



การวัดผลลัพธ์และสถิติที่ใช้บ่อย

พญ.หทัยทิพย์ ธรรมวิริยะกุล

พบ. วว.เวชศาสตร์ครอบครัว

Msc. Epidemiology LSHTM,UK

- ชนิดของตัวแปร

- ชนิดของสถิติ

- **สถิติเชิงพรรณนา (Descriptive Statistics)**

- ตารางแจกแจงความถี่
- การวัดค่ากลาง
- การวัดการกระจาย

- **สถิติเชิงอนุมาน (Inferential Statistics)**

- Confidence interval
- การทดสอบสมมติฐาน (Hypothesis Testing)
- Student 's T test ,Chi-Squared Test



'The Blind-men and the Elephant' (John Godfrey Saxe, 1849)

สถิติ...

ไม่สามารถบอกความจริง "truth"

ไม่สามารถทดแทนจุดอ่อนของ
งานวิจัยที่ออกแบบไม่ดี

ไม่สามารถบอก clinical
significant

ตัวแปรเชิง คุณภาพ

Qualitative

- Categories/Groups :
 - Binary 2 กลุ่ม
 - Nominal เช่น เพศ
Blood gr
 - Ordinal เช่น
pain score 1 2 3 4

ตัวแปรเชิง ปริมาณ

Quantitative

- Continuous
:uninterrupted มีทศนิยม
ได้
- Discrete : Numerical
value ไม่มีทศนิยม
- Interval ไม่มีศูนย์แท้ เช่น
อุณหภูมิ คะแนนสอบ
- Ratio มีศูนย์แท้ เช่น
น้ำหนัก อายุ ส่วนสูง

คุณสมบัติของตัวแปร

Independent variable

ตัวแปรอิสระ

ตัวแปรต้น

มักมีความถาวร
มากกว่า

เช่น การสูบบุหรี่

Dependent variable

- ตัวแปรผล
- ตัวแปรตาม
- ตัวแปรที่ผู้วิจัยคาดว่าจะเกิดผลเช่นไร
- เช่น การเกิดมะเร็งปอด

ชนิดของตัวแปร



สถิติเชิงพรรณนา (Descriptive Statistics)

- Categorical variable

- ตารางแจกแจงความถี่

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	12	28.6	28.6	28.6
	AB	10	23.8	23.8	52.4
	B	14	33.3	33.3	85.7
	O	6	14.3	14.3	100.0
	Total	42	100.0	100.0	

- Continuous variable

- ค่ากลางของข้อมูล เช่น mean median mode

- การกระจายของข้อมูล (dispersion) variance SD

(Inferential Statistics)

- เพื่อสรุปลักษณะบางอย่างของประชากรโดยใช้ทฤษฎีความน่าจะเป็นมาประยุกต์
- Confidence interval
- Hypothesis testing
- p value
- Student t test and chi-square test
- Non parametric test

ชนิดของสถิติ

Mean • Median • Mode

mean

The mean is the average of a set of numbers.

To find the mean:

- Add together all of the numbers in the set.
- Divide the total by how many numbers were added.

$$1 + 3 + 6 + 9 + 11 + 12 + 14 = 56 \rightarrow 56 \div 7 = 8$$

The mean is 8.

median

The median is the middle number in a sequence.

To find the median:

- Organize the set of numbers from smallest to largest.
- Locate the middle number.

2, 3, 5, 6, 7, 8, 10, 13, 14

The median is 7.

mode

The mode is the number that occurs most frequently.

To find the mode:

- Organize the list of numbers from smallest to largest.
- Locate the number that appears in the list most often.

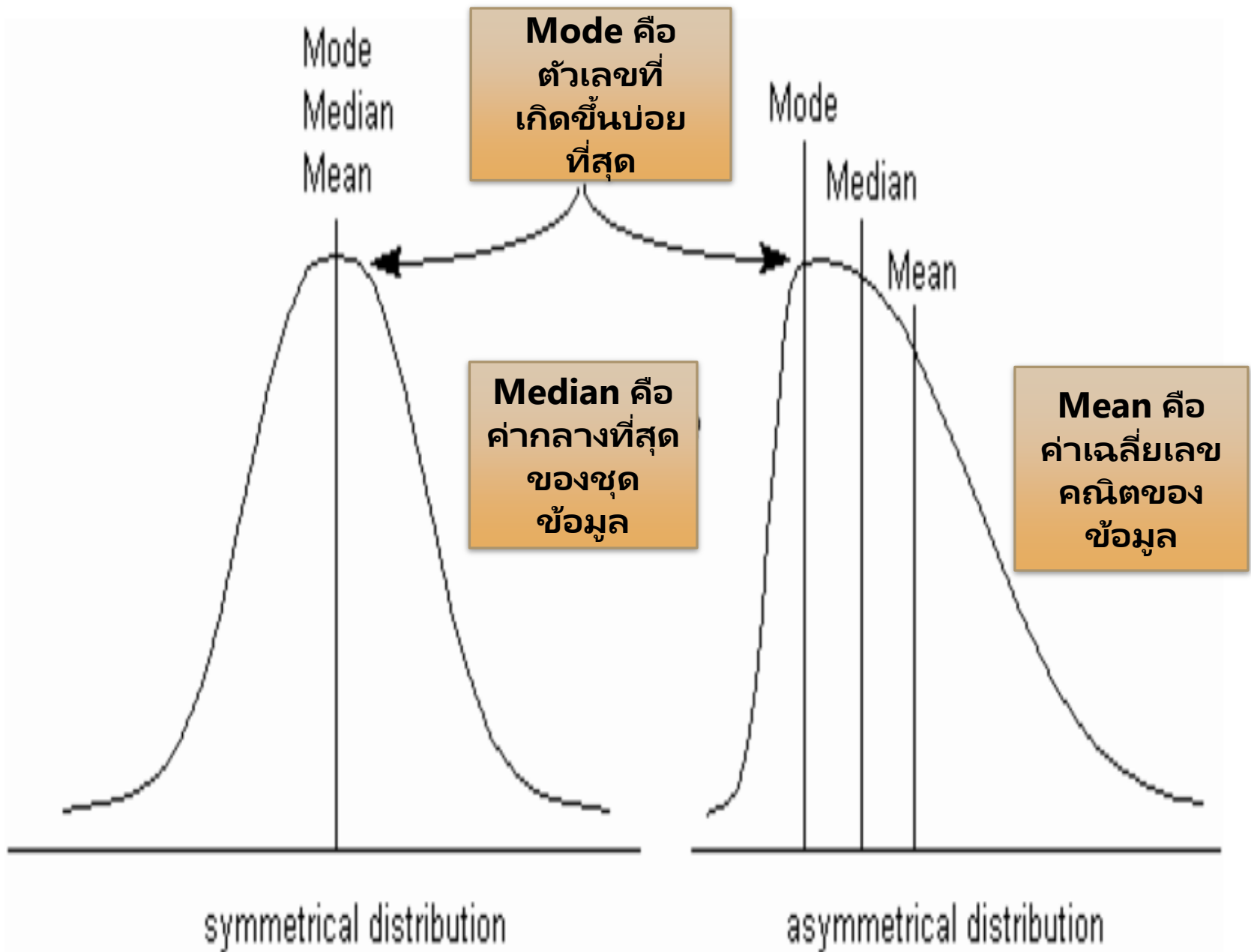
1, 2, 2, 3, 4, 4, 4, 5, 7, 7, 8

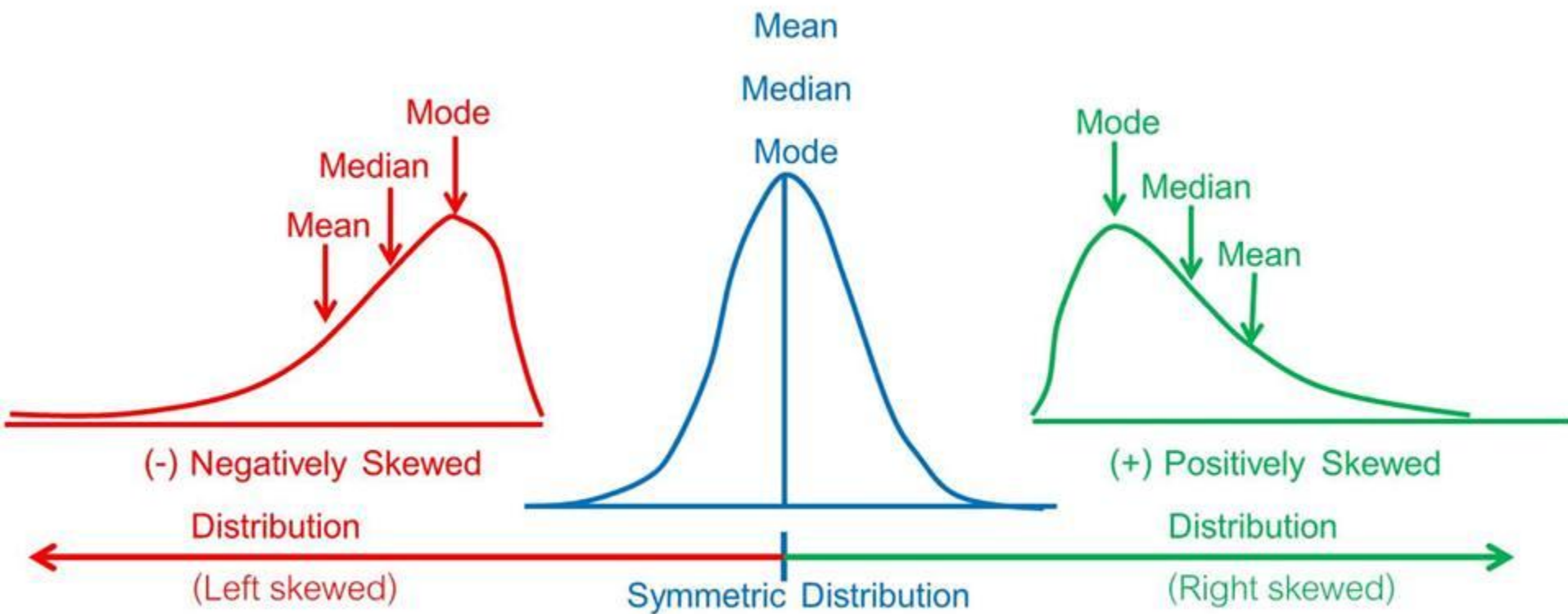
The mode is 4.

Math Matters!

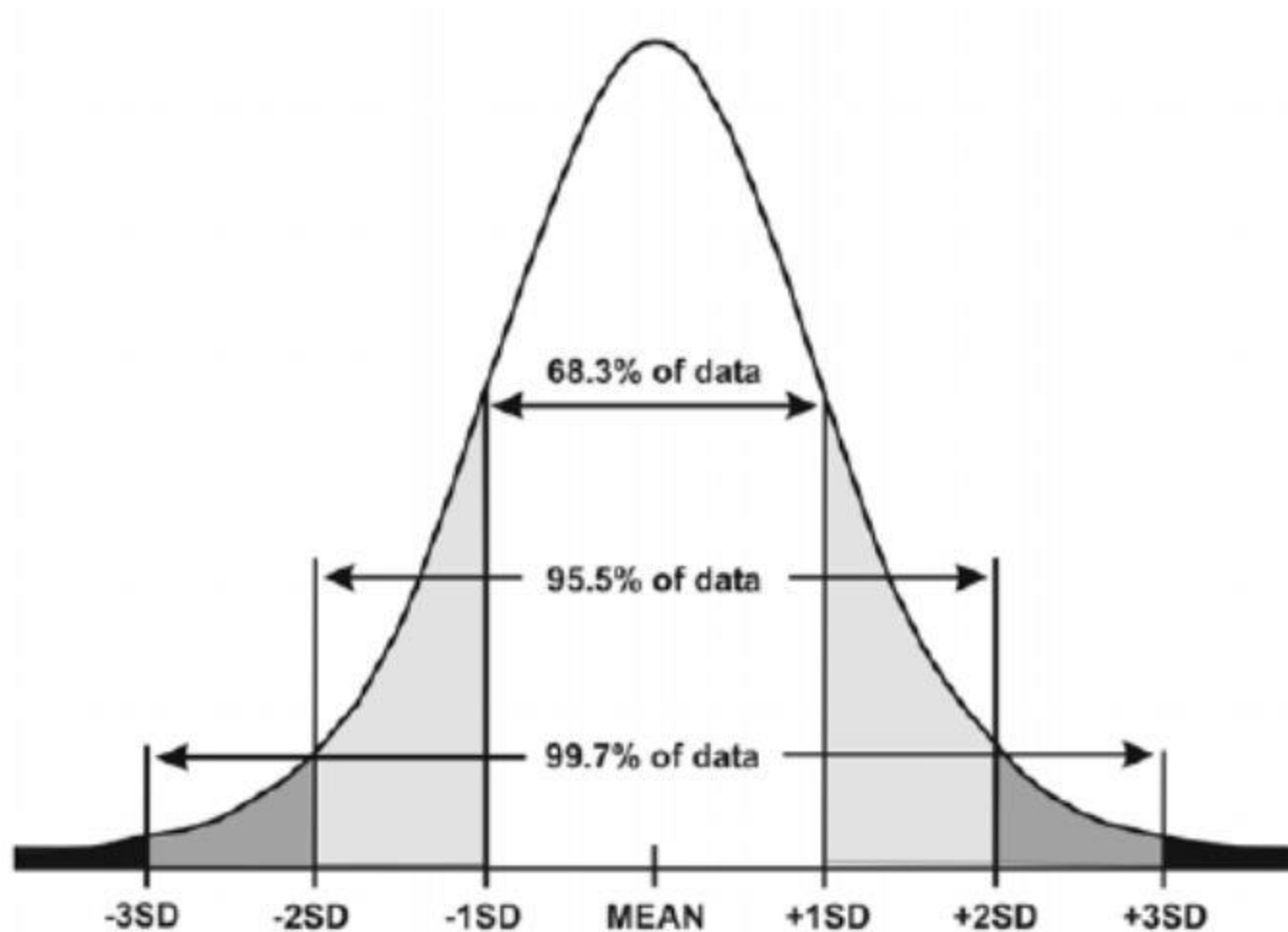
If there are two values located in the middle, find the mean of those numbers to solve for the median. (Example: The median for 2, 3, 4, 6, 11, 15 is 5.)

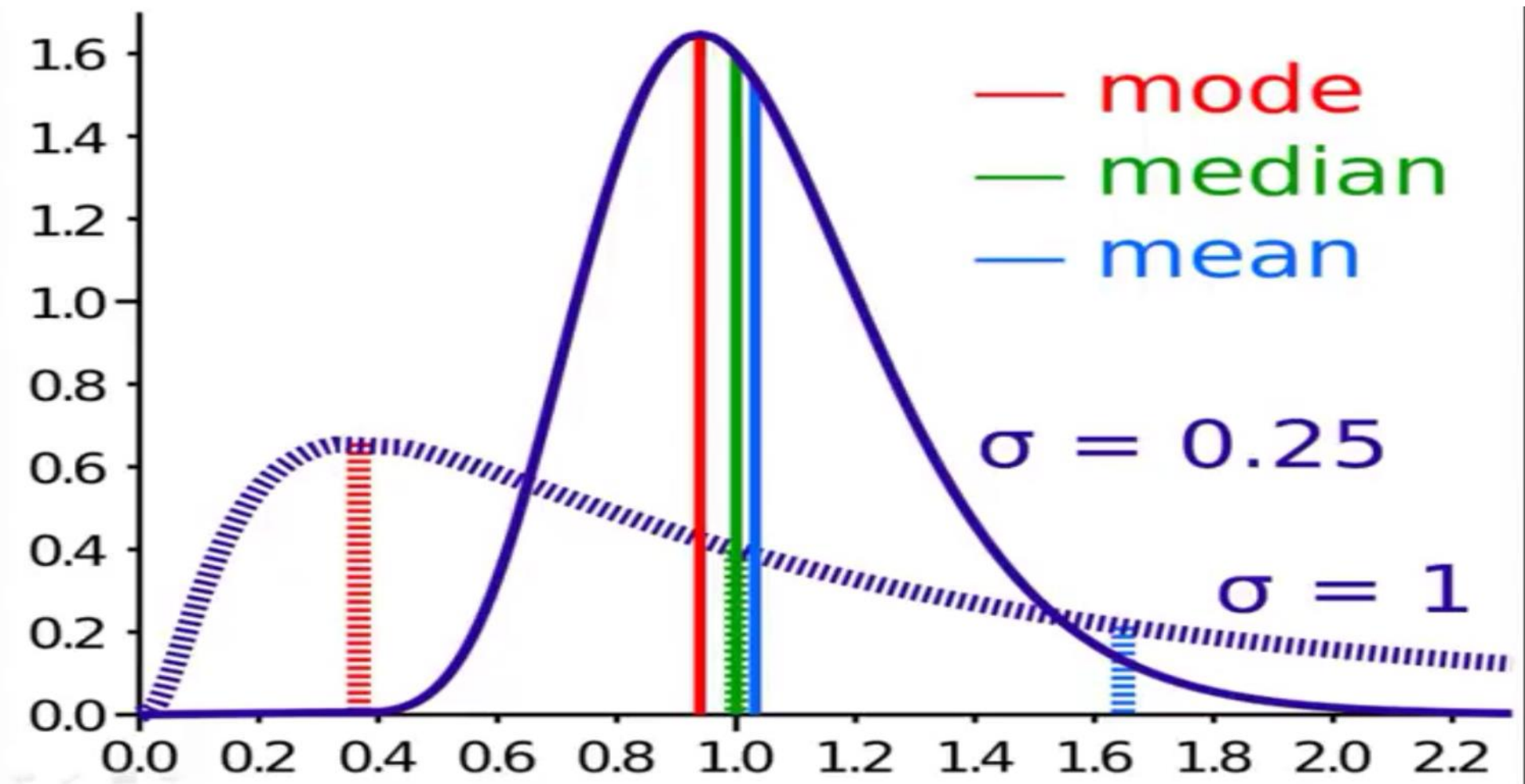
It is possible to have no mode when there are no repeated numbers or when there is more than one mode within a set of numbers.

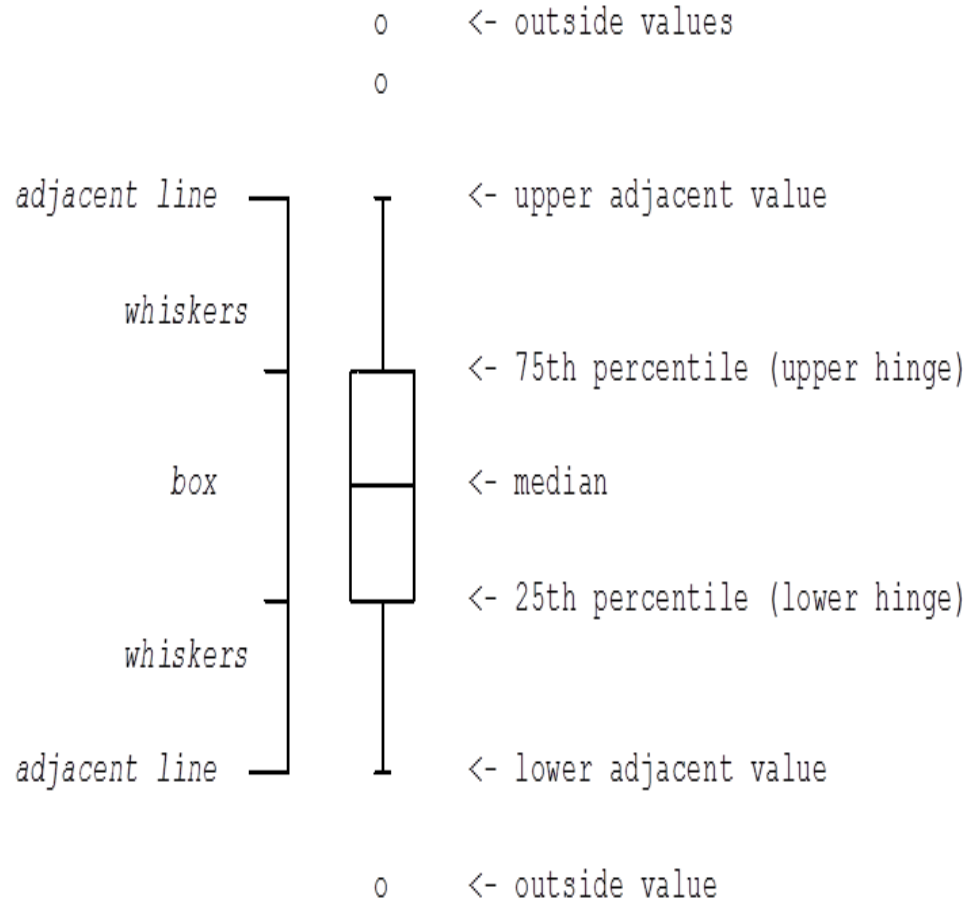




การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจายโดยใช้ Mean และ Standard Deviation







การกระจายของข้อมูล : **box plot**

ดูด้วยตา : Histogram & Box plot

สถิติ : Swilk test /sfrancia test

- **swilk** :Shapiro-Wilk W test for normality
(n 4-2000)
- **sfrancia** :Shapiro-Francia W'test for normality (n 5-5000)

Ho : Normal Distribution

test for normal
distribution

ตารางที่ 2 ความถี่ร้อยละข้อมูลทั่วไปของผู้เข้าร่วมวิจัย

ลักษณะกลุ่มตัวอย่าง		เขตเมือง (n=27)		เขตชนบท (n=28)	
		จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ
เพศ	ชาย	2	7.41	1	3.57
	หญิง	25	92.59	27	96.43
	รวม	27	100.00	28	100.00
อายุ	20 – 30 ปี	2	7.41	1	3.57
	31 – 40 ปี	6	22.22	2	7.14
	41 – 50 ปี	14	51.85	19	67.86
	> 50 ปี	5	18.52	6	21.43
	อายุเฉลี่ย mean(SD)	41.11 (6.90)		47.04 (5.91)	
ประสบการณ์	< 5 ปี	7	25.93	4	14.29
	5 – 10 ปี	12	44.44	11	39.28
	> 10 ปี	8	29.63	13	46.43
	ประสบการณ์เฉลี่ย mean(SD)	9.00 (5.57)		10.25(5.49)	

ตัวอย่าง : การพรรณนาในรูปแบบตารางแจกแจงความถี่

- การนั่งมณีเวชกับการลดการปวดระหว่างเจ็บครรภ์คลอดระยะที่2 เปรียบเทียบคะแนนปวดก่อนและหลังนั่ง
- จงบอก ตัวแปรต้นและตัวแปรตาม



Quiz Time

Let's have
some fun!

Quiz 1

- งานวิจัยเรื่อง การผ่าตัดทอนซิลแบบเย็บปิดแผลและ
ไม่เย็บปิดมีผลต่อระดับความปวดของแผล
(VAS)
- จงบอกมาตรวัดและชนิดของข้อมูลคะแนนระดับ
ความปวดเป็นชนิดใด



Quiz Time

Let's have
some fun!

Quiz 2

- งานวิจัยเรื่อง ผลการใช้น้ำจืดในการลดอาการท้องอืด
หลังส่องกล้องลำไส้ใหญ่ วัดผลลัพธ์ เป็นการหาย
และไม่หายของอาการท้องอืด
- การวิเคราะห์ข้อมูลควรนำเสนอด้วยค่าใด



Quiz Time

Let's have
some fun!

Quiz 3

สถิติเชิงเปรียบเทียบที่ใช้บ่อย

- **Ratio** : (A/B) กลุ่มสัมผัส / กลุ่มไม่ได้สัมผัส ,
กลุ่มรักษา / กลุ่มควบคุม
 - OR, RR ,Prevalence ratio
 - **Difference** : $(A-B)$ กลุ่มรักษา-กลุ่มควบคุม
 - Risk difference, Rate difference
-

Risk vs. Rate

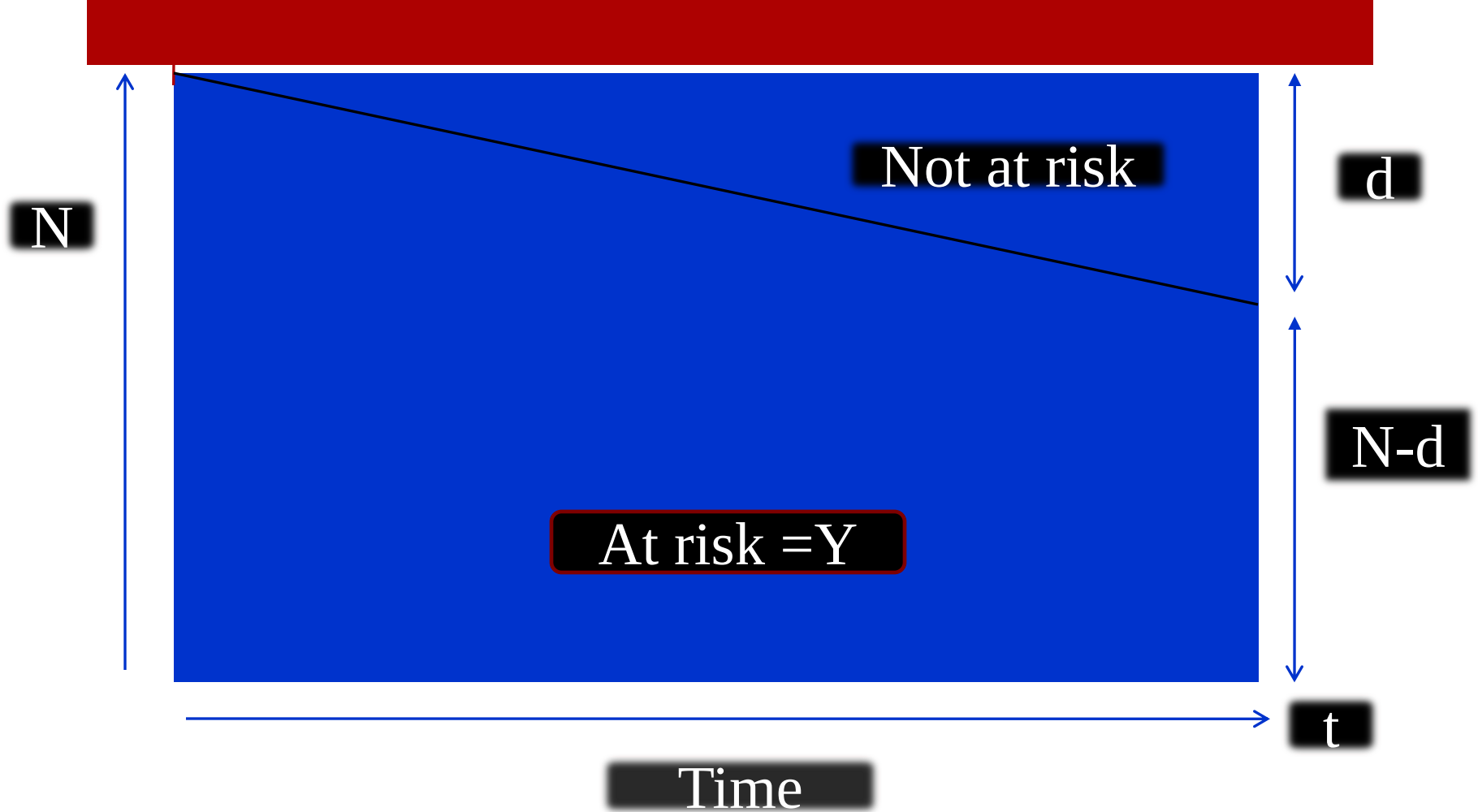
- $\text{risk} = d/N$
- d = จำนวนcase ใหม่
- N = จำนวนประชากรที่มีความเสี่ยงเมื่อตอนเริ่มต้นสนใจ

$$\text{rate} = d/Y$$

d = จำนวนcase ใหม่

- Y = person time at risk

เวลาทั้งหมดที่มีความเสี่ยงของแต่ละคนในช่วงเวลาที่ศึกษา มีการเปลี่ยนแปลงขนาดของความเสี่ยงตามเวลา



Incident Risk over time $t = d/N$
Incident Rate over time $t = d/Y$

- **Person time at risk**
 - 1คนเข้าทำงานวิจัยและเสียชีวิตเมื่อเวลาผ่านไป 6 ปี
 - คนคนนี้มีความเสี่ยงต่อการเสียชีวิตเท่าไร
- **d** คืออะไร ? **Y** คืออะไร ?
- **rate** ของการเสียชีวิตคืออะไร ?
-



Quiz Time

Let's have
some fun!

QUIZ 4 : Rates

- Person time at risk
 - 3 เข้าสู่งานวิจัย
 - 1 คนเสียชีวิตหลังจากติดตามไปเป็นเวลา 4 ปี, 1 คนเสียชีวิตหลังจากติดตามไปเป็นเวลา 8 ปี และที่เหลือมีชีวิตอยู่ต่ออีก 10 ปีจนสิ้นสุดการวิจัย
- d คืออะไร ? Y คืออะไร ?
- **rate** ของการเสียชีวิตคืออะไร ?
-

QUIZ 5 : Rates

ตัวอย่าง

ศึกษาRate การเกิด strokeกับความชอบทำวิจัย

	Per year at risk.	Cases	rate/1000 per yr
• ไม่ชอบทำวิจัย	6848	26	3.8
• ชอบทำวิจัย	8384	6	0.7

$(26/6848)*1000$

rate ratio ของการเกิด strokeกับความชอบทำวิจัย $= 3.8/0.7 = 5.43$

ตัวอย่าง

ศึกษา **Rate** การเกิด **stroke** กับความชอบทำวิจัย

Person year at risk. Cases. rate/1000 per yr

- ไม่ชอบทำวิจัย 6848 26 3.8
- ชอบทำวิจัย 8384 6 0.7

$(26/6848) \times 1000$

rate difference ของการเกิด CHD กับความชอบทำวิจัย

$$= 3.8 - 0.7 = 3.1/1000 \text{ py}$$



ODDS

จำนวนคนที่เกิดโรคในช่วงเวลาที่ศึกษา

- $\text{odds} = \frac{\text{จำนวนคนที่เกิดโรคในช่วงเวลาที่ศึกษา}}{\text{จำนวนคนที่ไม่เกิดโรคในช่วงเวลาที่ศึกษา}}$

ศึกษานักเรียนในห้องเรียนจำนวน 200 คน เป็นเวลา 1 ชั่วโมง พบว่ามี การจามอย่างน้อย 1 ครั้งในห้องเรียน

ถ้า risk ของ sneezing คือ

$$50/200=25\% \text{ or } 0.25$$

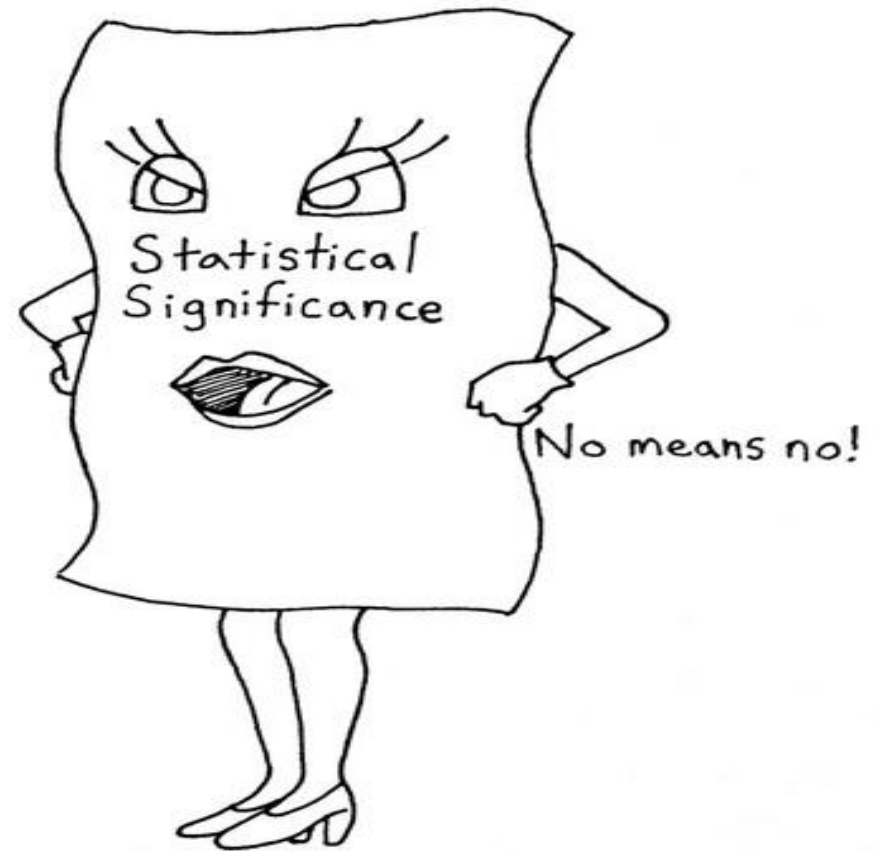
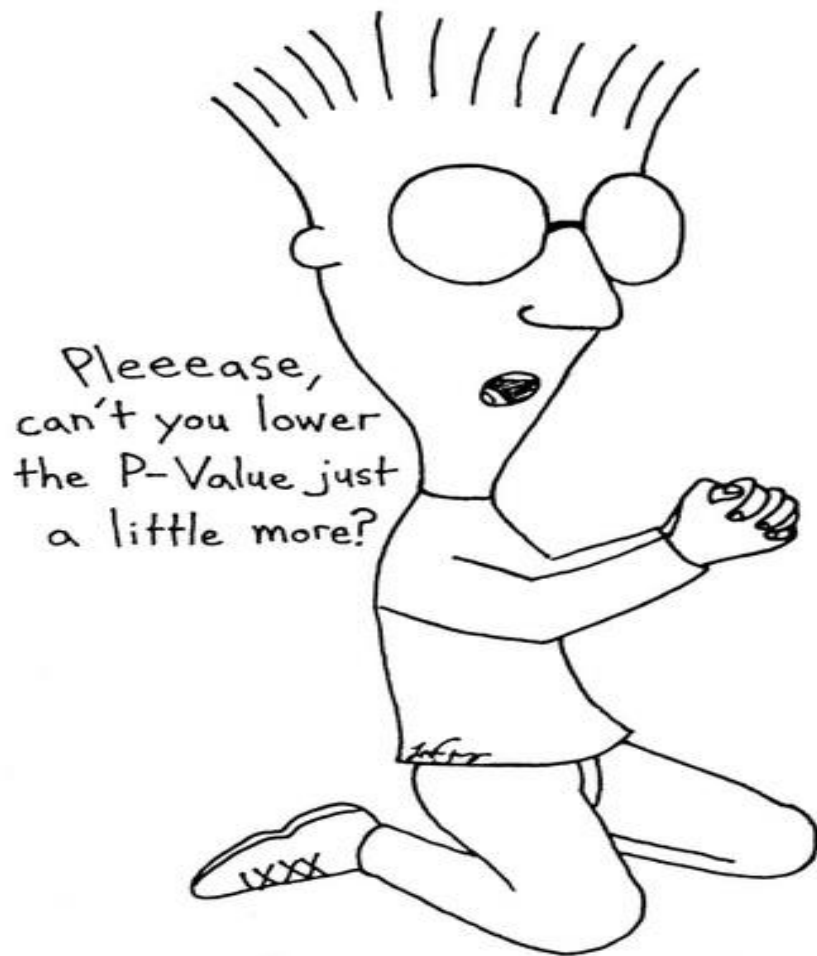
The odds ของ sneezing คือ

QUIZ 6

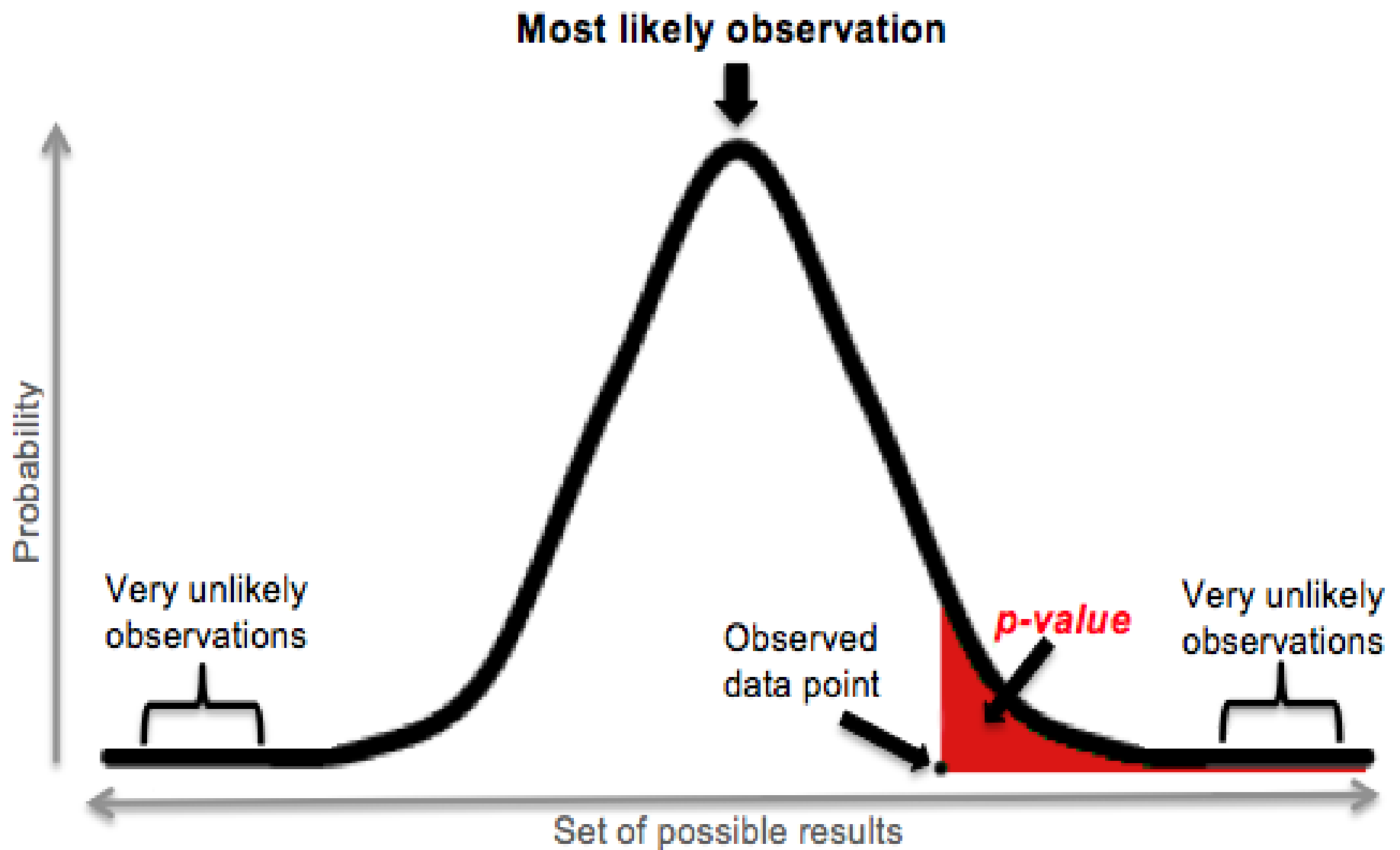


Quiz Time

Let's have
some fun!



Hypothesis testing



A *p-value* (shaded red area) is the probability of an observed (or more extreme) result arising by chance

Values of p	Inference
$p > 0.10$	No evidence against the null hypothesis.
$0.05 < p < 0.10$	Weak evidence against the null hypothesis
$0.01 < p < 0.05$	Moderate evidence against the null hypothesis
$0.05 < p < 0.001$	Good evidence against null hypothesis.
$0.001 < p < 0.01$	Strong evidence against the null hypothesis
$p < 0.001$	Very strong evidence against the null hypothesis



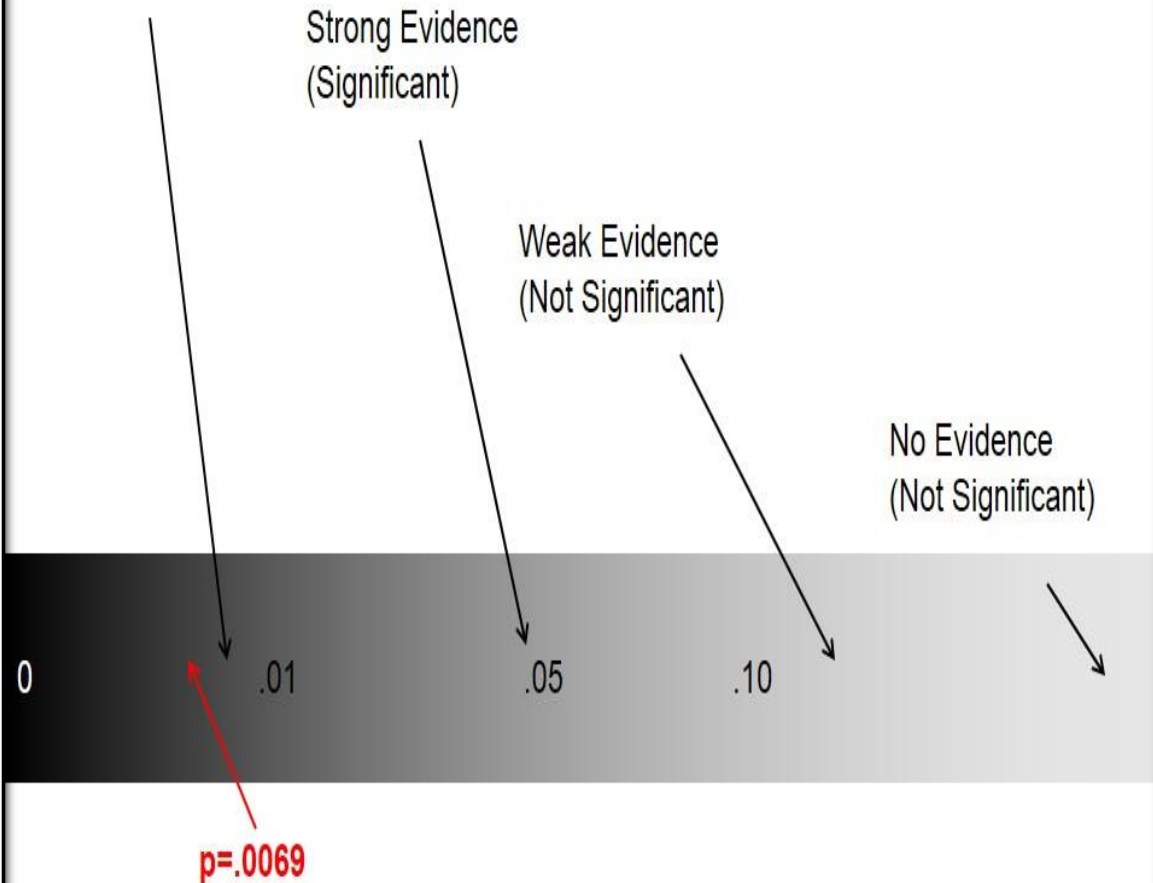
Interpreting the p-value...

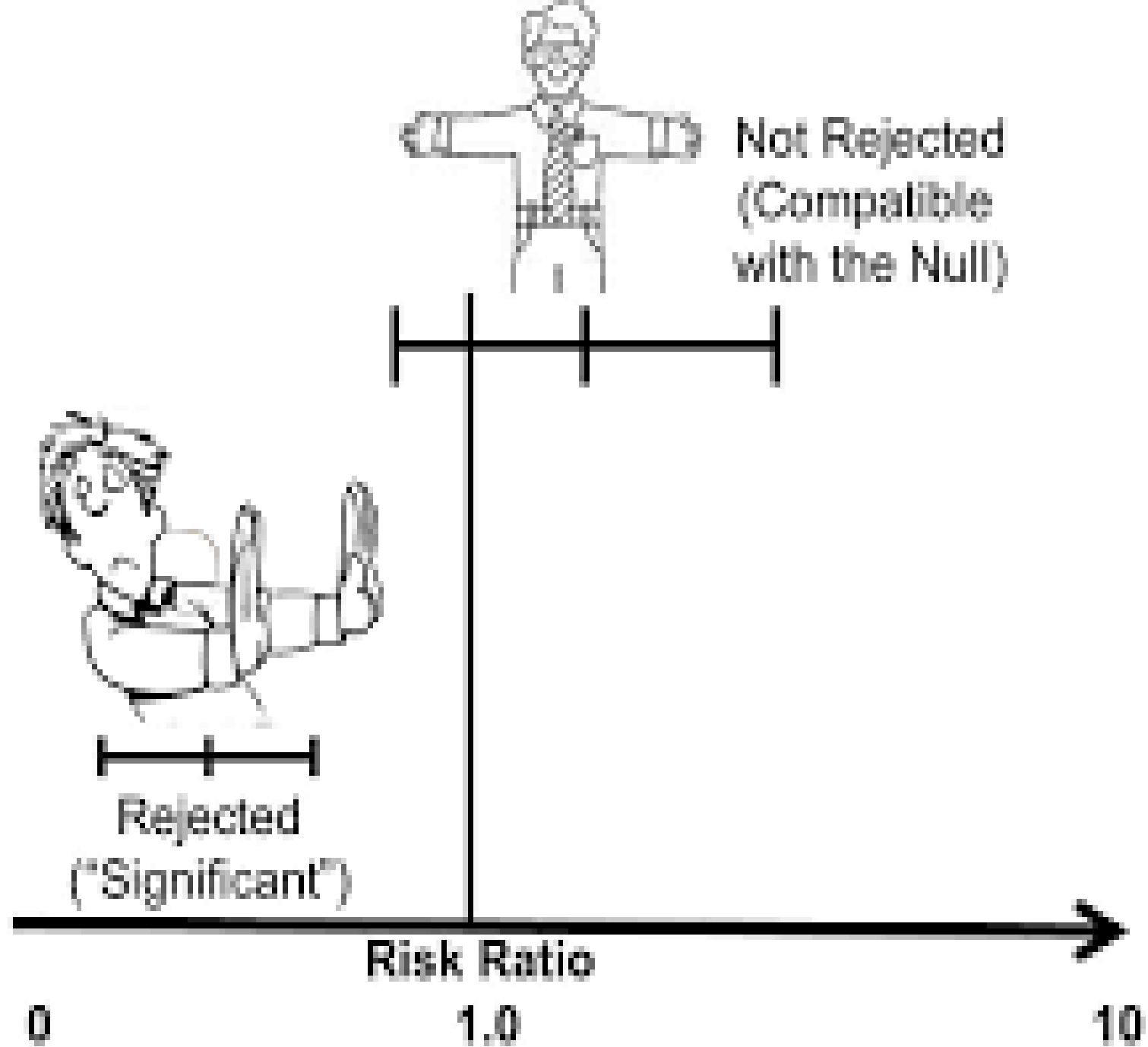
Overwhelming Evidence
(Highly Significant)

Strong Evidence
(Significant)

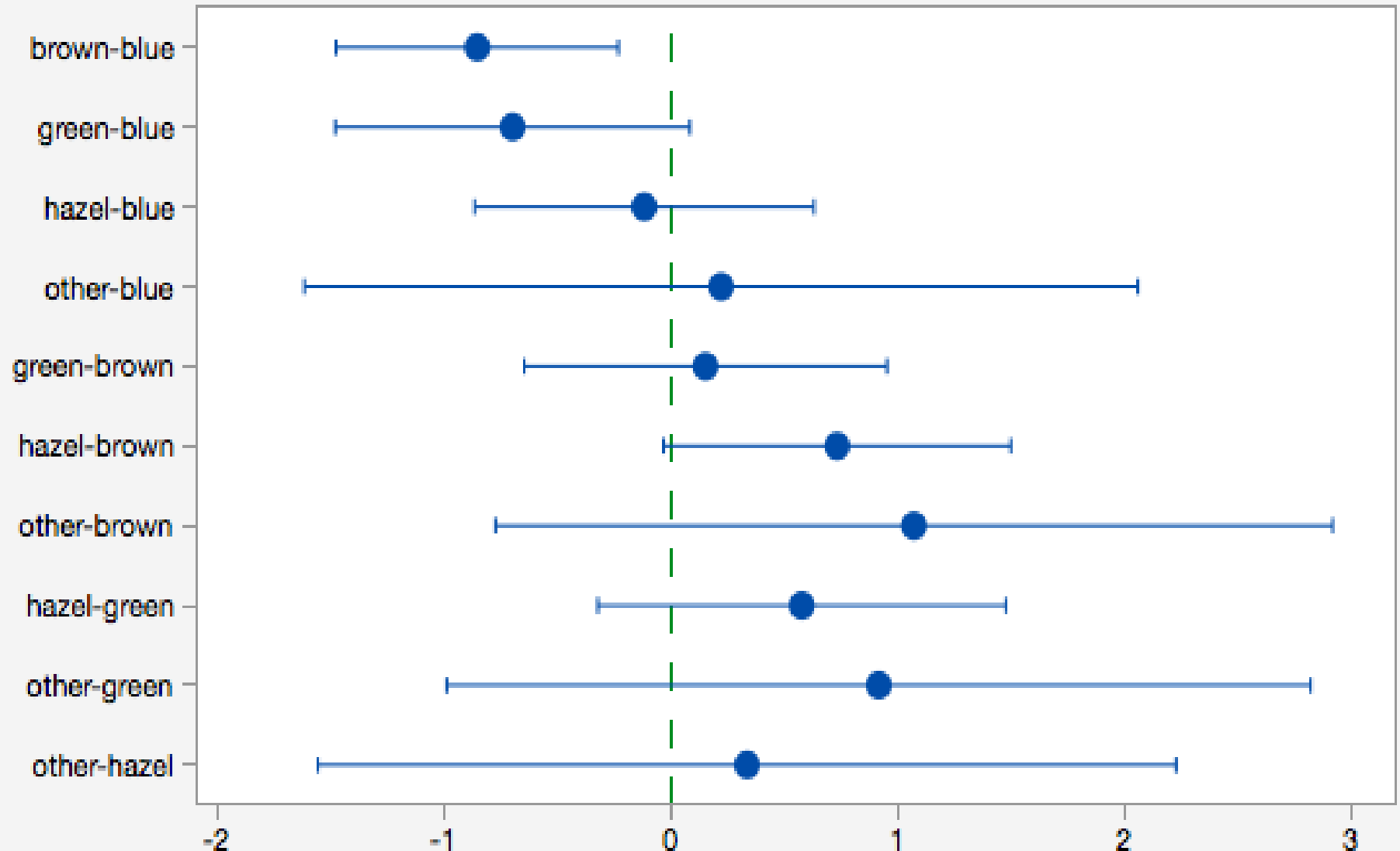
Weak Evidence
(Not Significant)

No Evidence
(Not Significant)





Tukey Simultaneous 95% CIs Differences of Means for height



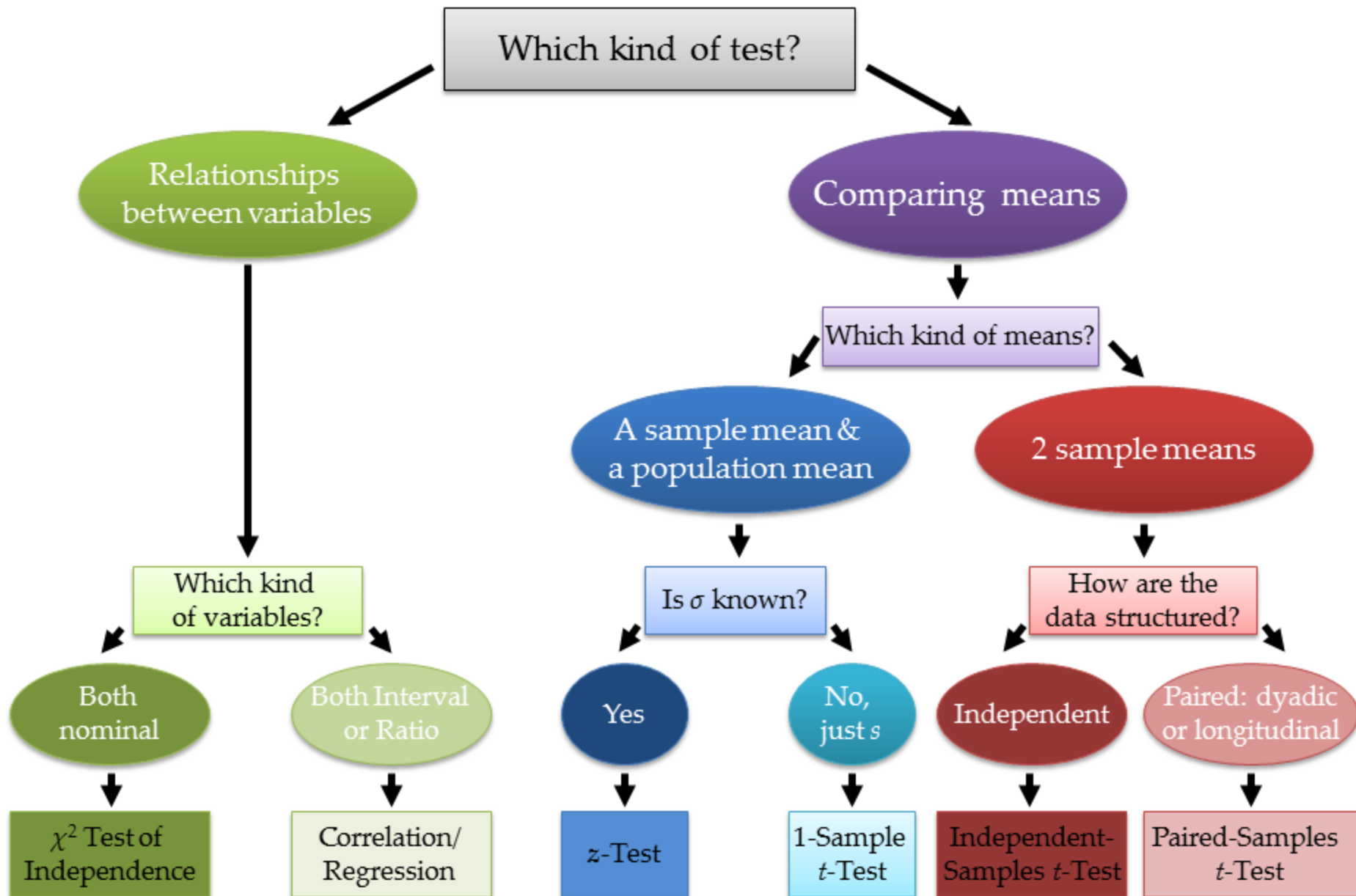
If an interval does not contain 0, the corresponding means are significantly different.

ANOVA

2 Samples 1

1 Sample 1

Decision Tree

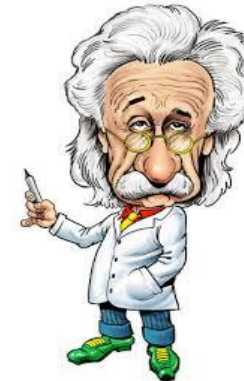


ความแตกต่างของการใช้ Anti-clotting drugs ระหว่าง 2 โรงพยาบาล

		Medication			Total
		Aspirin	Clopidogrel	Warfarin	
Hospital	a	3	1	11	15
	b	1	11	3	15
Total		4	12	14	30

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.905 ^a	2	.001

ตัวอย่าง : Chi-2 test



- Ho โอกาสที่จะเสียชีวิตในกลุ่มที่ได้ Antiserum = โอกาสที่จะเสียชีวิตในกลุ่ม Control
- H1 โอกาสที่จะเสียชีวิตในกลุ่มที่ได้ Antiserum $\neq$ โอกาสที่จะเสียชีวิตในกลุ่ม Control

Observe table

	Dead	Alive	รวม	% Dead
Antiserum	19	65	84	23
ควบคุม	18	22	40	45
รวม	37	87	124	

Expected Table

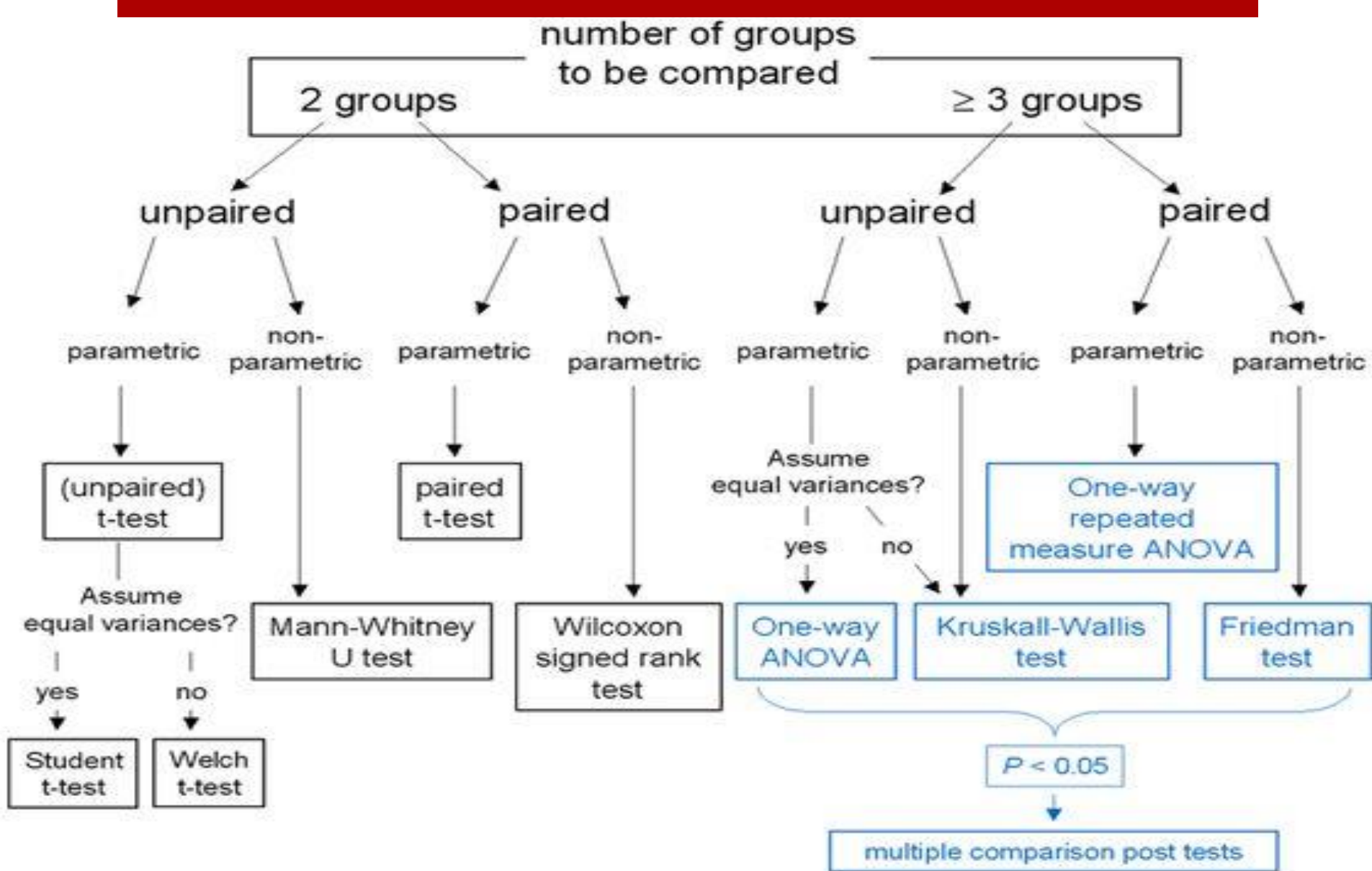
	Dead	Alive	รวม	% Dead
Antiserum	25.1	58.9	84	29.9
ควบคุม	11.9	28.1	40	29.8
รวม	37	87	124	

$$(37 \times 84) / 124$$



ตัวอย่าง Chi-2 test ทดสอบความต่าง
Expected vs. Observe

Expected count ของ
แต่ละ cell คำนวณได้จาก
 $E = (\text{Row total}) \times (\text{Column total}) / (\text{table total})$



Summary Statistic test for quantitative data

Descriptive Statistics				
Group		N	Mean	Std. Deviation
Weight	a	10.0	71.9	17.4
	b	10.0	90.7	8.1

Student's T-test for Equality of Means				
	t	df	Sig. (2-tailed)	Mean Difference
Weight	-3.1	18.0	0.006	-18.8

กลุ่ม a : การให้รับประทานอาหาร Low fat

กลุ่ม b : กลุ่มควบคุม

ตัวอย่าง : t test เปรียบเทียบน้ำหนักของ 2
กลุ่มหลังให้การรักษา

Independent variable (Predictor):ตัวแปรต้น	Dependent variable (outcome):ตัวแปรตาม	วิธีทางสถิติ
Category	Category	Chi-square
Dichotomous (2 categories)	Dichotomous (2 categories)	Odds ratio, Mantel-Haenszel OR, Relative risk, difference in proportions
Dichotomous (2 categories)	Continuous	t-test
Category	Continuous	ANOVA
Multivariate	Dichotomous (2 categories)	Logistic regression
Category	Time-to-event	Kaplan-Meier curve/ log-rank test
Multivariate	Time-to-event	Cox proportional hazard model
Continuous	Continuous	Simple linear regression
Multivariate	Continuous	Multiple linear regression

สรุปสถิติเชิงเปรียบเทียบ

- ข้อใดต่อไปนี้อีกกล่าวถูกต้องเกี่ยวกับสถิติเชิงอนุมาน
 - A. t Test ใช้เปรียบเทียบตัวแปรต้นและตัวแปรตามที่เป็นข้อมูลรายกลุ่ม
 - B. Chi 2 test ใช้เปรียบเทียบตัวแปรต้นที่เป็นข้อมูลต่อเนื่องและตัวแปรตามรายกลุ่ม
 - C. Chi 2 test ใช้ได้เหมาะสมถ้ามี จำนวนประชากรมากพออย่างน้อย 40 คนขึ้นไป
 - D. Fischer's exact test ใช้แทน Chi 2 test กรณีที่มีจำนวนประชากรมาก



Quiz Time

Let's have
some fun!

QUIZ 7

- จงบอกชื่อสถิติที่เหมาะสมในการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลต่อไปนี้

1. ตัวแปรต้น Category ตัวแปรตาม Category
2. ตัวแปรต้น 2 Category ตัวแปรตาม Continuous
3. ตัวแปรต้น Category ตัวแปรตาม Continuous



Quiz Time

Let's have
some fun!

QUIZ 8



PresenterMedia

