



รายวิชา 568352 สารสนเทศศาสตร์สำหรับบุคลากรสุขภาพ

DATA ANALYSIS WITH STATISTICAL SOFTWARE (บรรยาย)

รศ.ดร.ลาวัลย์ ศรีธราพุทธร

ภาควิชาสารสนเทศศาสตร์ทางสุขภาพ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร

วัตถุประสงค์



- ให้มีความรู้ความเข้าใจเกี่ยวกับ
 - หลักการสถิติเบื้องต้น (ทบทวน)
 - เครื่องมือช่วยวิเคราะห์ข้อมูลเชิงสถิติ PSPP
- ให้มีทักษะด้าน
 - การประยุกต์โปรแกรมสถิติ PSPP ในการวิเคราะห์ข้อมูลสุขภาพ

โครงสร้างเนื้อหา



- Part I: Basic Statistics (ทบทวน)
- Part II: Introduction to PSPP
- Part III: PSPP: Data Preparation and Transformation
- Part IV: PSPP: Descriptive Statistics
- Part V: PSPP: Compare Means

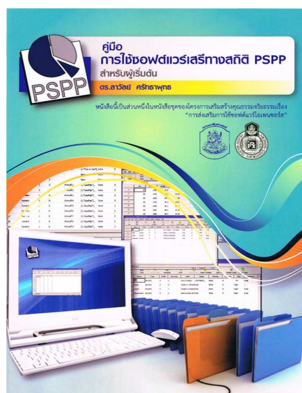
รายการอ้างอิง



- รศ.ศศิธร สุวิรัชวิทยกิจ สถิติสำหรับวิทยาศาสตร์และนักวิทยาศาสตร์ประยุกต์ เล่ม 1, เล่ม 2 มหาวิทยาลัยศิลปากร
- Elementary Statistics: A Step by Step Approach, 8th edition, Allan G. Bluman, McGraw-Hill, 2009.

รายการอ้างอิง

- ผศ. ดร.ลาวัณย์ ศรีธราพุทธร คู่มือการใช้ซอฟต์แวร์สำเร็จทางสถิติ PSPP สำหรับผู้เริ่มต้น, โรงพิมพ์มหาวิทยาลัยศิลปากร, 2012.



5



PART I

Basic Statistics

6

ความหมายของสถิติ

- คือ **ศาสตร์** ที่เรียนเกี่ยวกับ
 - การเก็บรวบรวมข้อมูล (Collection of data)
 - การวิเคราะห์ข้อมูล (Analysis of data)
 - การให้ความหมายข้อมูล (Interpretation of data)
 - การนำเสนอข้อมูล (Presentation of data)
- คือ **การวิเคราะห์** เพื่อนำผลสรุปไปใช้ในการตัดสินใจ
- คือ **ตัวเลข** ที่แสดงลักษณะสำคัญของข้อมูลชุดหนึ่งๆ

7



ประเภทของข้อมูลสถิติ

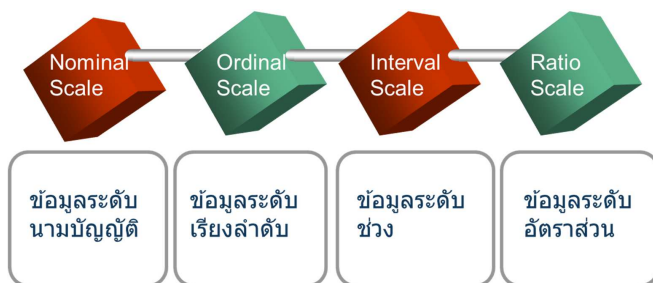
- แบ่งตามแหล่งที่มาของข้อมูล
 - ข้อมูลปฐมภูมิ, ข้อมูลทุติยภูมิ
- แบ่งตามลักษณะของข้อมูล
 - ข้อมูลคุณภาพ, ข้อมูลปริมาณ
- **แบ่งตามระดับการวัดของข้อมูล**
 - ข้อมูลระดับนามบัญญัติ, ข้อมูลระดับเรียงลำดับ/จัดอันดับ, ข้อมูลระดับช่วง/ภาคชั้น, ข้อมูลระดับอัตราส่วน

8



ระดับการวัดของข้อมูล

9



ข้อมูลระดับนามบัญญัติ

10



• ข้อมูลระดับนามบัญญัติ (Nominal Scale)

คือ ข้อมูลที่บอกการจำแนกข้อมูลออกเป็นกลุ่ม โดยอาศัยลักษณะที่มีร่วมกัน
จำแนก เช่น

- จำแนกตามเพศ
- จำแนกตามศาสนา

- ถึงแม้จะมีการกำหนดตัวเลขให้ตัวแปรก็ตาม จะนำตัวเลขเหล่านี้มา บวก ลบ คูณ หาร
กันไม่ได้เพราะผลลัพธ์ที่ได้ไม่ให้ความหมายอะไร
- การกำหนดตัวเลขเพื่อเข้า computer หรือใช้นับจำนวนได้อย่างเดียว

ข้อมูลระดับเรียงลำดับ

11



• ข้อมูลระดับเรียงลำดับ/จัดอันดับ (Ordinal Scale)

คือ ข้อมูลที่บอกการจำแนกข้อมูลออกเป็นกลุ่ม และบอกความแตกต่างระหว่าง
กลุ่มได้ แต่ไม่บอกว่าแตกต่างกันเท่าไร เช่น

- ความชอบ, ความสวย
- เกรด

- ตัวเลขที่แสดงไว้มีความหมายเพียงแสดงลำดับที่เท่านั้น ไม่นิยมนำไป บวก ลบ คูณ
หรือหาร เพราะผลลัพธ์ที่ได้ไม่ให้ความหมายอะไร
- ระยะห่างระหว่างตัวเลขแต่ละช่วงมีระยะเท่ากันหรือไม่ก็ได้

ข้อมูลระดับช่วง

12



• ข้อมูลระดับช่วง/ภาคชั้น (Interval Scale)

คือ ข้อมูลที่บอกการจำแนกข้อมูลออกเป็นกลุ่ม และบอกความแตกต่างระหว่าง
กลุ่มเป็นปริมาณหรือตัวเลขได้ แต่จุดเริ่มต้นของระดับข้อมูลไม่คงที่ สามารถเปลี่ยนได้ตาม
หน่วยของการวัด (ไม่มี '0' ที่แท้จริง คือค่า '0' ไม่ได้หมายถึงไม่มีสิ่งนั้น) เช่น

- อุณหภูมิ 0°C ไม่ได้แปลว่าไม่มีความร้อนเลย

- ระยะห่างระหว่างตัวเลขแต่ละช่วงมีระยะเท่ากัน

ข้อมูลระดับอัตราส่วน

13



- ข้อมูลระดับอัตราส่วน (Ratio Scale)

คือ ข้อมูลที่บอกการจำแนกข้อมูลออกเป็นกลุ่ม และบอกความแตกต่างระหว่างกลุ่มเป็นปริมาณหรือตัวเลขได้ และยังมีจุดเริ่มต้นที่คงที่ไม่เปลี่ยนแปลงตามหน่วยของการวัด (มี '0' ที่แท้จริง หรือมี absolute zero) เช่น

- น้ำหนัก
- ปริมาณคลอเรสเทอรอล

- สามารถเปรียบเทียบเป็นอัตราส่วนกันได้ด้วย

วิธีการทางสถิติ

14



- สถิติเชิงพรรณนาหรือสถิติบรรยาย (Descriptive Statistics) = สถิติสำหรับบรรยาย
- สถิติเชิงอนุมานหรือสถิติอ้างอิง (Inference Statistics)
= สถิติสำหรับอ้างอิง
- สถิติสำหรับทดสอบเครื่องมือ (Test Statistics)

สถิติเชิงพรรณนา

15



- สถิติเชิงพรรณนาหรือสถิติบรรยาย (Descriptive Statistics)
- เป็นสถิติที่ใช้ข้อมูลที่เกิดขึ้นทั้งหมดซึ่งอาจเป็นกลุ่มประชากรหรือกลุ่มตัวอย่าง มาทำการวิเคราะห์เพื่อมุ่งบรรยายลักษณะหรือสรุปผลเฉพาะของกลุ่มที่เก็บมา ไม่มีการนำไปใช้อ้างอิงกับกลุ่มอื่น

สถิติเชิงพรรณนา

16



- วิธีการบรรยายลักษณะข้อมูล มีดังนี้
 - การนำเสนอข้อมูล (Presentation)
 - บทความบรรยาย, ตาราง, กราฟ
 - การแจกแจงความถี่ (Frequencies)
 - จำนวนข้อมูล, ร้อยละ
 - การหาค่าแนวโน้มเข้าสู่ส่วนกลาง (Measure Central of Tendency) หรือการหาค่ากลาง
 - Mean, Mode, Median
 - การหาค่าการกระจายของข้อมูล (Dispersion)
 - Range, Standard Deviation, Variance, Quartile Deviation

สถิติเชิงอนุมาน

17



- สถิติเชิงอนุมานหรือสถิติอ้างอิง (Inference Statistics)
- เป็นสถิติที่ใช้ข้อมูลที่ได้จากกลุ่มตัวอย่างมาทำการวิเคราะห์โดยใช้วิธีการของสถิติบรรยายมาช่วย เพื่ออธิบายคุณลักษณะของสิ่งที่ต้องการศึกษาในกลุ่มนั้น แล้วอ้างอิงไปยังกลุ่มอื่น โดยใช้หลักของความน่าจะเป็นมาทดสอบสมมติฐาน
- เช่น การทดสอบสมมติฐาน
 - ค่าเฉลี่ย
 - ค่าสัดส่วน
 - ค่าความแปรปรวน

ประเภทของสถิติเชิงอนุมาน

18



- แบบอิงพารามิเตอร์ (Parametric Inference) สำหรับข้อมูลที่มีคุณสมบัติเป็นไปตามเงื่อนไขดังนี้
 - ระดับการวัดของข้อมูลควรอยู่ในระดับ interval หรือ ratio
 - ข้อมูลได้จากการสุ่มที่มีความเป็นอิสระต่อกัน
 - ข้อมูลมีการแจกแจงแบบปกติ (Normal distribution)
- แบบไม่อิงพารามิเตอร์ (Non-Parametric Inference) สำหรับข้อมูลไม่มีคุณสมบัติเป็นไปตามเงื่อนไขดังกล่าว หรือกลุ่มข้อมูลที่มีขนาดเล็กหรือมีจำนวนน้อย

ความน่าจะเป็น

19



- ความน่าจะเป็น (Probability) คือ อัตราส่วนระหว่างจำนวนสมาชิกของเหตุการณ์ที่สนใจ (n) กับจำนวนสมาชิกที่เป็นไปได้ทั้งหมด (N)

$$P(E) = \frac{n}{N}$$

การแจกแจงความน่าจะเป็น

20



- การแจกแจงความน่าจะเป็น (Probability distribution) หรือ ฟังก์ชันของความน่าจะเป็น (Probability function) คือ ค่าฟังก์ชันที่แสดงถึงค่าของตัวแปรสุ่มค่าหนึ่งว่าจะมีค่าความน่าจะเป็นเท่าใด

การแจกแจงความน่าจะเป็น

21



- เช่น การแจกแจงความน่าจะเป็นของการสำรวจจำนวนบุตรชายในครอบครัวที่มีบุตร 2 คน

X	0	1	2	total
P(X=x _i)	1/4	2/4	1/4	1

$$P(X=x_i) = \frac{{}^nC_x}{4} = \left(\frac{n!}{x!(n-x)!} \right) / 4; \dots \text{example} = \frac{2!}{1!(2-1)!} / 4 = \frac{2}{4}$$

Recall: $0! = 1$

การแจกแจงแบบปกติ

22



- การแจกแจงแบบปกติ (Normal Distribution) คือ การแจกแจงของตัวแปรสุ่ม x ที่มีค่าเฉลี่ยเลขคณิตคือ μ (mu) และค่าส่วนเบี่ยงเบนมาตรฐานคือ σ (sigma)

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2 \left(\frac{x-\mu}{\sigma} \right)^2}; -\infty < x < \infty$$

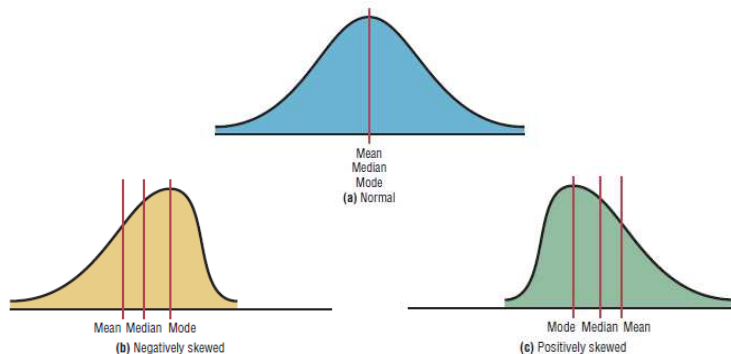
- เป็นการแจกแจงที่สำคัญที่สุดเนื่องจาก ทฤษฎีต่างๆทางสถิติ จะตั้งอยู่บนพื้นฐานการแจกแจงแบบนี้

คุณลักษณะของการแจกแจงแบบปกติ

23



- Normal and Skewed Distributions



คุณลักษณะของการแจกแจงแบบปกติ

24



- สมมาตรกันที่แกน $x = \mu$
- ที่ $x = \mu$ จะให้ค่าความถี่สูงสุดเพียงจุดเดียว
- มีค่า mean, median, และ mode เท่ากัน
- มีพื้นที่ใต้เส้นโค้งเหนือแกน x เท่ากับ 1 หน่วย (100%)

คุณลักษณะของการแจกแจงแบบปกติ

25



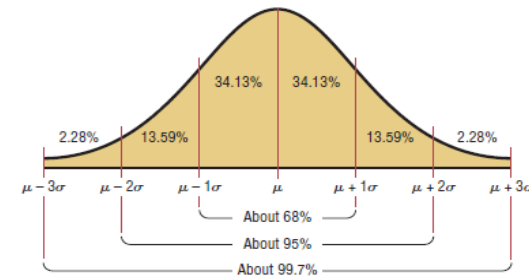
- ถ้าลากเส้นตรงตั้งฉากกับแกน x ที่ระยะห่างจาก μ ถึงเส้นตั้งฉากนี้ เป็นระยะเท่ากับ $\pm 1\sigma$ จะมีพื้นที่ใต้กราฟประมาณ 68% และ
- ถ้าห่างไปจาก μ เป็นระยะเท่ากับ $\pm 2\sigma$ หน่วย จะมีพื้นที่ใต้กราฟประมาณ 95% และ
- ถ้าห่างไปจาก μ เป็นระยะเท่ากับ $\pm 3\sigma$ หน่วย จะมีพื้นที่ใต้กราฟประมาณ 99.7%

คุณลักษณะของการแจกแจงแบบปกติ

26



- Areas Under a Normal Distribution Curve

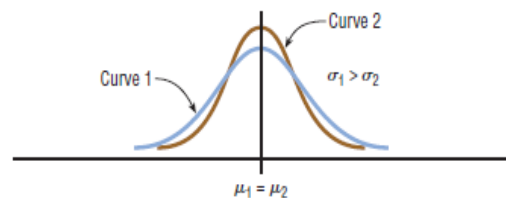


คุณลักษณะของการแจกแจงแบบปกติ

27



- การสร้างเส้นโค้งปกติโดยใช้ parameter μ และ σ
—ถ้าให้ μ มีค่าคงที่ แล้วเปลี่ยนแปลงค่า σ กราฟจะสูงต่ำ ดังรูป



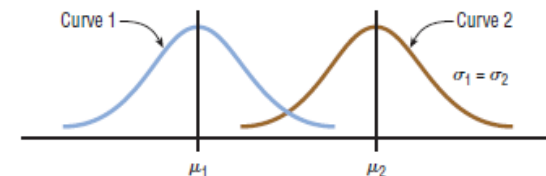
(a) Same means but different standard deviations

คุณลักษณะของการแจกแจงแบบปกติ

28



- การสร้างเส้นโค้งปกติโดยใช้ parameter μ และ σ
—ถ้าให้ σ มีค่าคงที่ แล้วเปลี่ยนแปลงค่า μ กราฟจะเลื่อนไปตามแกน x ดังรูป



(b) Different means but same standard deviations

คุณลักษณะของการแจกแจงแบบปกติ

29



- เมื่อเปลี่ยนเป็นค่ามาตรฐานแล้ว ค่าเฉลี่ยเลขคณิต(μ) และค่าส่วนเบี่ยงเบนมาตรฐาน (σ) ของค่ามาตรฐานมีค่า 0 และ 1 ตามลำดับ

การแจกแจงแบบปกติมาตรฐาน

30



- การแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution) คือ การแจกแจงของตัวแปรสุ่มปกติ x ที่มีค่าเฉลี่ยเลขคณิตเป็น 0 และค่าส่วนเบี่ยงเบนมาตรฐาน เป็น 1

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}; -\infty < z < \infty$$

การแจกแจงปกติมาตรฐาน

31



- All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

$$Z = \frac{x_i - \mu}{\sigma}$$

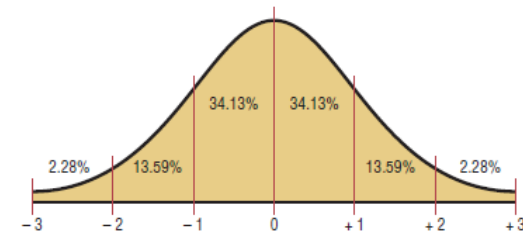
- ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน ของค่ามาตรฐาน Z จะมีค่า 0 และ 1 ตามลำดับ

การแจกแจงแบบปกติมาตรฐาน

32



- Standard Normal Distribution



การทดสอบสมมติฐาน

33



- การทดสอบสมมติฐาน (Hypothesis testing) คือกระบวนการตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมติฐาน (H_0) ที่ตั้งขึ้น เพื่อสรุปการอ้างอิงค่าตัวสถิติ (ของตัวอย่าง) ไปสู่ค่าพารามิเตอร์ (ของประชากร)
- สมมติฐานทางสถิติ (Statistical hypothesis) คือข้อสมมติเกี่ยวกับค่าพารามิเตอร์ของประชากร

ทบทวนนิยามศัพท์

34



- พารามิเตอร์ (Parameter) คือค่าคงที่ที่แสดงคุณลักษณะบางประการของประชากรที่ไม่ทราบค่าที่แท้จริง
- ตัวสถิติ (Statistic) คือ ค่าที่วัดมาจากหน่วยตัวอย่างต่างๆ ที่ถูกเลือกขึ้นมาเป็นตัวอย่าง ซึ่งมีค่าแตกต่างกันไปตามตัวอย่างที่สุ่มมาได้
 - ค่าของตัวสถิติที่คำนวณออกมาเป็นตัวเลขจะใช้เป็นค่าประมาณ (Estimate) ของพารามิเตอร์ต่อไป

สัญลักษณ์

35

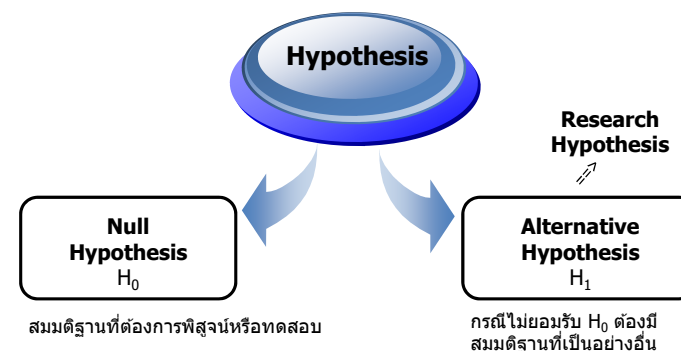


- การกำหนดสัญลักษณ์ที่แตกต่างกันของตัวสถิติและพารามิเตอร์

การวิเคราะห์	ตัวสถิติ	พารามิเตอร์
Mean	\bar{x}	μ
Proportion	p	π
Standard deviation	s	σ
Correlation	r	ρ
Regression	b	β

สมมติฐานทางสถิติ (Hypothesis)

36



การตั้งสมมติฐานทางสถิติ

37



- สมมติฐานว่าง (Null Hypothesis, H_0) คือสมมติฐานที่บอกถึงค่าของพารามิเตอร์ของประชากรที่มาจากความเชื่อที่ต้องการพิสูจน์
- สมมติฐานทางเลือก (Alternative Hypothesis, H_1 หรือ H_a) คือสมมติฐานที่ตรงกันข้ามกับ H_0 ซึ่งเป็นสมมติฐานของความเชื่อที่ผู้วิจัยสนใจ

เช่น

$$H_0 : \mu = 200 \quad H_1 : \mu \neq 200$$

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2 \text{ หรือ } \mu_1 > \mu_2 \text{ หรือ } \mu_1 < \mu_2$$

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0 \text{ หรือ } \mu_1 - \mu_2 > 0 \text{ หรือ } \mu_1 - \mu_2 < 0$$

ชนิดของความผิดพลาด (ความคลาดเคลื่อน)

38



- ชนิดของความผิดพลาด (Type of Error)

การตัดสินใจ	สมมติฐาน H_0 จริง	สมมติฐาน H_0 ไม่จริง
ไม่ปฏิเสธสมมติฐาน H_0	ไม่มีความผิดพลาด ($1-\alpha$)	ความผิดพลาดชนิดที่ 2 (β -error)
ปฏิเสธสมมติฐาน H_0	ความผิดพลาดชนิดที่ 1 (α -error)	ไม่มีความผิดพลาด ($1-\beta$)

บริเวณวิกฤต (Critical Region)

39



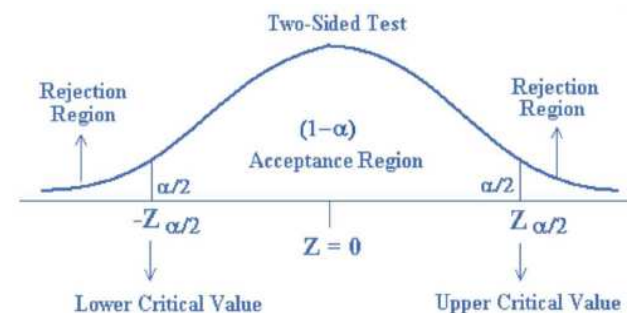
- บริเวณที่ปฏิเสธ H_0 (Rejection region, Critical region) คือพื้นที่ใต้โค้งที่มีค่าเท่ากับ α
- บริเวณไม่ปฏิเสธ H_0 (Non-rejection region, Noncritical region) คือพื้นที่ส่วนที่เหลือซึ่งมีค่าเท่ากับ $1-\alpha$
- ค่าวิกฤต (Critical value) คือ ค่าที่แบ่งบริเวณที่ปฏิเสธ H_0 กับบริเวณไม่ปฏิเสธ H_0 นี้

Two-tailed test

40



- Acceptance region and Rejection regions for the hypothesis test

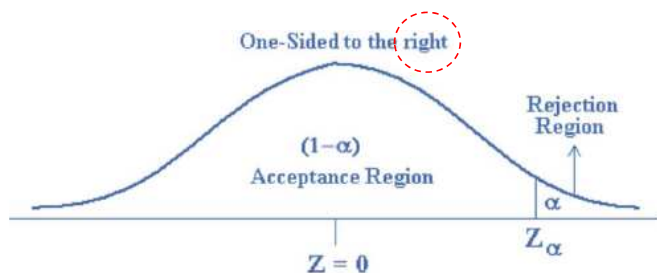


Right-tailed test

41



- Acceptance region and Rejection regions for the hypothesis test

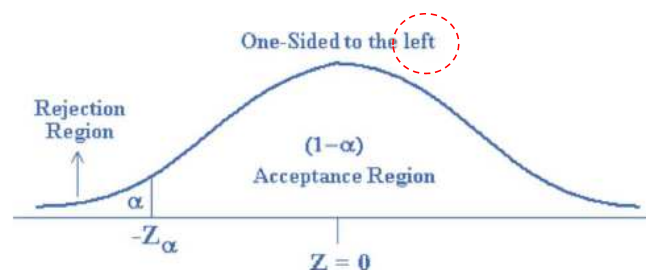


Left-tailed test

42



- Acceptance region and Rejection regions for the hypothesis test



ทบทวนนิยามศัพท์

43




- ค่าสถิติทดสอบ** (Test statistic, Test value) คือค่าสถิติที่ได้จากการคำนวณจากตัวสถิติ ซึ่งมีสูตรคำนวณที่แตกต่างกันไปขึ้นอยู่กับพารามิเตอร์ที่ต้องการทดสอบ และค่านี้มีผลต่อการตัดสินใจว่าจะปฏิเสธ H_0 หรือไม่ เช่น $Z_{\text{calculate}}$
- ค่าวิกฤติ** (Critical value) คือค่าสถิติที่ได้จากการเปิดตารางสถิติ เช่น Z_{table}

วิธีการทดสอบสมมติฐาน

44



- วิธีการทดสอบสมมติฐาน (Hypothesis testing method) มี 3 วิธี ได้แก่
 - Statistic value method
 -  P-value method
 - Confidence interval method

ระดับนัยสำคัญที่สังเกตได้ (p-value)

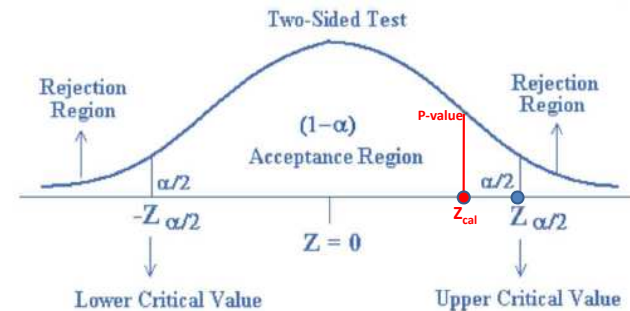
45



- ระดับนัยสำคัญที่สังเกตได้ (Observed significance level, p-value) คือ **ความน่าจะเป็น**ซึ่งได้จากค่าสถิติทดสอบที่ได้จากการคำนวณ => ค่าสถิติที่ได้จากการคำนวณจากตัวอย่าง ไปเปิดตารางหาค่าความน่าจะเป็น (p-value)
- ค่า p-value
 - คือโอกาสที่เกิดขึ้นได้ของ Null hypothesis
 - คือโอกาสที่ Null hypothesis เป็นจริง
 - ใช้ค่า p-value เป็นหลักฐานที่จะยอมรับหรือปฏิเสธสมมติฐาน H_0

Hypothesis testing

46



P-value Method

47



1. กำหนดสมมติฐาน
2. กำหนดระดับนัยสำคัญ (significance level) α
3. คำนวณค่าสถิติทดสอบ (test value) ของวิธีสถิติที่เลือก
4. นำค่าสถิติทดสอบไปเปิดตารางหาค่า p-value
5. ตัดสินใจว่าจะ reject H_0 หรือไม่โดยเปรียบเทียบค่า p-value กับค่า α
6. สรุปและแปลผล

P-value Method

48



- Decision Rule
- กรณี Two-tailed Test
 - ปฏิเสธ H_0 เมื่อ $p\text{-value} \leq \alpha/2$ หรือ
 - ปฏิเสธ H_0 เมื่อ $p\text{-value} \times 2 \leq \alpha$กรณีนี้ p-value คือ p-value(1-tailed) (หนังสือสถิติมักให้ค่านี้)
- ปฏิเสธ H_0 เมื่อ $p\text{-value} \leq \alpha$
- กรณีนี้ p-value คือ p-value(2-tailed) (โปรแกรมสถิติมักให้ค่านี้)

P-value Method

49



- Decision Rule

- กรณี One-tailed Test

- ปฏิเสธ H_0 เมื่อ $p\text{-value} \leq \alpha$

กรณีนี้ p-value คือ p-value(1-tailed) (หนังสือสถิติมักให้ค่านี้)

- ปฏิเสธ H_0 เมื่อ $p\text{-value} / 2 \leq \alpha$ หรือ

- ปฏิเสธ H_0 เมื่อ $p\text{-value} \leq \alpha \times 2$

กรณีนี้ p-value คือ p-value(2-tailed) (โปรแกรมสถิติมักให้ค่านี้)

- Note: กรณี one-tailed test จะต้อง check เครื่องหมายของ test value ให้สอดคล้องกับ H_1 ด้วยเนื่องจากมีแบบ Right-tailed (+) และ Left-tailed (-)

ตัวอย่างตาราง T-test

50



Table F		The t Distribution	
d.f.	Confidence intervals	80%	90%
	One tail, α	0.10	0.05
	Two tails, α	0.20	0.10
1		3.078	6.314
2		1.886	2.920

การทดสอบสมมติฐานของค่าเฉลี่ย

51



- เป็นการทดสอบความแตกต่างของค่าเฉลี่ยระหว่างค่าที่ได้จากตัวอย่างกับค่าคาดหวังของประชากร หรือระหว่างตัวอย่างกับตัวอย่าง ว่ามีนัยสำคัญหรือไม่

ข้อตกลงเบื้องต้น

52



- การทดสอบค่าเฉลี่ยที่จะใช้สถิติเชิงอนุมานแบบอิงพารามิเตอร์ได้ก็ต่อเมื่อข้อมูลที่นำมาทดสอบจะต้อง
 - เป็นข้อมูลระดับช่วงหรืออัตราส่วน
 - เป็นข้อมูลที่ได้จากการสุ่มที่มีความเป็นอิสระต่อกัน
 - เป็นข้อมูลที่มีการแจกแจงแบบปกติหรือใกล้เคียง (ถ้าไม่เป็นตามนี้จะต้องเก็บข้อมูลมากขึ้นตามทฤษฎี Central Limit Theorem)
- Central Limit Theorem คือถ้าข้อมูลจำนวนมากๆ (มากกว่า 30 ตัวอย่าง) จะทำให้ข้อมูลตัวอย่างมีการแจกแจงใกล้เคียงแบบปกติ

การทดสอบสมมติฐานของค่าเฉลี่ย

53



- สำหรับ 1 ประชากร
 - เป็นการศึกษาเพื่อตรวจสอบคุณลักษณะของข้อมูล โดยการหาค่าเฉลี่ยแล้วนำมาเปรียบเทียบกับค่าที่คาดหวังว่ามีความแตกต่างกันหรือไม่
- สำหรับ 2 ประชากร
 - เป็นการศึกษาเพื่อตรวจสอบคุณลักษณะของข้อมูล 2 กลุ่ม โดยการหาค่าเฉลี่ยแล้วนำมาเปรียบเทียบกับว่ามีความแตกต่างกันหรือไม่
- สำหรับ ≥ 3 ประชากร
 - เป็นการตรวจสอบว่าคุณลักษณะใดของข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป มีความแตกต่างกันหรือไม่ และถ้าแตกต่างกันจะแตกต่างกันอย่างไร

การทดสอบค่าเฉลี่ยของ 1 ประชากร

54



- กรณีทราบค่าความแปรปรวนของประชากร (σ^2) ใช้ z-test สูตรดังนี้

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

- กรณีไม่ทราบค่าความแปรปรวนของประชากร (σ^2) ใช้ t-test สูตรดังนี้

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \text{ df} = n - 1$$

- ค่า μ_0 คือค่าคาดหวังหรือค่าตัวเลขที่กำหนดไว้ในสมมติฐาน

การทดสอบค่าเฉลี่ยของ 2 ประชากร

55



- การทดสอบความแตกต่างของค่าเฉลี่ยสำหรับข้อมูล 2 กลุ่มซึ่งจำแนกออกเป็น 2 กรณี คือ
 - กรณีตัวอย่างเป็นอิสระต่อกัน (Independent Sample)
 - กรณีตัวอย่างไม่เป็นอิสระต่อกัน (Dependent Sample)

Independent Sample

56



- กรณีทราบค่าความแปรปรวนของประชากรทั้ง 2 กลุ่ม ใช้ z-test

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Independent Sample

57



- กรณีไม่ทราบค่าความแปรปรวนของประชากรทั้ง 2 กลุ่ม ใช้ t-test ซึ่งมี 2 กรณีคือ
 - กรณีความแปรปรวนของทั้ง 2 กลุ่มแตกต่างกัน
 - กรณีความแปรปรวนของทั้ง 2 กลุ่มไม่แตกต่างกัน

Independent Sample

58



- เมื่อความแปรปรวนของทั้ง 2 กลุ่มแตกต่างกัน ใช้สูตร

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p} \quad \text{โดย} \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2} \right)^2} - 2$$
$$S_p = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Independent Sample

59



- เมื่อความแปรปรวนของทั้ง 2 กลุ่มไม่แตกต่างกัน ใช้สูตร

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p} \quad \text{โดย} \quad d.f. = n_1 + n_2 - 2$$
$$S_p = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right]}$$

การทดสอบค่าความแปรปรวน

60



- การทดสอบค่าความแปรปรวน $H_0 : \sigma_1^2 = \sigma_2^2$
 $H_1 : \sigma_1^2 \neq \sigma_2^2$
- สำหรับ 2 กลุ่มตัวอย่าง ใช้ F-test

$$F = \frac{S_1^2}{S_2^2}$$
$$df = n_1 - 1, n_2 - 1$$
$$S_1^2 > S_2^2$$

การทดสอบค่าความแปรปรวน

61



- จากค่า F ที่ได้ หาค่า P-value แล้วสรุปผลเมื่อกำหนด α ดังนี้
 - ถ้า $P\text{-value} \leq \alpha$ แสดงว่าปฏิเสธ H_0
หมายถึง 2 กลุ่มตัวอย่างมีความแปรปรวนแตกต่างกัน
 - ถ้า $P\text{-value} > \alpha$ แสดงว่าไม่ปฏิเสธ H_0
หมายถึง 2 กลุ่มตัวอย่างมีความแปรปรวนไม่แตกต่างกัน

Dependent Sample

62



- เป็นการทดสอบความแตกต่างเป็นคู่ๆ จึงอาจเรียกว่า การทดสอบความแตกต่างแบบจับคู่ (Paired Difference Tests)

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, d.f. = n-1$$

$$\bar{d} = \frac{\sum d_i}{n} \quad s_d = \sqrt{\frac{1}{n-1} \left[\sum d_i^2 - \frac{(\sum d_i)^2}{n} \right]}$$

เมื่อ n = จำนวนคู่

การทดสอบค่าเฉลี่ยของหลายประชากร

63



- การทดสอบมี 2 วิธี ได้แก่
 - โดยการวิเคราะห์ความแปรปรวนแบบทางเดียว (One-Way Analysis of Variance)
 - เป็นการวิเคราะห์ความแปรปรวนที่ใช้กับข้อมูลที่ได้จากการแบ่งกลุ่มโดยใช้หลักเกณฑ์ปัจจัยเดียว
 - โดยการวิเคราะห์ความแปรปรวนแบบสองทาง (Two-Way Analysis of Variance)
 - เป็นการวิเคราะห์ความแปรปรวนที่ใช้กับข้อมูลที่ได้จากการแบ่งกลุ่มโดยใช้หลักเกณฑ์สองปัจจัย

ANOVA additional assumption

64



- ข้อตกลงในการใช้ ANOVA เพิ่มเติมจากข้อตกลงของ Parametric Inference ได้แก่
 - ความแปรปรวนเท่ากันทุกกลุ่ม (Homogeneity of variance)
(All variances are not necessarily equal, but not significantly different)

One Way ANOVA

65



- แผนผังข้อมูลเพื่อการวิเคราะห์หมีดังนี้

	กลุ่มที่					
	1	2	3	...	k	
	x_{11}	x_{21}	x_{31}	...	k_{k1}	
	x_{12}	x_{22}	x_{32}	...	k_{k2}	
	
	
	
	x_{1n}	x_{2n}	x_{3n}	...	k_{kn}	
ผลรวม	T_1	T_2	T_3	...	T_k	T
ค่าเฉลี่ย	x_1	x_2	x_3	...	x_k	\bar{x}

One Way ANOVA

66



- เปรียบเทียบค่าเฉลี่ยทั้ง k กลุ่ม

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

H_1 : มี μ_i อย่างน้อย 1 คู่ แตกต่างกัน

- ตารางวิเคราะห์ความแปรปรวนดังนี้

ความแปรปรวน	SS	DF	MS	F
ระหว่างกลุ่ม	SSC	k-1	MSC	MSC/MSE
ภายในกลุ่ม	SSE	n-k	MSE	
รวม	SST	n-1		

One Way ANOVA

67



- การคำนวณ ดังนี้
- Correction term (CT)

$$CT = \frac{T^2}{n} = \frac{(\sum x_{ij})^2}{n}$$

- Total sum of squares (SST)

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - CT$$

One Way ANOVA

68



- การคำนวณ ดังนี้
- Sum of squares for treatments (SSC)

$$SSC = \sum_{i=1}^k \frac{T_i^2}{n_i} - CT$$

- Error sum of squares (SSE)

$$SSE = SST - SSC$$

One Way ANOVA

69



- การคำนวณ ดังนี้
- Mean square of treatment (MSC)

$$MSC = SSC/(k-1)$$

- Mean square of error (MSE)

$$MSE = SSE/(n-k)$$

One Way ANOVA

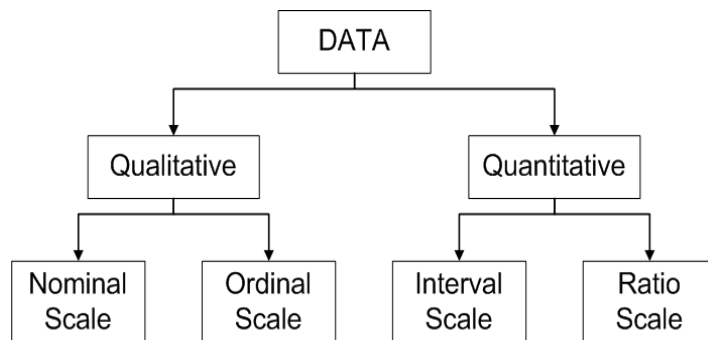
70



- $F = MSC/MSE$
- จากค่า F ที่ได้ หาค่า p-value แล้วสรุปผลเมื่อกำหนด α ดังนี้
 - ถ้า $p\text{-value} \leq \alpha$ แสดงว่าปฏิเสธ H_0 (มีค่าเฉลี่ยแตกต่างกันอย่างน้อย 1 คู่)
 - ถ้า $p\text{-value} > \alpha$ แสดงว่าไม่ปฏิเสธ H_0 (มีค่าเฉลี่ยเท่ากันทุกกลุ่ม)
- ถ้าผลของการวิเคราะห์พบว่ามีความแตกต่างอย่างน้อย 1 คู่ ที่แตกต่างกัน ก็สามารถหาคู่ที่แตกต่างกันได้

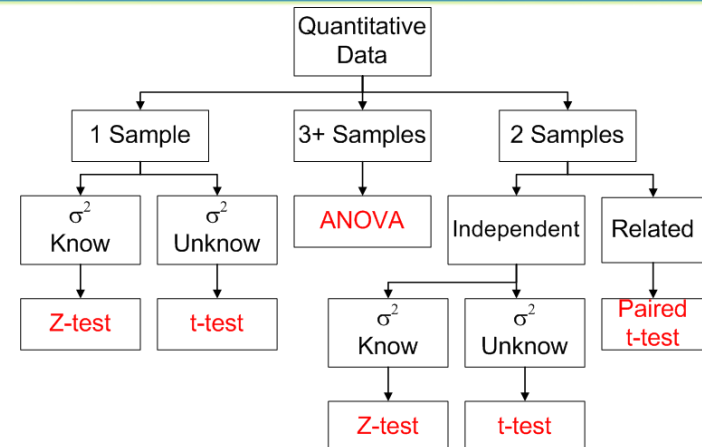
สรุป

71



สรุปการทดสอบค่าเฉลี่ย (Parametric Inference)

72



การทดสอบค่าเฉลี่ยด้วยโปรแกรมสถิติ

73



- โปรแกรมสถิติ SPSS สำหรับการทดสอบค่าเฉลี่ยดังนี้
- เมนู Compare Mean ประกอบด้วยคำสั่ง
 - One Sample T Test
 - Independent Sample T Test
 - Paired Sample T Test
 - One Way ANOVA

ข้อสังเกตการทดสอบค่าเฉลี่ย

74



1. ในทางทฤษฎี จะใช้ t- test เมื่อไม่รู้ค่าความแปรปรวนของประชากร (σ^2) หรือกลุ่มตัวอย่างมีขนาดเล็ก ($n < 30$) แต่ในทางปฏิบัติจะใช้ t-test กับกลุ่มตัวอย่างขนาดใดก็ได้ขอเพียงแต่ให้กลุ่มตัวอย่างที่สุ่ม มาจากประชากรที่มีการแจกแจงปกติ หรือเข้าใกล้การแจกแจงปกติ (Weiss. 1995:537)

2. ในทางปฏิบัติ t-test มีโอกาสใช้มากกว่า z-test เนื่องจากการในความเป็นจริงเราไม่มีโอกาสรู้ค่าความแปรปรวนของประชากร (σ^2) ดังนั้นจึงต้องประมาณค่า σ^2 ด้วยความแปรปรวนของตัวอย่าง (s^2) ดังนั้น ค่าสถิติทดสอบจะมีการแจกแจงแบบ t มากกว่าการแจกแจงแบบ z นั่นคือ ถ้าแทนค่าความแปรปรวนด้วย s^2 แล้วควรใช้ t-test (Homell. 1989:191)



Case Study

ตัวอย่าง 1

76



- ข้อมูลของซีรัมอะมิเลสในคนปกติ 16 ราย ได้ค่าเฉลี่ยเท่ากับ 110 units/100 ml และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 35 unit/100 ml อยากทราบว่าค่าจะแตกต่างจากค่าเฉลี่ยทั่วไป ซึ่งมีค่า 120 units/100 ml หรือไม่ ($\alpha = .05$)

$$H_0 : \mu = 120$$

$$H_1 : \mu \neq 120$$

$$\bar{x} = 110, \mu_0 = 120, s = 35, n = 16$$

ตัวอย่าง 1



$$t = \frac{110 - 120}{35/\sqrt{16}} = -1.1429, d.f. = 15$$

- จากตาราง t ได้ค่า p-value (two-tailed) = 0.2710
- กำหนด $\alpha = 0.05$
- p-value > α
- สรุป ไม่ปฏิเสธ H_0
- แปลว่า ค่าเฉลี่ยซีรัมอะมีเลส 16 ราย ไม่มีความแตกต่างจาก 120 units/100 ml อย่างมีนัยสำคัญเชื่อมั่นได้ถึง 95%

ตัวอย่าง 2



- ในการตรวจเชื้อแบคทีเรีย โดยปกติใช้เวลาเฉลี่ย 50 วินาที ได้ส่งเครื่องตรวจมาใหม่ ทดลองตรวจตัวอย่าง 12 ตัวอย่าง ใช้เวลาโดยเฉลี่ย 42 วินาที ส่วนเบี่ยงเบนมาตรฐาน 11.9 วินาที อยากทราบว่าเครื่องตรวจใหม่นี้ตรวจได้เร็วกว่าการตรวจแบบเก่าหรือไม่ ($\alpha = .05$)

$$H_0 : \mu = 50$$

$$H_1 : \mu < 50$$

$$\bar{x} = 42, \mu_0 = 50, s = 11.9, n = 12$$

ตัวอย่าง 2



$$t = \frac{42 - 50}{11.9/\sqrt{12}} = -2.33, d.f. = 11$$

- จากตาราง t ได้ค่า p-value (two-tailed) = 0.0399
- กำหนด $\alpha = 0.05$
- p-value/2 < α
- สรุป ปฏิเสธ H_0
- แปลว่า เครื่องตรวจใหม่ใช้เวลาน้อยกว่า 50 วินาที อย่างมีนัยสำคัญเชื่อมั่นได้ถึง 95%

ตัวอย่าง 3



- ในการทำวิจัยโรคภูมิแพ้ ต้องการเปรียบเทียบค่าเฉลี่ยอีโอซิโนฟิลในเลือดของผู้ชายและผู้หญิงว่าแตกต่างกันหรือไม่ ($\alpha = .05$) โดยพบว่า

เพศ	n	\bar{x}	s
ชาย	14	584	225
หญิง	16	695	185

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

ตัวอย่าง 3

81



- ขั้นตอนที่ 1: เปรียบเทียบความแปรปรวน ด้วย F-test

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

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

ตัวอย่าง 3

82



$$F = \frac{s_1^2}{s_2^2} = \frac{(225)^2}{(185)^2} = 1.4792, d.f. = 13, 15$$

- จากตาราง F ได้ค่า p-value = 0.2322
- กำหนด $\alpha = 0.05$
- p-value > α
- สรุป ไม่ปฏิเสธ H_0
- แปลว่า ค่าความแปรปรวนทั้ง 2 กลุ่มไม่แตกต่างกันอย่างมีนัยสำคัญเชื่อมั่นได้ 95%

ตัวอย่าง 3

83



- ขั้นตอนที่ 2: เปรียบเทียบค่าเฉลี่ย เมื่อข้อมูลอิสระต่อกัน และความแปรปรวนของทั้ง 2 กลุ่มไม่แตกต่างกัน ด้วย t-test

$$s_p = \sqrt{\left(\frac{1}{14} + \frac{1}{16}\right) \left[\frac{13(255)^2 + 15(185)^2}{14+16-2}\right]} = \sqrt{5603.4758} = 74.8564$$

$$t = \frac{584 - 695}{74.8564}, d.f. = 14 + 16 - 2$$

$$t = -1.4828, d.f. = 28$$

ตัวอย่าง 3

84



- จากตาราง t ได้ค่า p-value (two-tailed) = 0.1493
- กำหนด $\alpha = 0.05$
- p-value > α
- สรุป ไม่ปฏิเสธ H_0
- แปลว่า ค่าเฉลี่ยอีโอซินอฟิลในเลือดไม่แตกต่างกันระหว่างชายและหญิงอย่างมีนัยสำคัญทางสถิติเชื่อมั่นได้ 95%

ตัวอย่าง 4

85



- ผู้ป่วยทั่วไป 20 คน พักรักษาตัวในโรงพยาบาลเฉลี่ย 7 วันและส่วนเบี่ยงเบนมาตรฐาน 2 วัน ผู้ป่วยมะเร็ง 24 คน พบว่าค่าเฉลี่ยจำนวนวันพักรักษาตัว 36 วัน และส่วนเบี่ยงเบนมาตรฐาน 10 วัน ผู้ป่วยมะเร็งพักรักษาตัวนานกว่าหรือไม่ ($\alpha = .05$)

$$H_0 : \mu_G = \mu_C$$

$$H_1 : \mu_G < \mu_C$$

ตัวอย่าง 4

86



- ขั้นตอนที่ 1: เปรียบเทียบความแปรปรวน ด้วย F-test

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

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

ตัวอย่าง 4

87



$$F = \frac{s_1^2}{s_2^2} = \frac{(10)^2}{(2)^2} = 25, \text{d.f.} = 23, 19$$

- จากตาราง F ได้ค่า p-value < 0.0001
- กำหนด $\alpha = 0.05$
- p-value < α
- สรุป ปฏิเสธ H_0
- แปลว่า ค่าความแปรปรวนทั้ง 2 กลุ่ม แตกต่างกันอย่างมีนัยสำคัญเชื่อมั่นได้ 95%

ตัวอย่าง 4

88



- ขั้นตอนที่ 2: เปรียบเทียบค่าเฉลี่ย เมื่อข้อมูลอิสระต่อกัน และ ความแปรปรวนของทั้ง 2 กลุ่มแตกต่างกัน ด้วย t-test

$$s_p = \sqrt{\frac{2^2}{20} + \frac{10^2}{24}} = 2.0897$$

$$df = \frac{(4.3667)^2}{\frac{1}{21} \left(\frac{4}{20} \right)^2 + \frac{1}{25} \left(\frac{100}{24} \right)^2} - 2 = 25.38 \approx 25$$

ตัวอย่าง 4

89



$$t = \frac{7 - 36}{2.0897} = -13.8776, d.f. = 25$$

- จากตาราง t ได้ค่า p-value (two-tailed) < 0.0001
- กำหนด $\alpha = 0.05$
- $p\text{-value}/2 < \alpha$
- สรุป ปฏิเสธ H_0
- แปลว่า ผู้ป่วยมะเร็งจะพักฟื้นตัวนานกว่าผู้ป่วยทั่วไปอย่างมีนัยสำคัญทางสถิติด้วยความเชื่อมั่นได้ 95%

ตัวอย่าง 5

90



- ต้องการเปรียบเทียบปริมาณคอเลสเตอรอลของผู้ใช้โปรแกรมอาหารชนิดหนึ่งว่าทำให้คอเลสเตอรอลลดลงหรือไม่ ($\alpha = 0.05$)

คนที่	ปริมาณคอเลสเตอรอล		d_i	d_i^2
	ก่อน	หลัง		
1	201	200	-1	1
2	231	236	5	25
3	221	216	-5	25
4	260	233	-27	729
5	228	224	-4	16
6	237	216	-21	441
7	326	296	-30	900
8	235	195	-40	1600
9	240	207	-33	1089
10	267	247	-20	400
11	284	210	-74	5476
12	201	209	8	64
			-242	10766

ตัวอย่าง 5

91



- สมมติฐาน

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d < 0$$

หมายถึง $H_1: \mu_{\text{หลัง}} - \mu_{\text{ก่อน}} < 0$

$$\bar{d} = \frac{-242}{12} = -20.17$$

$$s_d = \sqrt{\frac{1}{11} \left[10766 - \frac{(-242)^2}{12} \right]} = 23.13$$

ตัวอย่าง 5

92



$$t = \frac{-20.17}{23.13/\sqrt{12}} = -3.02, d.f. = 11$$

- จากตาราง t ได้ค่า p-value (two-tailed) = 0.0117
- กำหนด $\alpha = 0.05$
- $p\text{-value}/2 < \alpha$
- สรุป ปฏิเสธ H_0
- แปลว่าโปรแกรมอาหารทำให้ปริมาณคอเลสเตอรอลลดลงจริงอย่างมีนัยสำคัญทางสถิติด้วยความเชื่อมั่นได้ 95%

ตัวอย่าง 6

93



- ต้องการเปรียบเทียบอัตราการเต้นของหัวใจก่อนที่ของผู้ป่วย 4 โรค ดังนี้ (โดยความแปรปรวนทั้ง 4 กลุ่ม ไม่แตกต่างกัน) ($\alpha = 0.05$)

	A	B	C	D	Total
	83	81	75	61	
	61	65	68	75	
	80	77	80	78	
	63	87	80	80	
	67	95	74	68	
n_i	5	5	5	5	
$\sum x_i$	354	405	377	362	1498
$\sum x_i^2$	25468	33309	28525	26454	113756
$(\sum x_i)^2$	125316	164025	142129	131044	
\bar{x}_i	70.8	81	75.4	72.4	
s^2	101.2	125.89	24.8	61.31	

ตัวอย่าง 6

94



- เปรียบเทียบค่าเฉลี่ยทั้ง 4 กลุ่ม

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_1 : \text{มี } \mu_i \text{ อย่างน้อย 1 คู่ แตกต่างกัน}$$

ตาราง ANOVA

ความแปรปรวน	SS	DF	MS	$F_{3,16}$
ระหว่างกลุ่ม	302.6	3	100.87	1.2878
ภายในกลุ่ม	1253.2	16	78.325	
รวม	1555.8	19		

ตัวอย่าง 6

95



- การคำนวณ

$$CT = \frac{(1498)^2}{20} = 112200.2$$

$$SST = 113756 - 112200.2 = 1555.8$$

$$SSC = \frac{(354)^2}{5} + \frac{(405)^2}{5} + \frac{(377)^2}{5} + \frac{(362)^2}{5} - 112200.2 = 302.6$$

$$SSE = 1555.8 - 302.6 = 1253.2$$

ตัวอย่าง 6

96



- จากตาราง F ได้ค่า p-value = 0.3125
- กำหนด $\alpha = 0.05$
- p-value > α
- สรุป ไม่ปฏิเสธ H_0
- แปลว่า อัตราการเต้นของหัวใจของคนที่เป็นโรคทั้ง 4 โรคนี้ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติด้วยความเชื่อมั่นได้ 95%