Malaysian Journal of Mathematical Sciences 8(2): 289-309 (2014)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

How to Tell the Truth with Statistics

[1*]A.H.M. Rahmatullah Imon and [2]Keya Das

[1]*Department of Mathematical Sciences, Ball State University, Muncie, IN 47306, USA*

[2]*Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Salna, Gazipur 1706, Bangladesh*

E-mail: imon_ru@yahoo.com

*Corresponding author

ABSTRACT

Statistic, as a subject, does a hugely challenging job. It tries to quantify the uncertainty of this world. Dealing with uncertainty may often result in wrong conclusion. Sometimes it is inevitable but sometimes it is intentional. Some people may use statistics to suppress the fact or just to lie. This type of wrong practices goes with the ethical issues. We also observe that that even when we are trying to give a fair treatment to this subject, lack of knowledge could result in erroneous, funny and nonsense conclusions. In this paper, we would like to discuss these issues seriously. Firstly, we will talk about abuse of statistics in real world problems. Then, we try to find the factors which lead to the origination of nonsense statistics. Our main and foremost objective has been always tell the truth with statistics. This paper is written to focus on the problems and to suggest remedial measures at the same time from the misconception that 'there are three kind of lies: lies, damn lies, and statistics'.

Keywords: Abuse of statistics, misleading and nonsense statistics, randomization, sampling, modelling, inference, Lurking variable and Simpson's Paradox, judgment, diagnostics, multi collinearity, outliers, robust statistics.

1. INTRODUCTION

According to Chernoff and Moses (1959), Statistics is the science of decision making in the face of uncertainty. For this reason there is a risk that decisions based on sophisticated statistical techniques may not hold in reality and few incidences may push back Statisticians and Statistics teachers to the unfair impression that Statistics teaches how to lie with data. To quote the eminent writer Twain (1924), "The remark attributed to Disraeli would often apply with justice and force: 'there are three kind of lies: lies, damn lies, and statistics'." This may be the most unfair quotation ever made against Statistics. The American humorist Esar (1943) commented "Statistics is the only science that enables different experts using the same figures to draw different conclusions." According to Paul Velleman (2008), "Those who believe incorrectly that Statistics is solely a branch of Mathematics (and thus algorithmic), often see the use of judgment of Statistics as evidence that we do indeed manipulate our results." We must realize the fact that Statistics lives on the empirical rather than the theoretical side of science. The availability of high speed computers and statistical software have freed statisticians from the grip of mathematicians to a greater extent, but it has created major problems the other way around, researchers who have a little knowledge about statistical methods can go to a computer and can create lots of senses and even more nonsense with the data. This problem has become so serious that many text books now contain sections on 'Nonsense Statistics'. Conclusions drawn from a study are trust worthy only when appropriate design and correct sampling methods are used. We cannot rely on the results of hypothesis testing unless the validity of all underlying assumptions such as independence, normality and purity of observations (free from outliers) are met. In this paper we will discuss all these issues in a very non-technical fashion with lots of interesting examples showing the abuses of statistics in various areas of research. We also discuss the consequences of using these nonsense statistics in research. But statistics is not meant for it. Our main objective has been always to tell the truth with statistics.

In Section 2, we give some real world examples of the abuse of statistics. This includes misleading and nonsense statistics. We often observe that this problem is caused because of lack of randomization and faulty sampling techniques which are discussed in Section 3. Another source of nonsense statistics is the inappropriate choice of models. The issue of modelling is discussed in Section 4. In Section 5, we discuss how the wrong design of inferential procedure can result in misleading conclusions. The entire inferential procedure is based on a series of standard assumptions. But in real world problems we often see that those assumptions do not hold and deviations from the standard assumption is a huge source of nonsense statistics. In Section 6, we discuss how to apply diagnostics to find problems with the assumptions. In Section 7, we discuss some possible remedies so that the subject itself can regain its prestige as a subject of searching the truth.

2. ABUSE OF STATISTICS: MISLEADING AND NONSENSE STATISTICS

Disbelief in statistical techniques has been in existence since the development of this subject and the growing abuse of statistics has increased over the years and a number of books are now available (see Campbell (1979), Hooke (1983), Jaffe and Spirer (1987)) on this topic. In some cases, the misuse may be accidental. In others, it is purposeful and for the gain of the perpetrator. Vellman (2008) pointed out that a Google books search of "lies, damn lies, and statistics" turns up 495 books, and a general Google search finds "about 207,000" hits. A small (nonrandom) sample of these references shows that most are meant to suggest dishonest manipulations and interpretations. For example, at Bangladesh, there are lots of doubts regarding the figures related to food production, literacy rate, foreign investment etc. The politicians often mention some figures which are entirely baseless.

In 1978 the government claimed that there are several areas in Bangladesh where not a single crime occurred. These areas were known as the so-called 'zero crime zone'. In the late 80's few 'zero population growth' zones were created and later it was found that most of the births in those areas are not registered there, they are registered in neighboring areas. It is generally believed that in 1990 census millions of people were uncounted just to show that 'the population control program in Bangladesh is working successfully based on that the president of the country got an award from the United Nations. In a recently leaked document from 'WikiLeaks' shows that the US Ambassador in New Delhi suspected that about 40 million Muslim population were deliberately undercounted in 2000 census. The Election Commission of Bangladesh allegedly enlisted 12 million fake voters in 2006 voter list which was corrected later after the verdict of Bangladesh High Court. These are a few examples of deliberate abuse of statistics but all of us must accept the fact that politicians, not the statisticians, are to blame for this.

Sometimes statistical techniques can be used in such a way that they may hide some facts or could mislead common people. Unlike the previous examples there is nothing wrong in the data but they are presented in an ambiguous way. Here we present a couple of examples taken from Sullivan (2011). The first example shows the price of power in the US from 2001 to 2007. The first graph shows that there is a dramatic increase in the price of power in years 2006-07 in comparison with the price in 2001-02. But the second graph shows that the increase is steady. It's unbelievable, but true that both of the above graphs are based on exactly the same data but in the first graph its scale is changed to give a false impression that the price hike grows exponentially while the reality is that the growth is steady.

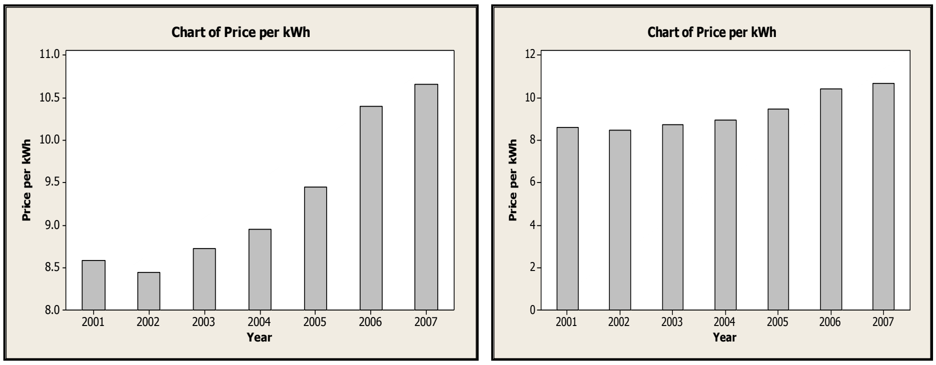


Figure 1: The price of power in the US plots in different scales

The second example shows deaths in road accidents in the US from 2001-05. The first graph puts lots of question marks to the road safety management in the US as the number of deaths tends to increase over the years. But the first graph is misleading because the number itself does not have any meaning unless we know how many motor vehicles are on the road or how many accidents occur every year. There will be no death if there were no car on the road! So the real information that we have to look at is the death rate as shown in the second graph and we feel now that the things are getting improved.

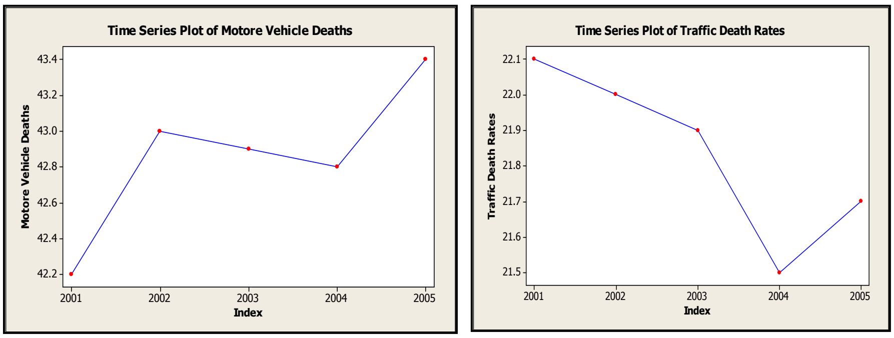


Figure 2: Plots of the number of deaths and death rates in the US

With the growing applications of statistics we often see that statistical works are not being done by statistics practitioners, not the professionals, and in many occasions they are blindly applying statistical techniques without knowing what they are doing. This kind of practice might lead the following nonsense findings:

- Smoking reduces the risk of heart attacks
- High calorie foods reduce age
- Mental deformity in the U.K depends on who were the presidents of the U.S at that time
- Watching television makes you live longer
- Suicides of women fit a Poisson distribution

Lurking variable and Simpson's Paradox

Here we show how a lurking variable could lead to a misleading conclusion. Lurking variable is a missing variable that contains very important and relevant information. Let us consider the following example. The University of California at Berkley was charged with having discriminated against women in their graduate admissions process for the fall quarter of 1973. Table 1 shows the respective numbers of men and women accepted and denied for two of the University's Graduate Program (see Bickel, Hammel and O'Connel (1975)).

TABLE 1: Overall acceptance of male and female students at the University of Berkley

	Men	Women
Accepted	533	1
Denied	665	3
Total	1198	4

From Table 1, we observe that 44.49% men are accepted in this program while the acceptance rate for women is only 25.17% and statistical test for the equality of two proportions yields a z value of 7.73 (p-value = 0.000) and Fisher's exact test also yields a p-value = 0.000 supporting the claim of female rights activists claim that this University's admission policy is biased to men. To understand this problem more clearly we now look at the respective numbers of men and women accepted and denied for two programs (denoted by A and F) separately as given in Table 2.

TABLE 2: Program-wise acceptance of male and female students at the University of Berkley

	Men		Women	
	Accepted	**Denied**	**Accepted**	**Denied**
Program A	511	314	89	19
Program F	22	351	24	317
Total	533	665	113	336

From Table 2, we compute the acceptance rate of both men and women in these two program and the stunning finding are presented Table 3.

TABLE 3: Program-wise acceptance rate of male and female students at the University of Berkley

	Program A	Program F
Men	61.94	5.90
Women	82.40	7.04

The above results clearly show that the women has a higher acceptance rate then men in both of the graduate programs though their acceptance rate is significantly less overall. This is a paradox which is known as Simpson's paradox. This paradox occurs in the absence of a lurking variable (a missing variable that contains very important and relevant information). In our example the lurking variable is 'the program applied to'. Further analysis will show that women attempted to enroll in a very higher rate (75.95% in comparison with men's rate 31.47%) to a program which is much harder (having the acceptance rate only 6.89% while the other program has the acceptance rate 64.30%) and that is the main reason of their overall low acceptance rate. In similar situation the acceptance rates should be calculated by the weighted average not by simple average as done before. If the proportion of men and women applied for both of these two programs were the same, there would be no paradox.

3. RANDOMIZATION, SAMPLING AND COLLECTION OF DATA

In 1975, the Pepsi Beverages Company organized a challenge that takes the form of a taste test and was popularly known as Pepsi Challenge 1975.

Shoppers were encouraged to taste both colas, and then select which drink they prefer. Then the representative revealed the two bottles so the taster can see whether they preferred Coke or Pepsi. The results of the test leaned toward a consensus that Pepsi was preferred by more Americans. Despite this claim, the market showed a different scenario. Americans used to buy Coke much more than Pepsi. Popular sources criticized the so called Pepsi challenge for the methods used. In Pepsi challenge, coca-cola was always served earlier and in a bit warmer than Pepsi. In general human being remembers more the last thing he/she tastes and most of the Americans like chilled cola. Although the shoppers were blind about the cola they tasted but the whole process lacked radomization and had a clear bias towards Pepsi.

Many people may not realize that the randomness of the sample is very important. In practice, many opinion polls are conducted by phone, which distorts the sample in several ways, including exclusion of people who do not have phones, favoring the inclusion of people who have more than one phone, favoring the inclusion of people who are willing to participate in a phone survey over those who refuse, etc. Non-random sampling makes the estimated error unreliable. For this reason a random (probability) sampling is always welcome. Design of sampling and the determination of sample size are the most challenging steps of a random sampling procedure. Sample size determination takes into account several points:

- What sampling is being used?
- How much precision the experimenter wants?
- How much margin of error one would allow in the inferential procedure?

For a very large population (nationwide survey) a sample size between 1200 and 1300 (e.g. Gallup polls with 1000 samples for a country like the USA) could be enough (Newport, Saad and Moore, 1997) in a simple random sampling to infer within 3% margin of error (for 1% margin of error the required sample size for the USA is 10000) if the sampling can be done very carefully and efficiently.

Another question might come in our mind, what to do if we fail to obtain a random sample? Can we not use any statistical techniques there? Researchers can use exploratory data analysis (EDA) techniques if the sample is not random. Researchers should employ very recent and advanced graphical displays to make the data more representative. If the data set itself is interesting, simple statistics like frequency distribution, percentage, proportion, rates etc could be interesting. For example, a nonrandom sample showing that in the USA, African Americans have 3 times higher unemployment rate than Whites could be very useful information. When samples are not random we have to be a bit careful in drawing conclusions. We should use words like 'more or less likely' instead of using the word significant or insignificant.

In social science study it is a very common practice to prepare questionnaire in such a way that the answers from the respondents are qualitative. It is also a common fashion to collect qualitative data even when the quantitative data are available. The advantage of this type of practice is ease of collecting data. But the main disadvantage of this practice is that it causes information loss and may end up with misleading inference. Let us consider the following example as given in Table 4. We are interested to see whether students' GPA depends on the income of their parents.

Table 4: Students' GPA and parents' income data

Parent's income GPA	Low	Mid	High	Total
Low	4 (2.4)	3 (3.6)	1 (2)	8
Mid	1 (2.1)	5 (3.15)	1 (1.75)	7
High	1 (1.5)	1 (2.25)	3 (1.25)	5
Total	6	9	5	20

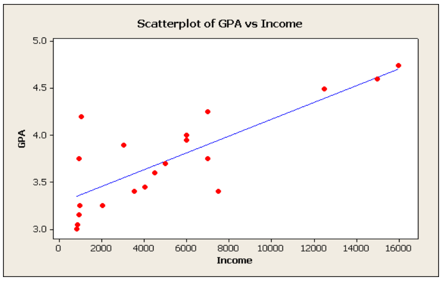
The calculated value of χ^2 this data is 6.96 (p-value 0.276) and we may conclude that student's performance in exam has no significant relation with their parents income.

Now we replicate the data twice. The replicated data are presented in Table 5. The calculated value of χ^2 for this data is 13.92 (p-value 0.015) and we may conclude that student's performance in exam has a significant relationship with their parents income. Since both of the variables are ordinal (not qualitative only) we should analyze the relationship by ordinal concordances or discordances as suggested by Agresti (1984, 2002), Simonoff (2003) or others.

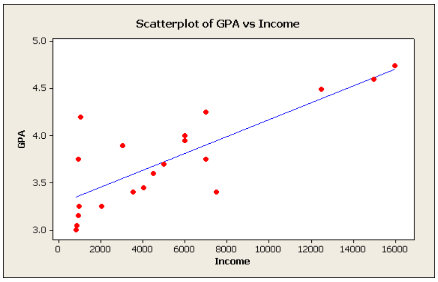
TABLE 5: Students' GPA and parents' income replicated data

Parent's income GPA	Low	Middle	High	Total
Low	8 (4.8)	6 (7.2)	2 (4)	16
Mid	2 (4.2)	10 (6.3)	2 (3.5)	14
High	2 (3.0)	2 (4.5)	6 (2.5)	10
Total	12	18	10	40

But the interesting feature of this particular problem is that we know the exact numerical values of both of these two variables. Let us do the same analysis using the correlation coefficient.



n = 20, r = 0.797, t = 5.598 (0.000)



n = 40, r = 0.797, t = 8.134 (0.000)

Figure 3: Scatter plot of original and replicated students' GPA and parents' income data

Replication of data should not change the relationship of the variables and we see this in Figure 3.1 and in both of these two correlations. So this example shows the limitation of using simple contingency analysis although that is a hugely popular technique in social science research.

4. MODELLING

Modeling is an essential part of the entire statistical procedure. The importance of a correct model is huge as the entire process is wrong if the model is wrong. To quote Velleman (2008), John Tukey taught that Statistics is more a science than it is a branch of Mathematics. For a mathematics theorem to be elegant, it is sufficient that it be beautiful and true. But Statistics is held to the additional standard imposed by science. A model for data, no matter how elegant or correctly derived, must be discarded or revised if it doesn't fit the data or when new or better data are found and it fails to fit them. Our reliance on model often suffers a setback from the experience of Box (1976).

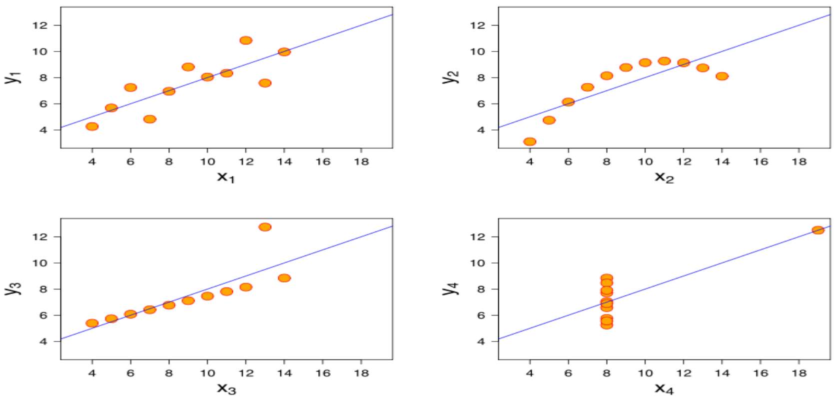


Figure 4: Scatter plot of Anscombe's quartet data

To illustrate this problem let us consider Anscombe (1973)'s quartet. Figure 4 clearly shows that the four graphs are totally different but if we model the relationship of Y on X by a linear regression model, the lease squares method produce exactly same results for all regression statistics as shown in Table 6.

TABLE 6: Regression summary statistics of Anscombe's quartet data

Statistics	Model I	Model II	Model III	Model IV
Intercept	3.0	3.0	3.0	3.0
Slope	0.5	0.5	0.5	0.5
R^2	0.667	0.667	0.667	0.667
t(p-value)	4.24 (0.002)	4.24 (0.002)	4.24 (0.002)	4.24 (0.002)
SST	41.23	41.23	41.23	41.23
SSR	27.49	27.49	27.49	27.49
MSE	13.74	13.74	13.74	13.74

Table 7 presents (see Kopits and Cropper (2003)) the change in traffic fatality risk in various developed and developing countries. Here the change in traffic fatality risk (deaths/10,000 persons) from 1975-1998 are presented.

TABLE 7: Change in traffic fatality risk data

Country	% Change	Country	% Change	Country	% Change
Canada	-63.4	France	-42.6	Malaysia	44.3
Hong Kong	-61.7	Italy	-36.7	India	79.3
Finland	-59.8	New Zealand	-33.2	Sri Lanka	84.5
Austria	-59.1	Taiwan	-32.0	Lesotho	192.8
Sweden	-58.3	United States	-27.2	Colombia	237.1
Israel	-49.7	Japan	-24.5	China	243.0
Belgium	-43.8			Botswana	383.8

From this table we see that the change is positive for Malaysia that means the fatality risk is increasing there. But Figure 5 based on data provided by Malaysian Institute of Road Safety shows that the fatality risk is declining.

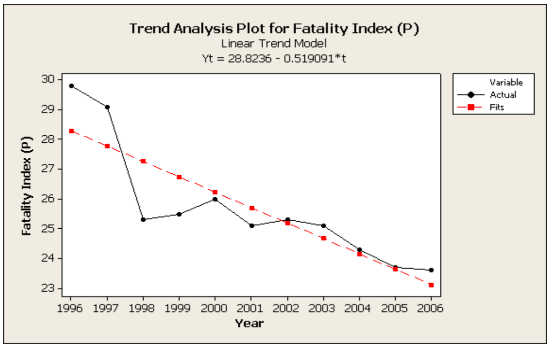


Figure 5: Malaysian traffic fatality risk data from 1996-2006

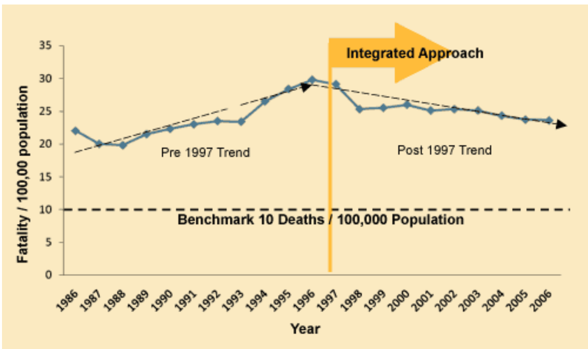


Figure 6: Malaysian traffic fatality risk data from 1986-2006

Figure 6 shows that both of the above claims are correct. Fatality risk in Malaysia increased up to 1996 and since then there is a declining pattern. A structural break has occurred in 1996 and hence the linear model is no longer appropriate for this data.

5. INFERENCE AND JUDGEMENT

Sometimes setting up null and alternate hypotheses could be a key inferential issue. Suppose a school has developed a new teaching method so that the students perform better in their exams. For such a problem, many text books set up hypotheses simultaneously in this way:

(a) $H_0 : \mu_{New} = \mu_{Old}$ vs $H_0 : \mu_{New} \neq \mu_{Old}$

(b) $H_0 : \mu_{New} = \mu_{Old}$ vs $H_0 : \mu_{New} > \mu_{Old}$

(c) $H_0 : \mu_{New} = \mu_{Old}$ vs $H_0 : \mu_{New} < \mu_{Old}$

To see the effectiveness of the new method they applied both old and new methods to two groups of students each of size 10 and obtained the following results: New teaching method: Average score 70.0 with a standard deviation 6.32 and Old teaching method: Average score 73.0 with a standard deviation 7.07. If we employ the equality of two mean tests, the test statistic is

$$t = \frac{\sqrt{n}\left(\bar{x}_{New} - \bar{x}_{Old}\right)}{\sqrt{\dfrac{\left(n_{New} - 1\right)s_{New}{}^2 + \left(n_{Old} - 1\right)s_{Old}{}^2}{\left(n_{New} + n_{Old} - 2\right)}}} \sim t_{n-2}$$

where $n = n_{New} + n_{Old}$. For this data the value of the test statistic t is – 2.0.

For the set of hypotheses (a), the critical region at the 5% level of significance is given by $t : t < -2.10 \cup t > 2.10.$ Since the value of the test statistic does not fall in the critical region, we may accept the null hypothesis and conclude that the new teaching method is similar to the old teaching method. For the set of hypotheses (b), the critical region at the 5% level of significance is given by $t : t > 1.73.$

Since the value of the test statistic does not fall in the critical region, we may accept the null hypothesis that the new teaching method is similar to the old teaching method. But someone may be skeptical about the new teaching method as there is empirical evidence that students with the new method is getting lower average score than the old method. So one may consider the set of hypotheses given in (c). In this case the critical region at the 5% level of significance is given by t: $t <$ 1.73.

Those who believe incorrectly that Statistics is solely a branch of Mathematics (and thus algorithmic), often see the use of judgment Statistics as evidence that we do indeed manipulate our results. It is in the area of hypothesis testing that we often see people apply statistics methods blindly, hoping for that statistically significant p <0.05, but neglecting to employ their judgment.

In general $r = 0.8$ is considered as a very high correlation coefficient, but when n = 3, the value of the t-statistic for testing the significance of the correlation coefficient is 1.333 (p-value = 0.410). So the correlation is not significant at all. Again r = 0.1 seems to provide very low correlation. But when n = 2000, the value of the t-statistic is 4.49 (p-value = 0.000) which shows that this correlation is highly significant at any level.

6. DIAGNOSTICS

Each and every simple step in statistical inference is guided by some kind of assumptions whose existences are essential for a valid inferential statement. For example, all four major test statistics z, t, χ^2 and F arc valid only when the sample observations come from a normal distribution. Tukey (1960) mentioned a tacit hope in ignoring deviations from the ideal model was that they would not matter; that statistical procedures which were optimal under the strict model would still be approximately optimal under the approximate model. Unfortunately, it turned out that this hope is often drastically wrong; even mild deviations often have much larger effects than were anticipated by most statisticians.'

Diagnostics are a set of measures which are designed to find problems with the assumptions. Most of the diagnostic techniques in statistics are guided to find outliers in the data. The term 'outlier' came from astrophysics to distinguish planets which are 'outlying' in our solar system. The term 'outlier' got popularity in statistics in the middle of the last century to solve a court case in England. In 1949, in the case of Hadlum vs Hadlum, Major Hadlum appealed against the failure of an earlier petition of divorce. His claim was based on an alleged adultery by Mrs. Radium, the evidence for which consisted of the fact that Mrs. Hadlum gave birth to a child which was 349 days later than when Major Hadlum had left the country to serve the nation during the World War II. The appeal judge rejected the appeal. In other similar cases conflicting views had prevailed. In Mr. T vs Mrs. T case also in 1949 the court had ruled that 340 days was impossible based on the fact that the average gestation period for the human female is 280 days. A much earlier case resurfaced at about the same time. In 1921, Mr. Gaskil failed in a petition for divorce on the grounds of adultery based on an absence of 331 days from home. In 1951, the House of Lords had ruled that the limit is 360 days based on a huge survey conducted by the British Medical Association for a sample of 13634 British Births. The concept of outliers in a data set is considered to be as old as the subject of statistics.

A more formal definition of outlier came from Barnett and Lewis (1994) said that we shall define an outlier in a set of data to be an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data. Hampel *et al.* (1986) claim that a routine data set typically contains about 1-10% outliers, and even the highest quality data set cannot be guaranteed free of outliers. Let us consider a set of climate data which was collected by the Indian Statistical Institute (ISI) Calcutta. The data contain rainfall update from 1990 to 2003 in Bihar, India. It also contains several variables, such as evaporation (mm), maximum temperature (°C) , minimum temperature (°C), humidity (%) at 8:30 am and humidity (%) at 4:30 pm. Figure 7 is taken from Imon *et al.* (2012). To get a better view, we use the brushing command in S-PLUS to select a portion of data where some kind of irregularity is visible more clearly. Figure 8 shows that on few days recorded minimum temperature was much higher than maximum temperature which is simply impossible.

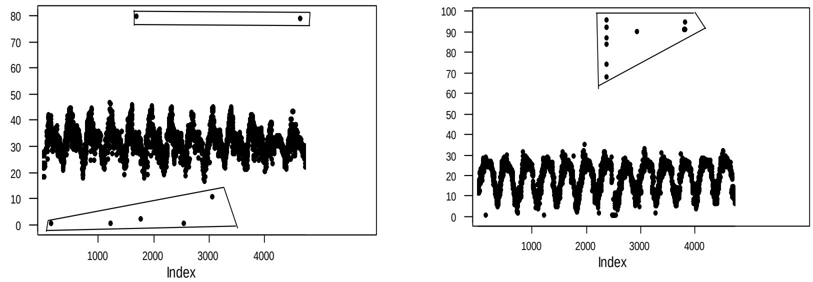


Figure 7: Scatter plot of maximum and minimum temperature of Bihar data

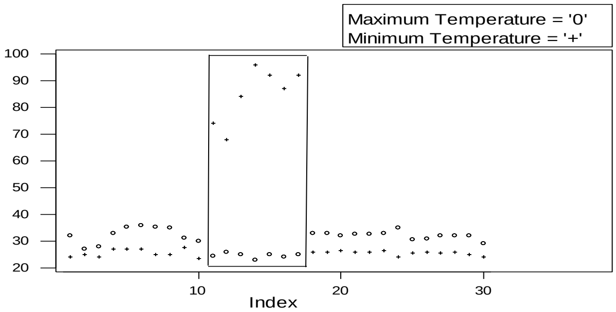


Figure 8: Brushed values of Bihar temperature data

The consequences of outliers are well-known to statisticians. It can create huge interpretative problems and that is why outlier detection is so important in Statistics. Let us consider the Belgian fire claim data taken from Rousseeuw and Leroy (1987).

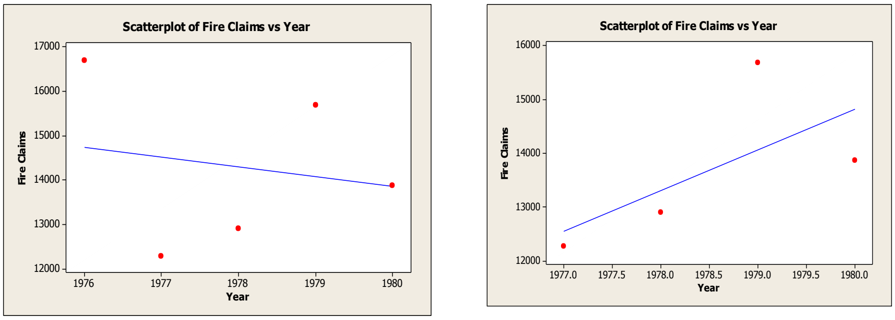
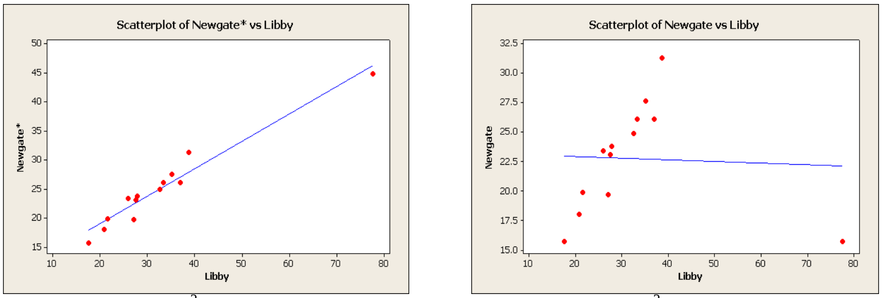


Figure 9: Scatter plot of Belgian fire claim data with and without outlier

There exists an outlier in this data. Figure 9 presents the scatter plot of Belgian fire claim data with and without an outlier. If we keep the observation in, we obtain a negative slope for the line. When we remove that point from the data, the least squares line show an upward slope. So it is a huge problem for the insurance company to set its target for the next year.

Our next example is the water flow data taken from Chattarjee and Hadi (2006). Figure 10 shows how a single outlier can destroy the goodness-of-fit of the least squares line.



$R^2 = 0.945$ $\qquad$ $R^2 = 0.002$

Figure 10: Scatter plot with goodness-of-fit for water flow data with and without outlier

The data presented in Table 8 is artificial in nature which is designed to show the effect of multicollinearity on the regression coefficients (Montgomery *et al*. (2006)).

TABLE 8: Montgomery *et al*. (2006) collinearity data

Y	X_1	X_2
1	2	1
5	4	2
3	5	2
8	6	4
5	8	4
3	10	4
10	11	6

When we fit Y on X_1 only we obtain the least squares line as

$$\hat{Y} = 1.835 + 0.463\, X_1$$

But this line becomes

$$\hat{Y} = 1.835 - 1.222 + 0.463\, X_1 + 3.649\, X_2$$

when Y is fitted on X_1 and X_2. The above two fits clearly show that the sign of the coefficient of X_1 change which is known as the wrong sign problem.

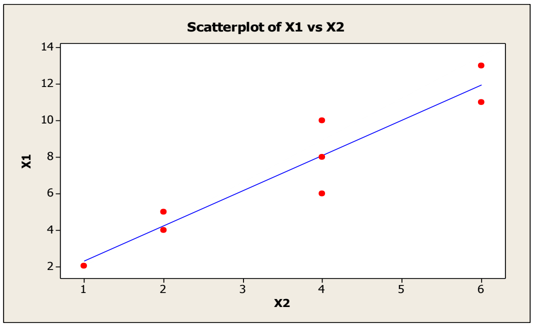


Figure 11: Scatter plot of X_1 versus X_2 for Montgomery *et al*. (2006) data

For a possible explanation we plot two explanatory variables against one another and it shows a linear relationship with a correlation coefficient 0.945 with p-value 0.000. This finding confirms our suspicion that multi collinearity causes wrong sign problem.

7. STATISTICS AND TRUTH

Our first inference was 'smoking reduces risk of heart attacks' (Mullet, 1976). In a regression fit for the prediction of heart disease of British people the coefficient of smoking had a significant negative meaning smoking helps reducing heart disease. But after a careful investigation it was observed that in the model there was another variable, 'drinking' which had a very high correlation with smoking. So the entire analysis suffered from 'multicollinearity' resulting in the so-called 'wrong sign' problem.

High calorie foods reduce age was our next finding obtained by M.Sc. project student in the Department of Statistics of an university in 2002. The researcher established a regression line in the form

$$\text{Age} = (\ldots) - (\ldots)\ \text{Calories}$$

we obtain

$$\text{Calories} = (\ldots) - (\ldots)\ \text{Age}$$

Now this makes sense. As we grow older we reduce taking high calorie foods.

The next inference was 'watching television makes you live longer.' This inference was drawn after observing that nations with many TV sets have higher life expectancy (see COMAP (2006)). Rich nations have more TV sets then poor nations. Rich nations also have longer life expectancy because they offer better nutrition, clean water and better health care. We ignored few lurking variables here and ended up with nonsense inference.

Montgomery *et al.* (2006) present an example where it shows that 'mental deformity in the U.K. depends on who were the presidents of U.S. at that time. A regression line of the number of certified mental defectives per 10,000 in the U.K. on the First name of the U.S. president from the years 1924-1937 produced the following output:

$$\text{Est. Mental Defectives in U.K.} = -26.4 + 5.9\ \text{U.S. President}$$

with $t = 8.996$ (p-value 0.000) and $R^2 = 0.871$.

Our final example is taken from Shil and Debnath (2001) as presented in Table 9 which contained number of suicides of 1096 women in 8 cities of a country during 14 years. The authors claimed that the data fit a Poisson distribution.

Table 9: Shil and Debnath's (2001) suicide data

No. of Suicides	0	1	2	3	4	5	6	7
Frequency	364	376	218	89	33	13	2	1

Velleman (2008) rightly said one can wield the tools of Statistics to mislead. After all, Statisticians do not claim to know things they cannot know. Instead, for example, they offer an interval of plausible values for an unknown parameter. Not satisfied with that, we spend more effort describing exactly how uncertain we are that even that interval covers the true value and just what we must assume about unknown and unknowable features of the world for those estimates to be correct. Statisticians are evidently taking great care to be honest, and readily admit their uncertainty. Liars usually assert their lies confidently in their striving to be believed and when a statistician's conclusion turns out to be wrong, the error is not seen as deliberate deception.

Another random sample may yield a different answer, but that isn't blamed on the statistician as a failure of ethical data collection or analysis. On the other hand we often see the life-or-death importance of data analysis and statistics. Hines (2007) showed that how one of the leading medical schools of the world, University of California San Francisco, ignored analyzing data that they regularly collect from their patients and later shockingly noticed that heart attack patients spent almost three hours on average at UCSF before their arteries were unblocked. Some had their electrocardiogram languish on a clipboard in the emergency room while doctors dealt with other patients. Although UCSF was faster than most hospitals, but its delays were still almost certainly killing some people and leaving others disabled because patients have the best chance of recovery if their arteries are opened within two hours, research has shown.

8. CONCLUSION

In practice we observe numerous occasions where statistical procedures give misleading conclusions. Sometimes it is done intentionally, so we must teach students regarding ethical issues while teaching statistics. But faulty design and sampling techniques, inappropriate modeling, selection of wrong inferential techniques and violation of standard assumptions are mainly responsible for nonsense statistics. If we are careful regarding these issues and use more nonparametric and robust statistical methods we can easily overcome many of these problems.

ACKNOWLEDGEMENTS

We gratefully acknowledge all valuable comments and suggestions made by the reviewer. The first author would like to thank Ball State University for funding a part of this research trough President's travel grant.

REFERENCES

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.

Agresti, A. (1984). *Categorical Data Analysis. 2nd ed*. New York: Wiley.

Anscombe, F. J. (1973). Graphs in Statistical Analysis. *Amer. Stat*. **27**: 17-21.

Barnett. V. and Lewis, T. (1994). *Outliers in statistical data*. 3rd ed. New York: Wiley.

Bickel, P. S., Hammel, E. A. and O'Connell, J. W. (1975). Sex bias in graduate admission: Data from Berkeley. **Science**. **187**: 398-403.

Box, G. E. P. (1976). Science and Statistics. *J. Amer. Stat. Assoc*. **71**: 791-799.

Campbell, S. K. (1974). *Flaws and Fallacies in Statistical Thinking*. New Jersey: Prentice Hall.

Chatterjee, S. and Hadi, A. S. (2006). *Regression Analysis by Examples*. 4th ed. New York: Wiley.

Chernoff, H. and Moses, L. E. (1959). *Elementary Decision Theory*. New York: Wiley.

COMAP. (2006). *For all practical purposes mathematical literacy in today's world*. 6th ed. New York: Freeman.

Esar, B. (1943). *Esar's Comic Dictionary*. New York: Harvest House.

Hadi, A. S., Imon, A. H. M. R. and Werner, M. (2009). Identification of outliers. Wiley Inter. Rev. *Comput. Statist*. **1**: 57-70.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P.J. and Stahel, W. (1986). *Robust Statistics: The approach based on influence function*. New York: Wiley.

Hines, A. (2007). *The life-or-death importance of data analysis.* New York Times.

Hooke, R. (1983). *How to tell the liars from the statisticians*. New York: Marcel Dekker.

Imon, A. H. M. R., Roy, M. and Bhattacharjee, S. K.B. (2012). Prediction of rainfall using logistic regression. *Pak. J. Statist. Oper. Res*. **8**: 655-667.

Jaffe, A. J. and Spirer, H. F. (1987). *Misused Statistics.* New York: Marcel Dekker.

Kopits, B. and Cropper, M. (2003). *World Bank Policy Research*. Working Paper No. 3035.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2006). *Introduction to linear regression analysis*. 4th ed. New York: Wiley.

Mullet, G. M. (1976). Why regression coefficient have the wrong sign?. *J. Qual. Tech.* **8**: 121-126.

Newport, F., Saad, L. and Moore, D. W. (1997). *How pools are conducted in where America stands*, Golay, M. (Ed.). New York: Wiley.

Peirce, B. (1852). Criterion for the rejection of doubtful observations. *Astro J.* **2**: 161-163.

Rousseeuw, P. Leroy, A. (1987). *Robust regression and oulier detection*. New York: Wiley.

Simonoff, J. S. (2003). *Analyzing categorical data*. New York: Springer.

Velleman, P. (2008). Truth, damn truth and statistics. *J. Statist. Edu*. **16**: 1-14.