REPORTING STATISTICS

The Common Errors Editors Won't Miss to pick-up

Dr.L.Jeyaseelan
Dept. of Biostatistics
Christian Medical College
Vellore – 632 002, India

JPGM WriteCon March 30-31, 2007, KEM Hospital,Mumbai

Decades of Misuse of Statistics

- 75% of 59 articles in ‘Transfusion” used an inappropriate statistical test or contained an error in calculation or interpretation. (Kanter & Taylor, 1994 Review of articles from 1992 - 1993)

- 19% of the 145 articles published in the American journal of obstetrics and Gynecology contained serious statistical errors. (Welch & Gabbe, 1996)

Current Scenario

- Medical Practitioners and researchers clearly find learning statistics difficult, not least in the exercise of judgment.

- A consensus needs to be reached to establish which tests……. are the most relevant ones to be adopted universally.

- This questionable aim indicates a ‘best buy’ approach to statistics

(Altman DG, 2000)

Reporting Descriptive Statistics

The Distribution of Data (Rule of thumb)

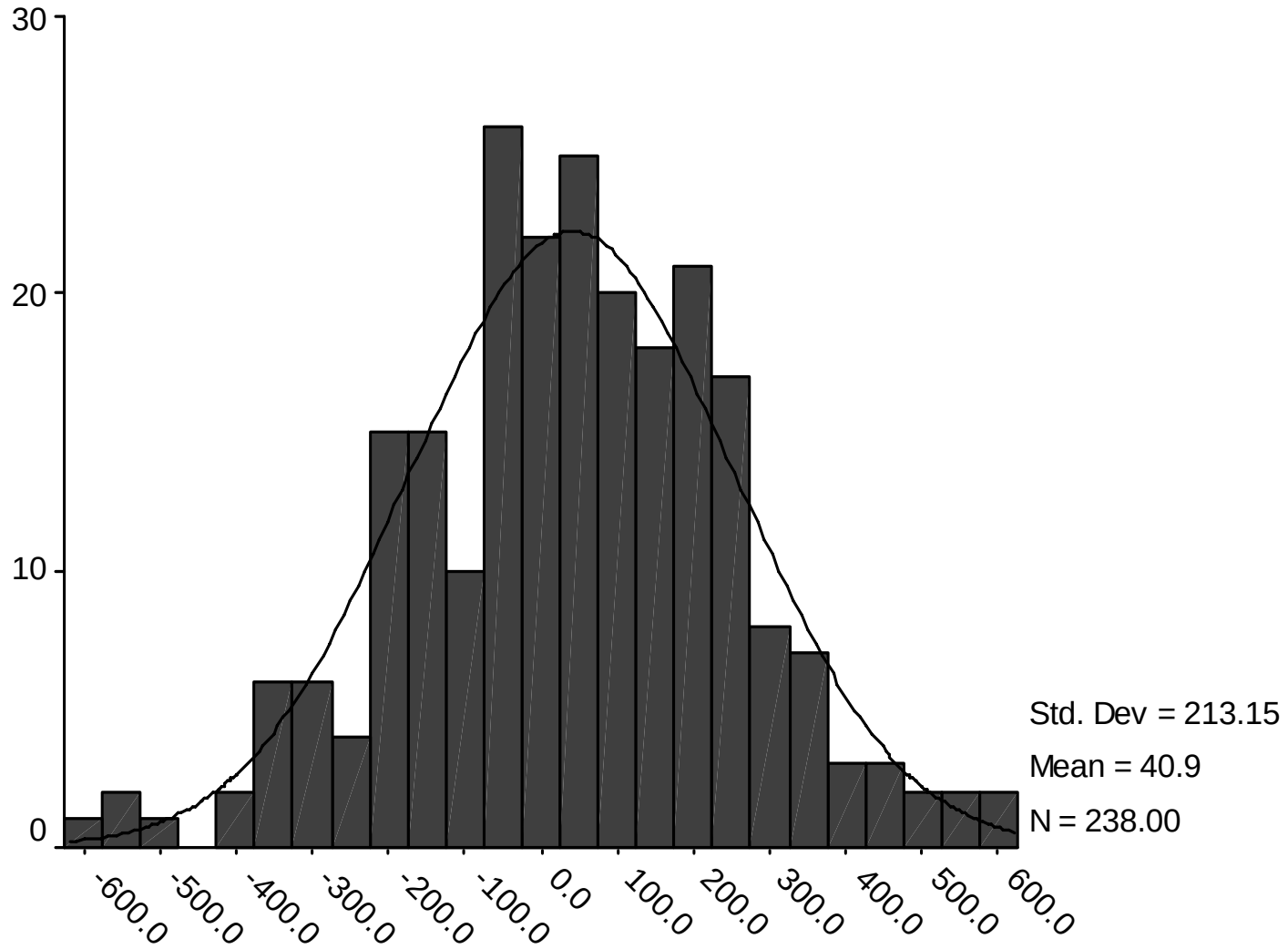
The statistical & clinical applications of the term "normal" are often confused and vague

SD>1/2 mean ⟶ Skewed/Non-normal data

Note: Applicable only for variable where negative values are impossible

Altman BMJ1991.

Comparison of TLV by Ultrasound and BSA (Western Population)



Difference between the Western formula using BSA and TLV

Contd..

- Data described with a SD that exceeds one-half the mean are non-normally distributed (assuming that negative values are impossible) and should be described with the median and range/interquartile range

- Subtracting the median from the mean produces a crude estimate of the skewness of the data:

The larger the difference, the greater the skewness

Presentation of Summary Statistics : SD or SE

- The terms “standard error” and “Standard deviation” are often confused.
- The contrast between these two terms reflects the important distinction between data description and precision/inference.
- **SD:** Is a measure of variability and explains how widely scattered some measurements are in a group.
- **SE:** Applicable for large samples & indicates the uncertainly around the estimate of the mean measurement.

Standard Deviation

Description of data:

Example:

If the mean weight of a sample of 100 men is 72kg and the SD is 8kg.

Assuming normal distribution 68% of the men are expected to weigh between

64 and 80kg.

Standard Error

72kg is also the best estimate of the mean weight of all men in the population.

How precise is the estimate 72kg?.

While testing hypothesis,

Difference in mean or proportions between groups.

Table1: Baseline characteristics of 2188 children with non-severe pneumonia randomised to 3 days or 5 days of treatment with amoxicillin. Values are numbers (percentages) of patients unless stated otherwise

Characteristic	3 day treatment (n=1095)		5 day treatment (n=1093)	
Mean (SD) Age (months)	17.0	(13.3)	16.9	(13.0)
Mean (SD) height (cm)	74.8	(10.98)	74.8	(10.75)
Mean (SD) weight (kg)	8.7	(2.49)	8.7	(2.4)
Mean (SD)duration of illness days)	4.7	(3.43)	4.5	(3.12)
Mean (SD) temperature (^{o}C)	37.1	(0.66)	37.2	(0.67)
Mean (SD) respiratory rate (breath / minute): 2 – 11 months old 12 – 59 months old	 56.4 47.3	 (5.02) (5.58)	 56.0 47.9	 (4.54) (6.1)
Male	685	(62.6)	676	(61.8)
Age (months): 2 – 11 12 – 59	 479 616	 (43.7) (56.3)	 475 618	 (43.5) (56.5)
Weight for height z score*: -2 to -1 -3 - 2	 300 188	 (27.4) (17.2)	 303 183	 (27.7) (16.7)

Table1 (Cont….)

Characteristic	3 day treatment (n=1095)		5 day treatment (n=1093)	
Duration of illness (days):				
< 3	538	(49.1)	540	(49.4)
≥ 3	557	(50.9)	553	(50.6)
Fever	833	(76.1)	850	(77.8)
Cough	1081	(98.7)	1078	(98.6)
Difficulty in breathing	417	(38.1)	387	(35.4)
Vomiting	135	(12.3)	141	(12.9)
Diahorrea	71	(6.5)	55	(5.0)
Excess respiratory rate (breaths / minute)				
≤ 10	903	(82.5)	881	(80.6)
> 10	192	(17.5)	212	(19.4)
Wheeze present	140	(12.8)	147	(13.4)
Adherence to treatment:	1031	(94.2)	1026	(93.9)
At day 3	937	(85.6)	928	(84.9)
At day 5				
RSV Positive	252	(23.0)	261	(23.9)

*Z score given as number of standard deviations from normal value.
†Rate above the age specific cut off
RSV=respiratory syncytial virus.

Comparison of outcome measures in 2188 children with non-severe pneumonia randomised to 3 days or 5 days of treatment with amoxicillin: intention to treat analysis. Values are numbers (percentages) of patients unless stated otherwise

	3 day treatment (n=1095)	5 day treatment (n=1093)	Difference (95% CI)
Primary outcome measures:			
Cure on day 5	980 (89.5)	983 (89.9)	0.4 (–2.1 to 3.0)
Relapse after day 5	58 (5.3)	48 (4.4)	1.0 (–1.0 to 3.0)
Secondary outcome measure:			
Cure on day 5 among wheezers	127/140 (90.7)	132/147 (89.8)	0.9 (–5.9 to 7.8)
Cure on day 5 among non-wheezers	853/957 (89.1)	851/946 (90.0)	0.7 (–2.1 to 3.4)

ISCAP Study Group BMJ 2004;328;791

Reporting Hypothesis Testing

Misuses of t-test

- **t-test for non-normal data.**

	Hospital 1		Hospital 2	
	Mean (SD)	**n**	**Mean (SD)**	**n**
Length of Stay (in days)	26 (17)	11	79 (57)	13

Heterogeneous data – SD > ½ (mean)

Correct Method: Non-parametric Mann-Whitney test with Median and Range values

Misuses of t-test

- **t-test for paired observations**

	Before intervention		After intervention	
	(n = 12)			
	Mean	SD	Mean	SD
BP Levels	142.0	30.5	120.5	31.5

Correct method: Paired t-test

Misuses of t-test (Contd. ..)

- **Multiple t-test**

Comparison of length of stays between three hospitals

	Hospital 1		Hospital 2		Hospital 3	
	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n
Length of Stay (in days)	25 (5)	12	75 (20)	13	30 (10)	14

Hospital 1 vs Hospital 2 ⟶ P- value = ?

Hospital 1 vs Hospital 3 ⟶ P- value = ?

Hospital 2 vs Hospital 3 ⟶ P- value = ?

The correct conclusion for 3 comparison is $(1-.05)^3 = .86$

Correct method: ANOVA with Bonferroni correction.

Chi Square Test - Misuses

1. Paired data

In a study to compare a new method with an established method for the culture of tubercle bacilli, each of 320 specimens was tested by the two methods. The new method detected 96(30%) positives and the established method 80(25%).

Chi square test = 1.76,

Conclusion: Difference in proportions of positives by the two methods was not significant.

Chi Square Test – Misuses (contd.)

		Established method	
		+ves	-ves
New	+ves	78	18
method	-ves	2	222

Appropriate test:

McNemar's test with the continuity correction.

$\chi^2=(|18-2| -1)^2 / (18 + 2)$, which yields a chi-square of 11.2 with 1 d.f. This is significant ($p<0.001$) and demonstrates that the new method is more sensitive than the established method in detecting tubercle bacilli.

Chi Square Test - Misuses

Chi-square test undertaken on mean values

In a study to assess the efficacy of an Ayurvedic drug in the treatment of bronchial asthma, each of 15 patients were treated with the drug for 7 days and with a lactose placebo for 7 days. Symptoms such as cough, wheezing attacks were scored according to their severity initially and at the end of the treatment.

To assess the efficacy of the Ayurvedic drug, the investigators computed mean scores and computed the 2×2 chi square test.

Chi Square Test – Misuses (Contd..)

Application of chi-square test: mean symptom score before and after treatment.

	Mean score (treatment)		
	Total	Before	After
Placebo	5.3	4.3	9.6
Ayurvedic drug	6.7	3.1	9.8

χ^2=0.36, they concluded that Ayurvedic drug had no effect in the treatment of bronchial asthma.

Chi Square Test – Misuses (Contd..)

Correct procedure:

Compute the difference between the initial score and the final score for each patient in both the drug and placebo group.

Independent t-test can be used to test whether the mean difference in drug is significantly different from placebo.

Chi Square Test – Misuses (Contd..)

Key points:

χ^2 must be actual number observed in each category

(neither proportions nor percentages)

If the expected number is < 5 in one cell, the ordinary chi square test should not be applied.

Yates correction for continuity must be applied when the numbers involved in the study are small.

Correlation Coefficient Vs Intra-class Correlation Coefficient

Real Life Example:

The mean arterial pressure for 40 adolescents was calculated for each of the instruments separately

Table : Arterial pressure by aneroid and mercury sphygmomanometer (mmHg)

S. No	Aneroid	Mercury	S. No	Aneroid	Mercury
1	85	86	21	90	91
2	102	101	22	92	92
3	91	101	23	100	100
4	82	82	24	103	101
5	96	96	25	93	93
6	101	100	26	93	95
7	86	90	27	89	89
8	97	101	28	96	95
9	114	114	29	91	91
10	86	90	30	89	90
11	93	95	31	92	91
12	91	100	32	94	94
13	100	100	33	100	100
14	101	102	34	90	91
15	92	88	35	88	86
16	90	92	36	91	91
17	105	103	37	97	97
18	86	89	38	100	100
19	91	91	39	92	92
20	92	95	40	86	86

Correlation Vs Intra-Class Correlation Coefficient (ICC)

- There are a number of common situations in which the correlation coefficient can be misinterpreted when comparing two methods of measurement.

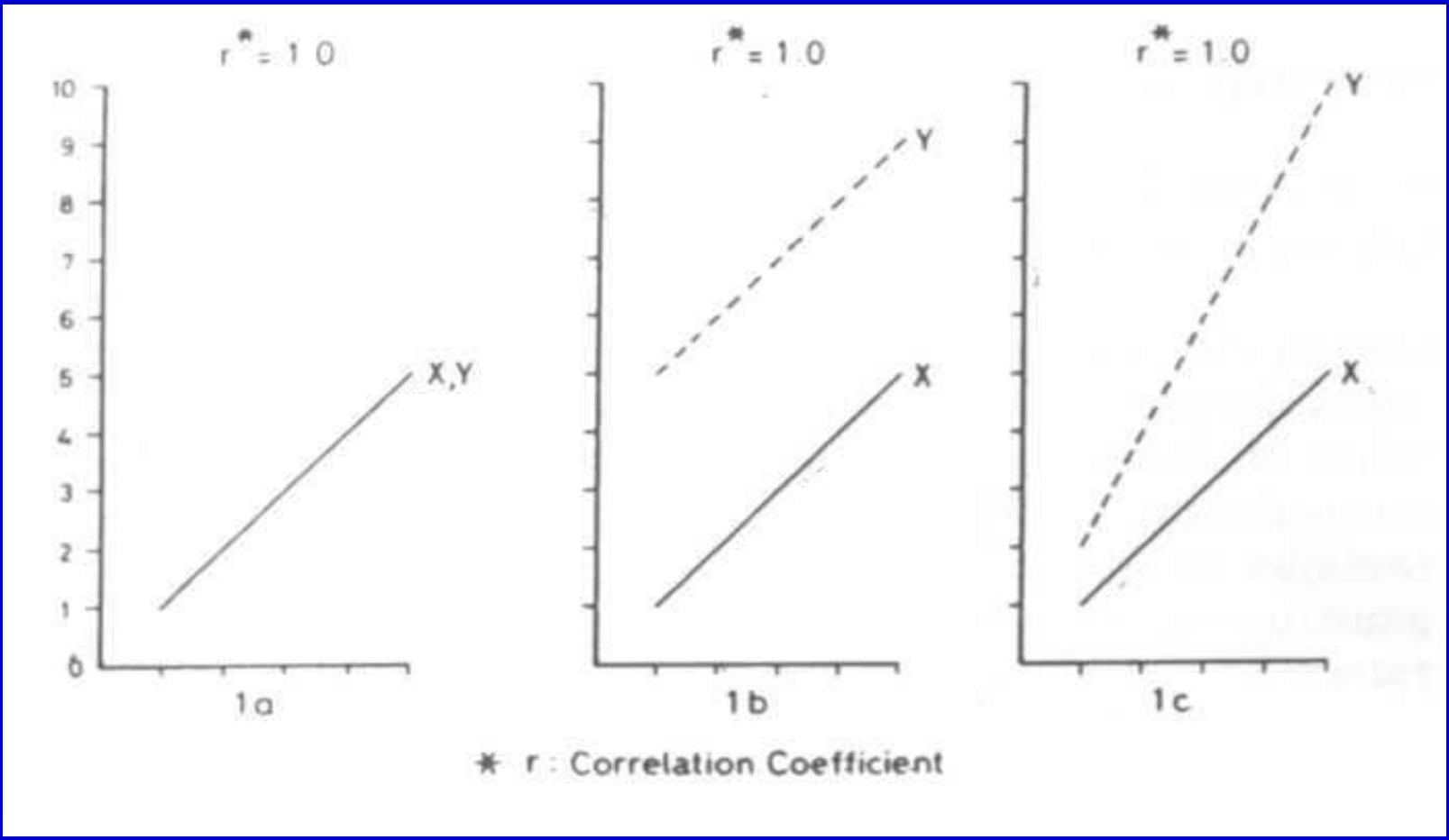
- The correlation coefficient is a measure of association and not of agreement

Three sets of ratings by two raters (X and Y)

Subjects	A		B		C	
	X	Y	X	Y	X	Y
1	1	1	1	5	1	2
2	2	2	2	6	2	4
3	3	3	3	7	3	6
4	4	4	4	8	4	8
5	5	5	5	9	5	10

Contd..

Plots of rater-table data



Summary

- Errors in reporting could be avoided, using check lists and guidelines.
- Journals should come out with teaching/tutorial articles in statistics and research methods, as Continuing Medical Education.
- ICMR and DBT should build capacity to teach Biostatistics and research methods for the PG courses.


Thank You