯獨p值因使用方便和直觀而被廣為接受。然而，近年關於p值的爭論一直未曾間斷；2016年美國統計協會（American Statistical Association, ASA）發表一份聲明，「The ASA’s Statement on p-Values: Content, Process, and Purpose」1，釐清各界對於p值概念的謬誤，尤其是部分非統計專業人士的質疑，並從多方面討論p值的意義和正確的使用方法，此亦為該協會創立177年來，首次針對特定統計檢定議題發聲，可見其受重視程度。此外，陸續有學術期刊公告禁用p值，如美國權威的政治學術期刊「政治分析」（Political Analysis）2018年1月22日在其官方社群網站-推特(Twitter)上宣佈，自2018年開始的第26期起，禁用p值2，此一消息更引起軒然大波，意味著學術界長久以來沿用的p值有可能將被廢除。	

- **Three types of testing problems**

Case 1: $H_0: \mu = \mu_0$ versus $H_a: \mu > \mu_0$ (單邊檢驗, one-sided test)

Case 2: $H_0: \mu = \mu_0$ versus $H_a: \mu < \mu_0$ (單邊檢驗, one-sided test)

Case 3: $H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$ (雙邊檢驗, two-sided test)

Remarks:

-You must specify the level of significance (顯著水準):

$$\alpha = \Pr(\text{reject } H_0 | H_0 \text{ is true})$$

Usually $\alpha = 0.05$

- You must state the “decision”: either “Reject H_0” or “Fail to reject H_0”.

- Another way of saying “Fail to reject H_0” → The evidence is not statistically significant at α (say 5%) level of significance.

- <u>Two methods for decision making</u> will be introduced and you need to learn both methods.

  * Method 1: apply the <u>formula</u> of the decision rule;

  * Method 2: compute the <u>p-value</u> and then compare it with α.

Situation 1: $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ $(i = 1, \ldots, n)$ and the value of σ^2 is known.

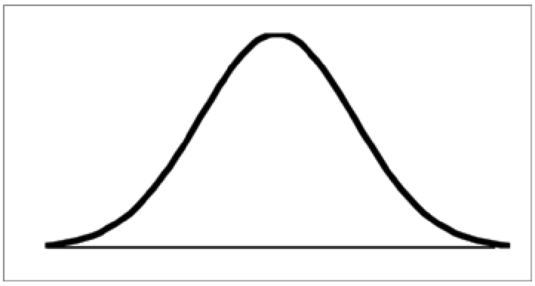
Case 1: $H_0: \mu = \mu_0$ *vs.* $H_a: \mu > \mu_0$

Note: Larger value of $\bar{x}_{obs}$ implies that H_0 is less plausible.

Formula: Reject $H_0: \mu = \mu_0$ if

$$\bar{x}_{obs} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad \frac{\bar{x}_{obs} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha$$

where $\Pr(Z > z_\alpha) = \alpha$ (z_α = z 切點)



Derivations:

$$\bar{X} \underset{H_0 \text{ is true}}{\sim} N(\mu_0, Var(\bar{X}) = \frac{\sigma^2}{n})$$

$$\Rightarrow \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \underset{H_0 \text{ is true}}{\sim} N(0,1)$$

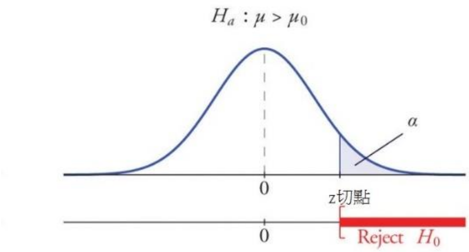
$$\Rightarrow \Pr(\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha \mid H_0 \text{ is true, } \mu{=}\mu_0) = \alpha$$

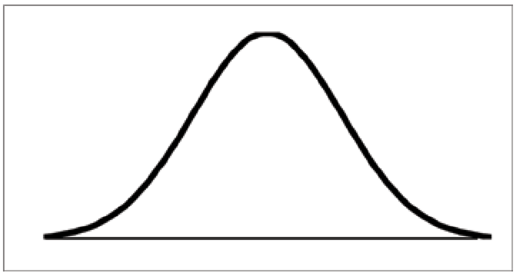
$$\Rightarrow \Pr(\bar{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \mid H_0 \text{ is true, } \mu{=}\mu_0) = \alpha$$

$$\Leftrightarrow \Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$$

Conclusion: Under Case 1, Reject $H_0: \mu = \mu_0$ if $\bar{x}_{obs} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$

https://youtu.be/pepD8ICGS70





Case 2: $H_0: \mu = \mu_0$ *vs.* $H_a: \mu < \mu_0$

Note: Smaller value of $\bar{x}_{obs}$ implies that H_0 is less plausible.

Formula: Reject $H_0: \mu = \mu_0$ if

$$\bar{x}_{obs} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \Leftrightarrow \frac{\bar{x}_{obs} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha$$

Derivations:

$$\bar{X} \underset{H_0 \text{ is true}}{\sim} N(\mu_0, Var(\bar{X}) = \frac{\sigma^2}{n})$$

$$\Rightarrow \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \underset{H_0 \text{ is true}}{\sim} N(0,1)$$

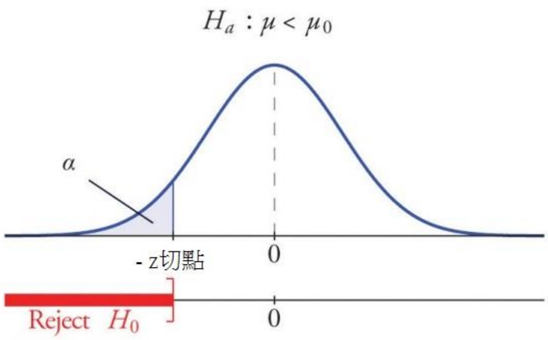
$$\Rightarrow \Pr(\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha \mid H_0 \text{ is true}) = \alpha$$

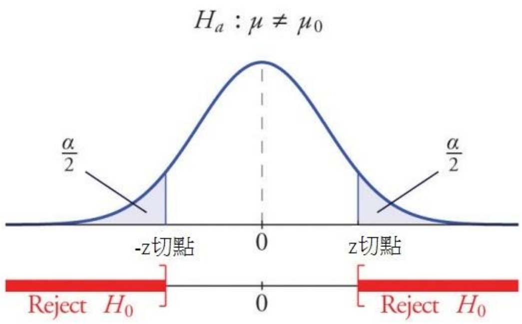
$$\Rightarrow \Pr(\bar{X} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \mid H_0 \text{ is true}) = \alpha$$

$$\Leftrightarrow \Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$$

Conclusion: Under Case 2, Reject $H_0 : \mu = \mu_0$ if $\bar{x}_{obs} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$ (下左圖)

https://youtu.be/ezTB2WJA7h4





https://youtu.be/TTsDukA4o6Q

Case 3: $H_0 : \mu = \mu_0$ *vs.* $H_a : \mu \neq \mu_0$ (下右圖)

Note: Larger or smaller value of $\bar{x}_{obs}$ both imply that H_0 is less plausible.

Formula: Reject $H_0 : \mu = \mu_0$ if

$$\bar{x}_{obs} > \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ or } \bar{x}_{obs} < \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \text{Reject } H_0 : \mu = \mu_0 \text{ if } \left| \frac{\bar{x}_{obs} - \mu_0}{\sigma / \sqrt{n}} \right| > z_{\alpha/2}$$

$\Pr(Z > z_{\alpha/2}) = \alpha / 2$（$z_{\alpha/2}$ = z 切點）

Derivations

$$\bar{X} \underset{H_0 \text{ is true}}{\sim} N(\mu_0, Var(\bar{X}) = \frac{\sigma^2}{n})$$

$$\Rightarrow \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \underset{H_0 \text{ is true}}{\sim} N(0,1)$$

$$\Rightarrow \Pr(\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2} \mid H_0 \text{ is true}) = \alpha$$

$$\Rightarrow \Pr(\bar{X} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \mid H_0 \text{ is true}) + \Pr(\bar{X} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \mid H_0 \text{ is true}) = \alpha$$

$$\Leftrightarrow \Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$$

Conclusion: Under Case 3, Reject $H_0 : \mu = \mu_0$ if

$$\bar{x}_{obs} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ or } \bar{x}_{obs} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow |z_{obs}| = \left|\frac{\bar{x}_{obs}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$$

Summary

Case 1: $H_0 : \mu = \mu_0$ *vs.* $H_a : \mu > \mu_0$

Reject $H_0 : \mu = \mu_0$ if $\bar{x}_{obs} > \mu_0 + z_{\alpha}\frac{\sigma}{\sqrt{n}}$

Case 2: $H_0 : \mu = \mu_0$ *vs.* $H_a : \mu < \mu_0$

Reject $H_0 : \mu = \mu_0$ if $\bar{x}_{obs} < \mu_0 - z_{\alpha}\frac{\sigma}{\sqrt{n}}$

Case 3: $H_0 : \mu = \mu_0$ *vs.* $H_a : \mu \neq \mu_0$

Reject $H_0 : \mu = \mu_0$ if $\bar{x}_{obs} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $\bar{x}_{obs} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Example of Z test → case 2 (one sided)

- Boys of a certain age are known to have a mean weight of μ = 85 pounds.
  ➔ $H_0: \mu = 85$ (a typical situation)
- A complaint is made that the boys living in a municipal children's home are underfed.
  ➔ $H_a: \mu < 85$ (the researcher's suspicion for kids in the municipal children's home)
- As one bit of evidence, n = 25 boys (of the same age) are weighed and found to have a mean weight of $\bar{x}_{obs}$ = 80.94 pounds. It is known that the population standard deviation σ is 11.6 pounds.
- Based on the available data, what should be concluded concerning the complaint?

Step 1: Write down the two hypotheses:

$$H_0: \mu = 85 \text{ vs. } H_a: \mu < 85$$

Step 2: From the direction of H_a, decide the rule of rejecting H_0

➔ Reject H_0 if $\bar{X}_{obs} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$

Step 3: Apply the formula and check whether $\bar{x}_{obs}$ = 80.94 is located in the rejection region: $\bar{X} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \Leftrightarrow \bar{X} < 81.18$

$$\mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = 85 - 1.645 \frac{11.6}{\sqrt{25}} = 81.18$$

Ans: Since 80.94 < 81.18, we reject $H_0: \mu = 85$ which implies that there is significance evidence to support $H_a: \mu < 85$.

Drawback of Method 1: If $\left| \frac{\bar{x}_{obs} - \mu_0}{\sigma / \sqrt{n}} \right|$ is very close to the cut-off value z_{table}, the conclusion can be totally different.

“勉強拒絕”或是“差一點拒絕”的結論完全相反，但“證據的強度”其實卻可能差不多.

14.11 Women's heights. Young American women aged 18 to 24 have an average height of 64.5 inches. You wonder whether the mean height of female students at your university is different from the national average. You find that the mean height of a sample of 78 female students on campus is $\bar{x} = 63.1$ inches. What are your null and alternative hypotheses?

14.12 Stating hypotheses. In planning a study of the birth weights of babies whose mothers did not see a doctor before delivery, a researcher states the hypotheses as

$$H_0: \bar{x} = 1000 \text{ grams}$$
$$H_a: \bar{x} < 1000 \text{ grams}$$

What's wrong with this?

14.11 $H_0: \mu = 64.5$ vs. $H_a: \mu \neq 64.5$; 14.12 Replace $\bar{x}$ by μ

Example of Z test → case 1 (one sided)

14.8 **Arsenic contamination.** Arsenic is a compound naturally occurring in very low concentrations. Arsenic blood concentrations in healthy individuals are Normally distributed with mean $\mu = 3.2$ micrograms per deciliter (μg/dl) and standard deviation $\sigma = 1.5$ μg/dl.[5] Some areas are known to have naturally elevated concentrations of arsenic in the ground and water supplies. We take two SRSs of 25 adults residing in two different high-arsenic areas.

H_0: $\mu = 3.2$

H_a: $\mu > 3.2$

(a) We seek evidence *against* the claim that $\mu = 3.2$. What is the sampling distribution of the mean blood arsenic concentration $\bar{x}$ of a sample of 25 adults if the claim is true? Sketch the density curve of this distribution. (Sketch a Normal curve, and then mark the axis using what you know about locating the mean and standard deviation on a Normal curve.)

(b) Suppose that the data from the first sample give $\bar{x} = 3.35$. Mark this point on the axis of your sketch. Suppose that the data from the second sample give $\bar{x} = 3.75$ μg/dl. Mark this point on your sketch. Using your sketch, explain in simple language why one result is good evidence that the mean blood arsenic concentration of all adults in one high-arsenic area is greater than 3.2 μg/dl and why the outcome for the other high-arsenic area is not.

Step 1: Write down the two hypotheses:

$$H_0: \mu = 3.2 \quad \text{vs.} \quad H_a: \mu > 3.2$$

Step 2: From the direction of H_a, decide the rule of rejecting H_0

➔ Reject H_0 if $\bar{X}_{obs} > \mu_0 + z_\alpha \dfrac{\sigma}{\sqrt{n}}$

Step 3-a: $\bar{x}_{obs} = 3.35$: Check whether $\bar{x}_{obs} = 3.35 > 3.2 + 1.645\dfrac{1.5}{\sqrt{25}} = 3.6935$??

Ans: No! Fail to reject $H_0: \mu = 3.2$ (證據不夠顯著)

Step 3-b: $\bar{x}_{obs} = 3.75$: Check whether $\bar{x}_{obs} = 3.75 > 3.2 + 1.645\dfrac{1.5}{\sqrt{25}} = 3.6935$??

Ans: Yes, so reject $H_0: \mu = 3.2$ (證據達 5%統計顯著性)

Example of Z test → case 2 (one sided)

14.7 **Anemia.** Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than 12 grams of hemoglobin per deciliter of blood (g/dl) are anemic. A public health official in Jordan suspects that the mean μ for all children in Jordan is less than 12. He measures a sample of 50 children. Suppose that the "simple conditions" hold: The 50 children are an SRS from all Jordanian children and the hemoglobin level in this population follows a Normal distribution with standard deviation $\sigma = 1.6$ g/dl.

(a) We seek evidence *against* the claim that $\mu = 12$. What is the sampling distribution of $\bar{x}$ in many samples of size 50 if in fact $\mu = 12$? Make a sketch of the Normal curve for this distribution. (Sketch a Normal curve, and then mark the axis using what you know about locating the mean and standard deviation on a Normal curve.)

(b) The sample mean was $\bar{x} = 11.3$ g/dl. Mark this outcome on the sampling distribution. Also mark the outcome $\bar{x} = 11.8$ g/dl of a different study of 50 children in another country. Explain carefully from your sketch why one of these outcomes is good evidence that μ is lower than 12 and also why the other outcome is not good evidence for this conclusion.

$H_0: \mu = 12$ vs $H_a: \mu < 12$

Reject if $\bar{X} - \mu_0 < z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

$\bar{X} - 12 < 1.645 \cdot \frac{1.6}{\sqrt{50}}$

Step 1: Write down the two hypotheses:

$$H_0: \mu = 12 \quad \text{vs.} \quad H_a: \mu < 12$$

Step 2: From the direction of H_a, decide the rule of rejecting H_0

➔ Reject H_0 if $\bar{X}_{obs} < \mu_0 - z_\alpha \dfrac{\sigma}{\sqrt{n}}$

Step 3-a: $\bar{x}_{obs} = 11.3$: Check whether $\bar{x}_{obs} = 11.3 < 12 - 1.645\dfrac{1.6}{\sqrt{50}} = 11.628$??

Ans: Yes! Reject $H_0: \mu = 12$ (證據達 5%統計顯著性)

Step 3-b: $\bar{x}_{obs} = 11.8$: Check whether $\bar{x}_{obs} = 11.8 < 12 - 1.645\dfrac{1.6}{\sqrt{50}} = 11.628$??

Ans: No, fail to reject $H_0: \mu = 12$ (證據不夠顯著)

Example of Z test → case 3 (two sided)

EXAMPLE 14.9 Executives' blood pressures

STATE: The National Center for Health Statistics reports that the systolic blood pressure for males 35 to 44 years of age has mean 128 and standard deviation 15. The medical director of a large company looks at the medical records of 72 male executives in this age group and finds that the mean systolic blood pressure in this sample is $\bar{x} = 126.07$. Is this evidence that the company's executives have a different mean blood pressure from the general population?

FORMULATE: The null hypothesis is "no difference" from the national mean $\mu_0 = 128$. The alternative is two-sided, because the medical director did not have a particular direction in mind before examining the data. So the hypotheses about the unknown mean μ of the executive population are

$$H_0: \mu = 128$$
$$H_a: \mu \neq 128$$

SOLVE: As part of the "simple conditions," suppose we know that executives' blood pressures follow a Normal distribution with standard deviation $\sigma = 15$. The one-sample z test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{126.07 - 128}{15/\sqrt{72}} = -1.09$$

To help find a P-value, sketch the standard Normal curve and mark on it the observed value of z. Figure 14.9 shows that the P-value is the probability that a standard Normal variable Z takes a value at least 1.09 away from zero. From Table B or software, this probability is

$$P = 2P(Z \geq 1.09) = 2P(Z \leq -1.09) = (2)(0.1379) = 0.2758$$

CONCLUDE: More than 27% of the time, an SRS of size 72 from the general male population would have a mean blood pressure at least as far from 128 as that of the executive sample. The observed $\bar{x} = 126.07$ is therefore not good evidence that male executives differ from other men.

Step 1: Write down the two hypotheses:

$$H_0: \mu = 128 \quad \text{vs.} \quad H_a: \mu \neq 128$$

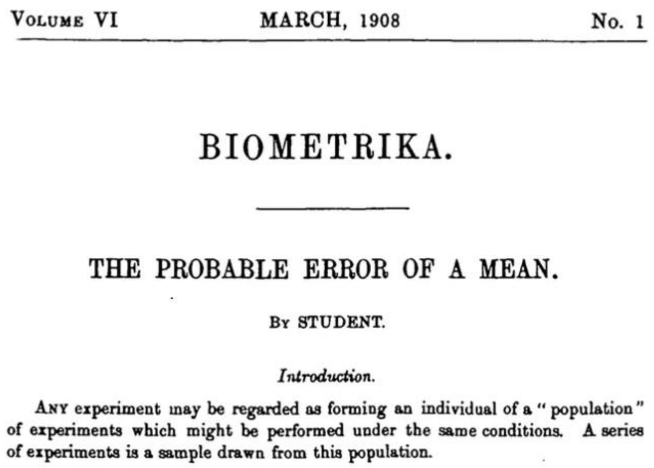
Step 2: From the direction of H_a, decide the rule of rejecting H_0

➔ Reject H_0 if $z_{obs} = \left| \dfrac{\bar{X}_{obs} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2} = 1.96$??

Step 3: $\bar{x}_{obs} = 126.07$, $z_{obs} = \left| \dfrac{126.07 - 128}{15/\sqrt{72}} \right| = |-1.09|$

Fail to reject $H_0: \mu = 128$ since $1.09 < 1.96$. (證據不夠顯著)

Gosset & Student's T Distribution



VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

The ultimate 'Student' author's journal paper. Biometrika (screen grab)

This was the origin of Student's t-test, a fundamental statistical method that is widely used to this day.

https://priceonomics.com/the-guinness-brewer-who-revolutionized-statistics/

The Guinness Brewer Who Revolutionized Statistics

One of the greatest minds in 20th Century statistics was not a scholar. He brewed beer. Guinness brewer William S. Gosset's work is responsible for inspiring the concept of statistical significance, industrial quality control, efficient design of experiments and, not least of all, consistently great tasting beer.

But Gosset is certainly no household name. Books and articles about him are sparse, and he is rarely discussed among history's most important statisticians. Because he used a pseudonym, his name isn't even familiar to most people who frequently use his most famous discovery. Gosset is the "student" of the Student's T-Test, a method for interpreting what can be extrapolated from a small sample of data.

How did a brewer of dry stout revolutionize statistics? And why is he so little-known? Gosset made his great innovations while working as a brewer for Guinness from 1899 to 1937.

A Gentleman Scientist

By all historical accounts, William S. Gosset was a pretty awesome guy. Both Karl Pearson and R.A. Fisher, the two most famous statistical thinkers of the early 20th Century, who were known to hate each other, found common ground in their fondness for Gosset.

Born in 1876 in Canterbury, England, Gosset entered a world of enormous privilege. His father was a Colonel in the Royal Engineers, and though he intended to follow in his footsteps, he was unable to due to bad eyesight. Instead, Gosset attended the prestigious Winchester College, and then Oxford, where he studied mathematics and natural sciences. Soon after graduating from Oxford, in 1899, Gosset joined the Guinness brewery in Dublin, Ireland, as an experimental brewer.

Gosset would spend the rest of his life working for Guinness, and it was through working on Guinness products that he would develop his great statistical innovations. The science that is part of the brewing process inspired Gosset's statistical innovations.

When Gosset began working at Guinness, it was already the world's largest brewery. Even compared to modern companies, Guinness was unusually focused on using science to improve its products. **They hired the "brightest young men they could find" as scientists, and gave them liberal license to innovate and implement their findings. Perhaps the equivalent of being a computer scientist at Bell Labs in the 1970s or an artificial intelligence researcher at Google today, it was a wonderful job for the inquisitive and practical minded Gosset.**

At that time, Guinness's primary focus was maintaining the quality of its beer, while increasing quantity and decreasing costs. Between 1887 and 1914, the output of the brewery doubled, reaching almost one billion pints. How could the company increase production, while keeping its beer tasting as consumers expected? Gosset was assigned as part of the team that would answer that question.

Gosset discovered Student's t-distribution; via Columbia University.

So Gosset set to work. His goal was to understand just how much less representative a sample is when the sample is small. In more technical terms, how much wider is the error distribution of an estimate when you only have a sample of two or ten, compared to when you have a sample of a thousand?

… His bosses at Guinness were thrilled with the findings. This would allow them to make intelligent decisions about which materials to use for their beer, in a way that no other business could. … The company obliged and sent Gosset to Karl Pearson's lab at the University College London. Pearson was one of the leading scientific figures of his time and the man later credited with establishing the field of statistics.

The original t-distribution table from Gosset's seminal work, "The Probable Error of the Mean."

How Gosset Became the "Student"
... It is the source of the concept of "statistical significance."
But why is it the "Student's" t-distribution rather than "Gosset's"?
Upon completing his work on the t-distribution, Gosset was eager to make his work public. It was an important finding, and one he wanted to share with the wider world. The managers of Guinness were not so keen on this. They realized they had an advantage over the competition by using this method, and were not excited about relinquishing that leg up. **If Gosset were to publish the paper, other breweries would be on to them.**

So they came to a compromise. Guinness agreed to allow Gosset to publish the finding, as long as he used a pseudonym. This way, competitors would not be able to realize that someone on Guinness's payroll was doing such research, and figure out that the company's scientifically enlightened approach was key to their success.
https://projecteuclid.org/download/pdf_1/euclid.ss/1177013437

Statistical Science
1987, Vol. 2, No. 1, 45–52

Guinness, Gosset, Fisher, and Small Samples

Joan Fisher Box

Abstract. The environment in which W. S. Gosset (Student) worked as a brewer at Guinness' Brewery at the turn of the century is described fully enough to show how it forced him to confront problems of small sample statistics, using the techniques he picked up from Karl Pearson. R. A. Fisher's interest in human genetics prompted biometrical applications of his mathematical training even as an undergraduate. As soon as he considered Student's work, he perceived its importance and began to extend its applications. Consequently, when he started work at Rothamsted Experimental Station in 1919, he was ready to respond to the experimental problems by developing statistical theory along with appropriate methods of experimental analysis and design.

Key words and phrases: Gosset, Student's *t*, Fisher, small samples, analysis of variance, correlation.