TUTORIAL

Statistics at a glance

Hugo ECTOR, MD, PhD, Editor
University Hospital Gasthuisberg, Leuven, Belgium.

I still remember my first book on statistics: “Elementary statistics with applications in medicine and the biological sciences” by Frederick E. Croxton. For me, it has been the start of pursuing understanding statistics in daily life and in medical practice. It was the first volume in a long row of books.

In his introduction, Croxton pretends that “nearly everyone involved in any aspect of medicine needs to have some knowledge of statistics”. The reality is that for many clinicians, statistics are limited to a “P < 0.05 = ok”. I do not blame my colleagues who omit the paragraph on statistical methods. They have never had the opportunity to learn concise and clear descriptions of the key features. I have experienced how some authors can describe difficult methods in a well understandable language. Others fail completely.

As a teacher, I tell my students that life is impossible without a basic knowledge of statistics. This feeling has resulted in an annual seminar of 90 minutes. This tutorial is the summary of this seminar. It is a summary and a transcription of the best pages I have detected.

Statistical Errors

Type 1 and type 2 errors are discussed in every textbook of statistics. I have learnt about type 3 and type 4 in a paper by Robin ED and Lewiston NJ.

Type 1 (alpha) errors are false positive errors that assign statistical benefit to a given modality, when such benefit does not exist. When we receive the famous message “p < 0.05” that treatment A is better than treatment B, we understand that the risk of a false positive result is less than 5/100! When we have a “p < 0.05”, we accept that the result is “statistically significant”.

Type 2 (béta) errors are false negative errors: the observed distinction is regarded as not significant (no benefit), whereas in reality a benefit exists. Type 2 errors can arise in circumstances in which the numbers of observations are too small: insufficient power of the study. The classical design of a trial requires a maximum risk of a type 2 error of 0.2 (i.e. 20/100), or, in other words, the chance of detecting a real difference (power of the study) must be at least 0.8 (80/100).

Type 3 errors are errors in which the risk of a given medical or public health approach is underestimated, undetected or not specifically sought, leading to an underestimate of the risk-benefit balance.

Type 4 errors arise because the risks of a given medical intervention are overestimated, leading to under-use or abandonment of a useful intervention.

It should be emphasized that type 3 errors appear to be much more frequent than type 4 errors. This stems, in part, from the failure of medicine to establish effective mechanisms for the rapid recognition and correction of errors. Most recorded medical history deals with the triumphs of medicine. The disasters are usually interred along with the victim-patients.

95% Confidence Limits of a proportion P when P is a percentage (%)

\[
P = P \pm 1.96 \sqrt{\frac{P(100 - P)}{N}}
\]

N = number of observations

When we are taking a sample from the entire population, we can pretend with 95% certainty that the true percentage for this population will be between these confidence limits.

When we want 99% certainty we have to use a factor 2.57 instead of 1.96 (99% confidence limits).
95% confidence limits of the difference between proportion \( P_1 \) & \( P_2 \) when \( P_1 \) & \( P_2 \) are percentages (%)

\[
(P_1 - P_2) \pm 1.96 \sqrt{\frac{P_1(100 - P_1)}{N_1} + \frac{P_2(100 - P_2)}{N_2}}
\]

\( N_1 \) = number of observations in group 1, \( N_2 \) = number of observations in group 2.

See Addendum: example (i).

The difference is significant at the 5% level when the 95% confidence limits of this difference do not include 0. The risk of a type 1 error is less than 5%.

When we want 99% certainty we have to use a factor 2.57 instead of 1.96 (99% confidence limits).

“Statistics with confidence” is a must in the library of every clinician.

Correlation of two variables

Correlation has to do with the relationship between two sets of data which vary together.

The correlation coefficient \( r \) is calculated with the help of a statistical package. If we square the correlation coefficient, we obtain the coefficient of determination \( r^2 \), which is

\[
r^2 = \frac{\text{Explained Variation}}{\text{Total Variation}}
\]

A correlation of 0.4 means that we can explain a proportion of (0.4² = 0.16 i.e. 16%) of the total variation.

This basic idea holds true for two-variable linear or non-linear correlation and for multiple correlation. The coefficient of determination \( r^2 \) tells us the proportion of the variation present in the dependent variable which has been explained by the use of the estimating equation. The statistical package will give a value of “\( r \)” and also a “p” value (risk of type 1 error: risk that the observed correlation is false). A p value of < 0.05 is considered as statistically significant.

The presence of correlation between two sets of data does not necessarily mean that causation is present, even though the correlation may be high.

Size and power of a clinical trial

A common statistical approach is to focus on a single outcome of patient response which is dichotomous: that is, each patient’s outcome on treatment can be classified either as a “success” or “failure” (e.g. death in a year = failure, survival = success).

One has has to decide on the following items:

(i) \( P_1 \) = percentage of successes expected on one treatment (usually the standard)

(ii) \( P_2 \) = percentage of successes on the other treatment which one desires to detect as being different from \( P_1 \)

(iii) alpha (\( \alpha \)) = type 1 error = the risk of a false-positive result

(iv) beta (\( \beta \)) = type 2 error = the risk of a false-negative result

(v) \( 1 - \beta \) = the power to detect a real difference of magnitude (\( P_1 - P_2 \)) = the degree of certainty that the difference, if present, would be detected.

\( P_1 \) and \( P_2 \) are hypothetical percentage successes on the two treatments that might be achieved if each were given to a large population of patients.

Example: Let us chose \( P_1 = 90\% \) for placebo; \( P_2 = 95\% \) for new treatment; \( \alpha = 0.05 \) (we accept 5% risk of a false-positive result); \( \beta = 0.1 \) (we allow 10% risk of a false-negative result, in other words we require a power of 90%).

The required number of patients on each treatment \( (N) \) is given by the following formula:

\[
N = \frac{P_1(100 - P_1) + P_2(100 - P_2)}{(P_2 - P_1)^2} \times F(\alpha, \beta)
\]

\( F(\alpha, \beta) \) is a function of \( \alpha \) and \( \beta \):

For \( \alpha = 0.05 \) and \( \beta = 0.2 \) \( F = 7.9 \) (Power = 80%)

For \( \alpha = 0.05 \) and \( \beta = 0.1 \) \( F = 10.5 \) (Power = 90%)

For our example we calculate:

\[
N = \frac{90 \times 10 + 95 \times 5}{(95 - 90)^2} \times 10.5 = 578 \text{ patients required on each treatment (total number } 1156)\]

See also Addendum: example (ii)

Risk reduction, statistical significance, clinical relevance

It is clear that statistical significance is not synonymous with clinical relevance. However in many papers, most emphasis goes to the relative risk reduction or to the odds ratio.

This obscures the clinical relevance, which is best reflected in the absolute risk reduction. When we reduce a risk from 50% to 25% or from 2% to 1%, in both instances, we obtain a relative risk reduction of 50%. However, the difference in clinical relevance is reflected in the absolute risk reduction: 25% versus 1%.

For saving one patient, the number-needed-to-treat is 4 (\( = 100/25 \)) versus 100 (\( = 100/1 \))

The number of patients, who have to be treated to prevent one outcome event, is directly related to the
Addendum: Examples

(i) Significance of difference between two percentages: 30 % VERSUS 20 %

\[(P_1 - P_2) \pm 1.96 \sqrt{\frac{P_1(100 - P_1)}{N_1} + \frac{P_2(100 - P_2)}{N_2}}\]

\[N_1 = 100 \quad N_2 = 100\]

\[(30 - 20) \pm 1.96 \sqrt{\frac{30(100 - 30)}{100} + \frac{20(100 - 20)}{100}}\]

\[(30 - 20) \pm 1.96 \sqrt{21 + 16} = 10 \pm (2 \times 6) = 10 \pm 12: \text{not significant}\]

The confidence limits include 0

\[N_1 = 200 \quad N_2 = 200\]

\[(30 - 20) \pm 1.96 \sqrt{\frac{30(100 - 30)}{200} + \frac{20(100 - 20)}{200}}\]

\[(30 - 20) \pm 1.96 \sqrt{10.5 + 8} = 10 \pm (2 \times 4.3) = 10 \pm 8.6: \text{significant}\]

\[p < 0.05 \text{ as the confidence limits do not include 0}\]

(ii) Calculation of Sample Size: Power 80 %

\[\frac{20(100 - 20) + 30(100 - 30)}{(30 - 20)^2} = \frac{(1600 + 2100)}{100} = 37\]

37 X 8 = 296 in each arm of study: total of ± 600 test cases
cost-benefit ratio of a treatment. The number-needed-to-treat is calculated by taking the reciprocal of the absolute risk reduction.

**Association and Prediction.**

Clinicians want to make predictions about outcomes for individual patients. Especially for diagnostic tests, positive predictive values are needed. I am always surprised that in most papers, for positive and negative predictive values, confidence limits are omitted.

It may escape our attention, that the more infrequent an event, the higher will be the negative predictive value of any test. In such instances, this parameter does not reveal anything about the quality of the test. Otherwise, the positive predictive value is often not 50%, thus not better than the result of tossing a coin.

Prediction is very often confounded with association. Between some parameters and outcome, there can be a significant association, while positive predictive values remain poor. I refer to some discussions on this issue7-10.

**Trials, Observational Studies, Data Dredging.**

For me the encounter with Prof. Dr. Alvan R. Feinstein -year 1970- has introduced me in the concept of “data dredging”11. A large number of statistical associations can be explored in an automated manner for diverse individual groups, agents, and outcomes. Whenever a “statistically significant” result emerges during the myriads of computations, the event may be proposed as a cause-effect relationship. When we examine 100 correlations, at least 5 will turn out to be significant!

We live in an era where clinical trials tend to dictate the law. What the physician thinks, suspects, believes, or has a hunch about, is assigned to “the not knowing” category. “Knowing” is defined on the basis of an arbitrary but accepted statistical test performed in a randomized clinical trial12. “We owe patients involved in the assessment of new therapies the best that science and ethics can deliver. Today, for most unproved treatments, that is a properly performed randomized clinical trial”13.

However, because of many limitations, most of our therapeutic decisions are not entirely the result of evidence based medicine14. Randomized trials are unfeasible for studying: (i) multiple therapeutic candidates; (ii) “instabilities” due to rapid technologic improvements; (iii) long-term adverse effects15. The basic science of patient care will also require our major attention to the events and observations that occur in the ordinary circumstances of clinical practice.

Valuable observational studies remain necessary15. Increased attention is needed to improve the basic science of patient-oriented research (clinical epidemiology), and to evaluate causal relations more rigorously16. A remarkable conclusion is reached by Vandenbroucke JP17. Restriction in research topics, design, and analysis helps observational research to attain the desired benefits of randomization, and gives observational research the chance to be as credible as randomized evidence.

**Logistic Regression.**

I have met few people being aware of the basics of logistic regression. I have found the best description, with example, in a SPSS manual18.

Predicting whether an event will or will not occur, as well as identifying the variables useful in making the prediction, is important.

In multiple linear regression, the regression coefficient tells you the amount of change in the dependent variable for a one-unit change in the independent variable.

The **logistic model** is rewritten in terms of the odds of an event occurring. The odds of an event occurring are defined as the ratio of the probability that it will occur \( \frac{\text{Prob(event)}}{\text{Prob(no event)}} \)

\[
\log \left( \frac{\text{Prob(event)}}{\text{Prob(no event)}} \right) = \frac{\text{Prob(event)}}{\text{Prob(no event)}} = \frac{1}{\frac{1}{\text{Prob(event)}}} \]

\[
\log \left( \frac{\text{Prob(event)}}{\text{Prob(no event)}} \right) = Z = B_1 + B_2 X_1 + B_3 X_2 + \ldots B_n X_n
\]

\[
\text{Prob(event)} = 1 - \text{Prob(no event)} \text{ if Prob(event)} = 0.4 \text{ then Prob(no event)} = 0.6
\]

\[ e \text{ is the base of the natural logarithms, approximately 2.718} \]

\[ B_1, B_2, B_3, \ldots \text{ are coefficients estimated from the data} \]

\[ X_1, X_2, \ldots \text{ are the independent variables} \]

**THE LOGISTIC COEFFICIENT CAN BE INTERPRETED AS THE CHANGE IN THE LOG ODDS ASSOCIATED WITH A ONE-UNIT CHANGE IN THE INDEPENDENT VARIABLE**

It is clear that a statement based on this genius concept must give more than a phrase as “logistic regression has demonstrated a statistically significant correlation between…””. For the sake of their readers, reviewers should require specification of Prob(event) for typical values of the independent variables. Such data remain absent in most papers and hide the “absolute” truth.

**The Kaplan-Meier Principle in Daily Life**

The Kaplan-Meier technique is used for the calculation of survival proportions or the so-called
“free-of-event proportions”. A concise and easy-to-understand explanation is given by Machin D and Garner MJ. A simple application for daily life is the estimation of a “free-of-event proportion” by hand. Suppose the risk of an event is 5% (0.05). For a constant risk of 0.05, after one year the chance of a free-of-event is 0.95, after two years 0.95², after three years 0.95³, …

Epilogue.

In this editorial, I have only made up a bunch of other men’s flowers, providing of my own only the string to tie them together. I feel obliged to refer to two more statistical flowers: one for the beginner, one for advanced study. One flower can embellish the infirmary. Having a good memory for a few statistical principles enlightens our clinical work and duty.

References

20. Quotation from Michel de Montaigne, Essais, 1580.