SERVIZIO SANITARIO REGIONALE
EMILIA-ROMAGNA
Azienda Ospedaliera di Reggio Emilia

Arcispedale S. Maria Nuova

HIPPOKRATIA
Quarterly Medical Journal
ISSN 1108-4189

Πολυμεταβλητές αναλύσεις

Κυργίδης Αθανάσιος MD, DDS, BOpt, PhD
MSc Medical Research, Μετεκπαίδευση ΕΠΙ ΕΚΑΒ
Γναθοπροσωπικός Χειρουργός
Ass. Editor, Hippokratia

Edward Norton Lorenz
(born May 23, 1917)

Θεωρία του Χάους

Deterministic Nonperiodic Flow[1]

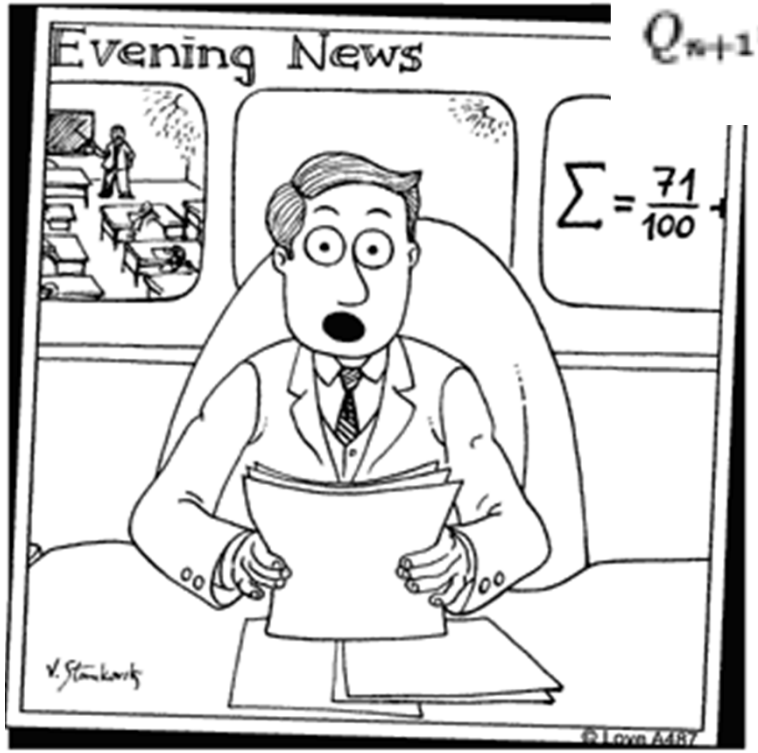
EDWARD N. LORENZ

Massachusetts Institute of Technology

(Manuscript received 18 November 1962, in revised form 7 January 1963)

$$Q_{n+1} = Q_n + (dQ/dt)_n \Delta t + \tfrac{1}{2} \sum_i F_i^2(P_n) \Delta t^2. \qquad (15)$$

Edward N. Lorenz, "Deterministic non-periodic flow," *Journal of the Atmospheric Sciences*, 1963



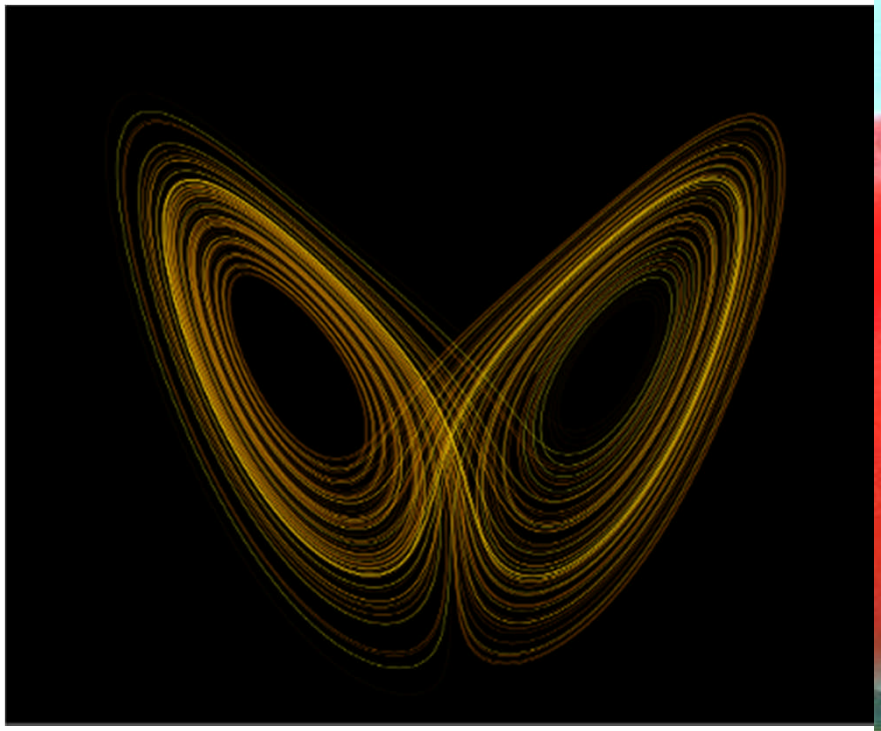
"Of students surveyed, 64% prefer English and 32% prefer math. The fact that these numbers do not add up to 100 may help explain why."

$$\alpha(s) \equiv \sup\left\{ \left| P(A \cap B) - P(A)P(B) \right| : -\infty < t < \infty, A \in X_{-\infty}^{t}, B \in X_{t+s}^{\infty} \right\}.$$

Stephen H. Kellert, *In the Wake of Chaos: Unpredictable Order in Dynamical Systems*, University of Chicago Press, 1993

“Παράξενοι ελκυστές” στο χώρο φάσεων

- **παράξενοι ελκυστές**
  - **στην καρδιά του Χάους υπάρχει <u>τάξη (ντετερμινισμός)</u>**



Stephen H. Kellert, *In the Wake of Chaos: Unpredictable Order in Dynamical Systems*, University of Chicago Press, 1993

Edward N. Lorenz, "Deterministic non-periodic flow," *Journal of the Atmospheric Sciences*, 1963

Narrow focus = limited number of endpoints

$p<0.05$ 1/20 → statistical significance

Multiple Comparisons (outcomes)

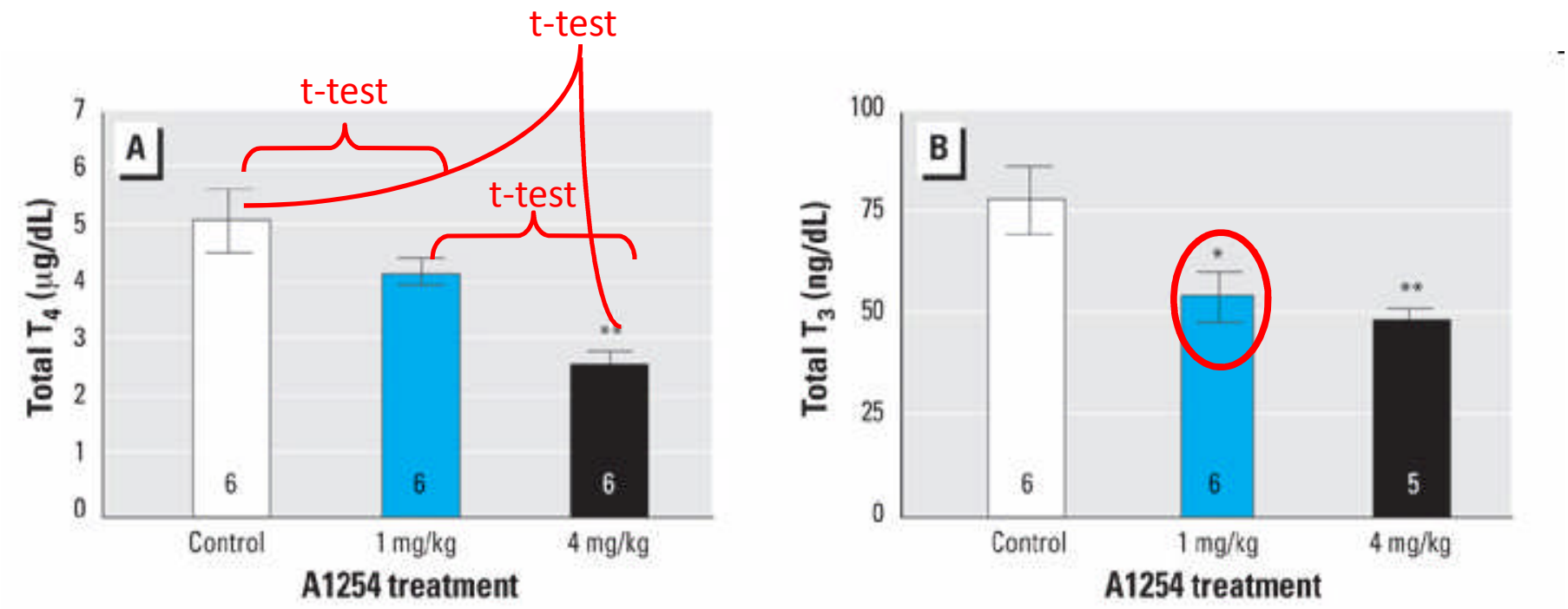


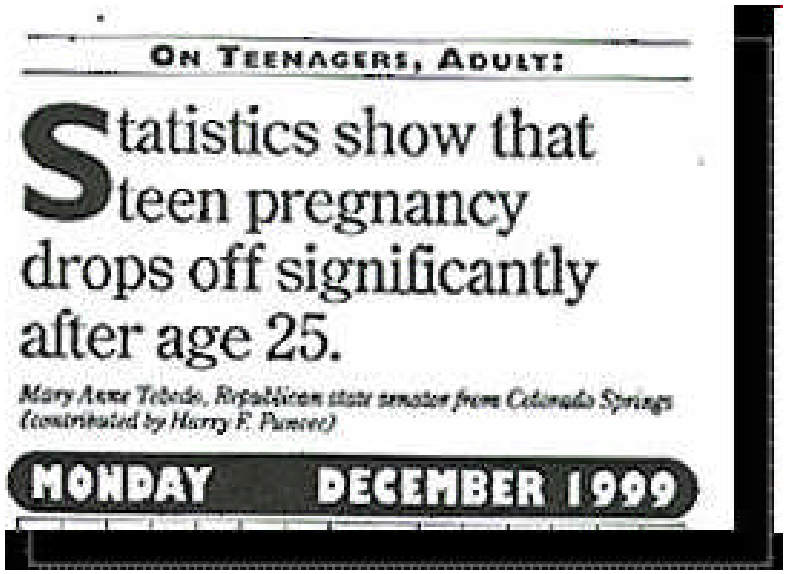
Figure 2. Effect of A1254 treatment on serum concentrations of total T_4 (*A*) and total T_3 (*B*) in dams at the time they were sacrificed on GD16. Bars represent mean ± SEM; number of animals in each group is shown within each bar. See Materials and Methods for treatment details.
*$p < 0.05$; **$p < 0.01$ (significantly different from control group using Bonferroni's *t*-test after one-way ANOVA).

Gauger et al. Environ Health Perspect 2004

Simpson's Paradox and Clinical Trials: What You Find Is Not Necessarily What You Prove

- RCT: brain injury following cardiac arrest
  - Myocardial Infarction
  - Drowning

  - Adverse effect: hypotension


ON TEENAGERS, ADULT:

Statistics show that teen pregnancy drops off significantly after age 25.

Mary Anne Tebedo, Republican state senator from Colorado Springs (contributed by Harry F. Punces)

MONDAY DECEMBER 1999

Abramson et al. Ann Emerg Med 1992

Kyrgidis et al. J Clin Oncol 2010

Kyrgidis et al. Lancet Oncol 2013

Placebo Therapy

200 Patients
95 Dead
45% Mortality

Hypotensive

20 Patients
15 Dead
75% Mortality

Normotensive

180 Patients
80 Dead
44% Mortality

MI

15 Patients
13 Dead
87%

Drowning

5 Patients
2 Dead
40%

MI

60 Patients
49 Dead
82%

Drowning

120 Patients
31 Dead
26%

Experimental Therapy

200 Patients
110 Dead
55% Mortality

Hypotensive

150 Patients
93 Dead
62% Mortality

Normotensive

50 Patients
17 Dead
34% Mortality

MI

60 Patients
54 Dead
90%

Drowning

60 Patients
39 Dead
43%

MI

6 Patients
5 Dead
83%

Drowning

44 Patients
12 Dead
27%

Confounders

Abramson et al. Ann Emerg Med 1992

Kyrgidis et al. J Clin Oncol 2010

Kyrgidis et al. Lancet Oncol 2013

Simpson's Paradox

- Example: 44% of male applicants are admitted by a university, but only 33% of female applicants
- Does this mean there is unfair discrimination?
- University investigates and breaks down figures for Engineering and English programmes

	Male	Female
Accept	35	20
Refuse entry	45	40
Total	80	60

Simpson's Paradox

- No relationship between sex and acceptance for either programme
  - So no evidence of discrimination
- Why?
  - More females apply for the English programme, but it it hard to get into
  - More males applied to Engineering, which has a higher acceptance rate than English
- Must look deeper than single cross-tab to find this out

Engineer-ing	Male	Female
Accept	30	10
Refuse entry	30	10
Total	60	20

English	Male	Female
Accept	5	10
Refuse entry	15	30
Total	20	40

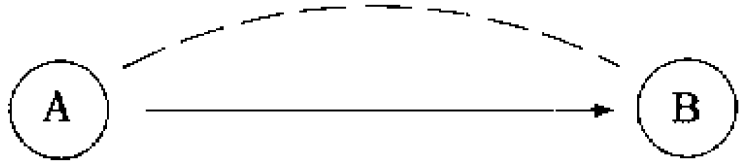
Another Example

- A study of graduates' salaries showed negative association between economists' starting salary and the level of the degree
  - i.e. PhDs earned less than Masters degree holders, who in turn earned less than those with just a Bachelor's degree
  - Why?
- The data was split into three employment sectors
  - Teaching, government and private industry
  - Each sector showed a positive relationship
  - Employer type was confounded with degree level

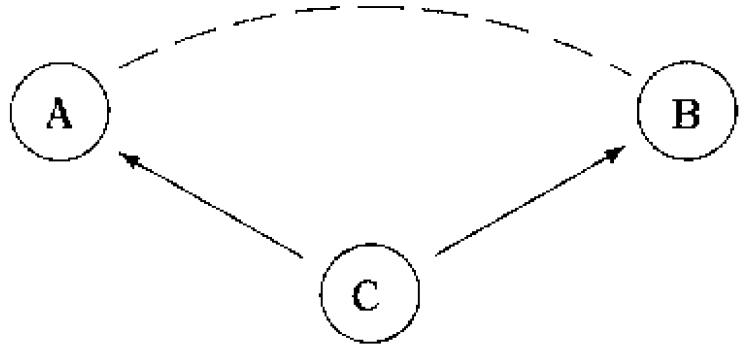
Abramson et al. Ann Emerg Med 1992

Kyrgidis et al. J Clin Oncol 2010

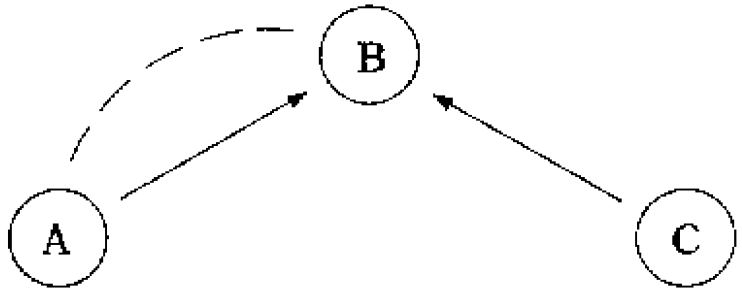
Kyrgidis et al. Lancet Oncol 2013



CAUSATION—Changes in A cause changes in B.



COMMON RESPONSE—Changes in both A and B are caused by changes in a third variable, C.



CONFOUNDING—Changes in B are caused both by changes in A and by changes in third variable C.

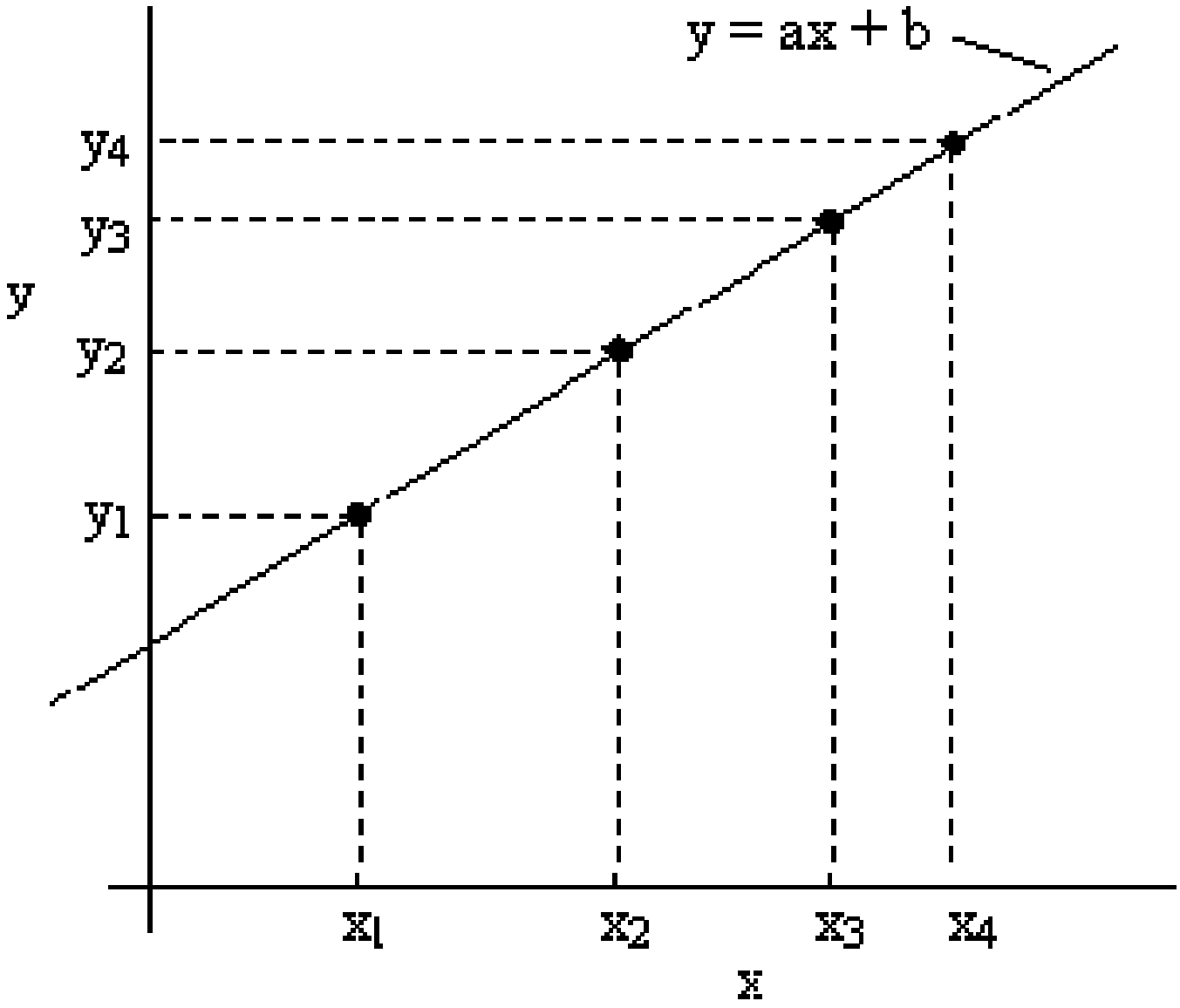
- Υπάρχει Στατιστική Σχέση για πρόβλεψη;
- Πόσο ισχυρή είναι αυτή η σχέση;
- Ποιο είναι το καταλληλότερο μοντέλο;
- Πώς εκφράζεται η καλύτερη προσαρμογή του μοντέλου;

ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ

Αναφέρεται στη γραμμική σχέση μεταξύ των δυο μεταβλητών

ΠΑΛΙΝΔΡΟΜΗΣΗ

Παραδεχόμαστε ή υποθέτουμε μια συναρτησιακή σχέση ανάμεσα στις μεταβλητές


y = ax + b
y4
y3
y
y2
y1
x1
x2
x3
x4
x

Selecting a Multivariate Technique

- Dependency
  - dependent (criterion) variables and independent (predictor) variables are present
- Interdependency
  - variables are interrelated without designating some dependent and others independent

Dependency Techniques

- Multiple regression
- Discriminant analysis
- Multivariate analysis of variance(MANOVA)
- Linear structural relationships (LISREL)
- Conjoint analysis

Uses for Multiple Regression

- Control for confounding variables to better evaluate the contribution of other variables

VOLUME 27 · NUMBER 32 · NOVEMBER 10 2009

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Longitudinal Cohort Study of Risk Factors in Cancer Patients of Bisphosphonate-Related Osteonecrosis of the Jaw

Konstantinos Vahtsevanos, Athanassios Kyrgidis, Evgenia Verrou, Eirini Katodritou, Stefanos Triaridis, Charalampos G. Andreadis, Ioannis Boukovinas, Georgios E. Koloutsos, Zisis Teleioudis, Kyriaki Kitikidou, Panagiotis Paraskevopoulos, Konstantinos Zervas, and Konstantinos Antoniades

Table 3. Adjusted Predictors for ONJ Development in Patients With Multiple Myeloma, Breast Cancer, and Prostate Cancer Treated With Zoledronate, Pamidronate, and Ibandronate (N = 1,621)

Adjusted Predictors for ONJ Development	Adjusted OR*	95% CI*	P†
Sex	0.72	0.38 to 1.37	.321
Age	0.99	0.97 to 1.03	.897
Smoking status	0.62	0.34 to 1.12	.115
Use of dentures	2.02	1.03 to 3.96	**.042**
Disease			
Breast cancer	1.341	0.63 to 2.85	.445
Prostate cancer	0.98	0.10 to 9.90	.986
History of dental work			
Extraction	32.97	18.02 to 60.31	**< .001**
Root canal treatment	0.76	0.39 to 1.47	.408
Ever received BPs			
Zoledronate	28.09	5.74 to 137.43	**< .001**
Pamidronate	12.32	2.04 to 74.47	**.006**
Ibandronate	0.26	0.07 to 0.95	**.042**
Each dose of BP administered			
All BPs	0.52	0.30 to 0.91	**.022**
Zoledronate	2.02	1.15 to 3.56	**.015**
Pamidronate	1.78	1.001 to 3.16	**.049**
Ibandronate	0.00	0.00	.992

Abbreviations: ONJ, osteonecrosis of the jaw; OR, odds ratio; BPs, bisphosphonates.

*ORs, 95% CIs, and *P* values were calculated with conditional multivariate logistic regression backward elimination model according to likelihood-ratio criteria. ORs were mutually adjusted for variables in the model. Logit for outcome variable "ONJ" is estimated for ONJ = no. Logit for independent nominal variable "disease" is estimated for disease = multiple myeloma. Logit for independent ordinal variables "Each dose of BP administered" is estimated for each dose of BP administered = 1. All other dichotomous variables = no.

†α level set to $P < .05$. Cutoff value set to 0.10. Statistically significant *P* values appear in **bold**.

Uses for Multiple Regression

- Control for confounding variables to better evaluate the contribution of other variables
- Predict values for a criterion variable by developing a self-weighting estimating equation

ORIGINAL ARTICLE

The International Criteria for Behçet's Disease (ICBD): a collaborative study of 27 countries on the sensitivity and specificity of the new criteria

International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD)
F. Davatchi,[1,*,†] S. Assaad-Khalil,[2,†] K.T. Calamia,[3,†] J.E. Crook,[4,†] B. Sadeghi-Abdollahi,[1,†] M. Schirmer,[5,†] T. Tzellos,[6,†] C.C. Zouboulis,[6,7,8,*,†] M. Akhlagi,[1] A. Al-Dalaan,[9] Z.S. Alekberova,[10] A.A. Ali,[11] A. Altenburg,[7,8] E. Avromdee,[12] M. Baltaci,[13] M. Bastos,[14] S. Benamour,[15] I. Ben Ghorbel,[16] A. Boyvat,[17] L. Carvalho,[18] W. Chen,[19] E. Ben-Chetrit,[20] C. Chams-Davatchi,[1] J. A. Correia,[18] J. Crespo,[21] C. Dias,[22] Y. Dong,[23] F. Paixão-Duarte,[24] K. Elmuntaser,[25] A.V. Elonakov,[10] J. Graña Gil,[26] A.-A. Haghdoost,[1] R.M. Hayani,[27] H. Houman,[16] A.R. Isayeva,[28] A.R. Jamshidi,[1] P. Kaklamanis,[29] A. Kumar,[30] A. Kyrgidis,[6] W. Madanat,[31] A. Nadji,[1] K. Namba,[32] S. Ohno,[32] I. Olivieri,[33] J. Vaz Patto,[34] N. Pipitone,[35] M.V. de Queiroz,[36] F. Ramos,[35] C. Resende,[37] C.M. Rosa,[36] C. Salvarani,[35] M.J. Serra,[38] F. Shahram,[1] H. Shams,[1] K.E. Sharquie,[27] M. Sliti-Khanfir,[16] T. Tribolet de Abreu,[39] C. Vasconcelos,[18] J. Vedes,[40] B. Wechsler,[41] Y.K. Cheng,[42] Z. Zhang,[43] N. Ziaei[1]

Table 5 International Criteria for Behçet's Disease – point score system: scoring ≥ 4 indicates Behçet's diagnosis

Sign/symptom	Points
Ocular lesions	2
Genital aphthosis	2
Oral aphthosis	2
Skin lesions	1
Neurological manifestations	1
Vascular manifestations	1
Positive pathergy test*	1*

*Pathergy test is optional and the primary scoring system does not include pathergy testing. However, where pathergy testing is conducted one extra point may be assigned for a positive result.

Uses for Multiple Regression

- Control for confounding variables to better evaluate the contribution of other variables
- Predict values for a criterion variable by developing a self-weighting estimating equation
- Test and explain causal theories
  - Path analysis

Accuracy of dermoscopic criteria for discriminating superficial from other subtypes of basal cell carcinoma

Aimilios Lallas, MD,[a] Thrassivoulos Tzellos, MD,[c] Athanasios Kyrgidis, MD,[d] Zoe Apalla, MD,[c]
Iris Zalaudek, MD,[a,e] Athanasios Karatolias, MD,[f] Gerardo Ferrara, MD,[g] Simonetta Piana, MD,[b]
Caterina Longo, MD,[a] Elvira Moscarella, MD,[a] Alexander Stratigos, MD,[h] and Giuseppe Argenziano, MD[a]
Reggio Emilia and Benevento, Italy; Thessaloniki, Volos, and Athens, Greece; and Graz, Austria

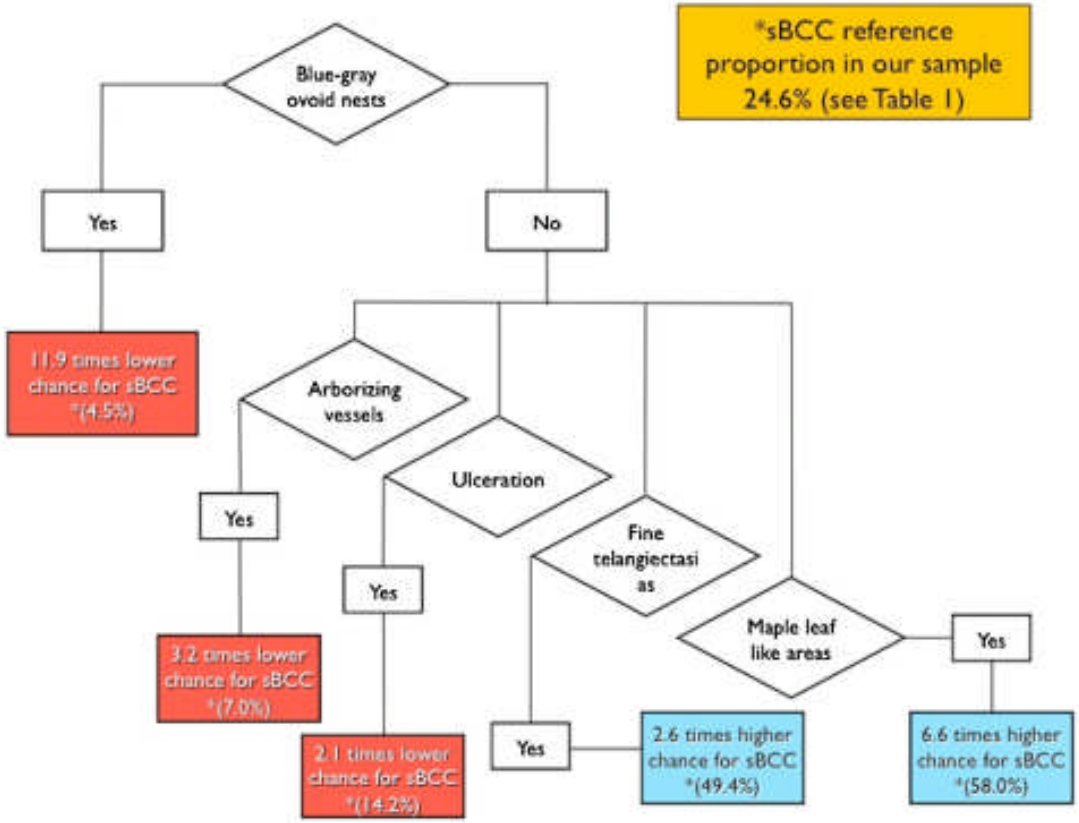


Fig 4. Flow chart for the dermoscopic diagnosis of superficial basal cell carcinoma (*sBCC*). In the parentheses after *asterisk*, we report the absolute proportion of tumors, showing each prognostic dermoscopic criterion, that were sBCC in our sample (reference proportion, 24.6%) (Table I). This information is only indicative and cannot be generalized (as it only applies to our sample), as opposed to reported adjusted odds ratios, which are universally valid and applicable.

Uses for Discriminant Analysis

- Classify persons or objects into various groups
- Analyze known groups to determine the relative influence of specific factors

BJD
British Journal of Dermatology

Accuracy of dermoscopic criteria for the diagnosis of psoriasis, dermatitis, lichen planus and pityriasis rosea

A. Lallas, A. Kyrgidis,* T.G. Tzellos,¶ Z. Apalla, E. Karakyriou, A. Karatolias, I. Lefaki, E. Sotiriou,† D. Ioannides,† G. Argenziano‡ and I. Zalaudek‡§

State Clinic of Dermatology, Hospital of Skin and Venereal Diseases, Delfon 124, 54643 Thessaloniki, Greece
*Department of Otolaryngology–Head & Neck Surgery and †First Dermatologic Department, Medical School, Aristotle University, Thessaloniki, Greece
‡Dermatology Unit, Medical Department, Arcispedale Santa Maria Nuova, Reggio Emilia, It
§Department of Dermatology, Medical University of Graz, Graz, Austria
¶Division of Evidence Based Dermatology, Departments of Dermatology, Venerology, Allergol

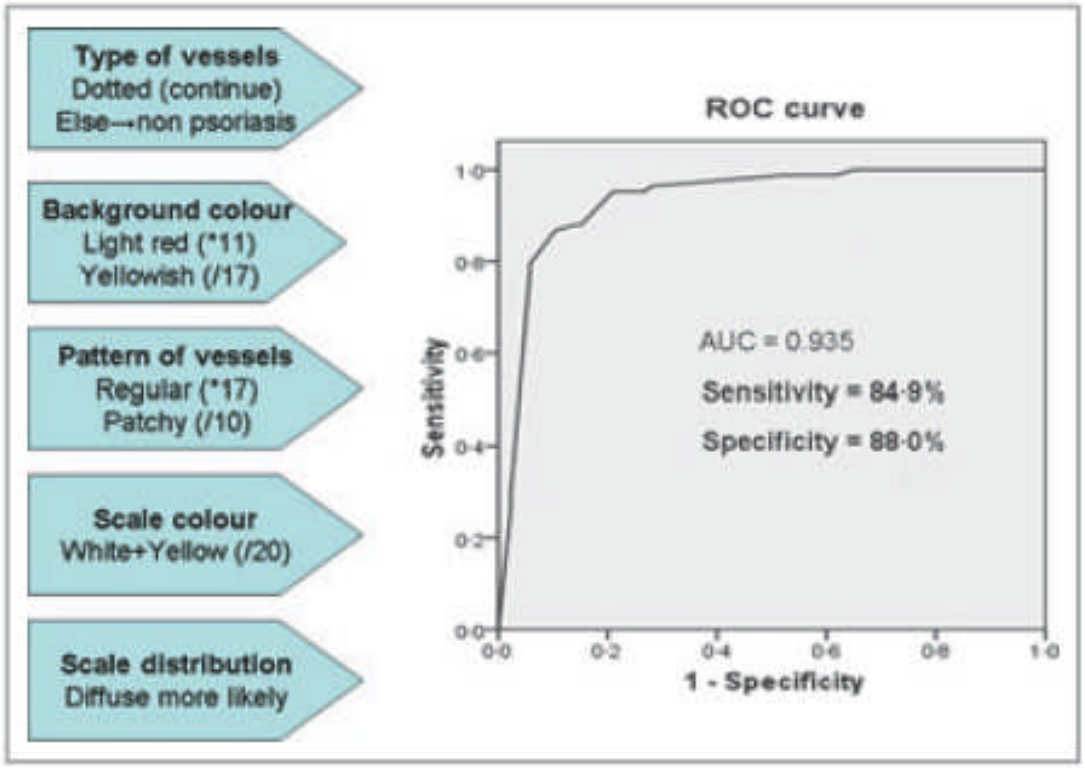


Fig 4. Receiver-operator characteristic (ROC) curve. Backward elimination according to likelihood ratio criteria. Area under curve (AUC) = 0·935, 95% confidence interval 0·896–0·974 ($P < 0·001$).

Use for MANOVA

- Assess relationship between two or more dependent variables and classificatory variables or factors samples
- E.g. . . . measure differences between
  - employees
  - customers
  - manufactured items
  - production parts

Uses of LISREL

- Explains causality among constructs not directly measured
- Two parts
  - Measurement model
  - Structural Equation model

Use for Conjoint Analysis

- Market research
- Product development

Binary Logistic Regression

- Anxious?

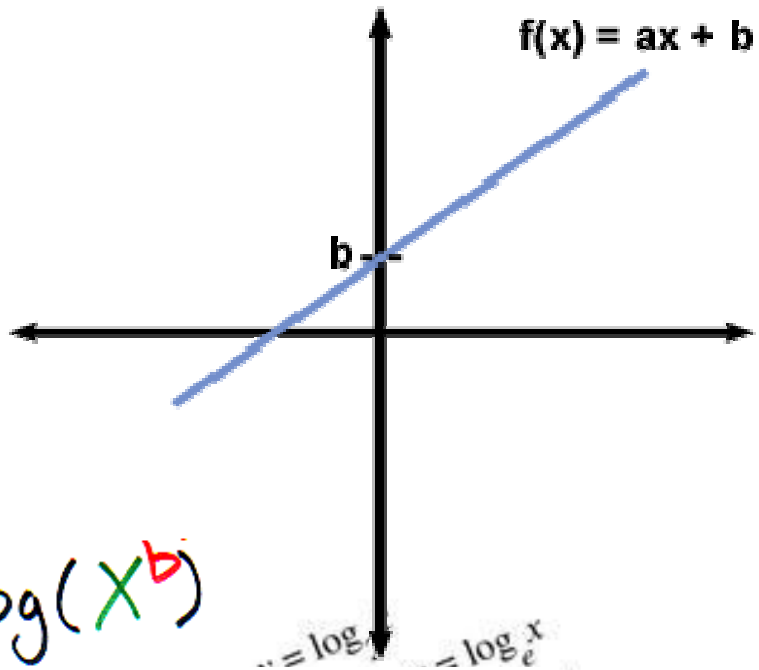
- Similarity to linear regression:
  1. Model fit
  2. Interpreting coefficients
  3. Inferential statistics
  4. Predicting Y for values of the independent variables

$$Y = aX^b$$

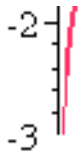
$$\log(Y) = \underline{\log(aX^b)}$$

what does this mean?

we have to remember our rules of logs!



$$\log(Y) = \log(aX^b) = \log(a) + \log(X^b)$$



$$\log(Y) = \log(a) + \log(X^b)$$

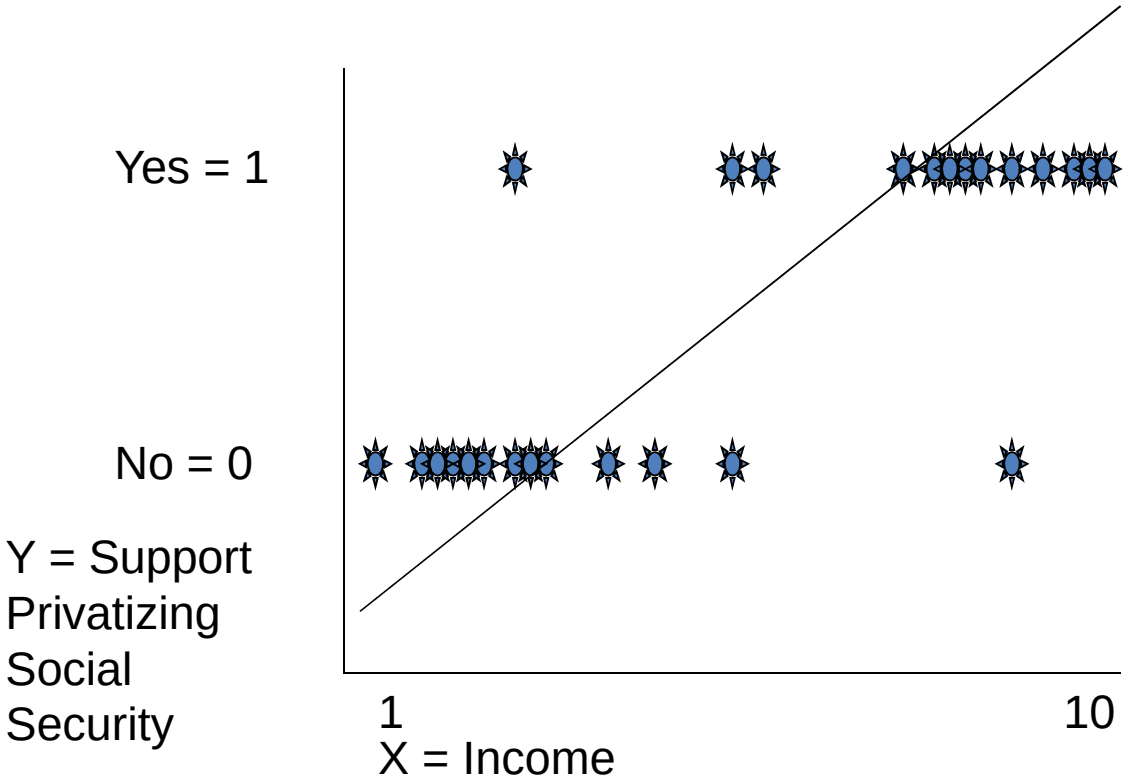
$$= \log(a) + b\log(X)$$

$$\log(Y) = b * \log(x) + \log(a)$$

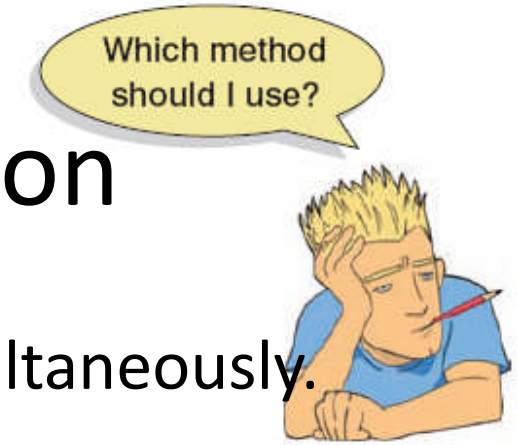
$$Y = b\,X + a$$

Binary Logistic Regression

Linear regression on a dichotomous dependent variable:



Methods of Regression



- Forced Entry: All variables entered simultaneously.

- Hierarchical: Variables entered in blocks.
  - Blocks should be based on past research, or theory being tested. Good Method.

- Stepwise: Variables entered on the basis of statistical criteria (i.e. relative contribution to predicting outcome).
  - Should be used only for exploratory analysis.

Equation for Step 1

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Intervention(1)	1.229	.400	9.447	1	.002	3.417	1.561	7.480
	Constant	-.288	.270	1.135	1	.287	.750		

a. Variable(s) entered on step 1: Intervention.

The number e is a famous irrational number, and is one of the most important numbers in mathematics. The first few digits are:**2.71828182845904523536028747135 27** (and more ...) e is the base of the Natural Logarithms (invented by John Napier).

times higher than those of a patient who is not treated, with a 95% CI of 1.561 to 7.480.

The impo
(both valu
than 1 me
this case)
predictor i

$$OR = \frac{e^{\alpha+\beta}}{e^{\alpha}} = e^{\beta} = e^{1,229} = 3,417$$

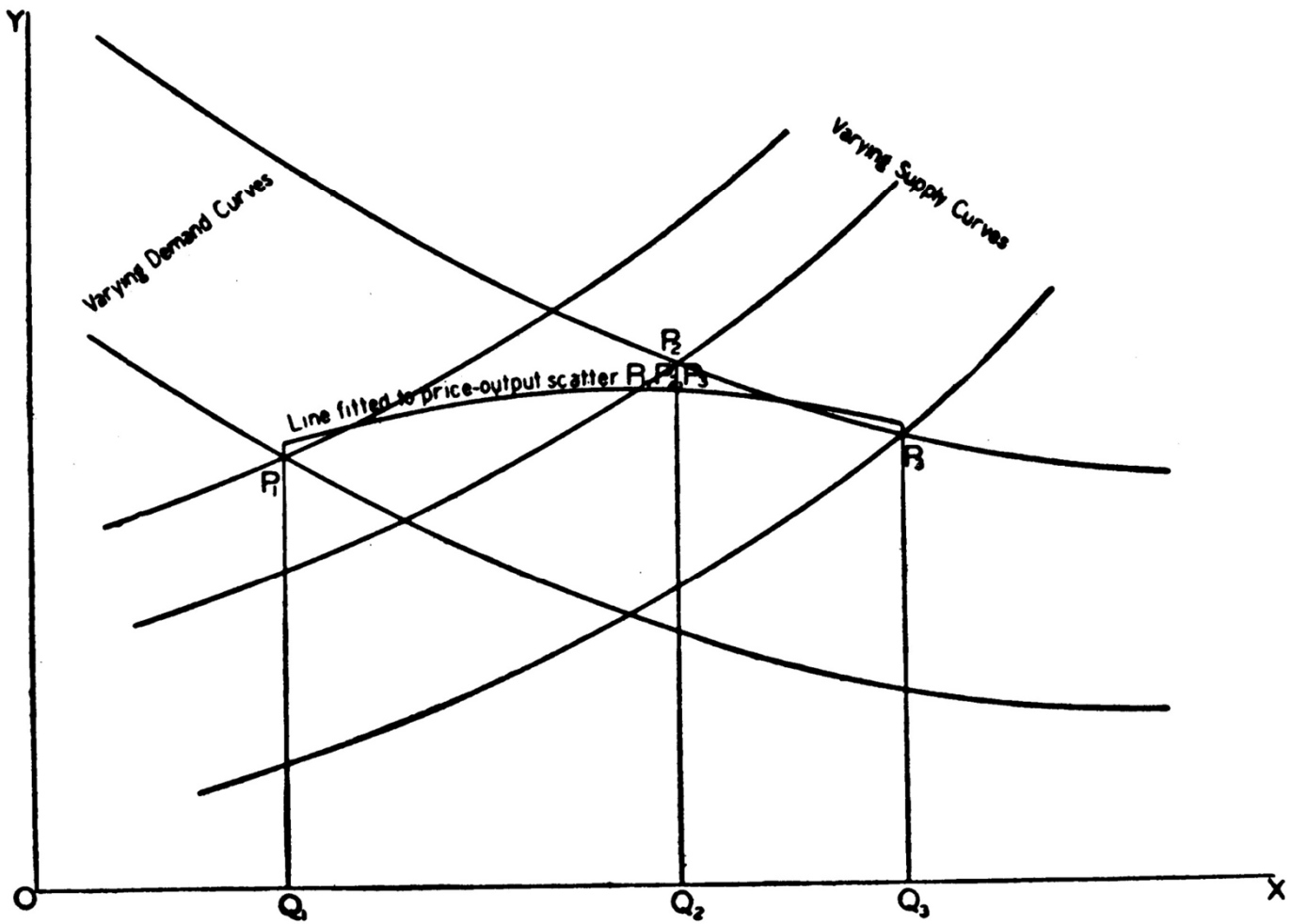
$$\ln(OR) = \beta = \ln(3,417) = 1,229$$

Μια πλήρης παρέκκλιση από την βασική μας ανάλυση: Η ιστορία της πολλαπλής παλινδρόμησης

- Πόσα χρήματα θα συγκεντρώνονταν από την επιβολή ενός εισαγωγικού δασμού στα ζωικά και φυτικά έλαια (βούτυρο, έλαιο σόγιας, κ.λ.π.);
- Ο υπολογισμός αυτός απαιτεί να γνωρίζουμε τις ελαστικότητες προσφοράς και ζήτησης, τόσο τις εγχώριες, όσο και των κρατών απ' όπου εισάγονται τα τα έλαια.
- Το πρόβλημα αυτό έλυσε πρώτος ο Wright το 1928 στο Παράρτημα Β του έργου του "**The Tariff on Animal and Vegetable Oils**".

Διάγραμμα 4, σελ. 296, Παράρτημα Β (1928):

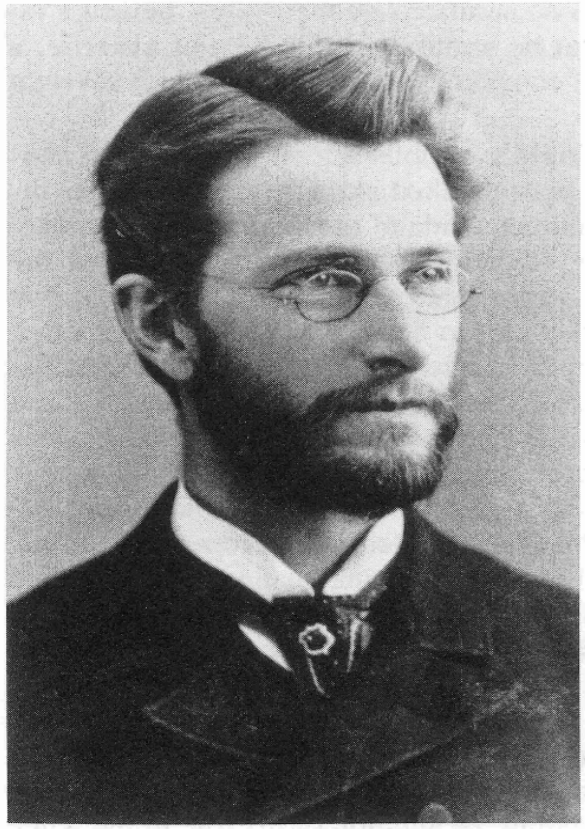
FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.



Ποιος, όμως, έγραψε το Παράρτημα Β…;

…το παράρτημα αυτό πιστεύεται ότι το έγραψε είτε ο ίδιος ο Philip Wright σε συνεργασία με το γιό του, Sewall Wright, που ήταν εξαίρετος στατιστικολόγος ή ο γιος του μόνος του.

Ποιοι ήταν οι δύο αυτοί άντρες και ποια η ιστορία τους;

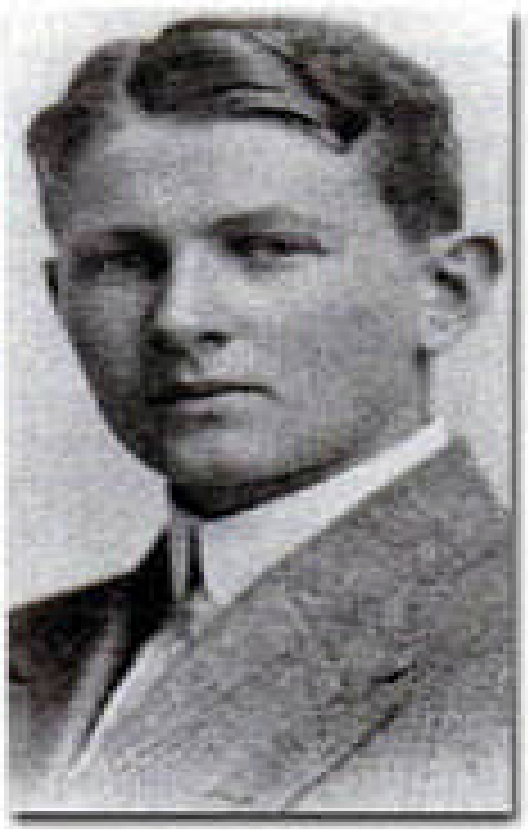


Philip Wright (1861-1934)

άσημος οικονομολόγος και ποιητής

MA Harvard, Econ, 1887

Lecturer,Harvard,1913-1917



Sewall Wright (1889-1988)

διάσημος γενετικός στατιστικολόγος

ScD Harvard, Biology, 1915

Prof., U. Chicago, 1930-1954

Παράδειγμα: Ζήτηση τσιγάρων

- Πόσο θα μειωνόταν η κατανάλωση τσιγάρων από την επιβολή ενός (υποθετικού) φόρου;
- Για να απαντήσουμε στο ερώτημα αυτό, χρειάζεται να γνωρίζουμε την ελαστικότητα της ζήτησης τσιγάρων, δηλαδή, το συντελεστή παλινδρόμησης β_1.

$$\ln(Q_i^{cigarettes}) = \beta_0 + \beta_1 \ln(P_i^{cigarettes}) + u_i$$

Logistic regression (1)

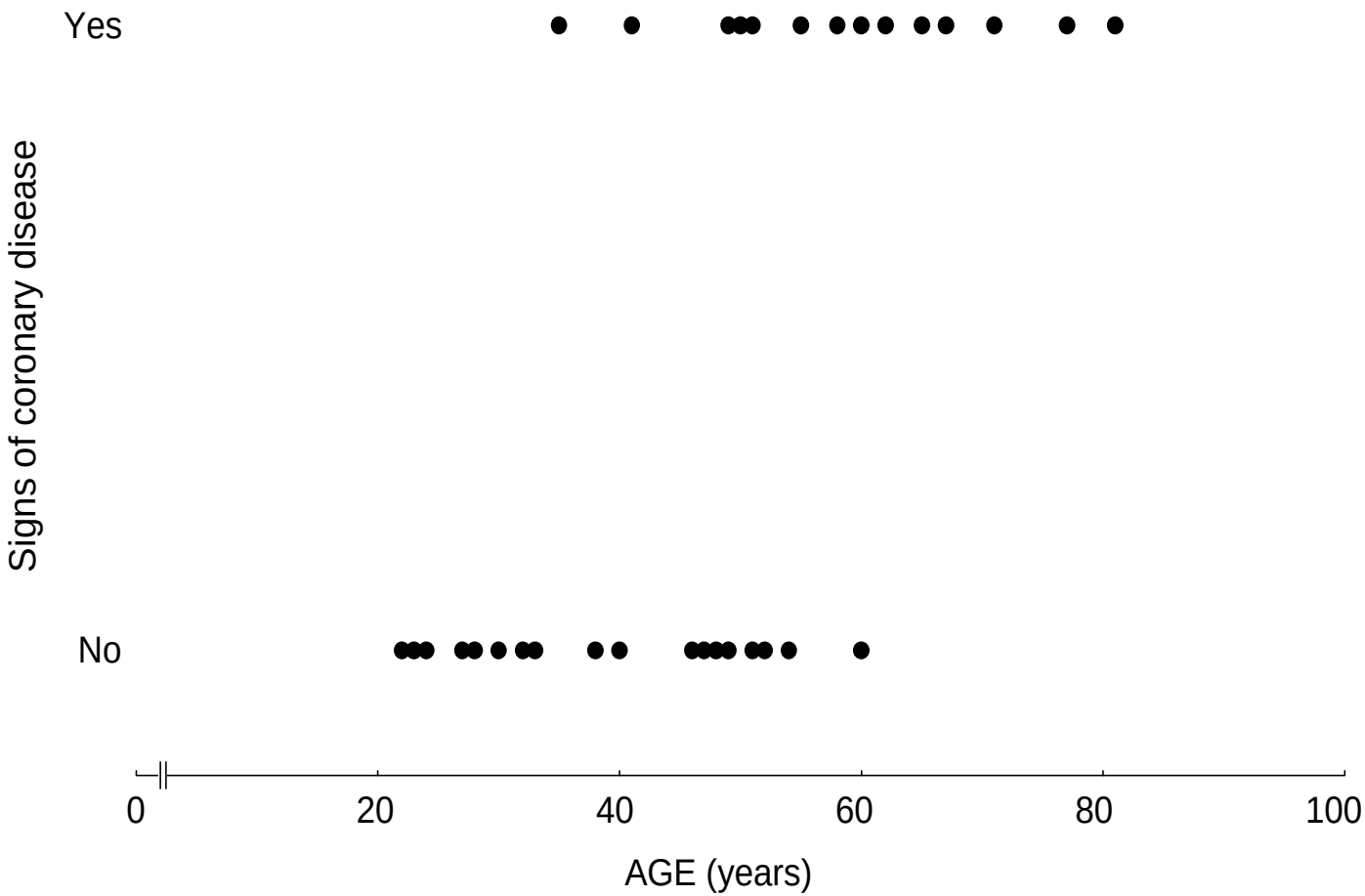
Table 2 Age and signs of coronary heart disease (CD)

Age	CD	Age	CD	Age	CD
22	0	40	0	54	0
23	0	41	1	55	1
24	0	46	0	58	1
27	0	47	0	60	1
28	0	48	0	60	0
30	0	49	1	62	1
30	0	49	0	65	1
32	0	50	1	67	1
33	0	51	0	71	1
35	1	51	1	77	1
38	0	52	0	81	1

How can we analyse these data?

- Comparison of the mean age of diseased and non-diseased women
  - Non-diseased: 38.6 years
  - Diseased: 58.7 years ($p<0.0001$)
- Linear regression?

Dot-plot: Data from Table 2

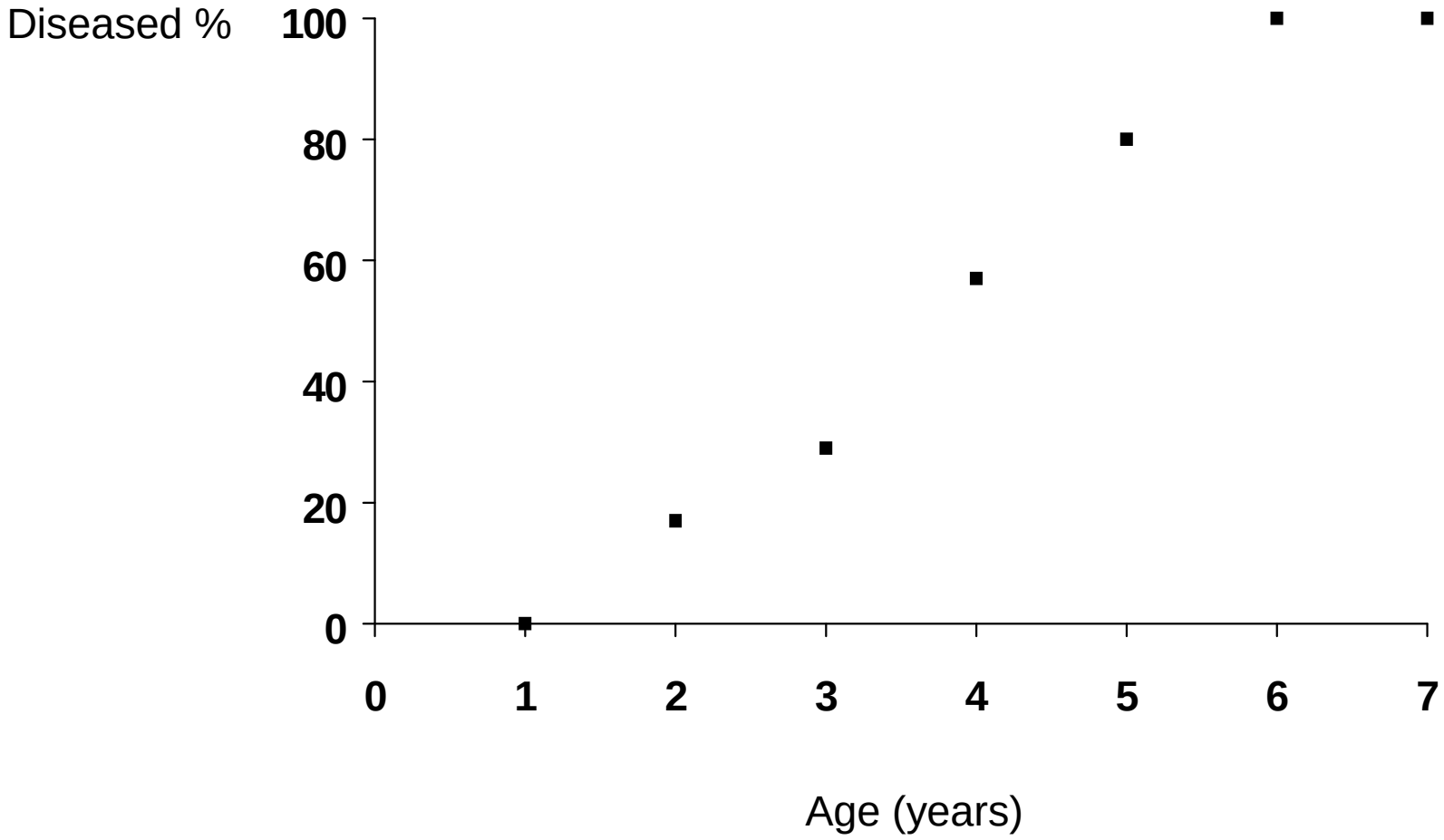


Logistic regression (2)

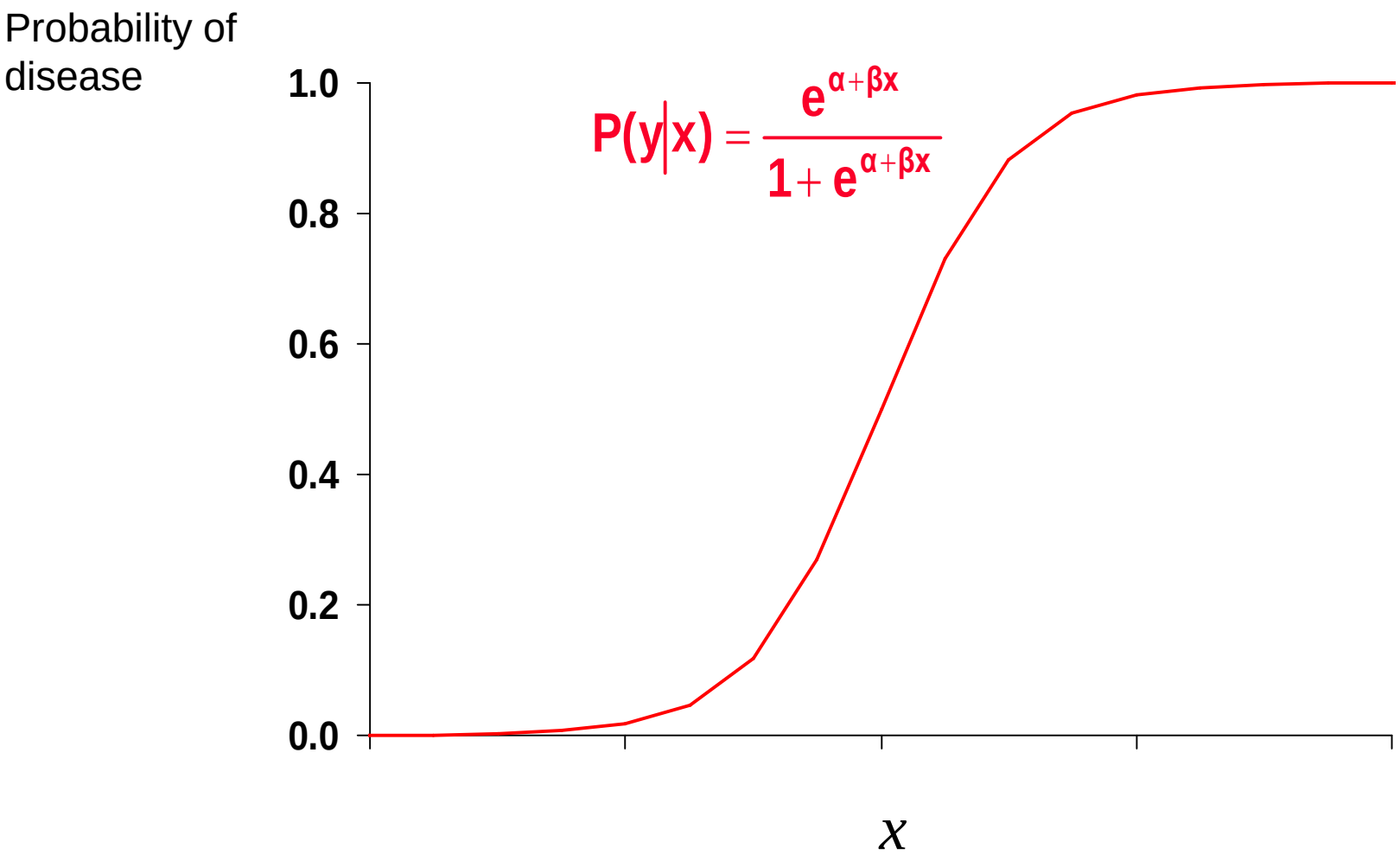
Table 3 Prevalence (%) of signs of CD according to age group

		Diseased	
Age group	# in group	#	%
20 -29	5	0	0
30 - 39	6	1	17
40 - 49	7	2	29
50 - 59	7	4	57
60 - 69	5	4	80
70 - 79	2	2	100
80 - 89	1	1	100

Dot-plot: Data from Table 3



The logistic function



multinomial logistic regression…

Case Processing Summary

	N
Histological_diag	
Superficial	
Nodular	
Micronodular	
Morpheaform	
Mixed	
Pinkus	
Infiltrating	
Flat	No
	Yes
Elevated	No
	Yes
Nodular	No
	Yes
Arborizingtelang	
	45,4%
	Yes

Histological_diagnosis[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
Nodular	Intercept	-5,953	322,441	,000	1	,985			
	[Flat=0]	9,147	160,712	,003	1	,955	9389,151	1,495E-133	5,896E140
	[Flat=1]	0[b]	.	.	0	.	.	.	.
	[Elevated=0]	8,310	160,712	,003	1	,959	4063,920	6,470E-134	2,553E140
	[Elevated=1]	0[b]	.	.	0	.	.	.	.
	[Nodular=0]	-,441	162,740	,000	1	,998	,643	1,921E-139	2,153E138
	[Nodular=1]	0[b]	.	.	0	.	.	.	.
	[Arborizingtelangiectasia=0]	-1,477	,481	9,421	1	,002	,228	,089	,586
	[Arborizingtelangiectasia=1]	0[b]	.	.	0	.	.	.	.
	[Largeblue_grayovoidnests=0]	-1,988	,608	10,682	1	,001	,137	,042	,451
	[Largeblue_grayovoidnests=1]	0[b]	.	.	0	.	.	.	.
	[Ulceration=0]	-,610	,439	1,932	1	,165	,544	,230	1,284
	[Ulceration=1]	0[b]	.	.	0	.	.	.	.
	[PseudoPigmentation=0]	,580	,504	1,325	1	,250	1,787	,665	4,801
	[PseudoPigmentation=1]	0[b]	.	.	0	.	.	.	.
	[Shortfinesuperficialtelangiectasia=0]	,788	,462	2,905	1	,088	2,199	,889	5,445
	[Shortfinesuperficialtelangiectasia=1]	0[b]	.	.	0	.	.	.	.
	[Multiplesmallerosions=0]	,164	,508	,104	1	,748	1,178	,435	3,190
	[Multiplesmallerosions=1]	0[b]	.	.	0	.	.	.	.
	[Shinywhite_redstructurelessareas=	-,296	,529	,313	1	,576	,744	,264	2,097

Journal of the American Academy of Dermatology

Volume 70, Issue 2, February 2014, Pages 303–311

Original article

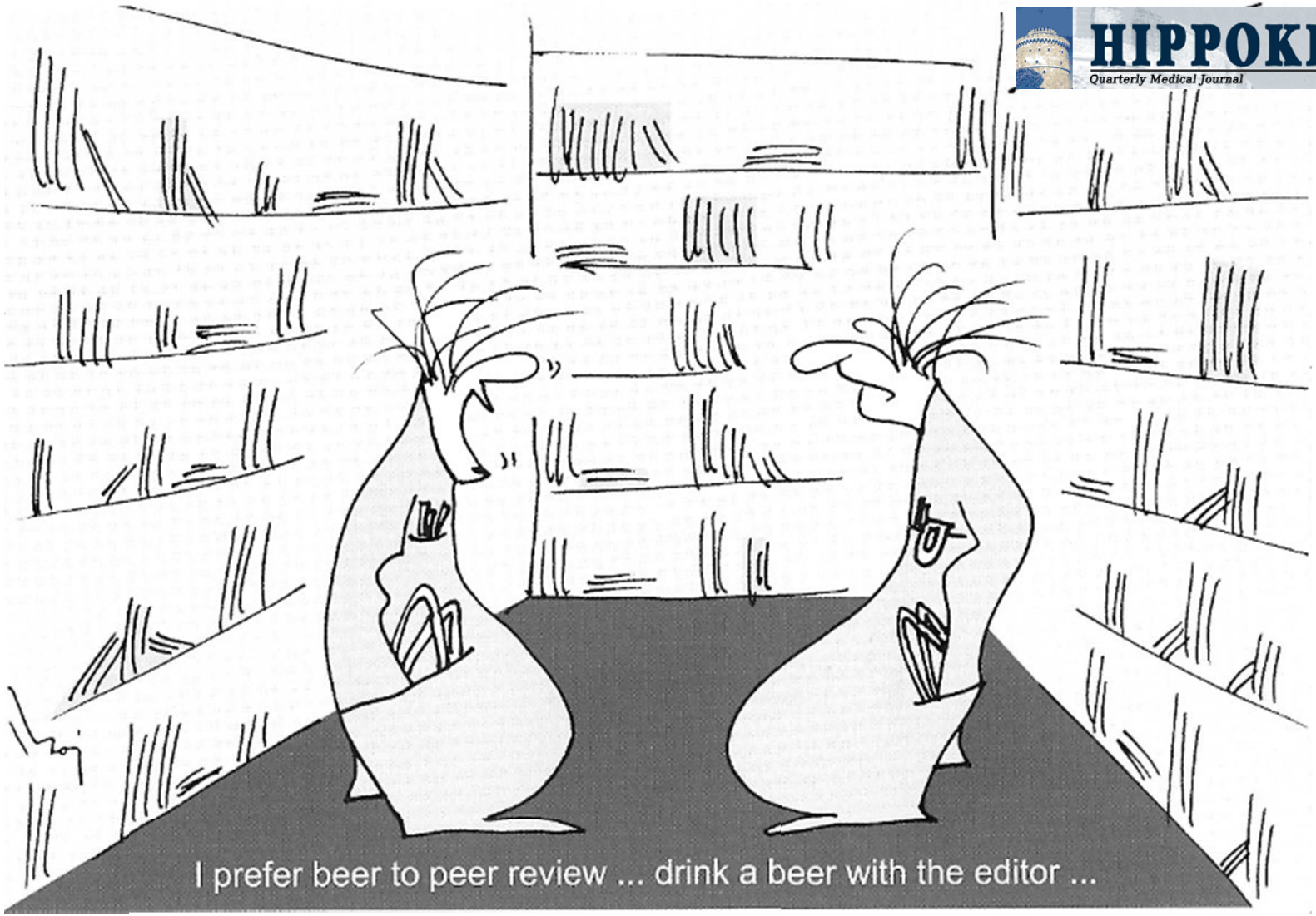
Accuracy of dermoscopic criteria for discriminating superficial from other subtypes of basal cell carcinoma

Aimilios Lallas, MD[a], Thrassivoulos Tzellos, MD[c], Athanasios Kyrgidis, MD[d], Zoe Apalla, MD[c], Iris Zalaudek, MD[a, e], Athanasios Karatolias, MD[f], Gerardo Ferrara, MD[g], Simonetta Piana, MD[b], Caterina Longo, MD[a], Elvira Moscarella, MD[a], Alexander Stratigos, MD[h], Giuseppe Argenziano, MD[a]

Multinomial (polytomous) regression

Using multinomial logistic regression, we created a model that classifies patients into 1 of 7 predefined BCC types, based on dermoscopic information from each patient. The results show that the model allows proper classification of sBCC and nodular BCC but not morpheaform, mixed, infiltrating, and fibroepithelioma types.

HIPPOKRATIA
Quarterly Medical Journal
ISSN 1108-4189



Ευχαριστώ για το χρόνο σας!

Kyrgidis Athanassios, MD, MSc, PhD
3 Papazoli St,
Thessaloniki, 546 30, Greece
Tel. +30-6947-566727
Fax. +30-2310-546701
E-mail: akyrgidi@gmail.com