*Common Statistical errors in clinical study*

# 의학학술지에서 흔히 보는 통계오류

연세대학교 의과대학 의학통계학과
남 정 모

*Common statistical errors in clinical study*

# Contents

# Errors in Descriptive Statistics

- **Not providing the level of measurement of each variable**

  - The levels of measurement of variables are important because they determine the type of statistical test.


| Team | A | | | | | B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Height | 174.6 | 176.3 | 178.1 | 178.8 | 185 | 170.1 | 174.8 | 178.1 | 179.6 | 191.8 | Continuous data |
| Rank | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | Ordinal data |

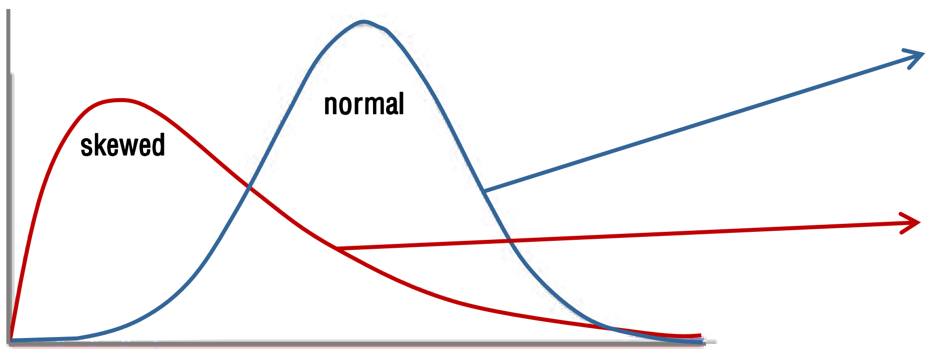
Nominal data (Team)

- **Dividing continuous data into nominal (or ordinal) categories without explaining how the categories were created**

[Ex] height : 170~175 ⇒ small, 176~180 ⇒ normal, 181~ ⇒ tall

# Errors in Descriptive Statistics

- Using the mean and standard deviation to describe continuous data that are not normally distributed



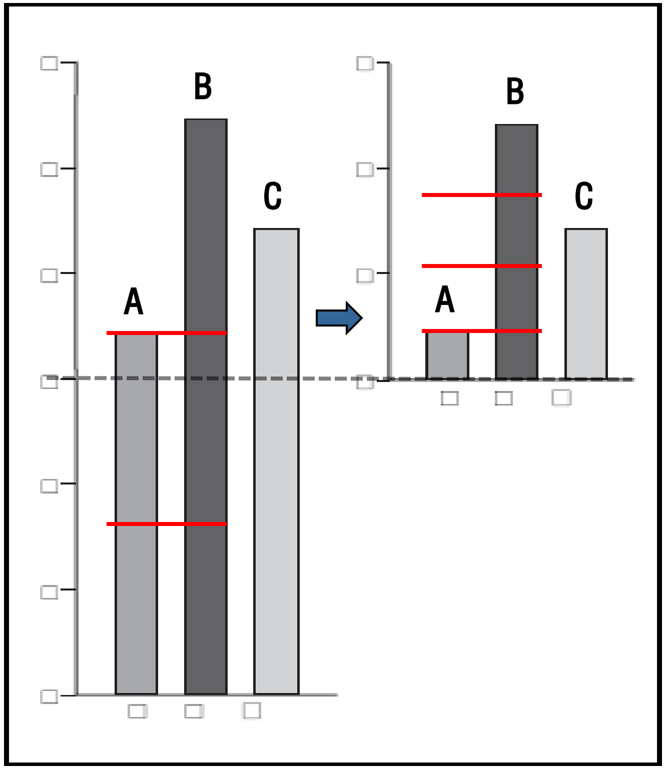

Data : Normal distribution
Mean ± SD

Data : Non-normal distribution
Median ± IQR (or Range)

- **Using the standard error (SE) as a descriptive statistic**
  - The SE(or SEM) is always smaller than the SD, and so its use makes measurements look more precise than they are.

# Errors in Data Displays

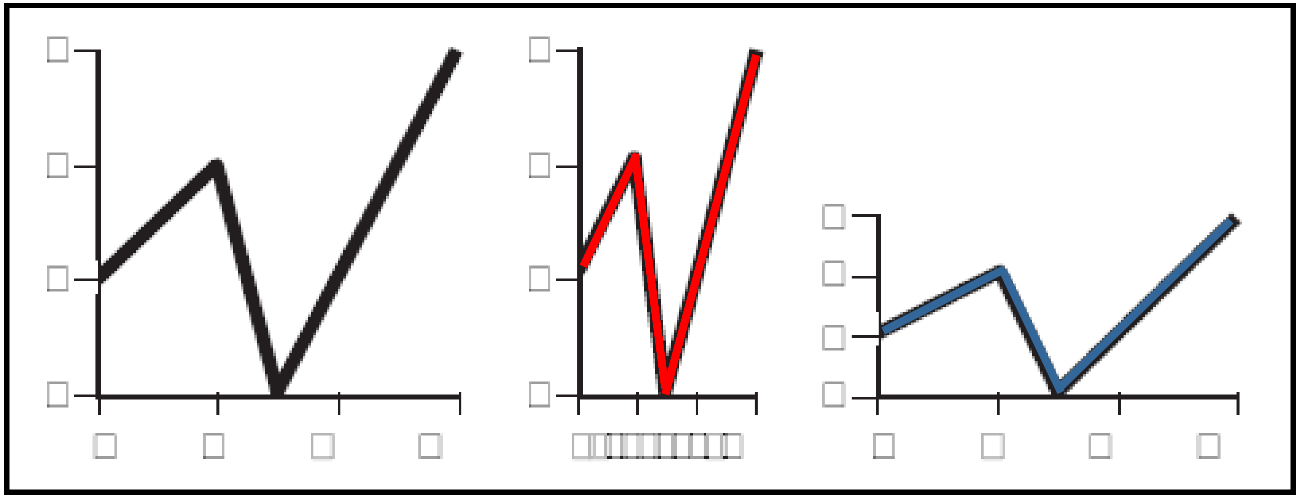
- Visually distorting relationships on a column chart by starting columns at a baseline value other than zero




- The suppressed zero visually distorts the relationships among quantities. Here, A is actually **two thirds** as large as B, but the suppressed zero makes A appear to be less than **one quarter** the size of B.

# Errors in Data Displays

- Visually distorting relationships among data by manipulating the relative scales on the X and Y axes



- A scale that seems to be unduly compressed or expanded may be a clue that the authors, intentionally or otherwise, are trying to minimize large differences or maximize small differences in the data.

# Errors in Expressing and Interpreting P-values

- **Reporting only P-values for analysis results**

  - AIM(Bailar JC et al. 1988) recommended reporting the 95% CI instead of, or in addition to, the p-value.

  Ex) A : The effect of the drug on lowering SBP was statistically significant ($P<0.05$).

  B : The mean SBP of the treatment group dropped from 110 to 92 mm Hg ($P=0.02$).

  C : The drug lowered SBP by a mean of 18 mmHg, from 110 to 92 (95% CI = 2 to 34; $P=0.02$).

- **Confusing statistical significance with clinical importance**

  - P-values have no clinical interpretation. The nature and size of the difference must be judged to determine clinical importance.
  - Statistical significant *vs.* clinical significant

# Errors in Expressing and Interpreting P-values

➢ **Interpreting non statistically significant results as "Negative" or "Similar " ?**

- Big effects may not be statistically significant if sample size is low.

| Group | Event | | Total |
|---|---|---|---|
| | Y | N | |
| Valsartan | 83(**5.5%**) | 1434 | 1517 |
| Non-ARB | 155(**10.2%**) | 1359 | 1514 |
| Total | 238 | 2793 | 3031 |

**p-value < 0.001**

| Group | Event | | Total |
|---|---|---|---|
| | Y | N | |
| Valsartan | 9(**5.9%**) | 144 | 153 |
| Non-ARB | 16(**10.5%**) | 136 | 152 |
| Total | 25 | 280 | 305 |

**p-value = 0.139**

- In adequately powered studies, statistically insignificant results are truly negative.

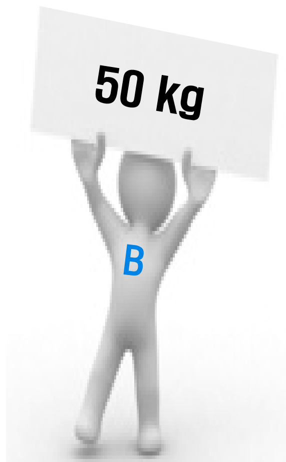
# Errors in Describing the Statistical Methods

- **Not naming or incorrectly naming the statistical tests used in the analysis**
  - Put general and correct reports of statistical methods in the Method section. Avoid describing the methods not used in data analysis.
- **Not reporting the statistical computer program (or software) used in the analysis**
  - General-use computer programs should be specified, because some programs are sometimes found to have errors.
- **Not describing the statistical decision rule**

# Errors in Interpreting Differences Between Groups

- Confusion of reference


\-




= 30 kg

Difference is always equal to 30

But, percent is different, depending on the **reference(denominator)**

① (80-50) / 50 × 100 = 60%
A is 60% heavier than B

② (80-50) / 80 × 100 = 37.5%
B is 37.5% lighter than A

- Not present reference of dummy variables

# Errors in Reporting of Diagnostic tests

- **Not reporting what the number of uncertain test results was**

  - For computing sensitivity and specificity, we do not include uncertain(non-positive or non-negative) results. So, reporting the number and proportion of uncertain results is important because such results affect the clinical usefulness of the test.

  - Even a highly sensitive or specific test may be of little value if much of the results are uncertain.

# Errors in Reporting of Diagnostic tests

- Confusing sensitivity; specificity; false-positive, and false-negative results; positive and negative predictive values.

| Test Result | Disease present | Disease absent | Total |
|---|---|---|---|
| Positive | a | b | a+b |
| Negative | c | d | c+d |
| Total | a+c | b+d | a+b+c+d |

Sensitivity = a/(a+c)

Specificity = d/(b+d)

*If the table reflects the prevalence of disease* :

Positive predictive value = a/(a+c)

Negative predictive value = d/(c+d)

- Misuse of PPV & NPV in case-control study

How could this data collect?
- Cohort
- Cross-sectional
- Case-control

# Errors in Reporting of Diagnostic tests

Population → Sampling

Population:

| 진단 \ 질병 | D | N | Total |
|---|---|---|---|
| Positive | 90 | 900 | 990 |
| Negative | 10 | 100 | 110 |
| Total | 100 | 1000 | 1100 |

$f_1 = 0.5$

$f_2 = 0.05$

Sampling:

| 진단 \ 질병 | D | N | Total |
|---|---|---|---|
| Positive | 45 | 45 | 90 |
| Negative | 5 | 5 | 10 |
| Total | 50 | 50 | 100 |

$$PPV = \frac{90}{990} = 0.091$$

$$PPV^{*} = \frac{45}{90} = 0.5$$

잘못된 PPV를 얻게 됨.

$$\text{sensitivity} = \frac{45}{50} = 0.9$$

$$\text{specificity} = \frac{5}{50} = 0.1$$

$$\text{prevalence} = 0.1$$

외부 자료에서 가져옴.

$$PPV = \frac{0.1 \times 0.9}{0.1 \times 0.9 + (1-0.1) \times (1-0.1)} = 0.1$$

모집단과 비슷한 PPV를 얻게 됨.

# Errors in Conclusions

| Multiplicity | Small sample size |
| --- | --- |
| ⬇ | ⬇ |
| Type I error ↑ | Type II error ↑ |
| ⬇ | ⬇ |
| False positive | False negative |

# References

John C.B. and Frederick M. (1998), Guidelines for Statistical Reporting in Articles for Medical Journals. *Annals of Internal Medicine*, 108;266~273.

Tom Lang (2003), Common Statistical Errors even you can find PART 1. *American Medical Writers Association Journal*, 18(2);67~71.

Tom Lang (2003), Common Statistical Errors even you can find PART 2. *American Medical Writers Association Journal*, 18(3);103~107.

Tom Lang (2004), Common Statistical Errors even you can find PART 3. *American Medical Writers Association Journal*, 19(2);9~11.

Tom Lang (2004), Twenty Statistical Errors even you can find in Biomedical Research Articles. *Croatian Medical Journal*, 45(4);361~370.

Tom Lang (2005), Common Statistical Errors even you can find PART 4. *American Medical Writers Association Journal*, 20(1);10~11.

Tom Lang (2005), Common Statistical Errors even you can find PART 5. *American Medical Writers Association Journal*, 20(2);50~51.

Tom Lang (2006), Common Statistical Errors even you can find PART 7. *American Medical Writers Association Journal*, 21(1);17~18.

Alan R.Z. and Jason T.C. (2008), Ten Common Statistical Errors and How to Avoid Them. *American Journal of Gastroenterology*, 103;262~266.

# 감 사 합 니 다.