

---

# **DESCRIPTIVE STATISTICS I**

---

**Sorana D. Bolboacă**

# OBJECTIVES

- Representation of qualitative data
- Absolute and relative frequencies
- Frequency tables
- Graphical representations
- Summary

# DESCRIPTIVE VS INFERENCEAL STATISTICS

- Descriptive statistics:
  - Summarize or describe a collection of data
- Inferential statistics:
  - Used to draw inferences about a population from a sample:
    - Estimation: parameter and confidence interval
    - Hypothesis testing (null and alternative hypotheses): determine whether the data are strong enough to reject the null hypothesis

# SUMMARIZING MEDICAL DATA

- Large amounts of medical data are compressed into more easily assimilated summaries
  - Provide the user with a sense of the content
- There a number of ways data can be presented depending by the type of variables:
  - Qualitative variables
  - Quantitative variables

# GOOD TABLES PRACTICES

1. Simple: it is preferred to have 2 or 3 small tables instead of one big table
2. Must be information without reading the associated text:
  - ❑ Abbreviations and symbols must be explained at the bottom of the table
  - ❑ Definitions of rows and columns with units of measurements in headings (if it is applied)
  - ❑ Brief descriptive heading: what? when? where?
  - ❑ Must not duplicate material in the text or in illustration
  - ❑ Synthesis (total) rows and columns
3. If data are taken from another research the source of data must be referred.

# ONE QUALITATIVE VARIABLE

## Raw data

Subject ID	Hypertension class
1	Stage I
2	Normal
3	Prehypertension
4	Stage II
...	...
1000	Stage II

Category	Blood Pressure (mm Hg)
Normal	SBP 90-119 and DBP 60-79
Prehypertension	SBP 120-139 or DBP 80-89
Stage 1 HTN	SBP 140-159 or DBP 90-99
Stage 2 HTN	SBP $\geq 160$ or DBP $\geq 100$
DBP = diastolic blood pressure; SBP = systolic blood pressure	

## FREQUENCY TABLES

What information can we extract from these data?

1. What % of subjects fall into each category
2. How the subjects are divided in the hypertension categories?



Numerical measures ...

# FREQUENCY TABLES

- Let's have the following incubation time expressed in days for a infectious diseases: 5, 6, 7, 7, 8, 8, 5, 7, 8, and 7. Which of the following values correspond to the ascending cumulative relative frequency of 0.7?
  - A. 8
  - B. 6
  - C. 5
  - D. 7
  - E. None

# FREQUENCY TABLES

- Let's have the following incubation time expressed in days for a infectious diseases: 5, 6, 7, 7, 8, 8, 5, 7, 8, and 7. Which of the following values correspond to 20%?
  - A. 8
  - B. 6
  - C. 5
  - D. 7
  - E. None

**Table-2:** Some Nutritional/Physical Activity Behaviours of the Children (N=531) and their Families.

Behaviours	N	%
<b>Frequency of child's consumption of junk food</b>		
Every day	49	9.2
Several times a days	98	18.5
Several times a week	161	30.3
Several times a month	181	34.1
Never	42	7.9
<b>Frequency of visiting fast-food restaurants with child</b>		
Never	45	8.5
Less than once a week	88	16.6
More than once a week	398	75
<b>Junk food consumption by child while watching television</b>		
Yes	121	22.8
No	410	77.2
<b>Breakfast eaten at home</b>		
Yes	307	57.8
No	224	42.2
<b>Meals eaten as a family</b>		
Breakfast	141	26.6
Lunch	57	10.7
Dinner	503	94.7

**Table 1 - Cumulative Radiation Exposure and Time.**

Month	Time (min.)	DDE (mrem)	LDE (mrem)	SDE (mrem)
April	30.49	9	31	30
May	43.09	5	17	16
June	17.04	0	0	0
July	18.35	4	14	13
August	39.01	7	22	22
September	36.39	15	51	51
October	38.53	21	70	66
November	11.16	12	41	38
December	18.38	14	47	46
<b>Sum</b>	<b>252.44</b>	<b>87</b>	<b>293</b>	<b>282</b>

**TABLE 3**  
Interobserver Variability in Lesion Detection

Time per bed position (min)	% Interobserver variability (all 3 readers)
1	36
2	34
3	30
4	22
5	15
6	15
7	15

## Table 1

### Participant demographics (N = 29)

Characteristic	N	Percent
<i>Age (in years, mean, SD)</i>	23	(39.0 ± 11.4; mean age)
<i>Country of origin</i>	26	
Cameroon	10	38.5
Kenya	12	46.1
Nigeria	4	15.4
<i>Years in the US</i>	28	
1 year	8	28.6
2–4 years	5	17.8
5 years	3	10.7
6–10 years	4	14.3
Over 10 years	8	28.6

**TABLE 1**  
Missed Lesions

Time per bed position (min)	Reader no.			Consensus
	1	2	3	
1	17	11	13	14
2	14	10	11	12
3	11	9	6	9
4	2	4	3	2
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0

**TABLE 1. Data from 59 Hypertensive Subjects and 52 Obese Subjects**

Subjects	Sex M/F	Age (yrs) (mean/range)	Arm circumference (cm) (mean/range)	Cuff size (cm)
Hypertensive (n = 59)	37/22	51.0 (24–70)	28.7 (22–34)	12 × 35
Obese (n = 52)	18/34	41.0 (18–73)	39.3 (35–49)	15 × 43

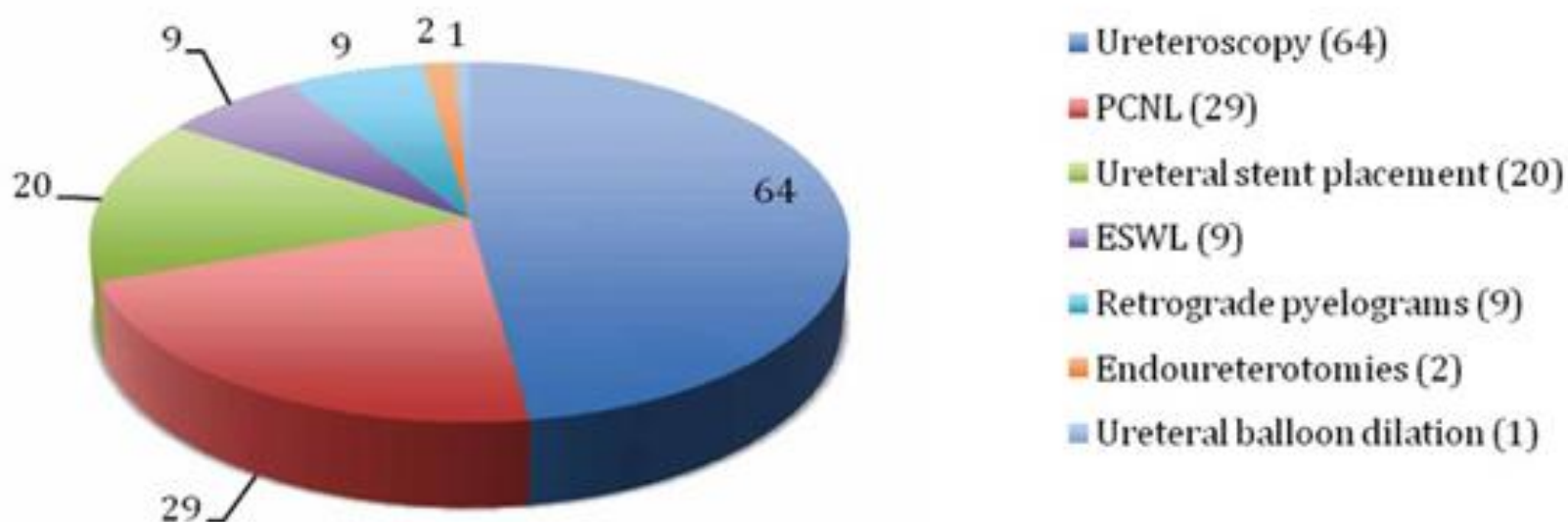
# GOOD GRAPHICAL PRACTICES

- Any graphical representation must to have:
  - Title
  - Definitions of axes
  - Units of measurements for each axe (if it is applied)
  - Legend (if it is applied)
- A good graphical representation must be as self-explanatory as possible!

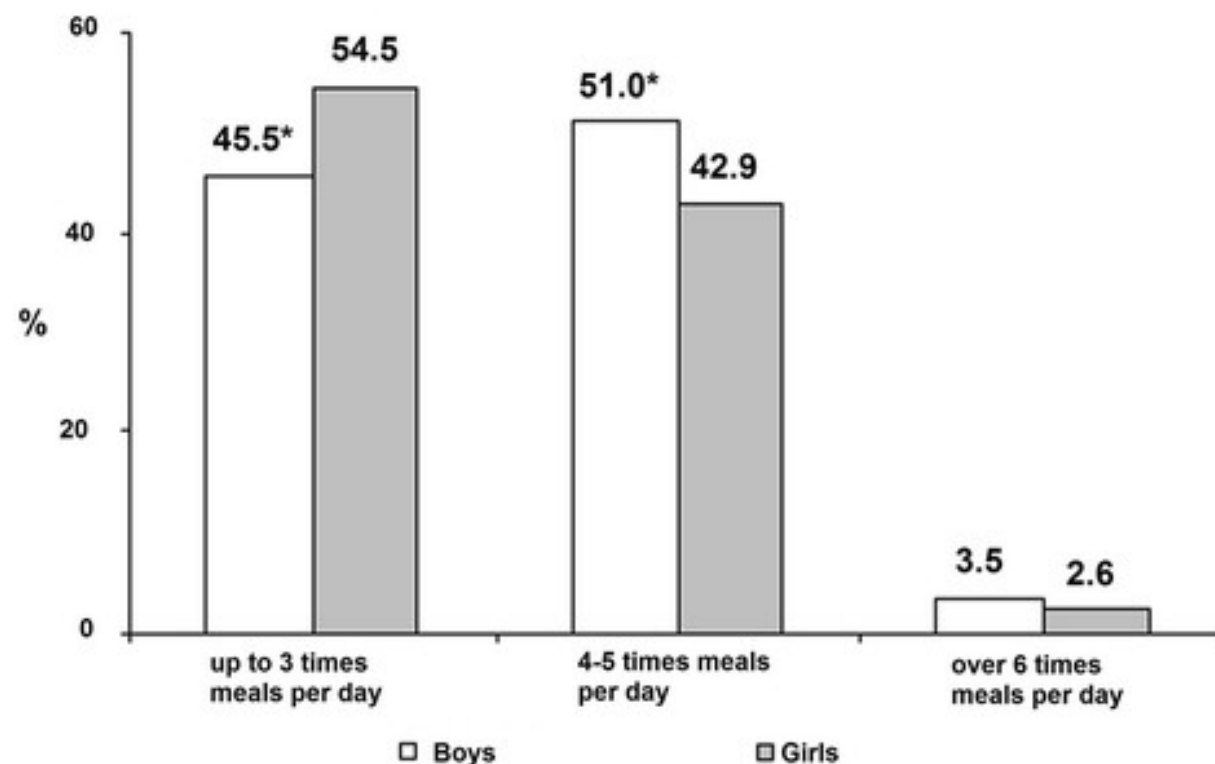
# GOOD GRAPHICAL PRACTICES

- The aim of a graphical representation is to transmit an information
- When drawing a graphical representation try to answer to the following question: Which is the aim of the graphical representation?
- Medical data must be represented graphically in a such a way in which to be useful for understanding the clinical phenomena
- Notice to:
  - The color composition (do not use color background)
  - The font size (it is suppose to be readable)

**Figure 1 - Endourological Surgeries Requiring Fluoroscopy Over 9 Months.**



**Fig. 2**



Distribution of meal frequency per day by gender. \* $P < 0.05$  in comparison with girls

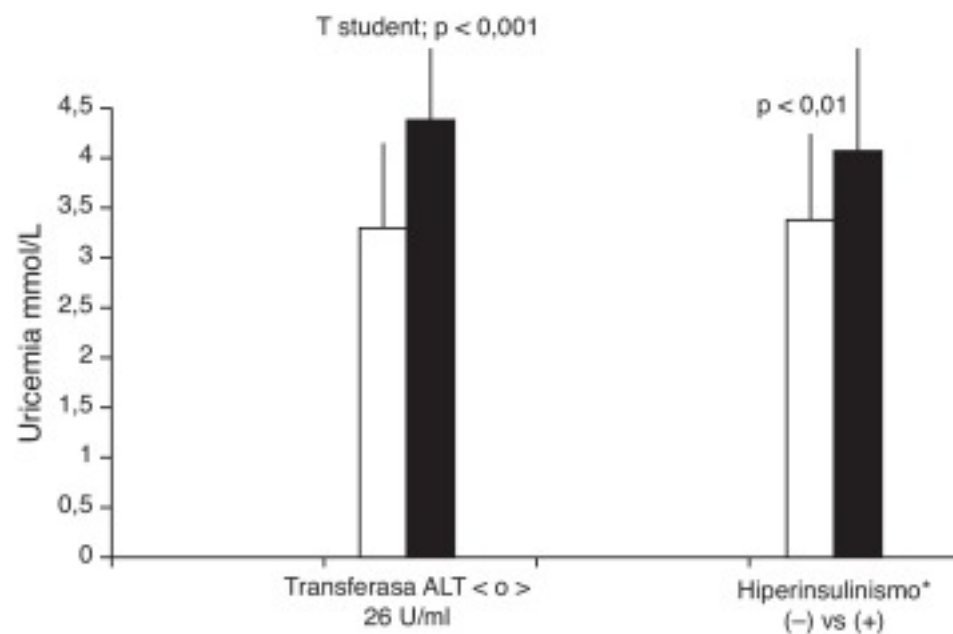
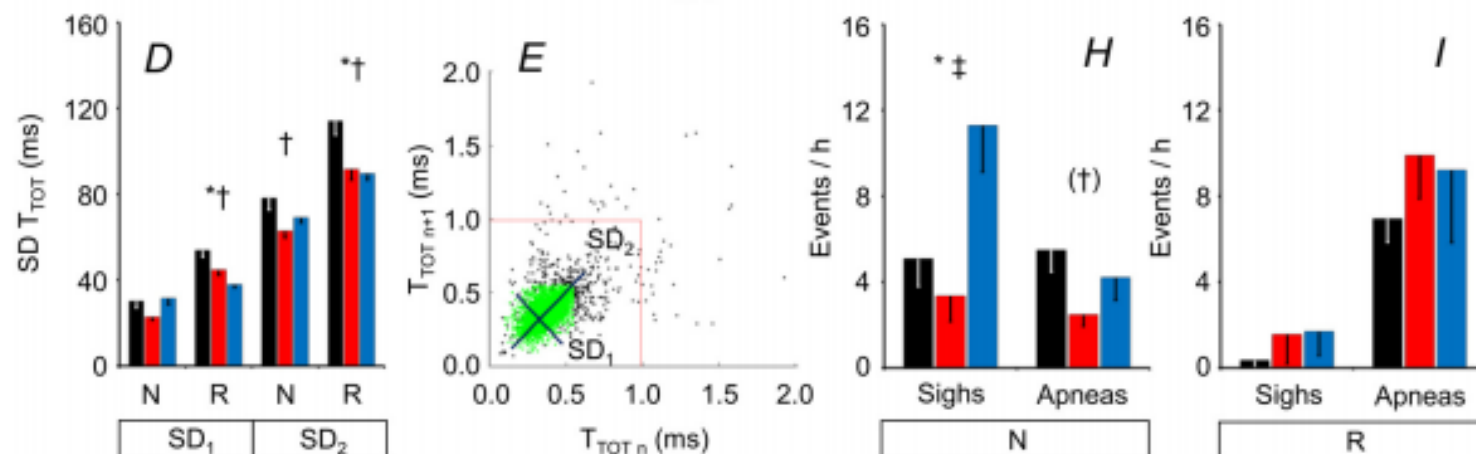


Figura 2.

Concentraciones séricas de ácido úrico (mmol/l) de acuerdo a transferasa ALT > or < 26 u/ml e hiperinsulinismo en niños obesos chilenos (n = 77).

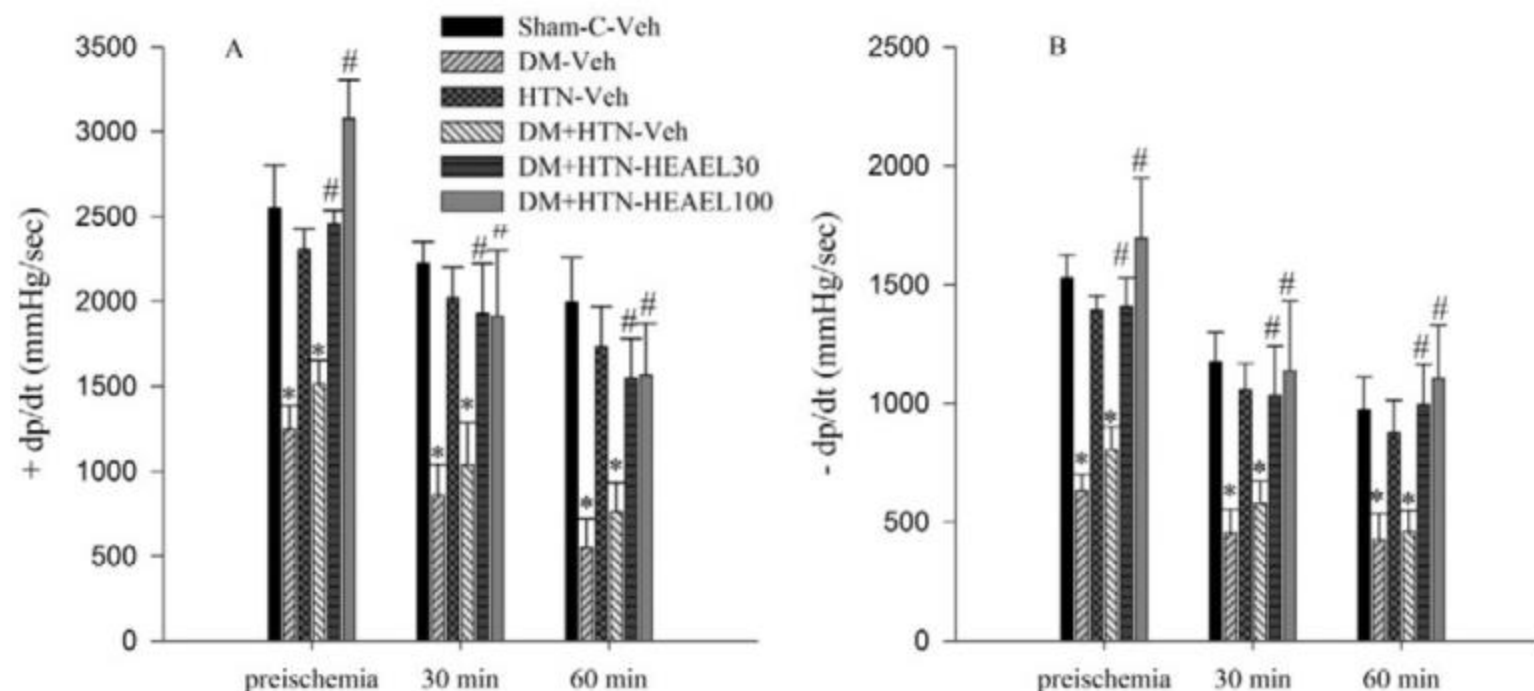
\*Hiperinsulinismo se analizó como concentraciones séricas  $\geq 10$  uU/ml en niños  $\leq 10$  años y  $\geq 15$  uU/ml en niños  $\geq 11$  años.

Figure options ▼



**Fig 5. Sleep-related changes in breathing.** (A) and (B), values of minute volume ( $V_E$ ) and breath duration ( $T_{TOT}$ ), respectively, during non-rapid-eye-movement sleep (N) and rapid-eye-movement sleep (R). In this and the other panels, data are means  $\pm$  SEM in HDC-KO ( $n = 11$  during N,  $n = 10$  during R), DM ( $n = 7$ ), and WT ( $n = 11$ ). (C) Representative tracing (plethysmographic ventilator signal, VENT; electroencephalogram, EEG; electromyogram, EMG) during a transition from N to R in a DM mouse. The decrease in  $T_{TOT}$  during R is evident from the occurrence of more closely spaced deflections of the VENT signal with individual breaths. The grey vertical bar shows the transition point between states. (D) Short-term (SD<sub>1</sub>) and long-term (SD<sub>2</sub>) variability of  $T_{TOT}$ . (E) Representative Poincaré plot during N in a DM mouse, in which abscissa and ordinate of each point indicate  $T_{TOT}$  of successive breaths. SD<sub>1</sub> and SD<sub>2</sub> correspond to the standard deviations around axes (black segments) oriented with or orthogonal to the line of identity of the Poincaré plot, respectively. SD<sub>1</sub> and SD<sub>2</sub> are computed excluding extreme values of  $T_{TOT}$  (black points). The red lines mark the threshold for apnea detection (i.e., three times the average  $T_{TOT}$  value). (F) and (G), representative tracings during N in a DM mouse showing augmented breaths (sighs) either isolated (F) or followed by breathing pauses (apneas; G, arrows). (H) and (I), frequency of occurrence of sighs and apneas during N and R, respectively. \*, †, and ‡,  $P < 0.05$ , WT vs. DM, WT vs. HDC-KO, and HDC-KO vs. DM, respectively (t-tests). In panel H, the symbol (†) indicates a statistical tendency for the difference between HDC-KO and WT ( $P = 0.051$ , ANOVA; HDC-KO vs. WT,  $P = 0.017$ , t-test).

**Fig. 1**



A: Rate of rise (+dp/dt), and B: rate of decrease (-dp/dt) of ventricular pressure of all groups (mean  $\pm$  SEM,  $n=6-8$  in each group) at preischemia (baseline), and after 30 and 60 min of reperfusion. \*: Significant difference ( $P \leq 0.05$ ) from Sham-C-Veh. #: Significant difference ( $P \leq 0.05$ ) from DM+HTN-Veh.

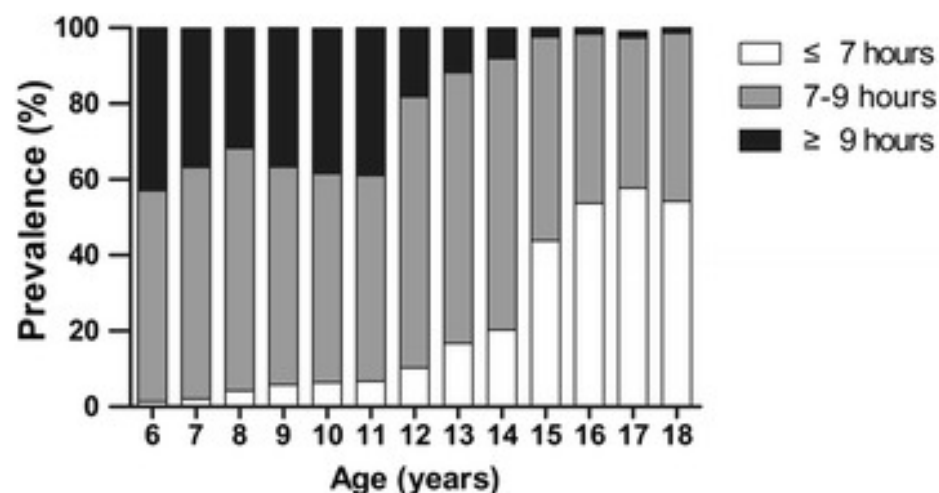
PMC full text: [BMC Public Health. 2015; 15: 1029.](#)

Published online 2015 Oct 7. doi: [10.1186/s12889-015-2359-0](#)

[Copyright/License](#) ►

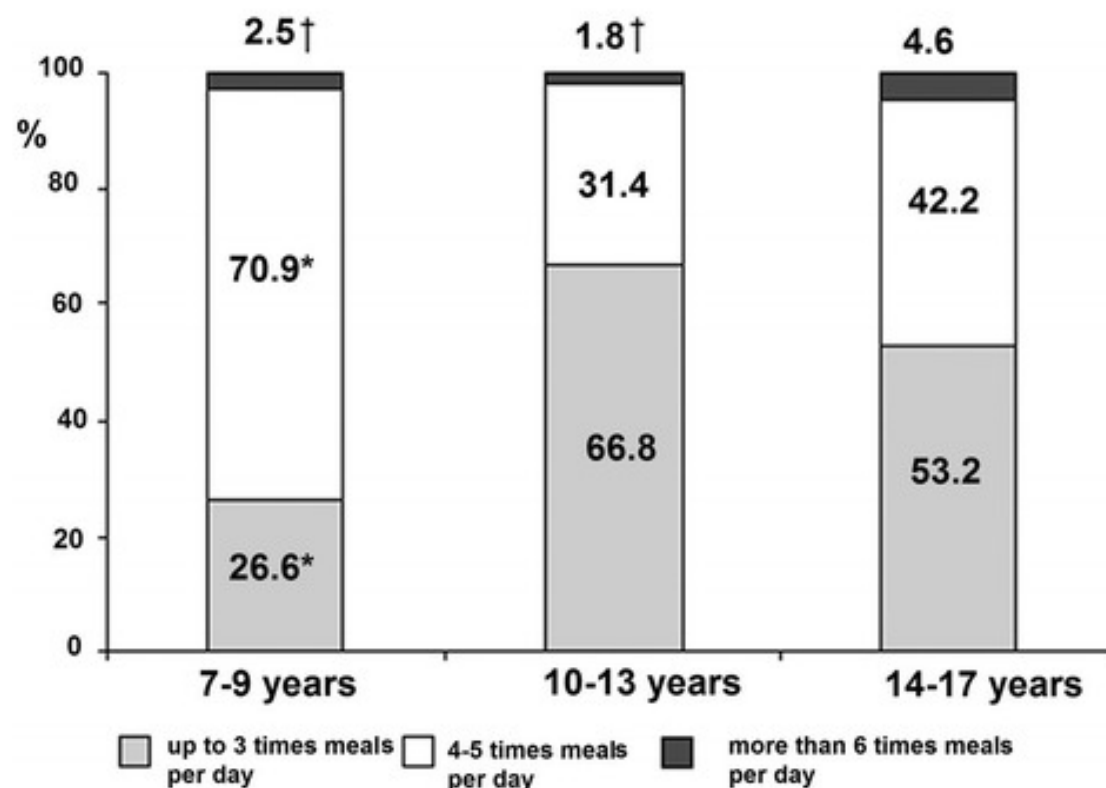
[Request permission to reuse](#)

**Fig. 1**



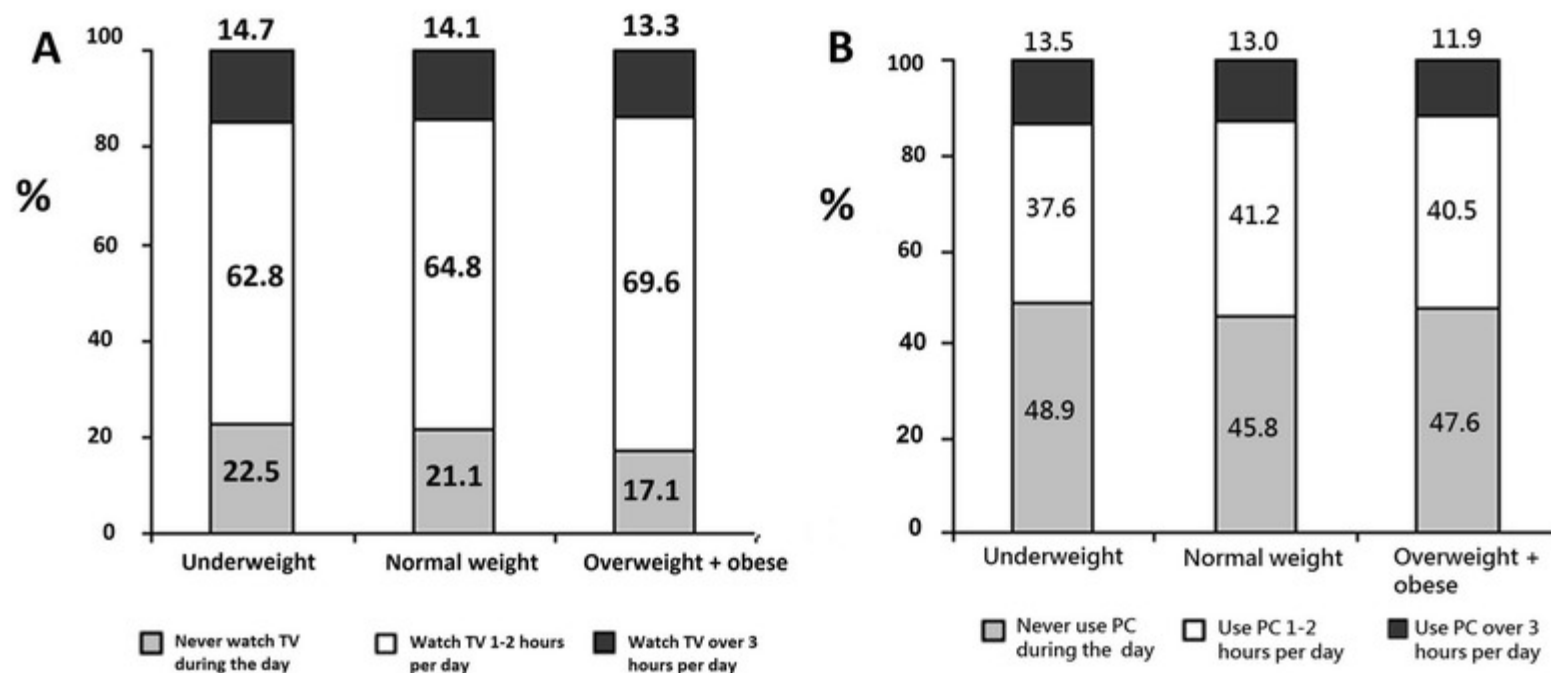
Distribution of sleep duration among children aged 6–18 years in urban area of Guangzhou, China

**Fig. 1**

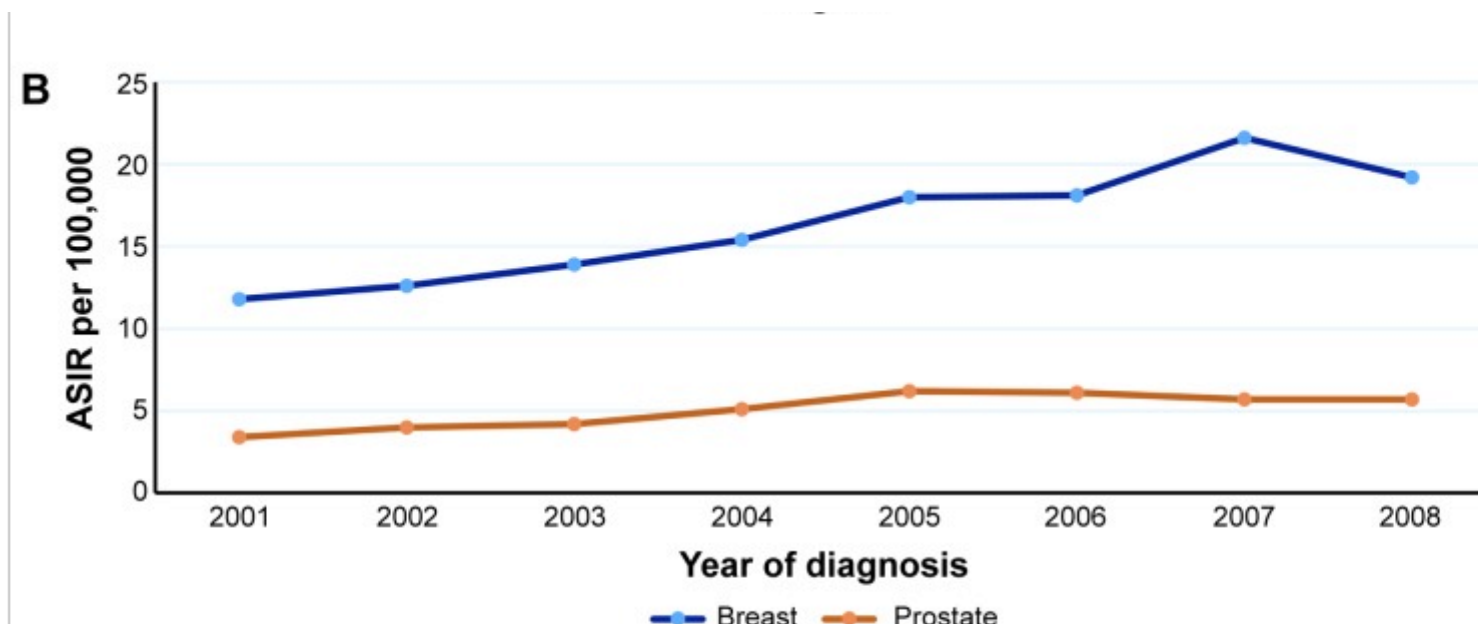


Distribution of meal frequency per day by the age groups. \* $P < 0.05$  in comparison with 10–13-year and 14–17-year age groups. † $P < 0.05$  in comparison with 14–17-year age group

**Fig. 4**



Distribution of time spent watching TV (a) and at the computer (b) during the day between schoolchildren by the BMI category



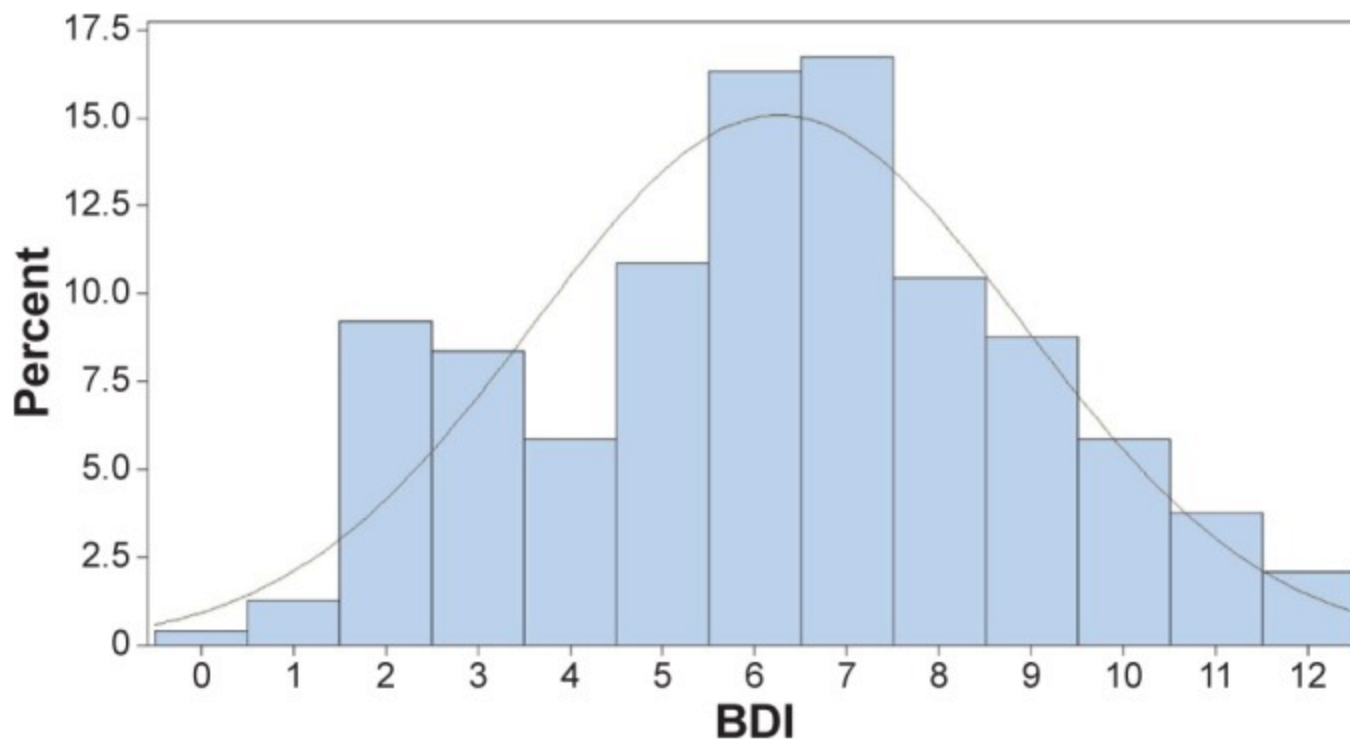
The ASIR of female breast cancer and prostate cancer adjusted by region and year of diagnosis in Saudi Arabia.

**Notes:** (A) The overall ASIR of female breast and prostate cancer, stratified by region in Saudi Arabia from 2001 to 2008. This method was tested on real data and proved its efficacy to find the best region for the gene link. The eastern region and Jazan were the best places to identify the association between female breast and prostate cancer. (B) The overall ASIR of female breast and prostate cancer in Saudi Arabia from 2001 to 2008. The trend of prostate and female breast cancer has the same direction from 2001 to 2008.

**Abbreviation:** ASIR, age standardized incidence rate.

PMC full text: [Int J Chron Obstruct Pulmon Dis. 2015; 10: 1663–1672.](#)  
Published online 2015 Aug 18. doi: [10.2147/COPD.S82408](#)  
[Copyright/License](#) ▶ [Request permission to reuse](#)

**Figure 1**



Distribution of the BDI score.

**Abbreviation:** BDI, Baseline Dyspnea Index.

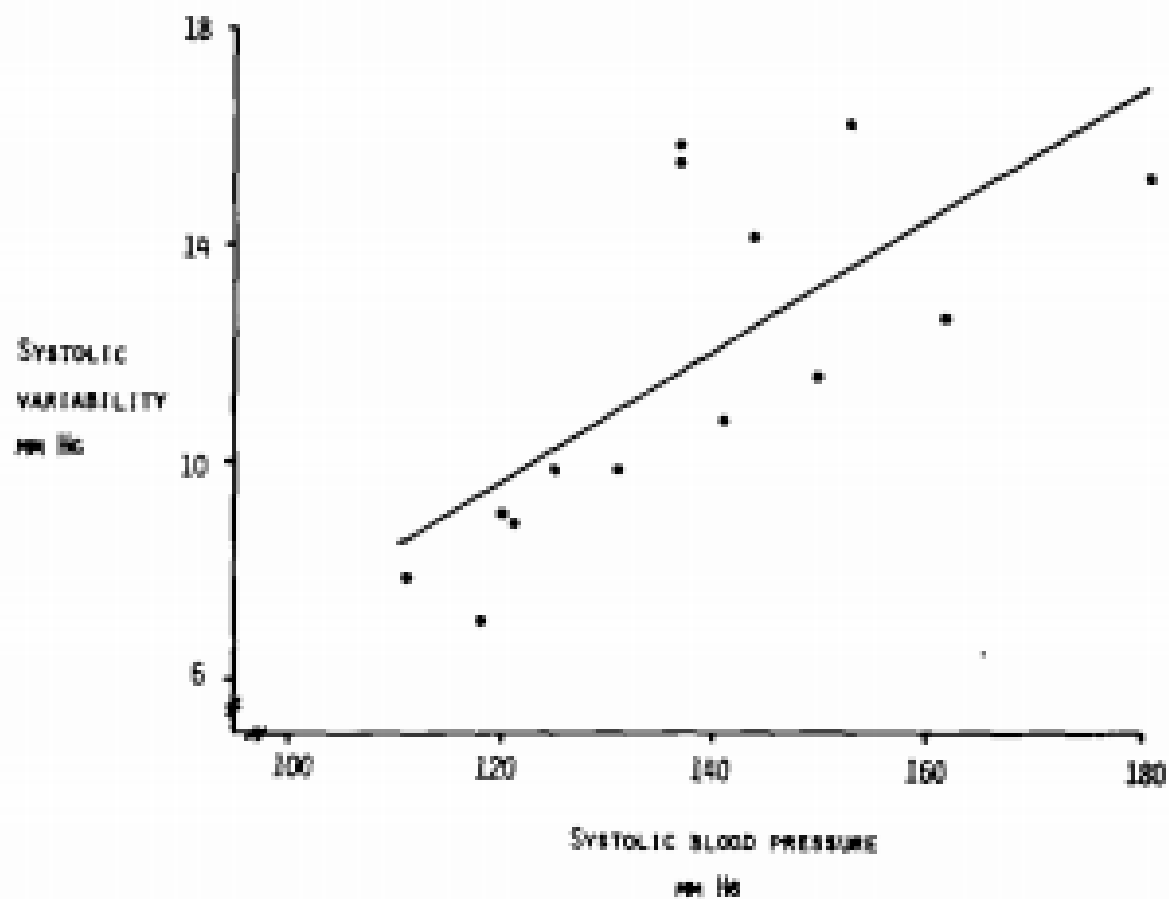


FIGURE 1. Relationship between average systolic pressure and variability of pressure in 14 patients studied during bed rest ( $r = 0.73$ ,  $p < 0.01$ ).

---

# GOOD TABLES PRACTICES: SUMMARY!

- Tables:
  - ❑ Capture: information concisely and display it efficiently
  - ❑ Provide information at any desired level of detail and precision
  - ❑ Number tables consecutively in the order of their first citation in the text and supply a brief title for each
  - ❑ Give each column a short or an abbreviated heading. Authors should place explanatory matter in footnotes, not in the heading
  - ❑ Explain all nonstandard abbreviations in footnotes
  - ❑ Identify statistical measures of variations
  - ❑ If you use data from another published or unpublished source, obtain permission and acknowledge that source fully

# GOOD GRAPHIC PRACTICES: SUMMARY!

- Figures should be made as self-explanatory as possible.
- Titles and detailed explanations belong in the legends-not on the illustrations themselves.
- Figures should be numbered consecutively according to the order in which they have been cited in the text.
- If a figure has been published previously, acknowledge the original source and obtain written permission from the copyright holder to reproduce the figure.
- Explain clearly in the legend each symbols, arrows, numbers, or letters used in a figure.
- Avoid 3D graphical representations!

# Thank you for your attention!

