

COMMON STATISTICAL ERRORS EVEN YOU CAN FIND*

PART I: ERRORS IN DESCRIPTIVE STATISTICS AND IN INTERPRETING PROBABILITY VALUES

TOM LANG, MA

TOM LANG COMMUNICATIONS, 2003

Presented by Sacha Rose Uritis

“Critical reviewers of the biomedical literature have consistently found that about half the articles that used statistical methods did so incorrectly.”

POOR STATISTICAL REPORTING IS...

- “Longstanding”
- “Widespread”
- “Potentially serious”
- “Almost unknown”
- BUT... “much can be **easily avoided** by following a few guidelines.”

TWO BROAD AREAS IN STATISTICS:

- 1) **Descriptive** (**DS**)
- 2) **Inferential** (**IS**)

DS ERROR #1

Not Defining Each
Variable in
Measurable Terms

Providing an Operational
Definition for Each
Variable.

Example : In our
StackOverFlow data, we
wanted to measure the
proportion of **“heavy”**
users of StackOverFlow.
Response var is
categorical and needs an
operational definition.

DS ERROR #2

Not Providing the
Level of Measurement
of Each Variable

How much info is collected
about the var (nominal,
ordinal, & continuous).

**Why? : different level of
measurements require
different statistical
tests and approaches...**

EXAMPLE IN ARTICLE - BLOOD PRESSURE

- **Nominal** - elevated or not elevated
- **Ordinal** - hypotensive, normotensive, hypertensive
- **Continuous** - systolic BP in millimeters of mercury
- Binary → Binomial
Assumption → Logistic regression → Log Odds
- Multinomial → Proportional Odds Model → Log Odds
- Normal → Regression

DS ERROR #3

Dividing Continuous Data
into Ordinal Categories w/o
Explaining Why or How the
Categories Were Created

Reduces precision. If not explained, it may appear that the choices were bias.

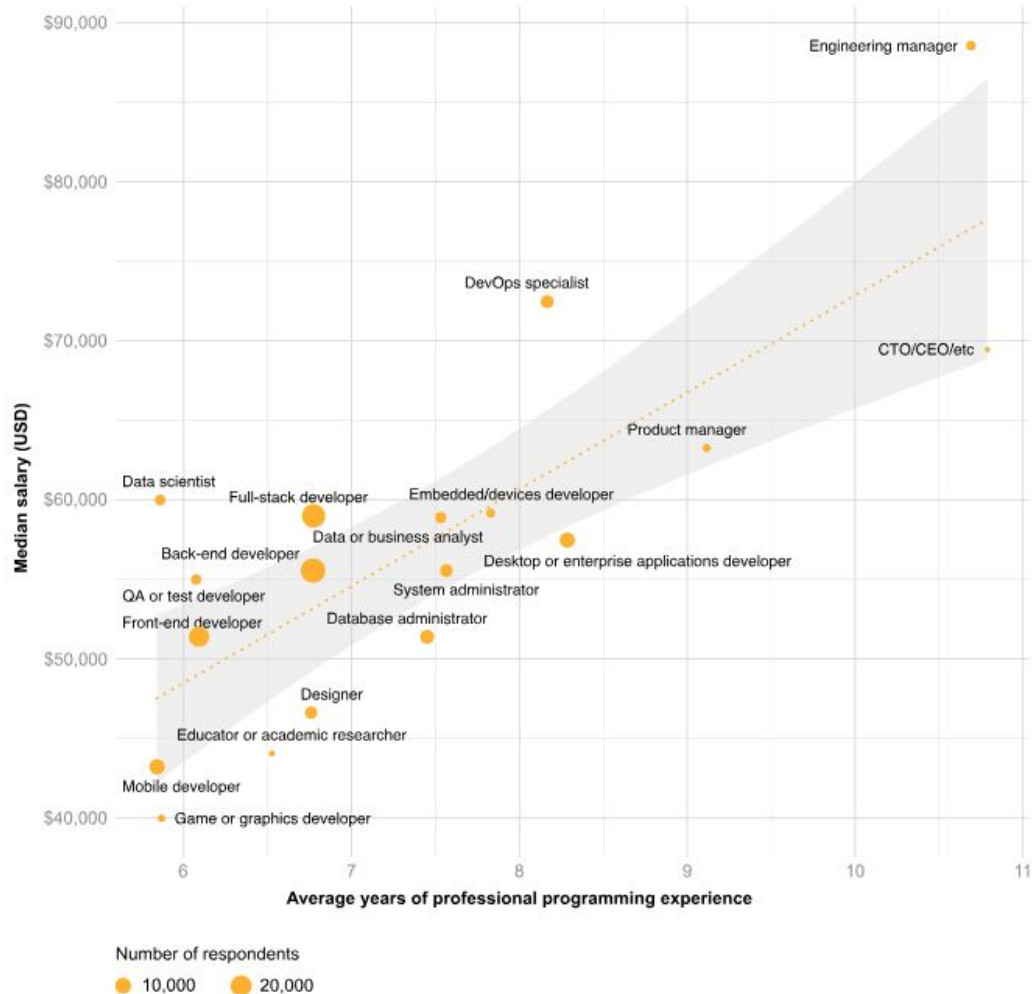
Example: taking continuous systolic blood pressure and reduce it to binary (1 for ≥ 140 mmHg to denote high blood pressure and 0 otherwise)

DS ERROR #4

Using the Mean and Standard
Deviation to Describe
Continuous Data That Are Not
Normally Distributed

Most Biologic data is not
normally distributed.





StackOverflow 2018

Descriptive Graph for Median Salary as it relates to Average years of professional programming experience.

Median Salary was the correct thing to do (right-skewed).

Weight of each data point shown.

Average years of professional programming experience - no report of how they calculated that (DS Error #3).

DS ERROR #5

Using the Standard
Error of the Mean
(SEM) As a Descriptive
Statistic

SEM is always less than
the SD.

Results may appear more
precise than they actually
are (Misleading).

Confidence Level (C): 95



Sample Size (n): 30



SAMPLE

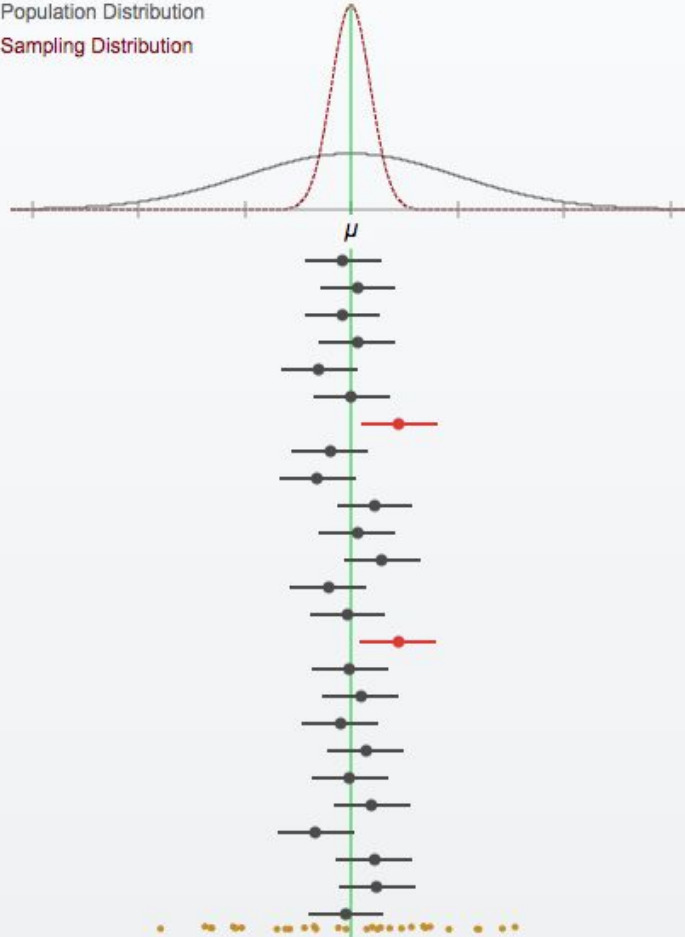
SAMPLE 25

Hit: 36 Total: 38

Percent hit: 95%

RESET

— Population Distribution
- - - Sampling Distribution



Descriptive Stats

- Mean + SD gives us a good idea on the distribution of X variable

Inferential Stats

- Purpose is to estimate the *true mean* using the *sample mean*.
- The sampling distribution by CLT :

$$\bar{X} \sim \text{approx. } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

IS ERROR #6

Reporting Only P
Values for Results

P-values and Significance
don't say much.

Instead we need an
estimate and confidence
interval in order to
answer the scientific
question + assess the
level of precision.

EXAMPLE STATEMENTS

The effect of Treatment A on reducing the average number of cigarettes smoked for patients per day was statistically significant ($P < 0.05$).



The average number of cigarettes smoked per day for patients in Treatment A dropped from 15 to 5 ($P=0.03$).



Treatment A decreased the average number of cigarettes smoked per day by patients by a mean of 10, from 15 to 5 (95% CI = 2 to 8).



IS ERROR #7

Not Confirming That
the Assumptions of
Statistical Tests
Were Met

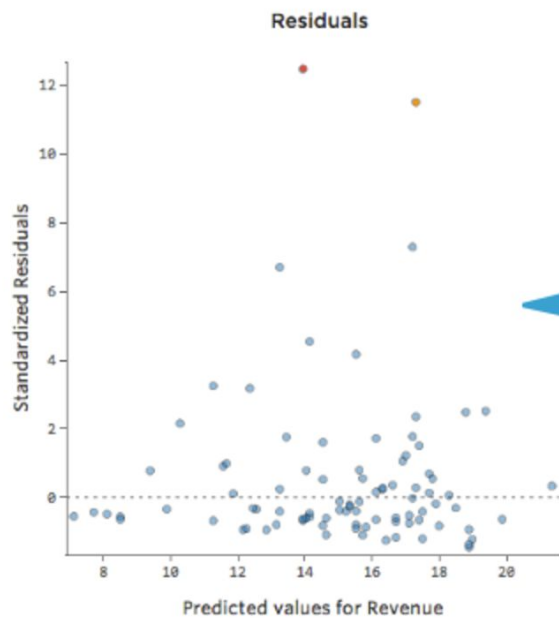
You must do diagnostic tests in order to make sure your assumptions are reasonable + your model is appropriate.

Example: Analysis of Residuals for linear regression.

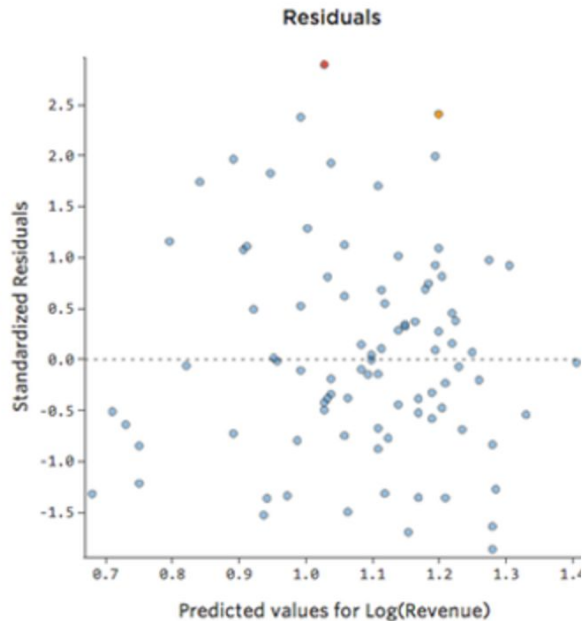


$$\text{Revenue} = .99 * \text{Temperature} - 11.5$$

$$\text{Log(Revenue)} = 0.05 * \text{Temperature} - 0.27$$



r-squared = 0.03



r-squared = 0.10

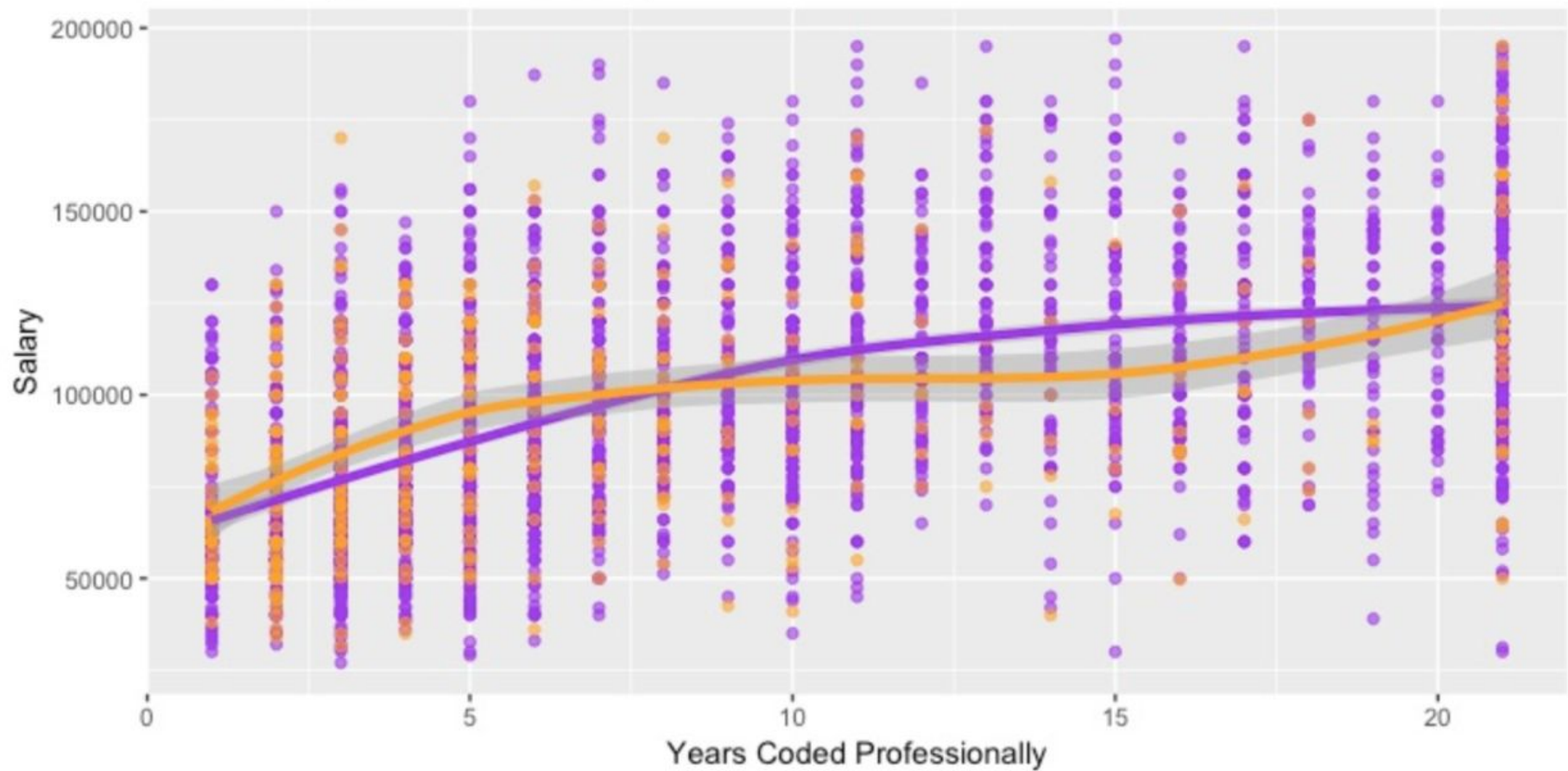
IS ERROR #8

Interpreting
Non-statistically
Significant Results As
“Negative” When They Are, in
Fact, Inconclusive

“Absence of proof is **not**
proof of absence.”

Failing to reject the
null. Doesn't mean the
null is true.

Mean Plot (Male vs. Non-Male)



IS ERROR #9

Not Reporting Whether or
How Adjustments Were Made
for Multiple Hypothesis
Tests

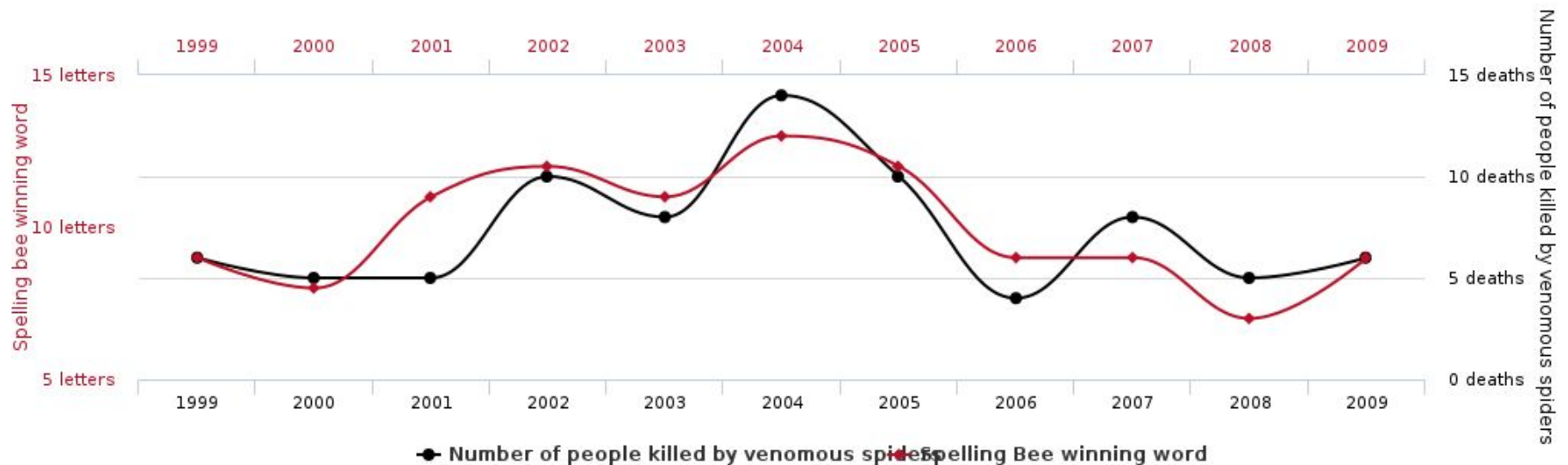
What adjustments were
made?

Example : Bonferroni
correction for P-Value.

Letters in winning word of Scripps National Spelling Bee

correlates with

Number of people killed by venomous spiders



tylervigen.com

THE MORE COMPARISONS MADE, THE MORE LIKELY **BAD** INFERENCE COULD BE MADE

IS ERROR #10

Confusing Statistical
Significance with
Biologic Importance

Basically, once you find something statistically significant, you will need to put it into context. In context, it may not be of importance.
