Modern Statistical Tendencies used in the Investigation in Medical Sciences

Scientific Session
November 29, 2018
CSMS- Jamaica

Professor Fidel Cathcart Roca
Dr./M.Sc. Yeiny Terry Pena

Objectives

- General:
  - To present the most relevant statistical techniques in the medical field.
- Specific:
  - To provide examples of the use of such techniques

Introduction
History of Statistics

- The term statistics was itself coined in 1749 in German.
- Connected with the development of Sovereign States.
- Designated the systematic collection of demographic and economic data by states

Introduction
History of Statistics

- In the early 19th century,
  - "statistics" : Discipline concerned with the collection, summary, and analysis of data.

- Today,
  - data is collected and statistics are computed
  - widely distributed in government, business, sports and sciences.

Introduction
Statistics in Sciences

- In the incept
  - Descriptive Statistics
    - Measures of frequency
    - Measures of Central Tendency
    - Measures of Dispersion and Variation
    - Measures of Position
- To characterize your data based on its properties

Introduction
Statistics in Sciences

- Modern era
  - Inferential Statistics
    - t-tests
    - ANOVA
    - Regression

- To predict outcome from known predictor variable(s)

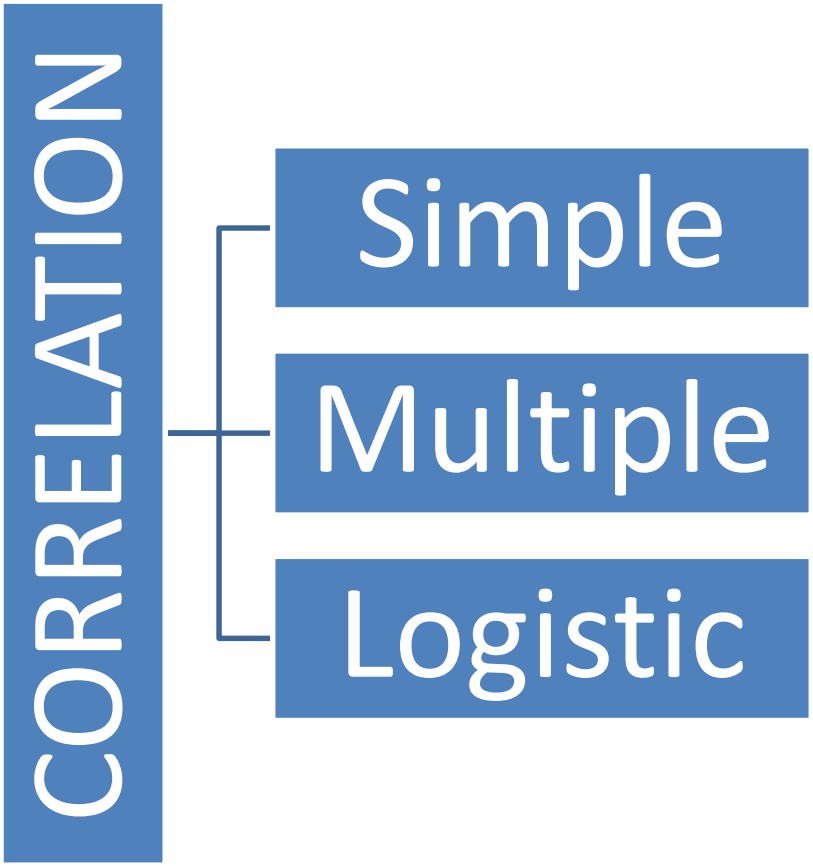
Statistics in Sciences

Professor Fidel Cathcart's words

- The fundament of Inferential Statistics relies on the Theory of Probability.

- Inferential Statistics allows you to infer or generalize the result of your data taking froma sample of a "probabilistic type"

STATISTICS IN SCIENCES



Statistical technique that can show whether and how strongly pairs of variables are related.

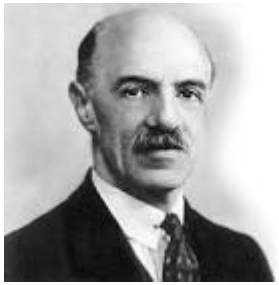
Statistics in Sciences

- **Simple** : Uses an independent variable.
- **Multiple**: Uses more than two independent variables.

- Quantitavie Variables:
  - Pearson
  - Fisher
  - Kendell
- Qualitative Variables
  - Spearman

Important contributors to and Founders of statistics

Sir Ronald Aylmer Fisher (17 February 1890 – 29 July 1962) was a British statistician and geneticist. For his work in statistics, he has been described as "a genius who almost single-handedly created the foundations for modern statistical science.



Charles Edward Spearman, (10 September 1863 – 17 September 1945) was an English psychologist known for work in statistics, as a pioneer of factor analysis, and for Spearman's rank correlation coefficient.

Statistics in Sciences

LOGISTIC REGRESSION

- Is a widely used statistical model.
- Dependent Variable: binomial, multinomial.
  - Dicotomous
    - Dummy
      - In Range
- Predictor variables that may be either continuous or categorical.

Pearson

Simple Lineal correlation (correlation)r= 0.791. [-1;+1]

LDL-Cholesterol
6.8
6.5
3.9
5.7
7.3
6.5
6.5
4.7
5.7
7.0

Total Lipids
7.2
7
5
6.5
9.5
7
6.8
5.2
7
9

Statistics in Sciences

Professor Fidel Cathcart's words

- Simple Lineal Correlation
- If r is close to 1, must perform a hypothesis test.
- You must prove, there is correlation.
- Without Correlation, you must not state Regression.

Statistics in Sciences

Professor Fidel Cathcart's work

- Equation of Linear Regression
- $\boldsymbol{Y^* = a + bX}$
- Substituted values
- Y*=-0.2+1.18x
- Particular value=8
- Y*= 9.2

Pearson

Simple Lineal correlation (correlation)r= 0.791. [-1;+1]

LDL-Cholesterol
6.8
6.5
3.9
5.7
7.3
6.5
6.5
4.7
5.7
7.0

Total Lipids
6
7
5
6.5
10
7
6.8
5.2
7
9

Statistics in Sciences

Professor Fidel Cathcart's words

Equation of Linear Regression

- It is used to estimate the values or the presence of the variable in study

- Ya sea en cualquier tipo de correlacion. Lineal multiple o logistica. De ahi sale una equacion

Pearson

Multiple correlation

X1	X2	X3
Memory Level	Age	Cognitive function
38.2	52	91
40	50	86
36	63	87
35.2	61	90
34	57	80
31	70	86.5
38	59	78
33	65	89
42.5	50	88
30	72	85

Pearson

Multiple correlation

- The Multiple Regression Equation is derived from a system of Normal equations.

- $X_1^* = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$
- $X_1^* = 62.113 - 0.464 X_2 + 0.017 X_3$

- $r = 0.918$
- $r^2 = 0.843$
- $F = 18.827$
- $p < 0.01$

Pearson

Multiple correlation

- $X_1^* = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$
- $X_1^* = 62.113 - 0.464\,X_2 + 0.017\,X_3$

- Regression for X_2=74 and X_3 = 80

- $X_1^* = 62,113 - 0.464\,(74) + 0.017\,(80)$
- $X_1^* = 62,113 - 34,336 + 1.36$
- **$X_1^* = 29.137$**

Statistics in Sciences
Logistic Regression

*Libro 2.sav [DataSet4] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

1 :

Visible: 16

	Discharge	Gender	Age	BMI	Hb	WBC	ESR	Glycemia	TGP	TGO	TGOTGP	Creatinine	Cholesterol	CD4	Length	Bactrim	var
1	1	1	42	22.7	11.8	8.4	55	5.0	85	106	1.3	72	2.4	320	8	0	
2	0	1	46	17.8	9.8	5.3	112	4.5	84	112	1.3	97	3.1	89	43	0	
3	1	1	35	23.3	8.9	3.4	6	4.1	23	9	.3	64	3.4	196	7	1	
4	1	1	29	19.8	13.4	9.4	40	3.4	25	45	1.8	73	4.2	106	11	1	
5	1	0	22	25.1	12.5	5.6	63	4.6	85	106	1.3	72	4.2	325	5	0	
6	1	1	29	21.0	15.3	5.7	39	3.8	511	686	1.3	79	6.2	73	31	0	
7	1	1	57	23.6	13.1	7.4	115	3.9	18	10	.5	80	4.4	27	36	0	
8	1	0	26	16.7	7.9	5.1	128	3.3	5	10	2.0	60	3.2	35	11	1	
9	1	1	30	21.6	13.8	9.6	3	3.9	139	20	.1	59	4.2	165	23	0	
10	1	1	44	25.3	9.2	6.5	92	4.7	10	23	2.3	52	3.2	54	36	1	
11	1	1	19	18.2	9.5	5.8	140	4.6	26	34	1.3	89	4.2	165	26	1	
12	1	1	44	25.6	9.9	4.2	92	10.8	12	18	1.5	50	4.1	57	36	1	
13	1	0	48	19.1	9.1	7.5	115	2.9	12	25	2.0	57	3.8	129	5	1	
14	1	1	55	21.3	13.5	10.4	25	6.6	19	15	.7	63	6.1	403	8	0	
15	1	0	30	19.3	9.9	6.6	67	4.0	307	399	1.2	72	4.2	81	23	1	
16	1	1	42	19.7	15.1	8.2	66	4.6	85	106	.	72	4.2	58	11	1	
17	1	1	39	22.1	12.2	6.9	13	4.0	116	247	2.1	130	6.5	165	13	0	
18	1	1	48	17.2	9.9	6.6	72	3.1	96	94	.9	87	3.0	158	18	1	
19	1	1	29	22.0	14.5	10.2	34	6.0	33	39	1.1	67	4.4	285	14	1	
20	1	1	38	21.3	9.9	11.2	33	3.7	16	24	1.5	47	5.5	206	29	0	
21	1	1	42	24.0	12.5	13.0	20	4.6	85	106	1.0	72	4.2	375	6	0	
22																	
23																	

Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON

Perfect Association

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.000[a]	.328	1.000

a. Estimation terminated at iteration number 19 because a perfect fit is detected. This solution is not unique.

Statistics in Sciences

Professor Fidel Cathcart's words

- Hay variables q no aportan pero su inclusion depende de la experiencia del investigador

- La inferencia es una teoria para ayudar a una decision y no determina la decision final.

THANK YOU

Any Questions/Suggestions ?