Statistics Without Probability (SWOP) - A New Paradigm

Dr Mithilesh Dronavalli dr.mit@me.com

Editor:

Abstract

Statistics Without Probability (SWOP)prescribes to the paradigm that each sample of patients is unique, forming its own population. Statistics without probability arises from the inductive argument of Analogy. By analogy, if a patient is sufficiently similar to the patients in the sample on who the study was done then that patient will also receive the benefit that the sample received.

Statistics without probability theory discards the need for probability distributions, standard errors, confidence intervals and P values. Statistics without probability does not assume the axioms of probability.

Point estimation is based on least squares. Two methods are proposed each for Hypothesis testing and interval estimation based on the influence statistic, which is the change in effect size after excluding a data point.

We show that interval estimation and hypothesis testing values do not contract with increasing sample size as seen in frequentist estimation, thereby overcoming the Jeffreys-Lindley Paradox. We also show that interval estimates contract with decreasing noise to signal ratio as they should.

SWOP can also carry out adjusting for confounders via the Corrected Treatment Effect (CTE) and develop and assess accuracy of Prediction modelling via the Standardised Mean Residual (SMR).

Keywords: Distribution-free Statistics, Jeffreys-Lindley Paradox, Sample Size invariant estimators, New Statistical Paradigm

1. Introduction

Recall from the philosophy of statistical science according to the frequentist perspective, that a sample is from a population and a patient is part of that population, that is the argument of *Generalisability* and *Statistical Syllogism*. From a Statistics Without Probability (SWoP) perspective a patient is sufficiently similar to the sample; that is the argument of Analogy. Both principles are present in inductive philosophy. Neither is more powerful, neither is more concrete and they both talk of likelihoods and increased or decreased gradients of effect sizes. Neither Generalisability (and Statistical Syllogisms) or Analogy are deductive arguments in that the outcomes are fixed and certain.

The inductive argument from analogy is as follows. A sample of patients shares some inclusion and exclusion criteria or some other set of properties that define the sample. For example, those over 55 years of age who do not smoke and do not have diabetes. These features of the sample are usually true for all of the members of that sample. A drug, say for heart disease is tested on this sample via a randomised control trial and is found to be effective in preventing heart attacks. Then, another individual with the same properties

(over 55, not diabetic, not a smoker), alive closely after that trial would benefit from that drug to prevent heart attacks. This is what happens in clinical medicine. That is if an intervention works for patient group X and my patient is similar to patient group X then the intervention would work for my patient.

One can argue that by analogy since patient "X" is sufficiently similar to the patients in the sample on who the RCT was done then patient "X" will also receive the benefit that the sample received.

So why then should we use SWOP theory over frequentest theory? With SWOP theory we discard the need for probability distributions, standard errors and confidence intervals and P values. We do not even need the axioms of probability for SWOP to work. We do not need to assume that any event has an apriori probability. SWoP theory has its own set of point estimation, interval estimation and hypothesis testing.

This paradigm removes the need for knowing probability distributions and standard deviations and standard errors in statistics.

This also means we don't have to worry about the independence of errors when calculating effects for longitudinal, correlated, multilevel and clustered data as we don't use P values or confidence intervals or standard errors in this paradigm of statistics without probability.

Also, a sample size determination method is discussed based on assessing the fluctuations of the point estimate and its associated statistics as the sample size increases using real data or simulations.

Furthermore, novel techniques for adjusting for single and multiple confounders under the SWoP paradigm are discussed along with prediction techniques for SWoP.

This makes Statistics Without Probability a fully fledged paradigm in Statistics.

The "Statistics Without Probability" paradigm removes the need for probability in statistics.

2. Methods

2.1 Point Estimation using Least Squares

SWoP uses least squares estimation (Legendre, 1805) for point estimation. Least Squares Estimation uses Calculus and not probability. Methods like maximum likelihood use Probability, but the least squares method does not.

This method of point estimation does not depend on the probability distribution of the outcome.

Because we are not concerned with the errors, or the distribution of the errors, or the variance, or the standard error or even the P value: The assumption that independence of error is not relevant.

By the above logic: The assumption that errors are independent of the predictor variable is not relevant.

Also by the above logic: The assumption that the errors are normally distributed does not hold. As we are not concerned with the errors because we are not calculating a P value.

Multicollinearity only effects the precision of the estimator. Precision again is concerned with the variance. If we do not need a variance as we do not need P values then: Multi-

collinearity or X having full column rank (i.e. any dependent variable must not be a linear combination of other dependent variables) is not relevant.

The assumption of linearity does hold if you use $y = mx + b$. If this linear relation did not hold you can always use a more appropriate equation to model the relationship as long as the least squares estimator is still valid. The least squares estimator is quite flexible.

This means that point estimation can be used for data that is correlated without any adjustment of the error terms because we are not calculating P values or 95% Confidence Intervals.

Therefore least squares method also works under the paradigm of "Statistics without Probability" for correlated data.

Similarly, logistic regression can be done by transforming the binary outcome using a logit transformation and reverting back to least squares regression.

To replace Poisson regression the data can be transformed into a dataset suitable for logistic regression with a binary outcome instead of a rate outcome. This also circumvents the need for probability.

Also for survival analysis; an application of multivariate least squares estimation with 2 outcome variables (follow-up time and occurrence of an event) can be used. A more simple method would be just to transform the survival analysis data set into a longitudinal dataset and analyzed using logistic regression with the event as a binary outcome.

Longitudinal data can be analyzed without worrying about correlated error terms as interval estimation in SWOP does not use standard error as it has its own unique method of interval estimation and hypothesis testing.

In conclusion besides the linearity assumption, there are no other assumptions for using least squares in the paradigm of "Statistics without probability". Note that the linearity assumption can be overcome by using a more appropriate equation to model the relationship between the predictor x and the outcome y.

2.2 Interval Estimation by exclusion of data

Interval Estimation can be done without assuming underlying probability distributions for the data. For example in Statistics we have to usually assume that the error term for the regression model is normally distributed to develop a confidence interval for any regression coefficient m in $y = mx + b$.

The interpretation for interval estimation in SWOP is this: In the discovery of knowledge based on observation effect sizes can be relatively loose or relatively tight. When measuring the gravitational pull of the Earth in physics we may have a relatively tight effect size. In contrast measuring effect sizes in certain topics in the humanities and social sciences we may have a relatively loose effect size. This is just for example. If the interval overlaps the null effect there may be little or no effect to quantify in the study. This interpretation differs from the frequentist interpretation that in 20 such similar studies, 19 studies will have a 95% confidence interval that contains the population effect size. This is a very important difference between the interpretation of SWOP and frequentist confidence intervals.

This method bypasses such a normality assumption by generating intervals based on each point (x, y) and its influence on m.

In $Y_i = mX_i + b$, Y_i and X_i are vectors of data with Real Entries. m is the regression coefficient and is calculated via least squares. This is well known as simple linear regression. $(x, y)_i$ denotes the i_{th} point of the data.

$$inf_i = m_U - m_{U-i} \tag{1}$$

In equation (1) inf_i is the influence of the point i in calculating m. This is defined as the following: m_U is the effect size m calculated for all points in the dataset, where U denotes the entire dataset. m_{U-i} is m calculated without the i_{th} data point.

We then sort the data by values of descending inf_i. The set of data comprising of the top 15% of inf_i is labeled as "UL" for upper limit and the bottom 15% of $inf_i$is labeled as "LL" or lower limit. By removing the UL set of data we calculate the total upper influence statistic

$$infTot_{UL} = \sum_{i=1}^{i=U-UL} (m_U - m_{U-i})$$

Excluding UL data subset.

$$infTot_{LL} = \sum_{i=LL}^{i=U} (m_U - m_{U-i})$$

Excluding LL data subset. The confidence intervals by excluding the influence from a 15% of data are:

$$upd_{15} = m_U + InfTot_{UL}$$

$$lpd_{15} = m_U + infTOT_{LL}$$

This method does not rely on the probability distribution of any variable X_i or Y_i. It does not rely on the equation between X_i or Y_i. It does not even require for the i subjects to be independent or uncorrelated. It just requires that m can be calculated reproducibly using some method like least squares.

Using this method we can generate confidence intervals at different levels from excluding less than 1% to anywhere below 50% of either end of data after sorting for inf_i.

2.3 Interval Estimation by exclusion of effect

By summating the positive and negative components of the influence statistic inf_i we get the Positive Sum Coefficient and the Negative Sum Coefficient.

$PosSumCoef = \sum (m_U - m_{U-i})$: for i where $m_U - m_{U-i} > 0$

$NegSumCoef = \sum (m_U - m_{U-i})$: for i where $m_U - m_{U-i} < 0$

The Upper and Lower Confidence Intervals by a 30% percentage effect are as below:

$$upe30coef = m_U + PosSumCoef \times 0.3$$

$$lpe30coef = m_U + NegSumCoef \times 0.3$$

2.4 Hypothesis Testing via q-percent and Q-value

Hypothesis testing can be done without the use of probability. Firstly we sort data by the influence statistic inf_i and remove the minimum $q\%$ data from either the highest or lowest extremity of data till the effect size $m = 0$ from $y = mx + b$.

$$inf_i = m_U - m_{U-i} \tag{2}$$

m_U is the effect size using all of the data. m_{u-i} is the effect size of all of the data minus is the i_{th} data point. inf_i represents the change in effect size due to the i_{th} data point The minimum percent of data to exclude so that the effect size reaches or passes zero is the $q\%$ of data. Note to determine the $q\%$ of data is rather cumbersome. This is because the most influential data point in a certain direction has to be determined and deleted one at a time. So potentially we have to recalculate inf_i a total of $n/2$ (where n is the sample size) at most to get the effect size to reach the null hypothesis in the fastest way. This way we are sure that we get the minimum percent of data to exclude to get a null-effect.

Because data can contribute to the effect size in an unequal way, it would be inaccurate just to report the minimum $q\%$ of data to be excluded so that the effect size is zero.

Hence we have the following statistic based on deriving interval estimation by excluding effect:

Recall that: $PosSumCoef = \sum (m_U - m_{U-i})$: for i where $m_U - m_{U-i} > 0$

$NegSumCoef = \sum (m_U - m_{U-i})$: for i where $m_U - m_{U-i} < 0$

The Upper and Lower Confidence Intervals by a 30% percentage effect are as below:

$$upe30coef = m_U + PosSumCoef \times 0.3$$

$$lpe30coef = m_U + NegSumCoef \times 0.3$$

Therefore the Q-value or percent of effect to be excluded is derived as below:

$Q - value = \frac{-m_u}{\sum (m_U - m_{U-i})}$: for i where: $m_U - m_{U-i} > 0$ if $m_U < 0$

$Q - value = \frac{-m_u}{\sum (m_U - m_{U-i})}$: for i where: $m_U - m_{U-i} < 0$ if $m_U > 0$ A large Q-value indicates a real result not due to chance with a high signal to noise ratio. Q-values are real values ranging from zero to +ve infinity.

2.5 Sample Size Determination

The following method can be used to determine whether more data is needed and this method works as data is collected rather than before data collection.

Sample Size-Method 1. Randomly pick subsets of data of increasing size. All possible combinations of subsets of data of increasing size can be done as well.

2. Measure fluctuations in the point estimate.

3. The point estimate and other statistics mentioned in this paper should increase in precision as the subsets of data increase in size.

4. Once the change in fluctuations decreases to below a cutoff as the sample size of subsets increase the sample size of total data is sufficient.

5. If fluctuations of point estimate do not settle then more data is needed.

You can repeat this method using simulated data with similar effect sizes to your data to find when the sample size is large enough so that the point estimator, interval estimator and hypothesis testing values does not fluctuate meaningfully.

Beware of selection bias

Note: Statistics Without Probability overcomes the Jeffreys-Lindley Paradox

The Jeffreys-Lindley Paradox states that as sample size increases greatly the P-value is always significant given any alpha. However as sample size increases greatly the Bayesian Factor becomes non-significant i.e. greater than 1. (Shafer, 1982)

In SWOP increased sample size is always better than less sample size provided the data is of sufficient quality with regards to measurement error and selection bias, but each additional data point leads to a reduced return in accuracy once a sufficient sample size is reached. However, the overall accuracy of the hypothesis test only increases with increasing sample size defeating the Jeffreys-Lindley Paradox. (Shafer, 1982)

2.6 Influence of Extremities

The effect size m can be considered unstable when much of the pushing or pulling of m come from the extremities of data. That is when a collection of extreme data points have a high influence on m in an equation such as $y = mx + b$ for that effect size is ***Unstable***. The Influence of Extremities Factor has been devised to measure the instability of the effect size m.

In $Y_i = mX_i + b$, Y_i and X_i are vectors of data with Real Entries. m is the regression coefficient and is calculated via least squares. This is termed as simple linear regression. $(x, y)_i$ denotes the i_{th} point of the data.

$$inf_i = m_U - m_{U-i} \tag{3}$$

In equation (3) inf_i is the influence of the point i in calculating m. This is defined as the following: m_U is m with all points in the dataset, where U denotes the entire dataset. m_{U-i} is m calculated without the i_{th} data point. We then sort the data by values of descending inf_i. The set of data comprising of the top 10% of inf_i is labeled as "UL" for upper limit and the bottom 10% of inf_i is labeled as "LL" or lower limit. By removing the UL set of data we calculate m_{U-UL}. By removing the LL set of data we calculate m_{U-LL}.

$$m_{U-UL(10\%)} \leq m_U \leq m_{U-LL(10\%)} \tag{4}$$

By removing data that is pushing up "m" (i.e. high inf_i) we get $m_{U-UL(10\%)}$ which is lower than m_U. By removing data that is pulling down "m" (i.e.low inf_i) we get $m_{U-LL(10\%)}$ which is higher than m_U.

Influence of Extremities Factor at 10% intervals would be:

$$IEF_{(n=10\%)} = \frac{(m_U - m_{U-UL(10\%)}) + (m_{U-LL(10\%)} - m_U)}{m_U} \tag{5}$$

$$IEF_{(n\%)} = \frac{m_{U-LL(n\%)} + m_{U-UL(n\%)}}{m_U} \tag{6}$$

A low IEF indicates a stable effect size m. A high IEF indicates an unstable effect size m. By stable, the value of m is not overly influence by a few data points and by unstable, the value of m is overly influenced by a few data points. Datasets that are stable are more representative of the entire cohort and datasets that are unstable are likely not representative of the entire cohort.

2.7 Adjusting in SWoP via the Corrected Treatment Effect

2.7.1 Univariate Corrected Treatment Effect

Corrected Treatment Effect (CTE) is an alternative way for adjusting for confounders without using multivariable regression. Here confounders are variables that predict both the predictor variable x and the outcome variable y. Multivariable regression for this purpose is to be avoided as the treatment effect is conditioned by the multivariable fit of that particular sample. This is problematic when the sample is not representative of the population. CTE on the other hand calculates a factor for each confounder separately. This factor when multiplied by the treatment effect gives the CTE.

CTE is to be interpreted as the effect of x on y had the confounder been balanced for all values of X. When the confounder is balanced for all values of X then Z is no longer a confounder and this is akin to balancing confounding in a randomized control trial or a causal analysis.

The CTE is the treatment effect adjusted for that confounder. These factors of confounding for particular samples are actually values that can be used in other datasets where that confounder has not been measured. These factors can be compared between different datasets to create a relative measure for confounding between datasets. This is where CTE is particularly useful and valuable.

The CTE can also be used in other paradigms of statistics, including frequentist, Bayesian, bootstrapping based statistics. All that is required is to calculate a standard error for the CTE and a distribution with hypothesis testing.

$$y = m_x x + b_x \tag{7}$$

$$\hat{X} = m_z z + b_z \tag{8}$$

$$y = m_x \hat{X} + b_x \tag{9}$$

Here we use $\hat{x}$ to approximate x. If z is not a predictor of x then we do not need to adjust for it. If the relationship of z onto x is not linear as per equation (8) then adaptation of the Corrected Treatment Effect technique is requires and can be generalised with simple algebraic manipulation.

$$y = m_x(m_z z + b_z) + b_x$$

$$\frac{y - b_x}{m_z z + b_z} = m_x$$

To calculate the effect of predictor x (independent of confounder z) on outcome y, we force $m_z = 0$ on the left hand side and to balance this effect we multiply m_x with a_z which

is a correcting factor for the confounding imparted by z on x. Then we have:

$$\frac{(y - b_x)}{b_z} = a_z m_x$$

$$\frac{(y - b_x)}{m_x b_z} = a_z$$

$$\frac{x m_x}{m_x b_z} = a_z$$

Note here $y - b_x$ is $m_x x$ from (7) not $m_x \hat{X}$.

$$\frac{x}{b_z} = a_z \tag{10}$$

$$y_{x \perp z} = a_z m_x x + b_x \tag{11}$$

For estimation purposes:

$$a_z = \frac{\bar{x}}{b_z} \tag{12}$$

2.7.2 Multivariable Product Method: Corrected Treatment Effect

To generalize this to the case of multiple confounders.

$$\hat{X} = m_q q + bq \tag{13}$$

$$y_{x \perp z} = a_z m_x (m_q q + b_q) + b_x$$

$$\frac{y_{x \perp z} - b_x}{m_q q + b_q} = a_z m_x$$

To calculate the effect of predictor x (independent of confounder z and q) on outcome y, we force $m_q = 0$ on the left hand side and to balance this effect we multiply m_x with a_z to correct confounding by z on x (as above) and a_q to correct confounding by as above.

Then we have:

$$\frac{y_{x \perp z} - b_x}{b_q} = a_{qz} m_x$$

Also

$$a_q \times a_z = a_{qz}$$

$$\frac{y_{x \perp z} - b_x}{m_x b_q} = a_{qz}$$

$$\frac{a_z m_x x + b_x - b_x}{m_x b_q} = a_{qz}$$

$$\frac{a_z m_x x}{m_x b_q} = a_{qz}$$

$$\frac{a_z}{b_q} = a_{qz}$$

From (3) we have:

$$\frac{x\frac{x}{b_z}}{b_q} = a_{qz}$$

$$\frac{x^2}{b_z b_q} = a_{qz} \tag{14}$$

For estimation purposes:

$$a_{qz} = \frac{\bar{x}^2}{b_z b_q} \tag{15}$$

In a similar fashion this can be generalised to the case of many confounders. Note that so far there is always one outcome and one predictor variable. All variables thus far have to be continuous or binary or discrete.

x is from the data. b_z is and b_q are calculated by applying the least squares method to equations (7) and (13)

2.7.3 Multivariable Regression Method: Corrected Treatment Effect

There is a multivariable regression version of the Corrected Treatment Effect, which is surprisingly simple. The method regresses the x the dependent variable on all confounders. It can strictly be argued that any variable that predicts x can be used as a confounder in this setting as we are just getting the best prediction of x via $\hat{X}$

$$y = m_x x + b_x \tag{16}$$

$$\hat{X} = m_{z1}z_1 + m_{z2}z_2 + m_{z3}z3 + b_{z123} \tag{17}$$

$$y = m_x \hat{X} + b_x$$

$$\frac{y - b_x}{\hat{X}} = m_x$$

$$\frac{m_x x}{\hat{X}} = m_x$$

$$\frac{m_x x}{m_{z1}z_1 + m_{z2}z_2 + m_{z3}z3 + b_{z123}} = m_x \tag{18}$$

In equation (18) we force m_{z1} and m_{z2} and m_{z3} etc... to equal 0 including any interaction variable coefficients on the left hand side. To balance the equation we multiply the right hand side with a_{z123}

$$\frac{m_x x}{b_{z123}} = a_{z123} m_x$$

$$\frac{x}{b_{z123}} = a_{z123} \tag{19}$$

Therefore the corrected treatment effect for the dependent variable x onto y is a_{z123}.

The crude effect of x onto y is m_x from equation (16). The unconfounded effect of x onto y correcting for the confounders in equation (17) is a_{z123} multiplied by m_x.

$$y_{x \perp z1,z2,z3} = a_{z123} m_x x + b_x \tag{20}$$

2.7.4 Note regarding CTE

There are multiple methods illustrated in this paper for adjusting for confounding. However for the method of Corrected Treatment Effect to work best, $\hat{X}$ has to best approximate x. To do this we have to model x and test the fit of the model using Prediction modeling and testing which is discussed in the next sub-section.This reasoning would make the multivariable regression method of the corrected treatment effect (2.6.3) the most effective method of adjusting for confounding. Since it is difficult to model $\hat{X}$ with one variable. Note also that the variables selected are confounders and here that means they predict both x and y, that is they predict both the predictor variable x and the outcome variable y.

2.8 Logistic Regression and CTE

Logistic regression is used when the outcome is dichotomous or binary (1 or 0):

Logistic regression is just a logit transformation of the outcome variable and the resulting regression remains linear so we can still use least squares estimation.

When adjusting for confounder in logistic regression an interesting concept arises. The regression equation has a logit outcome and if the predictor is binary the adjustment equation from the corrected treatment effect also has a logit outcome.

It can be derived that the Corrected Treatment Effect as an odds ratio OR_{CTE} is as follows:

$$OR_{CTE} = Exp(Coef * \frac{x}{inverse - logit(b_z)})$$

The inverse logit function is: $\frac{e^{b_z}}{1+e^{b_z}}$

Exp means the term in brackets is exponentiated.

Coef means coefficient. The coefficienthere is the coefficient of the predictor onto the outcome in the regression equation $logit(y) = m_x x + b_x$ where m is the coefficient.

x is the prevalence of exposure, which is merely the average of all x-values.

b_z is the constant of the adjustment equation: $logit(x) = m_z z + b_z$

2.8.1 Non-linear CTE

Where the association equation ($y = m_x x + bx$) is non-linear the following holds for the corrected treatment effect (a_z)in the various non-linear models listed below.

Polynomials - Quadratic, Cubic, Quartic Here for brevity we discuss the quadratic polynomial for the association equation in the form of:

$$y = m_{x1}x^2 + m_{x2}x + b_x$$

If we use the corrected treatment effect to change the leading term we can multiply m_{x1} by a factor of a_z, where:

$$a_z = \frac{\bar{x}^2 + \frac{m_{x2}}{m_{x1}}(x - b_z)}{(b_z)^2}$$

Logarithmic

$$y = a + b \ln x$$

Where k is the treatment effect and a_Z is the corrected treatment effect. Then $a_z = \frac{\ln \bar{x}}{\ln b_z}$.

Logistic

$$y = \frac{a}{1 + be^{kx}}$$

Where k is the treatment effect and a_Z is the corrected treatment effect. Then $a_z = \frac{\bar{x}}{b_z}$.

Exponential Decay - Decreasing Form

$$y = Ce^{-kx}$$

Where k is the treatment effect and a_Z is the corrected treatment effect. Then $a_z = \frac{\bar{x}}{b_z}$.

Exponential Decay - Increasing Form

$$y = C(1 - e^{-kx})$$

Where k is the treatment effect and a_Z is the corrected treatment effect. Then $a_z = \frac{\bar{x}}{b_z}$.

Exponential Growth

$$y = Ce^{kx}$$

Where k is the treatment effect and a_Z is the corrected treatment effect. Then $a_z = \frac{\bar{x}}{b_z}$.

Gaussian

$$y = ae^{-(x-c)/b)^2}$$

The Gaussian equation is different because here we are looking for a shift in the center of the Gaussian Curve rather than a change in the gradient as we have seen previously. So the treatment effect is c in the above equation and corrected treatment is a_z, where a_z is added to c to shift the center of the Gaussian curve. Then $a_z = \bar{x} + b_z$.

It is left to the reader to derive these results and check the solutions as the exercise in doing so is quite exhilarating.

2.9 Prediction modeling and testing without Probability

There are two ways we can use statistics. One is to determine the effect size of x onto y. The other is to predict y. We can predict y using multivariate prediction modeling. Least squares can handle multivariate linear and non-linear prediction modeling as well. Here the testing of the accuracy of prediction models will be discussed with out resorting to probability.

The observed outcome vector is y_i. The predicted outcome vector using the chosen prediction model is $\hat{y}_i$ Mean Residual:

$$\sum_{i=1}^{i=n} \frac{|y_i - \hat{y}_i|}{n} \tag{21}$$

To standardize the mean residual we divide by

$$\sum_{i=1}^{i=n} \frac{|y_i|}{n} \times \frac{1}{n}$$

Standardized Mean Residual(SMR):

$$\sum_{i=1}^{i=n} \frac{|y_i - \hat{y_i}|}{n} \div \sum_{i=1}^{i=n} \frac{|y_i|}{n^2} \tag{22}$$

$$\sum_{i=1}^{i=n} \frac{|y_i - \hat{y_i}|}{n|y_i|} \tag{23}$$

The SMR is comparable between datasets and in different fields of science as it is standardized by the mean. Note that the SMR also does not depend on the sample size (n). Small SMR means high accuracy and large SMR means low accuracy

3. Verifications

Verification of simulations were done for generated datasets of 50, 100 and 500. Coefficients used to generate the data ranged from (-10,+10) with increments of 1 and including 0.5 and 0.1. Error sizes included: (0.5,5,10,50,100,200,300,400,500). Constants or the y-intercept for the model was (-1,0,+1). Signal:noise ratio was defined as the fitted coefficient from the simple linear regression model divided by the error sizes.

3.1 Verification of Interval Estimation by percent data and percent effect

From Figure 1 interval estimation by percent effect is approximately equivalent regardless of sample size for samples of 50, 100 and 500. The association of length by percent effect and sample size returned a coefficient 0.0003 (95%CI-0.007:0.007). Also the length decreases with decreasing error sizes using various coefficients and constants. Length also decreased with models of increasing fitted coefficient (signal) (coefficient: 0.183; 95%CI: 0.008:0.358).

From Figure 2 the length of the percent data Confidence intervals follow a similar association to the percent effect confidence intervals with error. However the gradient of increasing length for increasing error is twice that for the length of percent data confidence intervals. This makes the percent data confidence intervals more conservative. Note that we are using 30% of data and 15% of effect in these simulations to calculate these results and this may explain why the gradient is double for percent of data.

3.2 Verification of Q-value and q-percent Hypothesis Testing Method

In Figure 3 below it can be seen that there is an increasing linear relationship between the Q-value and the signal:noise ratio that is symmetric around zero-signal. The Q-value from our result is independent of the fitted coefficient in the model (coef -0.04 95%CI: -0.1:+.1) and sample size (coef -0.00001 95%CI: -0.003:0.002).

Our simulations indicated that q-percent was useful and discriminatory only for datasets in the "Goldilocks zone". It lacks discriminatory ability for larger effect sizes (slope$>$ 5) where the sample size becomes larger (n=500).

3.3 Verification of the Corrected Treatment Effect

Dataset: Smoking in pregnancy and birth weight of infants Hypothesis: That mothers who smoke have smaller babies (less weight).

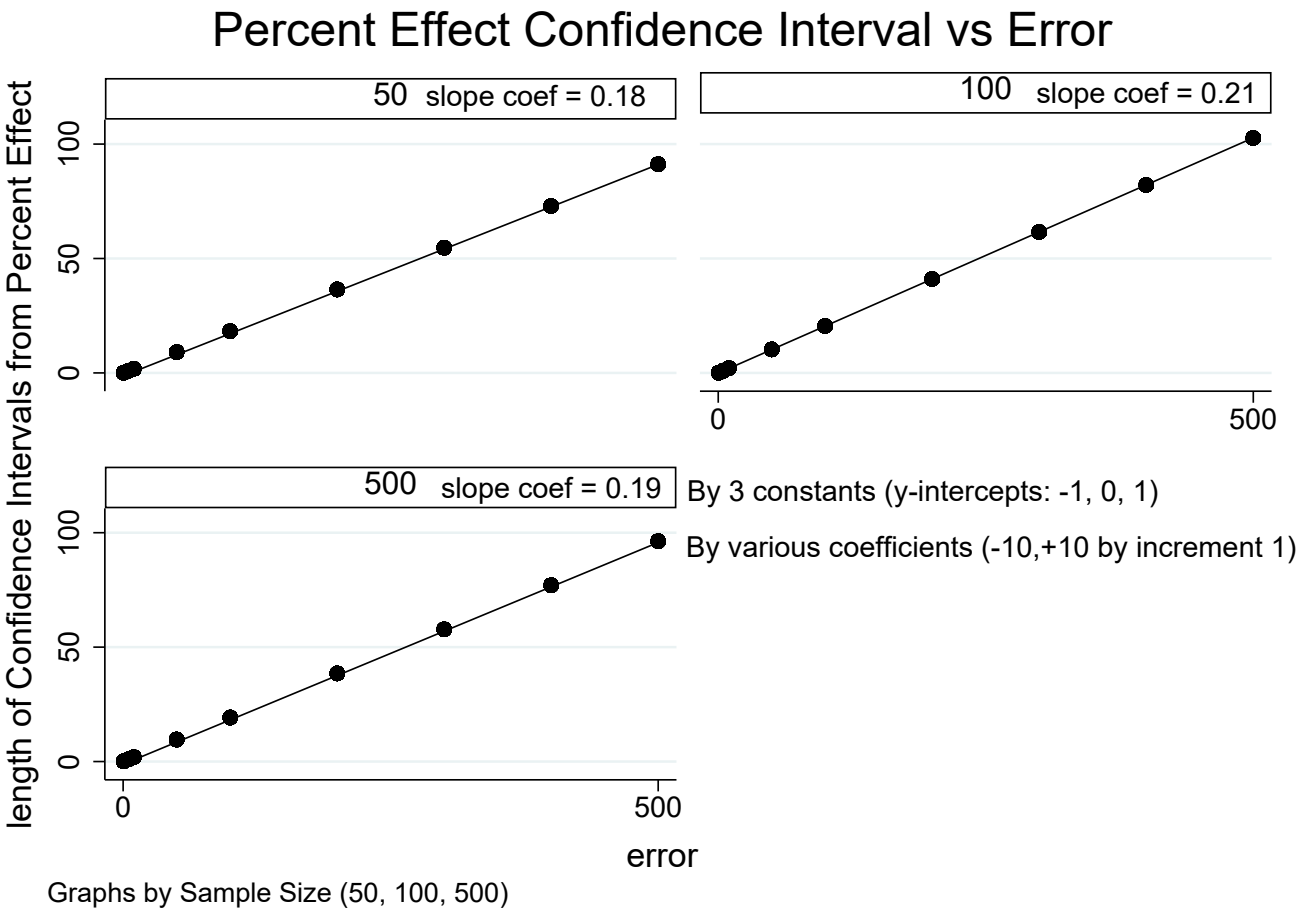

Figure 1: Confidence Interval Lengths for 30% of effect excluded by various samples sizes and signal:noise ratios.

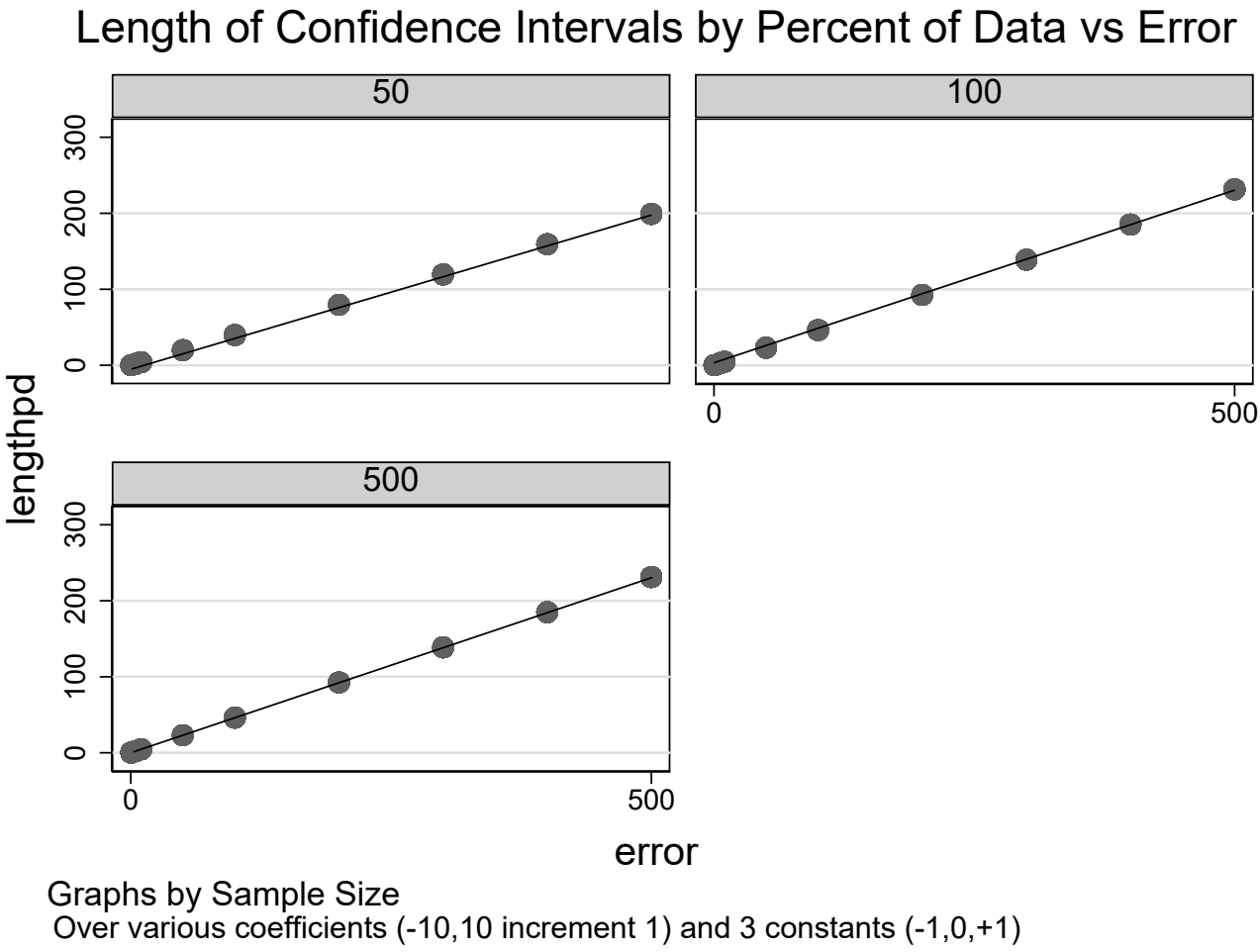

Figure 2: Confidence Interval Lengths for 30% of effect excluded by various samples sizes and signal:noise ratios.

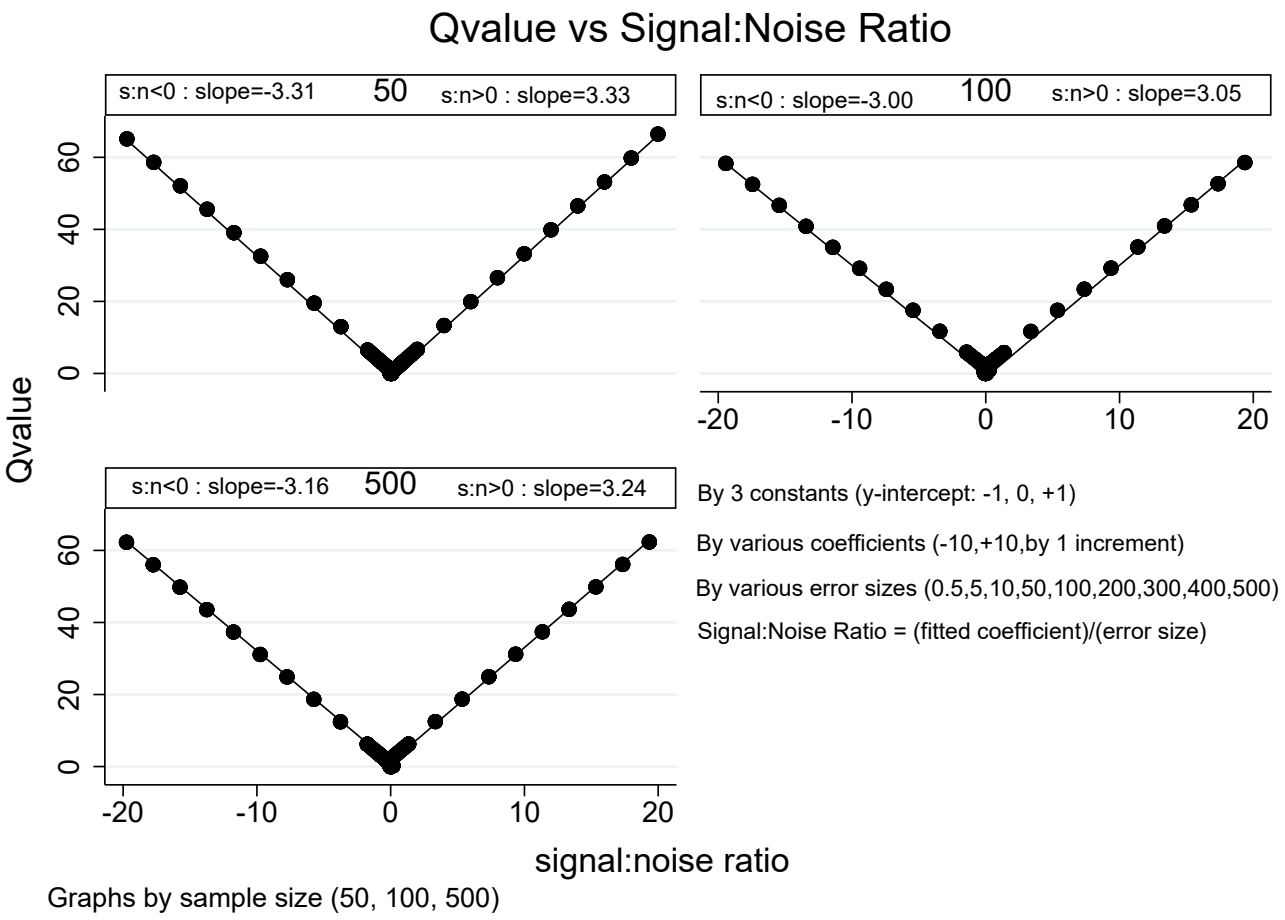

Figure 3: Q-value increases linearly with increasing signal:noise ratio and stay constant in the absence of signal.

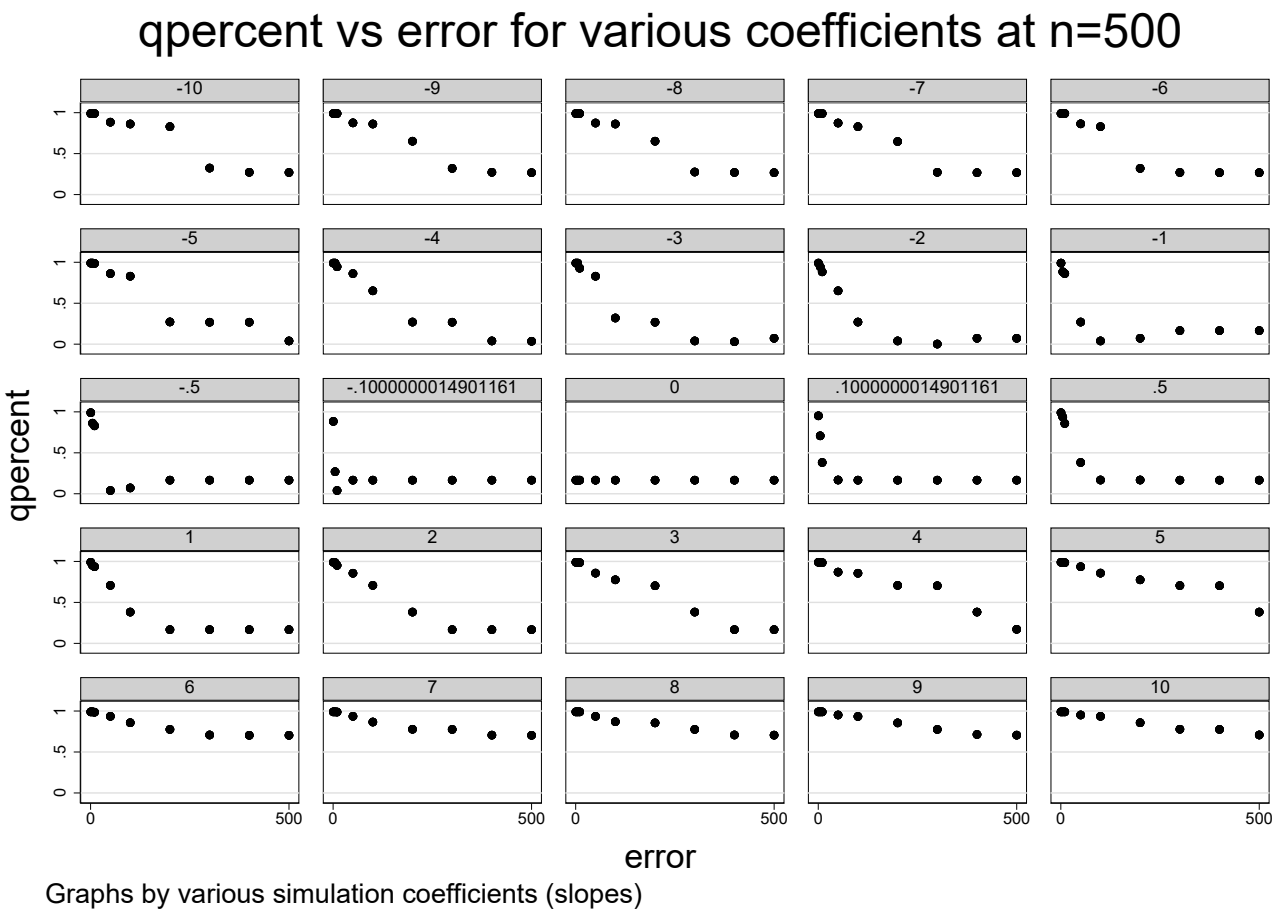

Figure 4: qpercent vs error for various simulation coefficients (slopes) at a sample size of 500.

Outcome: Birth Weight

Predictor: Smoking (Yes or No)

Confounders: Married at birth, Marriage Age, First Baby or Not, Years of Education at Marriage.

Obs: n=4,642

Using **Propensity Score Matching and treatment effects estimation**:

Stata Command:

.teffects psmatch (bweight) (mbsmoke mmarried c.mage##c.mage fbaby medu)

The average birth weight if all mothers were to smoke would be 211 grams(P<0.0001; 95%CI(-273.7284, -148.2083)) less than the average that would occur if none of the mothers had smoked. This assumes conditional probability and directed acyclic graph theory which arises from conditional probability.

An example of Corrected Treatment Effect: **multivariable confounding via the product method** using SWOP:

regress bweight mbsmoke -275.2519 coef

x-bar = .22869243

regress mbsmoke mmarried

-.2004752 coeff P<0.0001

.3263989 const

$a_{mmarried}$ = .70065319 = x-bar/$const_{mmarried}$=.22869243/.3263989

regress mbsmoke mmage -.0078884 coeff P<0.0001

.3952054 const

a_{mage} = .57866727 = x-bar/$const_{mage}$ = .22869243/.3952054

regress mbsmoke fbaby -.050231 coeff P<0.0001

.2081257 const

a_{fbaby} = 1.0988188 = x-bar/$const_{fbaby}$ = .22869243/.2081257

regress mbsmoke medu

-.0307855 coeff P<0.0001

.5767809 const

a_{medu} = .39649792 = x-bar/$const_{medu}$ = .22869243/.5767809

$a_{confounding}$=$a_{mmarried} \times a_{mage} \times a_{fbaby} \times a_{medu}$

$a_{confounding}$ = .70065319 × .57866727 × 1.0988188 × .39649792 = 0.17664405

regress bweight mbsmoke -275.2519 coef

m= -275.2519 : crude effect of maternal smoking on birth weight

m* = −275.2519 × 0.176664405 = −48.627213 : adjusted effect of maternal smoking on birth weight adjusted for the above mentioned confounders.

That is mothers who smoke have babies that are 49 grams lighter than mothers who do not smoke when adjusted for the above confounders.

Using the above example in **Corrected Treatment effect and multivariable regression method** via SWOP:

regress bweight mbsmoke

-275.2519 coef

x-bar = .22869243

regress mbsmoke mmarried mage medu fbaby

all variables P<0.0001 except mage P=0.863

const: .575807

$a_{confounding}$ = x-bar/$const_{mv}$ = .22869243/.575807

$a_{confounding}$ = .39716855

m= -275.2519 : crude effect of maternal smoking on birth weight

m* = $-275.2519 \times .39716855 = -109.3214$: adjusted effect of maternal smoking on birth weight adjusted for the same confounders as above.

That is mothers who smoke have babies that are 109 grams lighter than mothers who do not smoke when adjusted for the above confounders.

The multivariable regression method assumes that we have the best prediction of x via $\hat{X}$.

4. Summary of Advantages of Statistics Without Probability

- Does not assume axioms of probability, or probability distributions
- Does not assume existence of a population.
- Sample size required is calculable using simulations or the real data.
- More sample size is always better than less sample size –> Defeats Jeffreys-Lindley Paradox
- No need for standard errors or variances.
- No frequentist P-Values and confidence intervals. Replaced by asymptotically consistent interval estimation and hypothesis testing
- No need for probability based maximum likelihood estimators.
- No need for longitudinal data analysis, analysis of correlated data, multilevel modeling.
- Just one regression function for everything using generalized least squares.
- Hypothesis testing and Interval estimation values that contract with increasing signal to noise ratio, but stays stable across sample size, thus defeating Jeffreys-Lindley Paradox.
- Hypothesis testing (Q-value) is independent of both sample size and coefficient.
- SWoP can adjust for confounders by balancing confounding using the corrected treatment effect similar to an RCT or a causal analysis.
- SWoP can be used for predicting the outcome variable while checking for accuracy of the prediction using the Standardized Mean Residual

5. Conclusion

Statistics without Probability is a comprehensive statistical paradigm for the analysis of data without resorting to Probability. SWoP covers point and interval estimation, hypothesis testing, prediction, adjusting for confounding, sample-size determination. Also we show through simulations that interval estimation and hypothesis testing values are effective statistics.

Further research may focus on applying the principles of SWOP and Corrected Treatment Effect in the setting of differential equations. Also research may be directed on comparing CTE estimates of effect sizes that are measured in both RCTs and observational studies with measured confounding where the studies are carried out on similar populations. This will widen the scope of SWOP theory and establish the Corrected Treatment Effect as a gold standard for adjusting for confounding in observational studies.

Acknowledgments

I thank my close family for encouraging me every step of the way in writing this paper.

SUPPLEMENTARY MATERIAL

R code for SWOP R code file for Statistics containing code to generate simulation data and calculate interval estimation and hypothesis testing for the 2 different methods described in this article. (RSWOP.r)

R code for SWOP Dataset generated from simulation code from RSWOP.r (SWOP-data.csv)

Smoking in Pregnancy vs Birth weight in infants
Data From: `http://www.stata-press.com/data/r15/cattaneo2`
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)

References

Adrien Marie Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot, 1805.

Glenn Shafer. Lindley's paradox. *Journal of the American Statistical Association*, 77(378): 325–334, 1982.