

# **ANOVA**

Analyse of Variance

# So sánh trung bình nhiều nhóm ( $>2$ )

Để khảo sát địa điểm năng suất tốt khi trồng rừng, người ta tiến hành trồng thử ở 3 khu vực khác nhau, sau nhiều năm đo lại chiều cao của cây (tính bằng m). Người ta muốn biết chiều cao của cây ở 3 khu vực có khác nhau?

							Mean
Khu 1	23,4	24,4	24,6	24,9	25	26,2	24,75
Khu 2	18,9	21,1	21,1	22,1	22,5	23,5	21,53333
Khu 3	22,5	22,9	23,7	24	24	24,5	23,6

# TRƯỜNG HỢP 1

	$X_1$	$X_2$	$X_3$
	3	5	7
	3	5	7
	3	5	7
	3	5	7
	3	5	7
mean	3	5	7

- Không biến đổi bên trong mỗi nhóm
- Có biến đổi giữa các nhóm



# TRƯỜNG HỢP 2

	$X_1$	$X_2$	$X_3$
	5	6	7
	4	8	2
	5	4	7
	6	3	6
	5	4	3
mean	5	5	5

- Có sự biến đổi bên trong mỗi nhóm
- Không có sự biến đổi giữa các nhóm

# TRƯỜNG HỢP 3

	$X_1$	$X_2$	$X_3$
	3	6	3
	6	8	3
	5	7	2
	6	8	2
	5	6	5
mean	5	7	3

TH1			
	$X_1$	$X_2$	$X_3$
	3	5	7
	3	5	7
	3	5	7
	3	5	7
	3	5	7
mean	3	5	7

	$X_1$	$X_2$	$X_3$
	3	5	7
	3	5	7
	3	5	7
	3	5	7
	3	5	7
mean	3	5	7

	$X_1$	$X_2$	$X_3$
	5	6	7
	4	8	2
	5	4	7
	6	3	6
	5	4	3
mean	5	5	5

	$X_1$	$X_2$	$X_3$
	3	6	3
	6	8	3
	5	7	2
	6	8	2
	5	6	5
mean	5	7	3



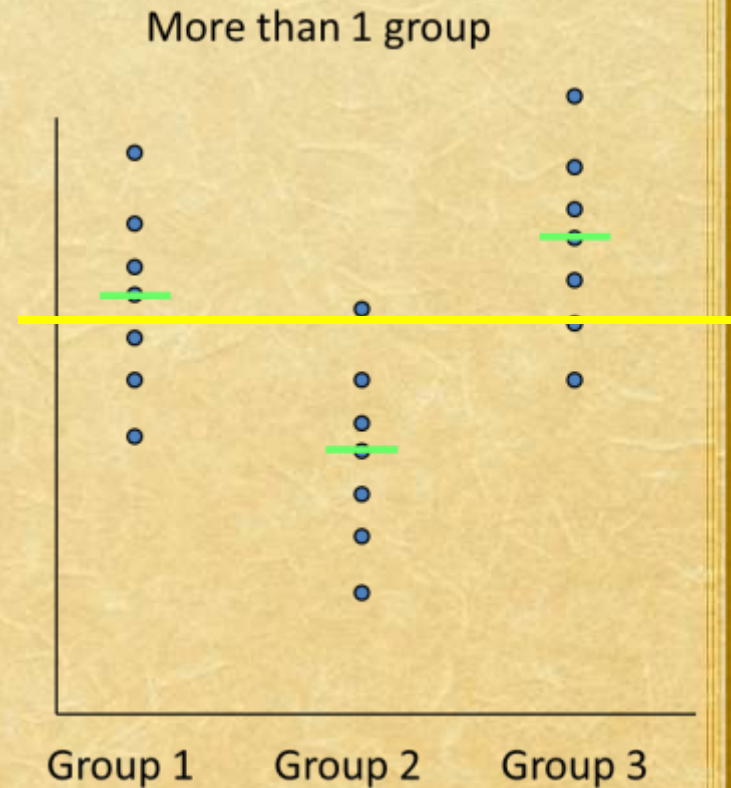
# Tại sao không dùng t-test so sánh từng cặp

- ♦ Với 3 group X ta sẽ phải so sánh X1 và X2; X2 và X3; X1 và X3
  - ♦ Với  $\alpha=0,05$  ta sẽ có độ tin cậy của t-test là 0,95
  - ♦ Nếu so sánh cặp thứ 2, độ tin cậy của t-test là  $0,95*0,95=0,9025$ , như vậy  $\alpha= 1-0,9025=0,0975$
  - ♦ Nếu ta có 4 group A,B,C,D ta sẽ so sánh 6 cặp: A và B; A và C; A và D; B và C, B và D; C và D
- $\alpha=1 - (1-0,05)^6=0,265$



# One way ANOVA

- ♦ Giả định khi phân tích phương sai
  - ♦ Các nhóm có phân phối chuẩn
  - ♦ Các phương sai của các nhóm bằng nhau
  - ♦ Các quan sát độc lập nhau
- ♦ Giả sử ta có 3 nhóm:  $k=3$
- ♦  $H_0: \mu_1 = \mu_2 = \mu_3 (= \dots = \mu_k)$
- ♦  $H_1: \exists \mu_i \neq \mu_j, \forall i \neq j$





# Các Bước tính ANOVA

1. Tính trung bình mẫu của các nhóm, tính trung bình chung của tổng thể
2. Tính các tổng chênh lệch bình phương (gọi tắt là tổng bình phương Sum of Square)
  1. Trong nội bộ nhóm SSW (SS within group): biến thiên do các yếu tố không phải nghiên cứu tạo ra
  2. Giữa các nhóm SSG (SS Between group): Tổng các chênh lệch bình phương giữa các trung bình mẫu của từng nhóm với trung bình chung của k nhóm
  3.  $SST (SS \text{ total}) = SSG + SSW$

# Các bước tính ANOVA (tiếp)

3. Tính các phương sai

i. Phương sai trong nội bộ nhóm (MSW)

ii. Phương sai giữa các nhóm (MSG)

4. Kiểm định giả thiết  $F = \frac{MSG}{MSW}$

Bác bỏ giả thiết  $H_0$  khi  $F > F_{\alpha, (k-1; n-k)}$



- ♦ Hàm lượng alkaloid (mg) trong 1 loại dược liệu được thu hái từ 3 vùng khác nhau

Vùng 1	7,5	6,8	7,1	7,5	6,8	6,6	7,8
Vùng 2	5,8	5,6	6,1	6	5,7		
Vùng 3	6,1	6,3	6,5	6,4	6,5	6,3	

- ♦ Hỏi hàm lượng alkaloid có khác nhau theo vùng không?



Một nghiên cứu được thực hiện nhằm xem xét năng suất lúa trung bình của 3 giống lúa. Kết quả thu thập qua 4 năm như sau:

Năm	Giống A	Giống B	Giống C
1	65	69	75
2	74	72	70
3	64	68	78
4	83	78	76

Hãy cho biết năng suất lúa trung bình của 3 giống lúa có khác nhau hay không?  $\alpha=0,05$

# **ANOVA 2-FACTOR**

- ♦ Chiết suất chất X từ 1 loại dược liệu bằng 3 phương pháp và 5 loại dung môi, ta có kết quả:

Dung môi (A)	Phương pháp chiết suất (B)		
	B1	B2	B3
A1	120	60	60
A2	120	70	50
A3	130	60	50
A4	150	70	60
A5	110	75	54

- ♦ Hãy xét ảnh hưởng của phương pháp chiết suất và dung môi đến kết quả chiết suất chất X với  $\alpha = 0,05$



- ♦ Hàm lượng saponin (mg) của cùng một loại dược liệu được thu hái trong 2 mùa (khô và mưa: trong mỗi mùa lấy mẫu 3 lần - đầu mùa, giữa mùa, cuối mùa) và từ 3 miền (Nam, Trung, Bắc) thu được kết quả sau:

Mùa	Thời điểm thu mẫu	Miền		
		Nam	Trung	Bắc
Khô	Đầu mùa	2,4	2,1	3,2
	Giữa mùa	2,4	2,2	3,2
	Cuối mùa	2,5	2,2	3,4
Mưa	Đầu mùa	2,5	2,2	3,4
	Giữa mùa	2,5	2,3	3,5
	Cuối mùa	2,6	2,3	3,5

- ♦ Hãy cho biết hàm lượng saponin có khác nhau theo mùa hay miền không? Nếu có thì 2 yếu tố mùa và miền có sự tương tác với nhau hay không với  $\alpha = 0,05$

# Bài tập

- ♦ Thử nghiệm phân bón dung trong thủy canh rau
- ♦ Đối tượng khảo sát: Rau muống
- ♦ Mục đích: nhà sản xuất phân bón muốn so sánh thử nghiệm trồng rau muống với dung dịch hữu cơ và vô cơ và sự tương tác giữa N và P có xảy ra không?
- ♦ N: dùng urê,  $\text{KNO}_3$ , 2 hợp chất hữu cơ HC1 và HC2
- ♦ P: dùng DAP và  $\text{H}_3\text{PO}_4$
- ♦ Chỉ tiêu được dùng để khảo sát là độ dài thân (cm)



# Trồng thủy canh rau muống





# Các nghiệm thức được thực nghiệm

- ☐ Nghiệm thức A:  $\text{H}_3\text{PO}_4 + \text{URE}$ .
- ☐ Nghiệm thức B:  $\text{H}_3\text{PO}_4 + \text{KNO}_3$ .
- ☐ Nghiệm thức C:  $\text{H}_3\text{PO}_4 + \text{HC1}$
- ☐ Nghiệm thức D:  $\text{H}_3\text{PO}_4 + \text{HC2}$
- ☐ Nghiệm thức E:  $\text{DAP} + \text{URE}$ .
- ☐ Nghiệm thức F:  $\text{DAP} + \text{KNO}_3$ .
- ☐ Nghiệm thức G:  $\text{DAP} + \text{HC1}$
- ☐ Nghiệm thức D:  $\text{DAP} + \text{HC2}$

## H3PO4

## DAP

URE	KNO3	HC1	HC2
20,5	18,5	21,5	23
14,5	16,5	23,5	30,2
19	13,5	21	26
21,3	22,5	26	23
13,5	15,5	15	19,6
21,5	14,5	28,8	29,8
18	9	21,3	29,5
16,5	14	10	26,5
17,5	21,5	22,8	27,5
21	13	23,5	24
14,5	21,5	26,2	31,3
21,9	9	23	24,6
9	21	22	21
19,2	16	27	26,4
13,3	16,5	23,4	25,6
15,6	12,5	22,6	25
14,1	18,5	28,1	22,3
15,5	11	17,4	27,5
15	21	28,4	28
16,5	21,5	25,6	26,5

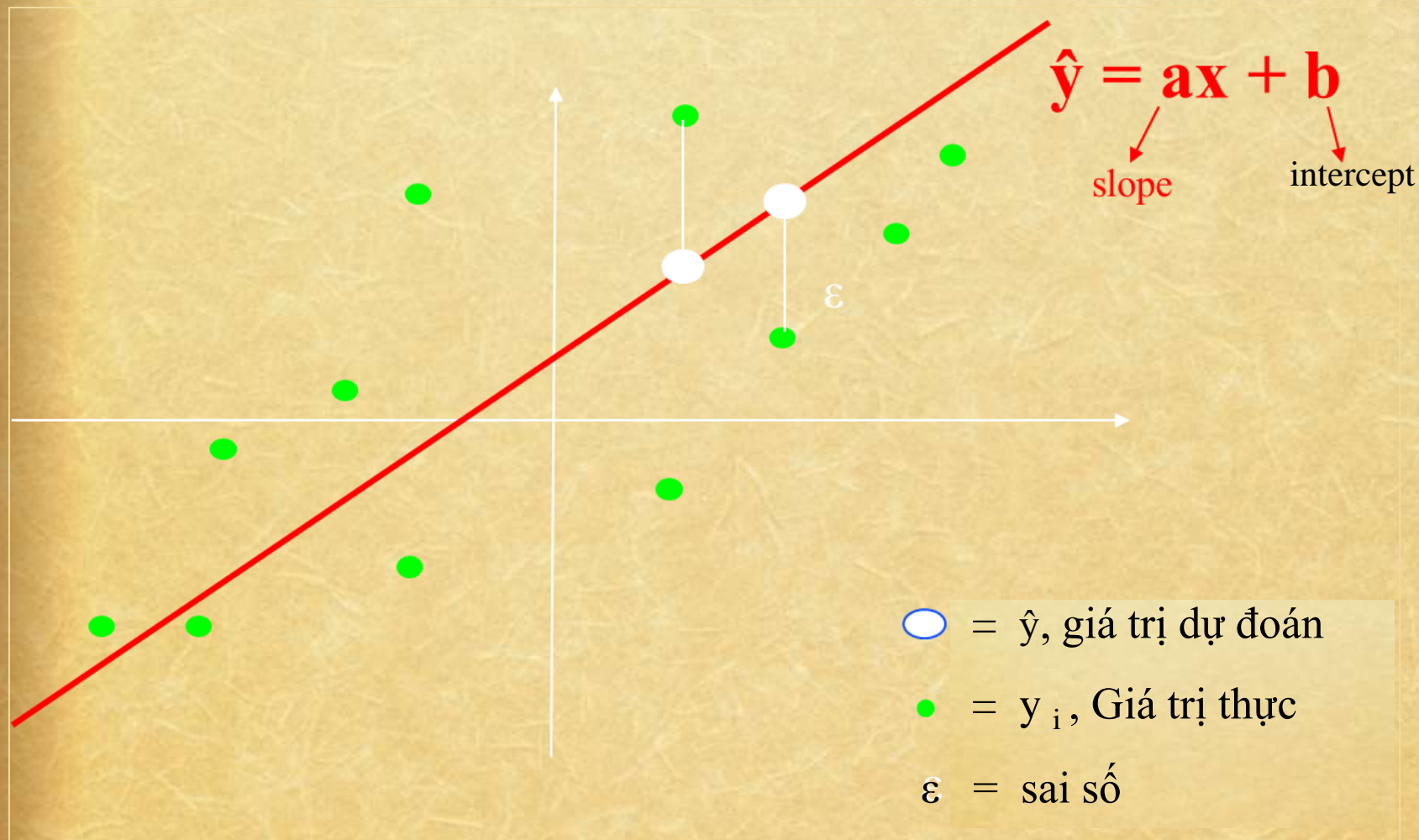
URE	KNO3	HC1	HC2
18,2	21,5	29,1	27,5
26	23,5	23,5	33
25	12,5	24,3	30,5
26,3	13,5	20,3	20,1
19,4	11,6	26,2	30
24,9	18,8	16,5	228,1
21	16	24,1	25,5
30,4	11,6	23,5	28,7
22	14,7	21,5	25,5
20,1	10,8	21,5	23,1
23	10	25,5	22,3
22,3	15,7	24,5	25,5
26,1	13,4	23,4	19,1
20,2	13,5	21,6	20,6
28,2	14,7	23	29,3
17,4	10,3	16,5	15,5
27	8,5	18,7	14,7
17,5	14,8	18,9	18,6
25,8	18,5	17,5	16
23,5	14,2	20,9	17,2

# Hỏi

- ♦ Có sự khác biệt giữa 4 nghiệm thức khi trồng với  $\text{H}_3\text{PO}_4$  hay không?
- ♦ Có sự khác biệt giữa 4 nghiệm thức khi trồng với DAP hay không?
- ♦ Có sự tương tác giữa N và P hay không?
- ♦ Những chú ý khi bố trí thí nghiệm là gì?



# Phương trình hồi quy



# Regression

	Length (cm)	Weight (g)
	60	136
	69	198
	66	194
	64	140
	54	93
	67	172
	59	116
	65	174
	63	145
Mean	63	152
SD	4.6	35.3

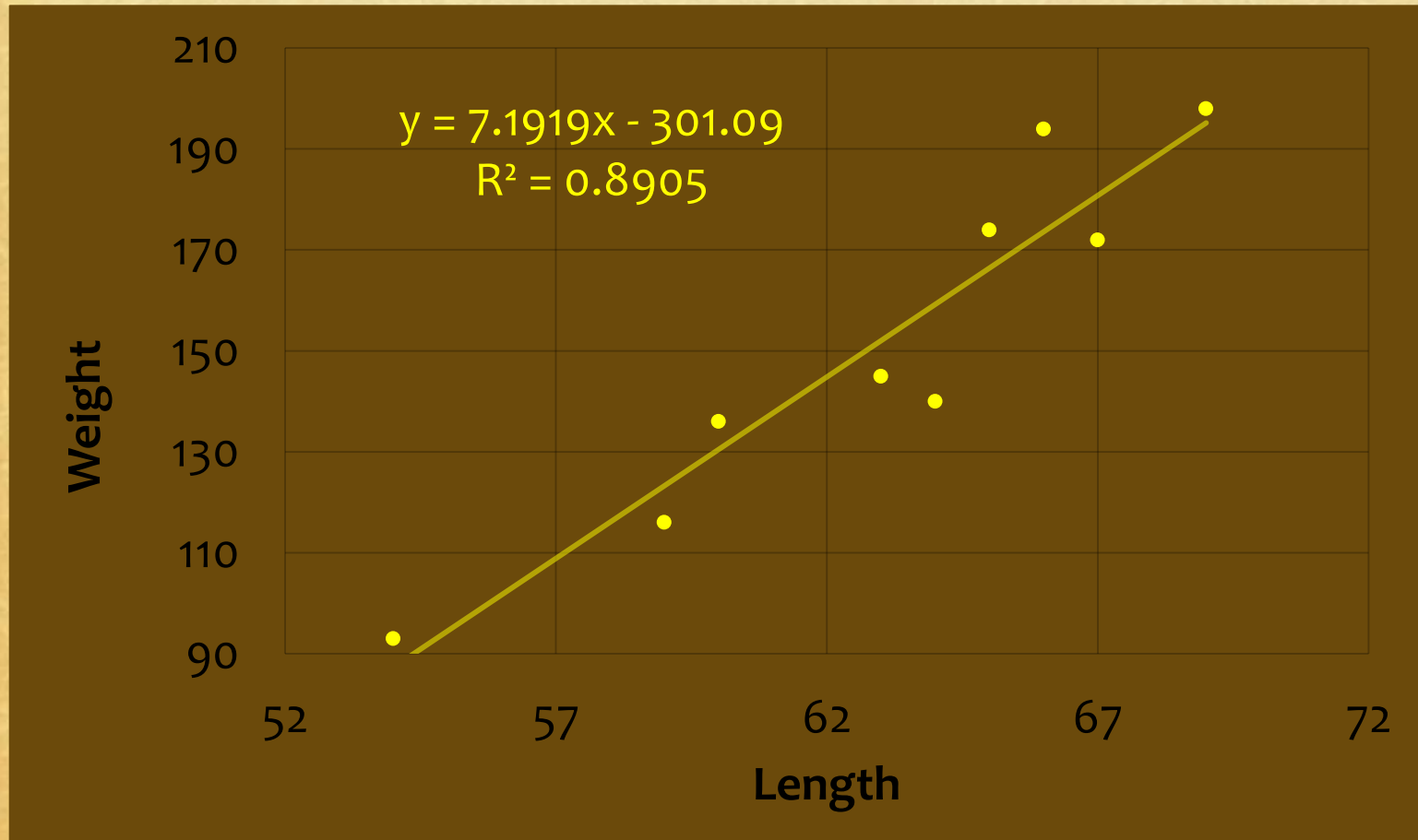
- ♦ Khi nghiên cứu quần thể rắn sống ngoài tự nhiên, các nhà nghiên cứu bắt và đo 9 cá thể cá theo 2 biến;
  - ♦ Chiều dài (X)
  - ♦ Trọng lượng (Y)

# Kết quả từ Excel

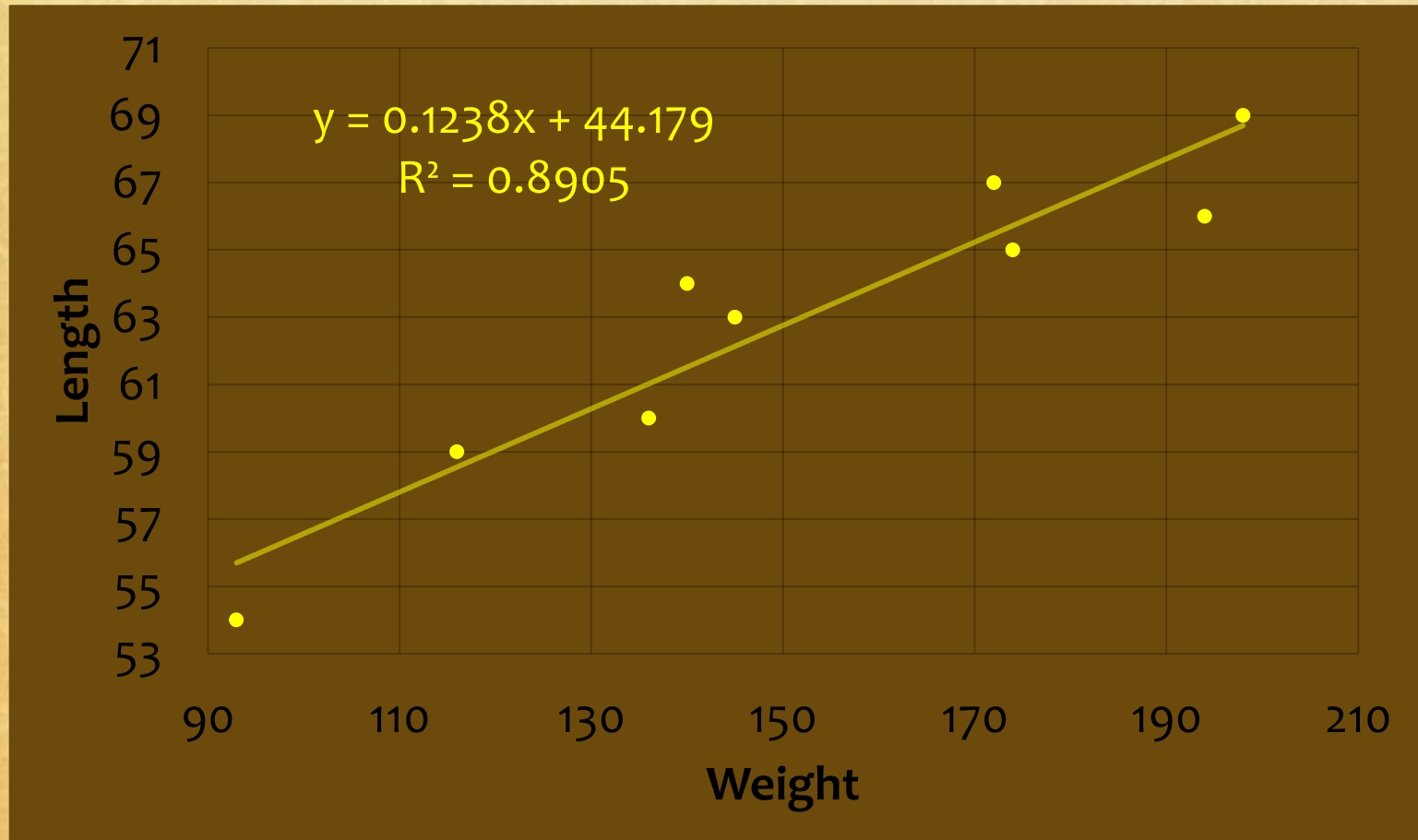
SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.943675613					
R Square	0.890523663	89% trọng lượng được giải thích theo chiều dài	100-89=11% biến thiên của khối lượng không giải thích được			
Adjusted R Square	0.874884187					
Standard Error	12.49953488					
Observations	9					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	8896.331395	8896.331395	56.94076	0.000132114	
Residual	7	1093.668605	156.2383721			
Total	8	9990				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept (b)	-301.0872093	60.18846311	-5.002407334	0.001561	-443.4103088	-158.764
Length (cm) (a)	7.191860465	0.953080599	7.545910043	0.000132	4.938182967	9.445538



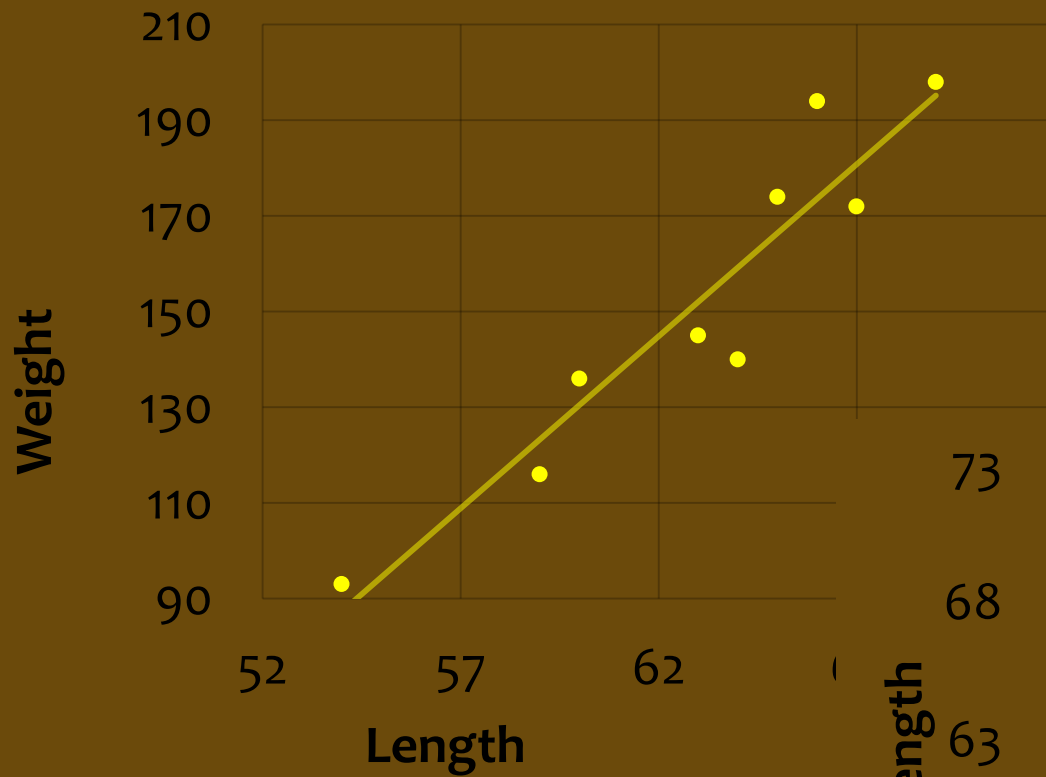
# Weight~Length



# Length~Weight



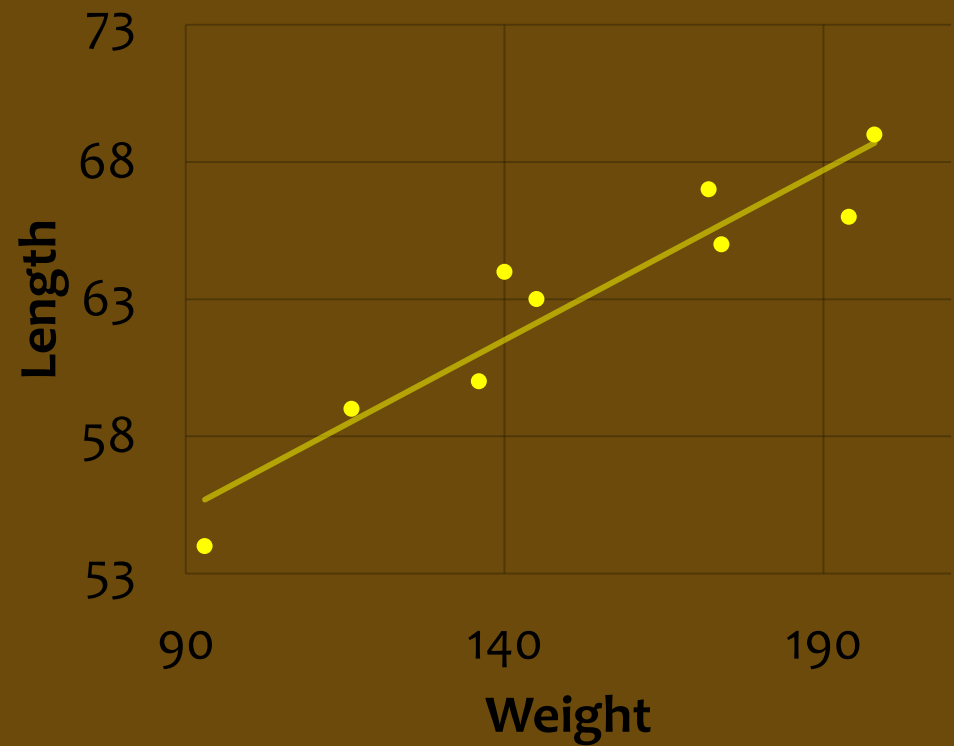
## Weight~Length



$$y = 7,1919x - 301,09$$
$$R^2 = 0,8905$$

## Length~Weight

$$y = 0,1238x + 44,179$$
$$R^2 = 0,8905$$





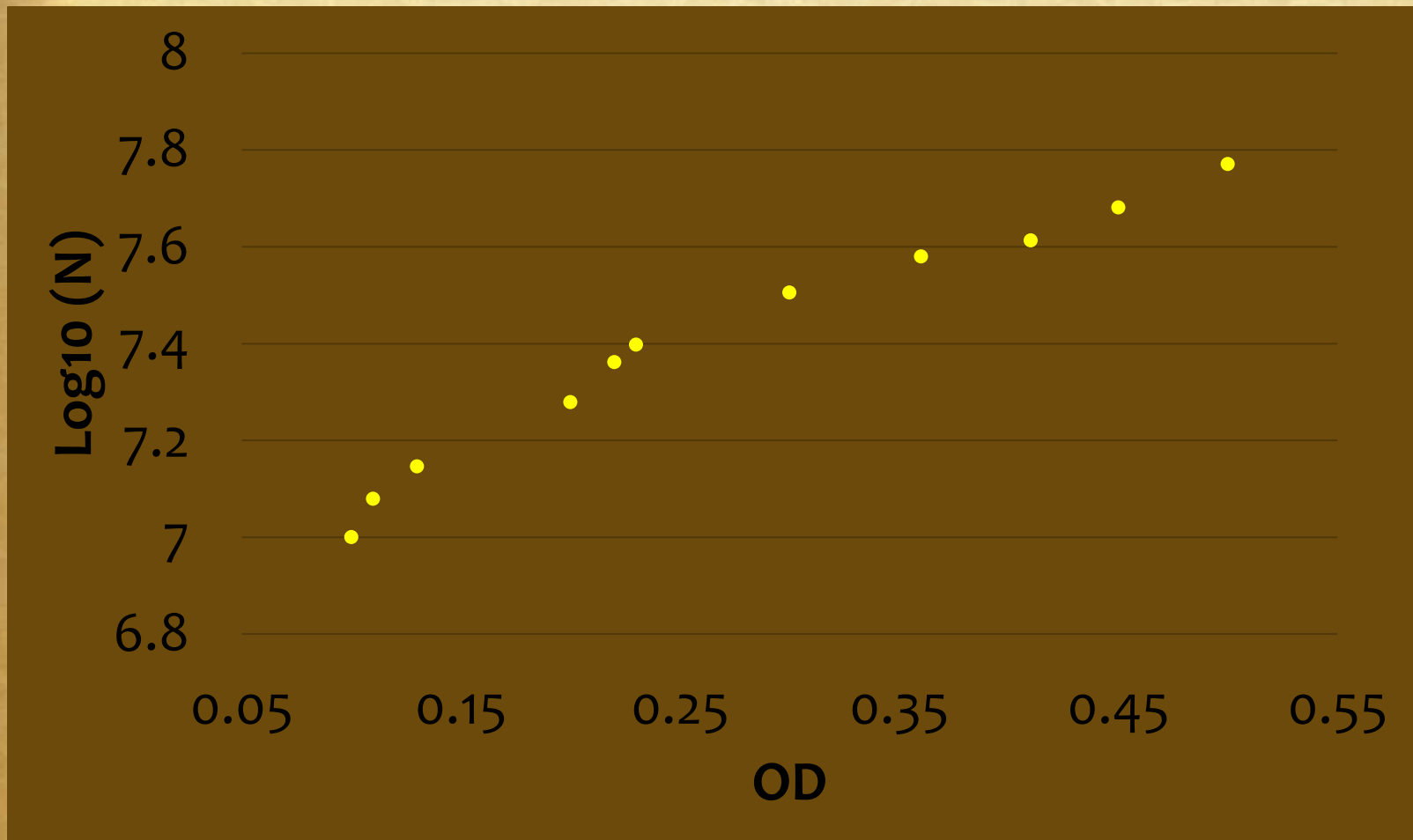
# OD vs. Log N

- ♦ Để khảo sát đường cong tăng trưởng của vi sinh vật bằng phương pháp đo OD
- ♦ Xây dựng đường chuẩn theo bảng sau:

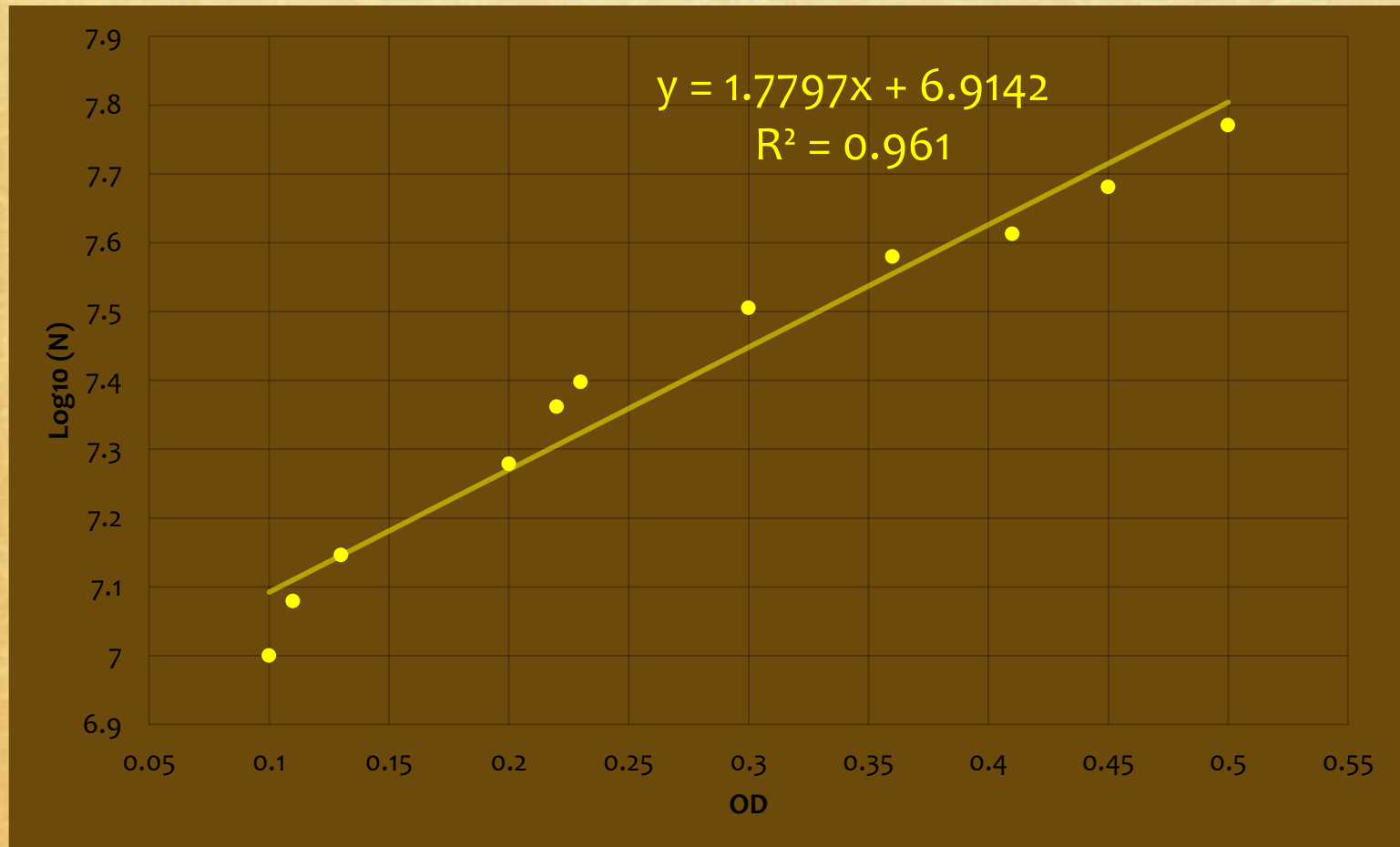
OD	0,1	0,11	0,13	0,2	0,22	0,23	0,3	0,36	0,41	0,45	0,5
log10	7,00	7,08	7,15	7,28	7,36	7,40	7,51	7,58	7,61	7,68	7,77

- ♦ Tìm phương trình đường chuẩn  $\log_{10} \sim \text{OD}$
- ♦ Cho biết phương trình có ý nghĩa?
- ♦ Bao nhiêu % thông tin được lý giải bởi phương trình này?

# Đám mây điểm



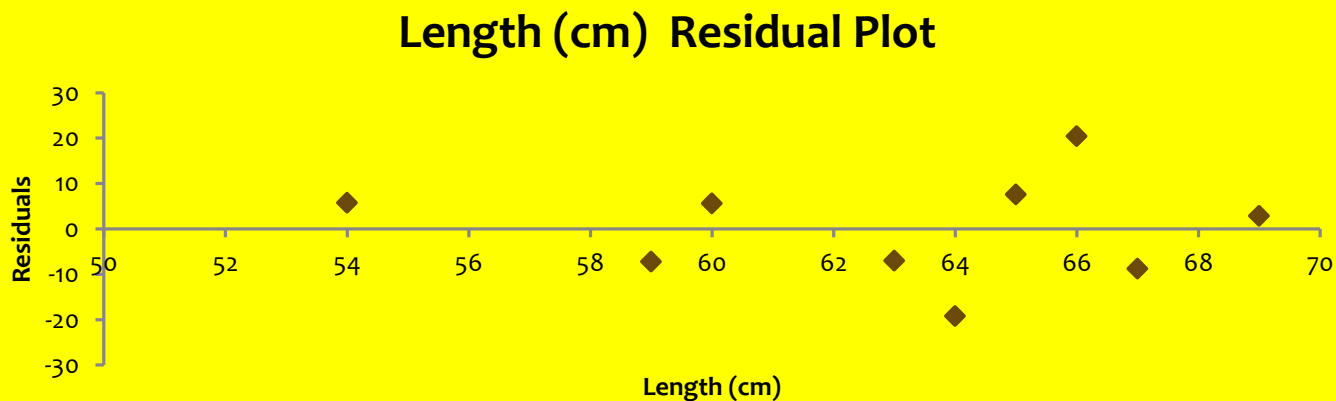
# Hồi quy



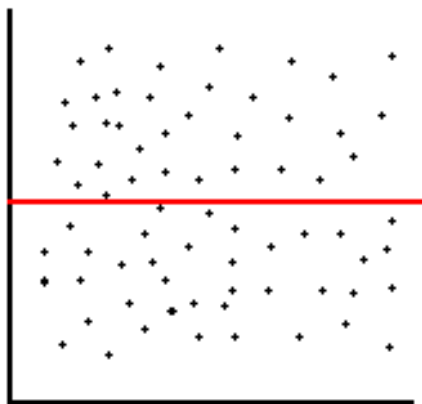


# Số dư

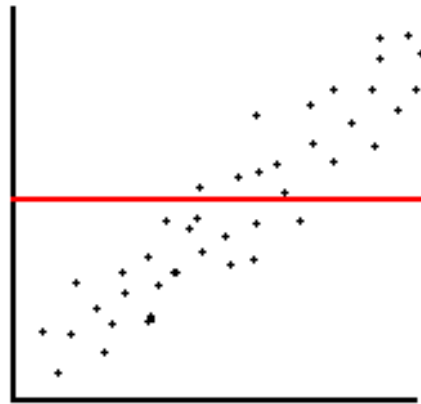
RESIDUAL OUTPUT			
Observation	Predicted Weight (g)	Residuals	Standard Residuals
1	130.4244186	5.575581395	0.476861523
2	195.1511628	2.848837209	0.243651873
3	173.5755814	20.4244186	1.746834754
4	159.1918605	-19.19186047	-1.641418025
5	87.27325581	5.726744186	0.489789989
6	180.7674419	-8.76744186	-0.74985107
7	123.2325581	-7.23255814	-0.618577408
8	166.3837209	7.61627907	0.651395823
9	152	-7	-0.598687459



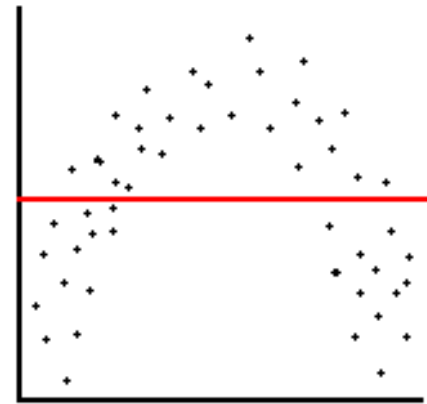
# Phần dư



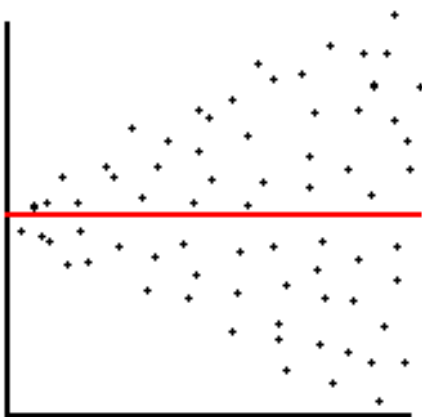
(a) Unbiased and Homoscedastic



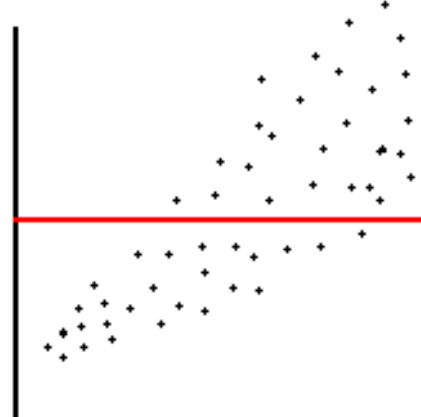
(b) Biased and Homoscedastic



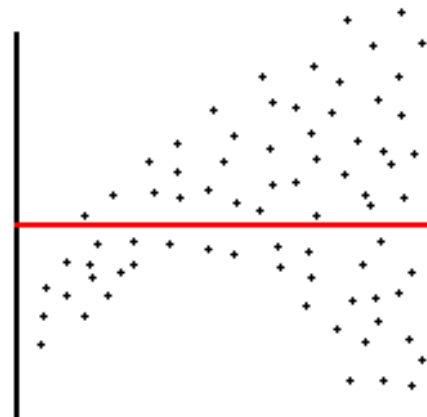
(c) Biased and Homoscedastic



(d) Unbiased and Heteroscedastic

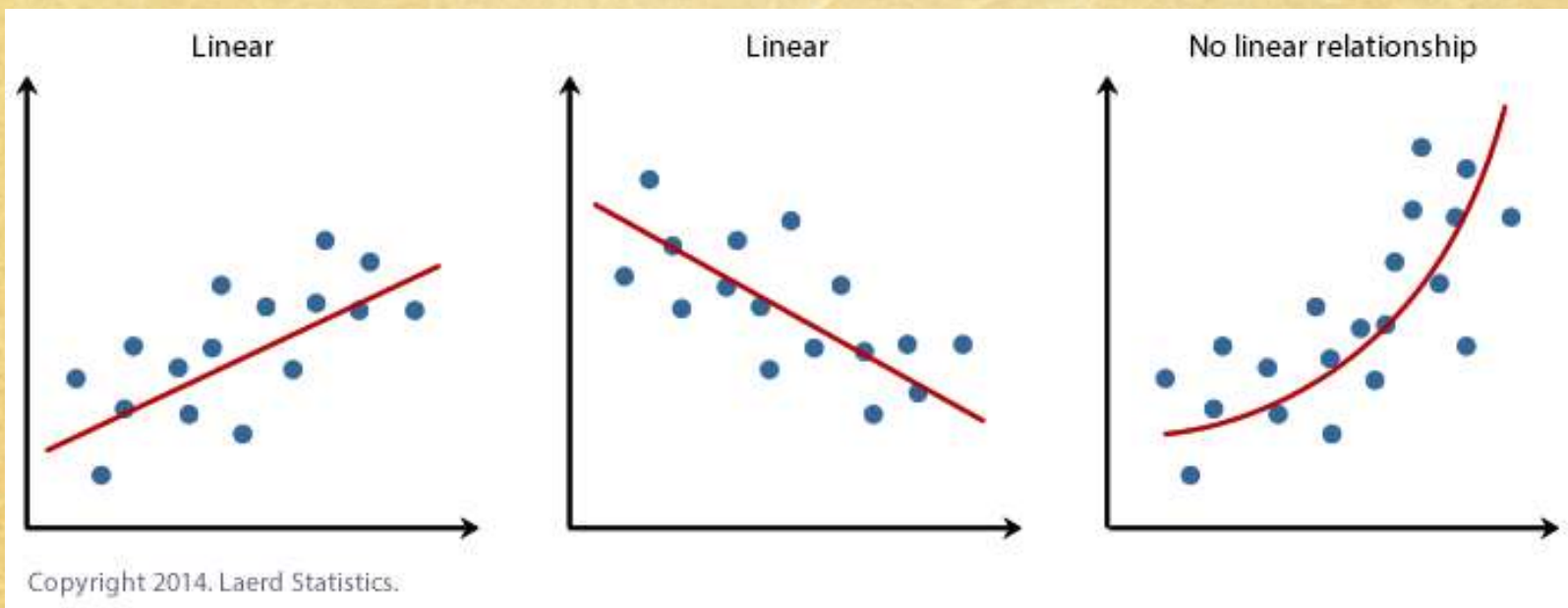
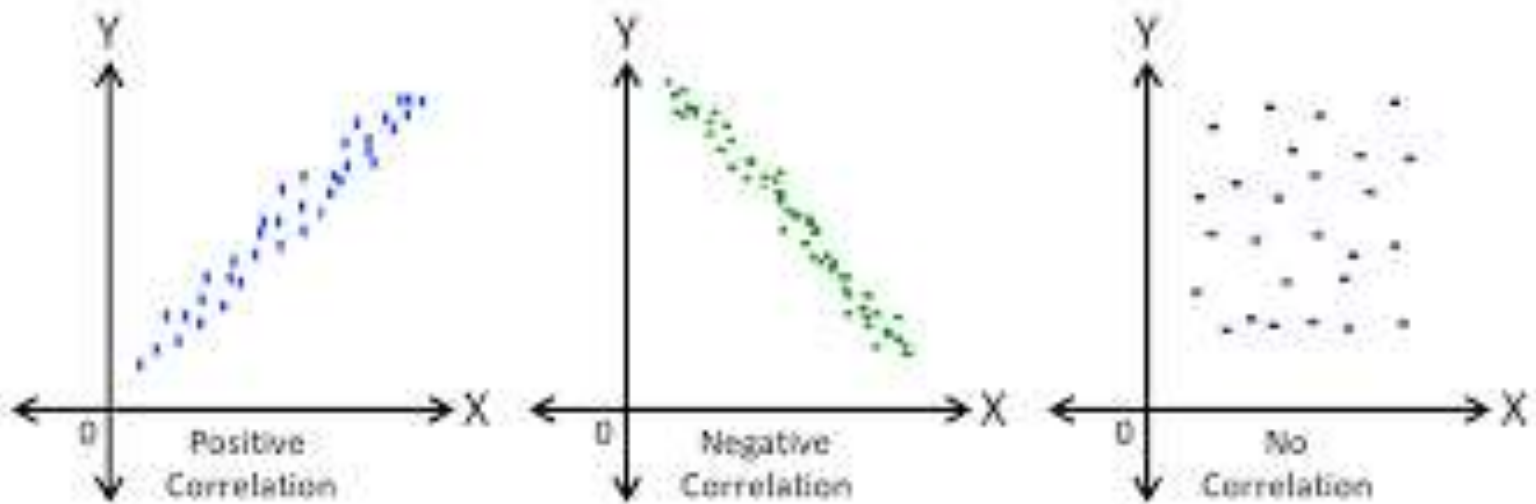


(e) Biased and Heteroscedastic



(f) Biased and Heteroscedastic

## Scatter Plots & Correlation Examples





# Giả định cho hồi quy

- ♦ Độc lập
- ♦ Tuyến tính
- ♦ Trung bình của các sai số phải bằng không
- ♦ Đồng phương sai
- ♦ Phần dư phải tuân theo phân phối chuẩn