2-2021

Visualizing Bivariate Data: What's Your Point of View?

Mamunur Rashid
DePauw University, mrashid@depauw.edu

Jyotirmoy Sarkar
Indiana University Purdue University Indianapolis

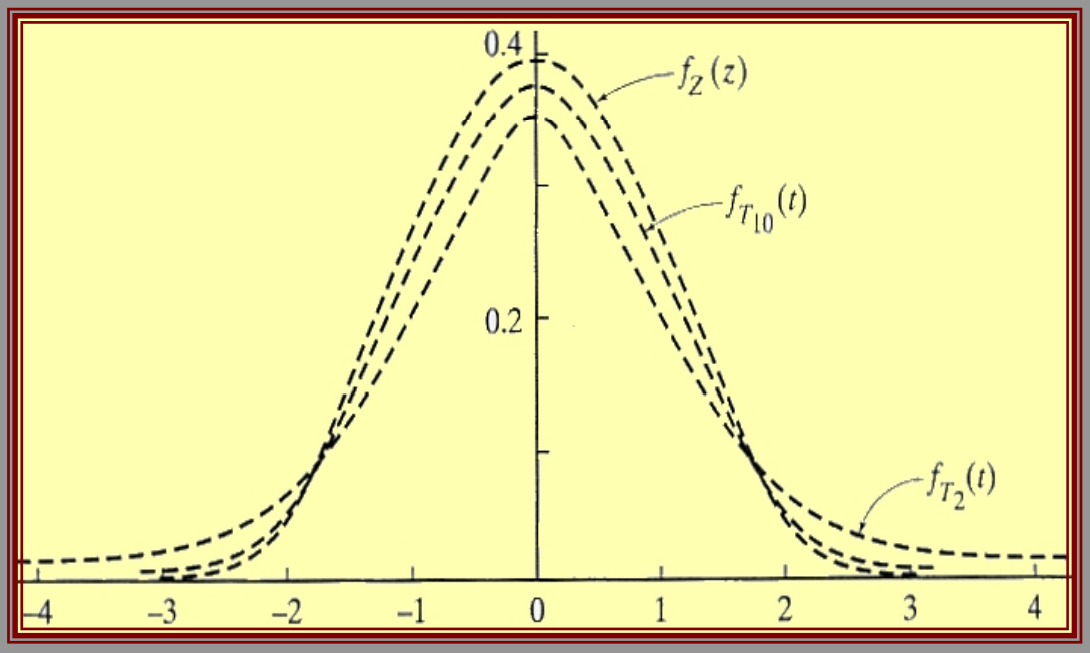
Recommended Citation

Rashid, M., Sarkar, J. "Visualizing Bivariate Data: What's Your Point of View?" Journal of Probability and Statistical Science 19(1), 83-95, Feb. 2021

ISSN 1726-3328

JPSS

A comprehensive journal of probability and statistics
for theorists, methodologists, practitioners, teachers, and others



JOURNAL OF PROBABILITY AND STATISTICAL SCIENCE

Volume 19 Number 1 February 2021

ISSN 1726-3328

JPSS

Journal of Probability and Statistical Science

Volume 19 Number 1 February 2021

Table of Contents

Published by: Susan Rivers' Cultural Institute, Hsinchu City, Taiwan, ROC.

Journal of Probability and Statistical Science
19(1), 83-95, Feb. 2021

Visualizing Bivariate Data: What's Your Point of View?

Jyotirmoy Sarkar *Indiana University-Purdue University Indianapolis*
Mamunur Rashid *DePauw University*

ABSTRACT A scatter plot shows the relationship between two continuous variables x and y. If the relationship is linear or if the two variables have a bivariate normal distribution, then the least squares regression lines of y on x and x on y can predict one variable as a linear function of the other. These two regression lines suffice to identify the mean vector, the coefficient of determination, Pearson's product moment correlation coefficient, and the ratio of the standard deviations (SD). So does a coverage ellipse! Additionally, we answer: In which direction must the points be projected to maximize (or minimize) the SD of the projections?

Keywords Coverage ellipse; Dot plot; Five number summary, Gaussian interval; Histogram; IVY plot.

1. Introduction

For efficient and impactful presentation of information, one must choose with care the most appropriate graph. The choice depends on the type of variable — qualitative variable (nominal/categoricalor ordinal), or quantitative variable (discrete or continuous). In this paper, we focus on continuous variables which take values over a continuum. In other words, even though for quick comprehension or for lack of sophisticated instruments we may report the value of a variable to the closest integer or to a few decimal places, we still recognize that finer intermediate values are possible. Some examples of continuous variables are temperature, pressure, speed, time, lung capacity, blood sugar level, heights, weights, exam scores, etc.

The purpose of this paper is to simplify the visual depiction of interrelation between two quantitative variables. Section 2 reviews some graphical depictions of one single quantitative variable. Section 3 reviews the scatter diagram and its projections in various directions to display two quantitative variables and their linear combinations. In Section 4, we model bivariate

Received October 2020, revised December 2020, in final form January 2021.
Jyotirmoy Sarkar is a Professor in the Department of Mathematical Sciences at Indiana University-Purdue University Indianapolis, IN 46202, USA; email: jsarkar@iupui.edu. Mamunur Rashid (corresponding author) is an Associate Professor in the Department of Mathematics at DePauw University, Greencastle, IN 46135, USA; e-mail: mrashid@depauw.edu.

 JPSS: ISSN 1726-3328

data either by a simple linear regression model with normally distributed errors or by a bivariate normal distribution. Section 5 summarizes bivariate statistics such as mean vector, standard deviations (SD), correlation coefficient, regression lines, and coefficient of determination. Section 6 explains how the two regression lines suffice to recover all bivariate statistics. Section 7 defines the coverage ellipse, and Section 8 explains how to recover all bivariate statistics starting from a coverage ellipse. Section 9 concludes the paper.

2. Graphical Presentation of Univariate Data

As a precursor to understanding the relationship between two continuous variables, let us first recall how to summarize and present each quantitative variable singly.

For a quantitative variable, when the data size is small, an appropriate unabridged graphical depiction is a **dot plot**: Along the number line, dots are placed exactly at the values taken by the variable. If the same value is taken multiple times, we stack the dots vertically. A dot plot gives a holistic look at the variable with no loss of precision since the raw data can be recovered from the dot plot. See Figure 1(a).

However, for ease of comprehension, we also compute important statistics such as the mean, SD, and the five-number-summary — the minimum, the first quartile, the median, the third quartile and the maximum. A boxplot is a graphical depiction of the five-number summary, showing a box stretching from the first quartile Q_1 to the third quartile Q_3, together with a vertical line at the median or the second quartile Q_2. The box plot also has two whiskers which are supposed to extend left from the first quartile Q_1 and right from the third quartile Q_3 for up to one-and-a-half times the interquartile range $Q_3 - Q_1$. All values farther away than the whiskers' intended expanse are flagged as potential outliers; and the whiskers are shortened to reach up to the most extreme value within their intended expanse. Thus, the minimum (also the maximum) is either flagged as an outlier or it is the end point of a whisker. See Figure 1(b). Furthermore, as in Rashid and Sarkar [3] and Sarkar and Rashid [4], an arrow may be drawn along the number line with the tail of the arrow at the mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and the length of the arrow representing the SD

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

(which is the positive square-root of the variance). This arrow is called the mean-SD arrow. See Figure 1(b).

Moreover, if the variable can be assumed to be roughly normally distributed, as an alternative to a box plot, one may draw the c-SD **Gaussian interval** around the mean, $\bar{x} \pm cs_x$,

and flag any value outside this interval as a potential outlier. See Figure 1(c). Some details are given below. For full details, see Sarkar and Rashid [6].

First, draw two two-headed, solid arrows spanning intervals $(\bar{x} - s_x, \bar{x})$ and $(\bar{x}, \bar{x} + s_x)$. Clearly, where the arrowheads meet is the location of the mean. Next, draw two two-headed, dotted arrows intending to span intervals $(\bar{x} - cs_x, \bar{x})$ and $(\bar{x}, \bar{x} + cs_x)$. These arrowheads also meet at the mean. Obviously, portions of the dotted arrows will be hidden behind the solid arrows; this is intentional. Any datum outside the expanse of the dotted arrows is flagged as an outlier. Moreover, each dotted arrow may be shortened to the most extreme value within their intended expanses. Additionally, the quartiles Q_1 and Q_3 may be shown using parentheses and Q_2 using a vertical notch; and the sample size may be printed alongside the *c*-SD Gaussian interval. Thus, a *c*-SD Gaussian interval is strictly more informative than the box plot; and it eliminates the non-informative thickness of the box.

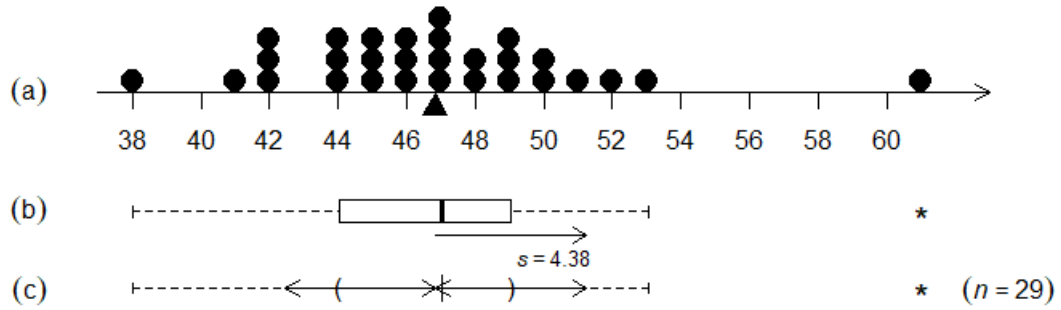


Figure 1 The speeds of 29 cars (in a 45 MPH zone) are shown using (a) a dot plot, (b) a boxplot together with a mean-SD arrow, and (c) a *c*-SD Gaussian interval with $c = 2.3263$.

How to choose the multiplier c? If one wishes to flag a fraction p of extreme values ($p/2$ on each side), under the assumption of normal distribution, one should choose the multiplier to be the $100(1 - p/2)$-th percentile of the standard normal distribution (using the code `qnorm(1-p/2)` in R). For example, if $p = .01$, then the multiplier is 2.5758; if $p = .02$, then the multiplier is 2.3263. Conversely, when the multiplier is 2, the associated $p = .0455$; and when the multiplier is 3, $p = .0027$. Throughout this paper, we have chosen $c = 2.3263$, wishing to flag roughly one percent of normally distributed data in each tail.

Another widely used graphical display of a univariate continuous variable is a **histogram**. It depicts the distribution of the values by drawing rectangles of a fixed bin-width and heights proportional to the frequencies of values within the bins. Most software packages have built-in algorithms to determine the number of bins and the common bin-width, typically using Sturges' rule [10]. Conventions may differ: R uses left-open, right-closed bins, except for the first bin which is closed; python uses left-closed, right-open bins, except for the last bin which is closed. Usually, the user can modify the bins by choosing the break points to ensure no datum falls on them and by allowing unequal widths. One shortcoming of a histogram is that the exact values of the variable are lost, only their membership within a bin is preserved.

Nonetheless, a histogram is very useful in depicting the overall shape of the distribution and it works well for large data.

While a dot plot works perfectly for a small data set, it becomes tedious to count the number of dots stacked at each distinct value as the data size becomes large. To mitigate the shortcomings of a dot plot and a histogram, Sarkar and Rashid [6] proposed an **IVY plot** which preserves the exact values and facilitates counting frequencies of all distinct values quickly and correctly. Figure 2 illustrates this advantage of the IVY plot. Interested readers may utilize an R package called `IVYplot` (see Nguyen *et al*. [1]) to construct IVY plots of quantitative variables.

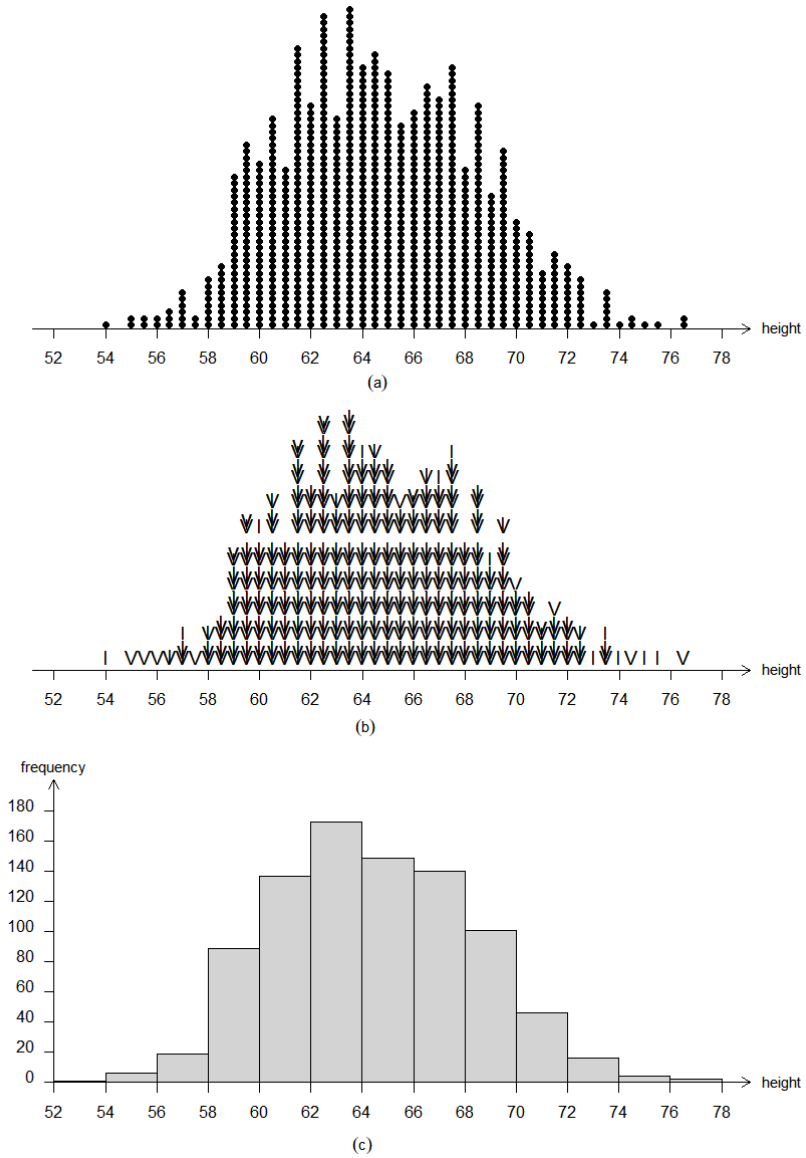
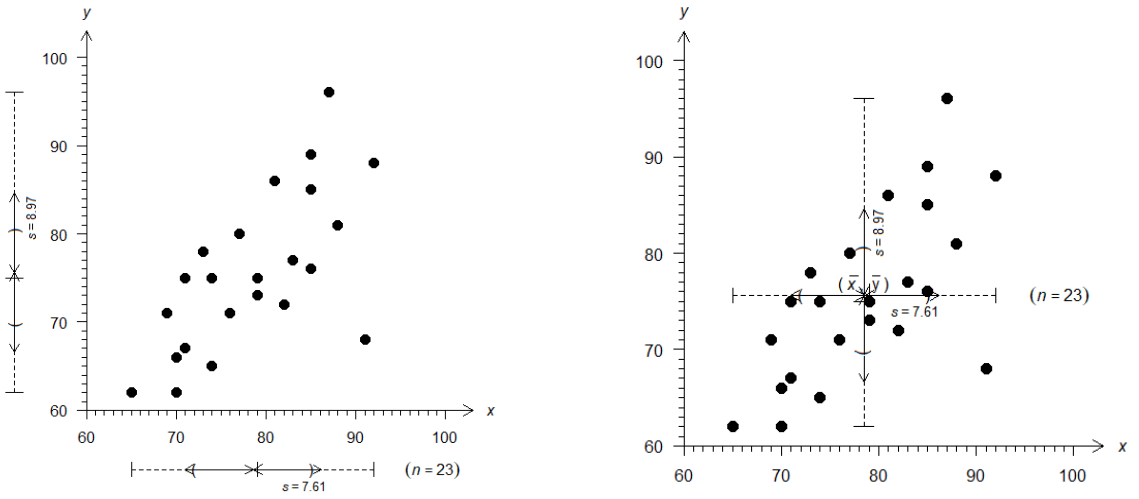


Figure 2 The heights (inch) of 883 high school seniors depicted using (a) a dot plot, (b) an IVY plot and (c) a histogram (with bin width 2 inches chosen according to Sturges' rule). Both the dot plot and the IVY plot show values correct to one half inch and hint at bimodality of the distribution. However, frequencies are harder to obtain from the dot plot, but easy from the IVY plot. The histogram has reduced precision of values and fails to capture bimodality.

3. Graphical Presentation of Bivariate Data

Having recalled the graphical displays of one quantitative variable, let us now focus on simultaneously visualizing two continuous variables measured on the same set of items. Above and beyond their individual characteristics (visualized through a dot plot, an IVY plot, a histogram, a boxplot, a mean-SD arrow and a *c*-SD Gaussian interval), we intend to reveal the interrelation between the two quantitative variables. An appropriate simultaneous depiction of the two variables is a **scatter plot**, in which the two axes represent the two variables, and each item's measurements are represented by a point whose coordinates are the values of the two variables. There ought to be as many points as the number of items. However, if two or more points coincide, stacking them is no longer an option, nor is a three-dimensional histogram helpful because a shorter stack may lurk behind a taller one. Instead, one may replace the point by writing down its multiplicity, perhaps in a single digit, or by color coding the dot and explaining the code in a legend. If so, then the original data can be reconstructed from the scatter plot. Unfortunately, most software packages do not automatically make this correction. Consequently, often the frequency of each dot (and hence the full accuracy of the data) is lost. This drawback of a scatter plot should be recognized and when possible corrected. For instance, the dots (use circles rather than bullets) may be made of various sizes with areas proportional to the frequencies, making sure no dot of a small size is hidden behind a dot of a larger size.

Of course, the scatter points can be projected onto either the horizontal or the vertical axis to reconstruct the dot plot or the IVY plot of the corresponding variable. Thereafter, these plots can be summarized into box plots, mean-SD arrows and *c*-SD Gaussian intervals, which may be depicted either along the margin or atop the scatter diagram parallel to the respective axis and intersecting at the point representing the mean vector $(\bar{x}, \bar{y})$. See Figure 3.



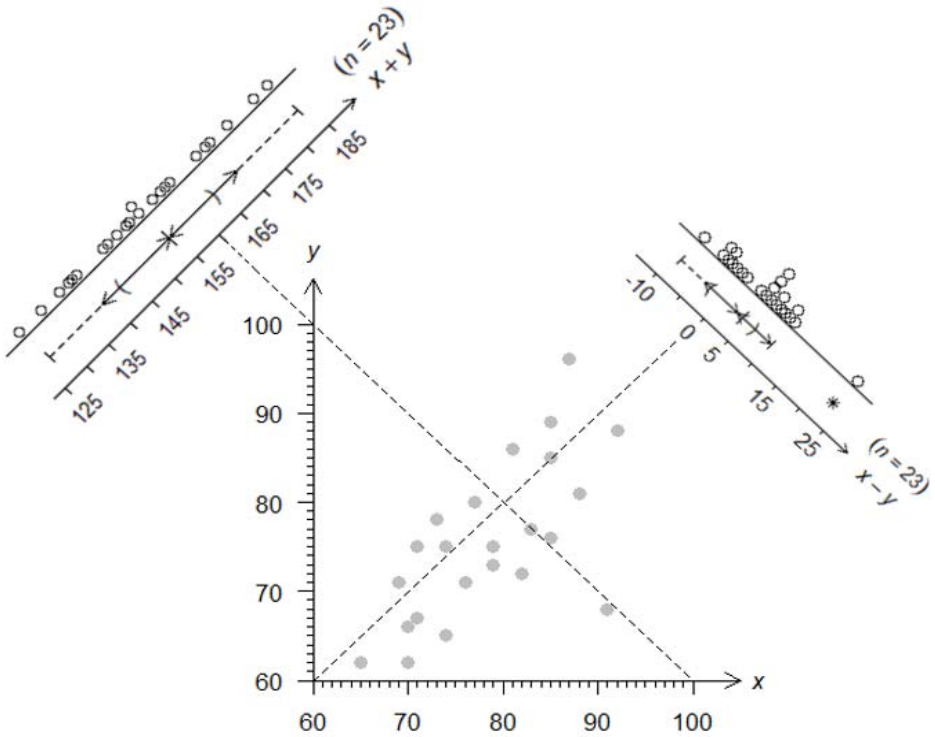
(a) (b)

Figure 3 A scatter plot of midterm score (x) and final exam score (y) of 23 students in an *Introduction to Statistics* course, together with the Gaussian interval of each score shown (a) in the margins, and (b) atop the scatter plot.

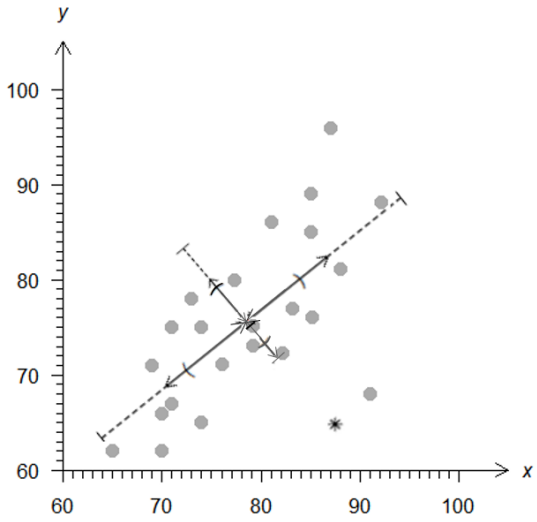
We can also project the points in a scatter plot along some other useful directions:

1) project on the line $y = x$ to obtain the dot plot of the sum $(x + y)$;
2) project on the line $y = -x$ to obtain the dot plot of the difference $(x - y)$; and
3) on the line $y = (b/a)x$ to obtain the dot plot of the linear combination $(ax + by)$.

See Figure 4 for projections of scatter points on the $\pm 45^{\circ}$ lines.



(a)



(b)

Figure 4 A scatter plot of midterm score (x) and final exam score (y) of 23 students in an *Introduction to Statistics* course, together with the summary statistics of the total score $x + y$ and the difference $x - y$ between midterm and final exam scores shown (a) after projecting on the peripheries, and (b) atop the scatter plot.

One may ask, "Projection in which direction causes the corresponding dot plot to exhibit the largest SD, and in which direction the smallest?" This is an important question whose answer will be useful when one aims to reduce the dimension of a multivariate data set. We will settle this problem towards the end of this paper. First, let us focus on predicting one variable as a linear function of the other.

4. Modeling Bivariate Data

The primary purpose of a scatter plot is to depict the relationship between two continuous variables with an aim to predict one variable as a function of the other. For example, the length of a person's shadow and his height are perfectly linearly related (having almost no error). In fact, shadow length is proportional to height, where the constant of proportionality depends on the angle of elevation of the sun at the time of measurement. Knowing the value of one variable, one can calculate the value of the other variable exactly. However, a person's weight and his height are not perfectly linear — they are at best "statistically linear," which according to the regression model means: "Variable *y* is a linear function of *x*, plus a random error or noise variable." Oftentimes, the error is assumed to have a **normal distribution** (in which case it is referred to as white noise). For example, a person's weight is a linear function of his/her height, plus a white noise. Sometimes one may assume that the two variables are distributed as **bivariate normal**. For example, a student's score in the final exam is not deterministically linear with her score in the midterm exam, though the two scores are somewhat linearly related. (More knowledgeable students tend to score higher in both exams; less knowledgeable students tend to score lower in both exams.) Moreover, the two scores may be jointly bivariate normally distributed. Using the score in one exam, we can predict (with some error) the score in the other exam. Furthermore, each score is univariate normally distributed, and any linear combination of the two scores is also univariate normally distributed.

5. Measuring Linear Relationship in Bivariate Data

For a linear regression model with normal error variable or for bivariate normal variables, the points in a scatter plot do not fall exactly on a line. Nonetheless, they seem to hover around a line. In fact, several items sharing the same *x*-value (or close enough *x*-values), will likely exhibit different observed *y*-values. In other words, the relation between *y* and *x* is partly linear and partly chaotic. The strength and direction of the linear relation between the two variables is measured by **Pearson's product moment correlation coefficient** given by

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}} = \frac{s_{xy}}{s_x s_y}, \quad (1)$$

where

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

is called the covariance between x and y; s_{xx} is called the variance of x, and s_{yy} the variance of y. It is well known that $-1 \le r \le 1$, which is typically proved by using the Cauchy Schwartz inequality. See Steele [9]. When $|r| = 1$, the two variables x and y have an exact linear relationship. When $0 < r < 1$ (or when $-1 < r < 0$), the linear regression of y on x, according to the least squares method, is the line $\hat{y}$ that minimizes the sum of squares of vertical distances of the points from the line, and is given by

$$\hat{y} - \bar{y} = r\frac{s_y}{s_x}(x - \bar{x}) \tag{2}$$

with a positive (or negative) slope $r(s_y / s_x)$; and the linear regression of x on y is the line

$$\hat{x} - \bar{x} = r\frac{s_x}{s_y}(y - \bar{y}) \tag{3}$$

with a positive (or negative) slope $r(s_x / s_y)$. Thus, the two regression lines pass through the bivariate mean vector $(\bar{x}, \bar{y})$. Finally, when $r = 0$, there is no linear relationship between x and y (although there may still exist a non-linear relationship), the $\hat{y}$ line is horizontal, and the $\hat{x}$ line is vertical. Thus, the correlation coefficient only measures the strength of a *linear* relationship between variables x and y.

If the purpose of documenting the two variables simultaneously is to predict the hard-to-measure (or yet-to-measure) y-value based on a more easily obtainable x-value, then the least squares regression line $\hat{y}$ is the answer. For example, if we know a student's mid-term score x, how do we predict his final exam score y? We use the $\hat{y}$ line, which shows the mean y-value (final exam score) of all students who share the same x-value (midterm score) as a linear function of x. Likewise, if the roles of the two variables are interchanged, then the least squares regression line $\hat{x}$ shows the mean x-value of all items sharing the same y-value. For example, if a teacher wishes to impute the midterm score of a student who missed it, she can use the $\hat{x}$ line, which shows the mean x-value (midterm score) of all students who share the same y-value (final exam score) as a linear function of y. (She can, of course, impose a suitable penalty to this imputed score.) Finally, the coefficient of determination r^2 (the square of the correlation coefficient) tells us what proportion of variation in y-values is attributed to its *linear* dependence on x-values (and vice versa). Thus, larger the r^2, the better the linear regression model. [One may work with the adjusted-r^2, which equals $r^2 - (1-r^2)/(n-2)$.]

Under both the regression model and the bivariate normal model, the error random variable $\varepsilon = y - (\alpha + \beta x)$ is normally distributed. However, the two models differ with respect to the associated variance: In the regression model, error variance is a constant (with

respect to *x*); but in the bivariate normal model, the error variance decreases as *x* moves away from $\bar{x}$ in either direction.

6. Recovering Bivariate Statistics from the Two Regression Lines

It is quite well known that when both $\hat{y}$- and $\hat{x}$- regression lines are superimposed atop a scatter diagram their point of intersection immediately identifies the mean vector $I = (\bar{x}, \bar{y})$. However, it is less known that the two regression lines together also depict the correlation coefficient r and the coefficient of determination r^2. See details in Sarkar and Rashid [7]. Here we briefly describe the method of extracting r^2 and r from the two regression lines: Combine (2) and (3) to see that r^2 is the ratio of the slope of the less steep regression line $\hat{y}$ to the slope of the steeper regression line $\hat{x}$. Specifically, see Figure 5, if one draws a horizontal line *IH* of *any* magnitude and then draws its perpendicular through *H* intersecting $\hat{y}$ at *P* and $\hat{x}$ at *Q*, then

$$r^2 = \frac{HP}{HQ}. \tag{4}$$

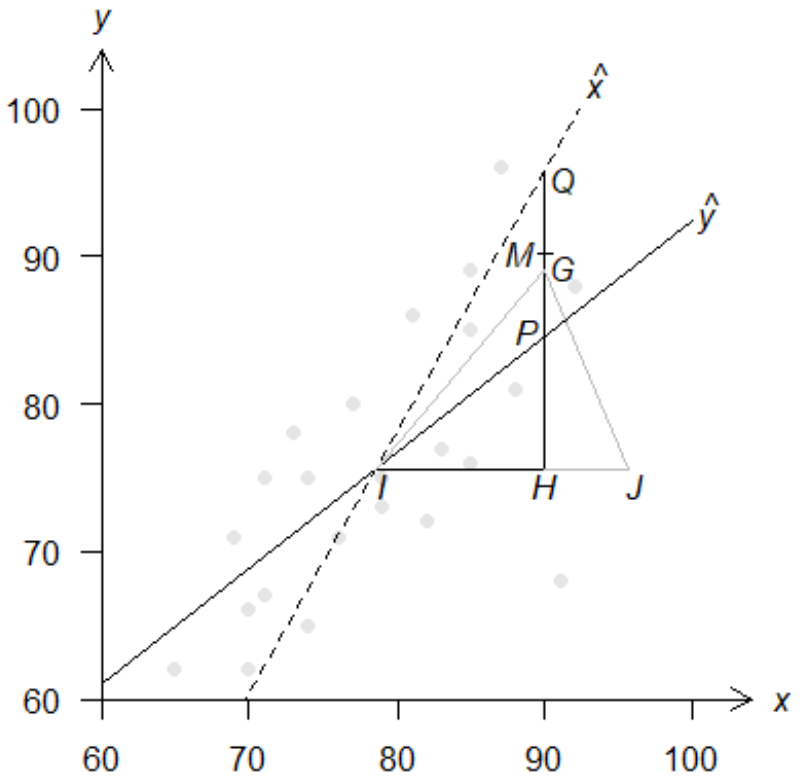


Figure 5 Given the two regression lines, the coefficient of determination r^2 can be obtained as the ratio *HP*/*HQ*, where *H* is any point on a horizontal line through *I* and *HQ* orthogonal to *IH*.

Having obtained r^2, it is straight-forward to find the correlation coefficient *r*. There are many ways to do it. Here is what we recommend: Let *M* be the midpoint of *PQ*; so that *HM* is

the (arithmetic) mean between *HP* and *HQ*. Extend *IH* to a point *J* such that $HJ = MQ = (HQ - HP)/2$. Next, find *G* on *PQ* such that $JG = HM = (HQ + HP)/2$. Then $HG = \sqrt{HP \cdot HQ}$; so that *HG* is the geometric mean between *HP* and *HQ*. Then

$$r = \sqrt{HP / HQ} = HG / HQ = HP / HG. \quad (5)$$

Of course, the sign of r is the same as the sign of the slope of either regression line. From expression (5), we note that *G* is the point on *HQ* such that $HG = \sqrt{HP \cdot HQ}$, the geometric mean between *HP* and *HQ*. The line *IG* with slope s_y / s_x is called the SD-line. Thus, given the two regression lines, we can compute the ratio of the two SDs as the slope of the SD-line *IG*; however, we cannot compute the SDs themselves.

7. The Coverage Ellipse

Instead of an arbitrary *IH*, let us choose $IH = \sqrt{6}s_x$. (We will explain the choice of this multiplier $\sqrt{6}$ at the end of this section.) Then $I = (\bar{x}, \bar{y})$, $H = (\bar{x} + \sqrt{6}s_x, \bar{y})$ and the corresponding $G = (\bar{x} + \sqrt{6}s_x, \bar{y} + \sqrt{6}s_y)$. On the SD line *IG*, let *F* be a point such that $GI = IF$. Clearly, $F = (\bar{x} - \sqrt{6}s_x, \bar{y} - \sqrt{6}s_y)$. See Figure 6.

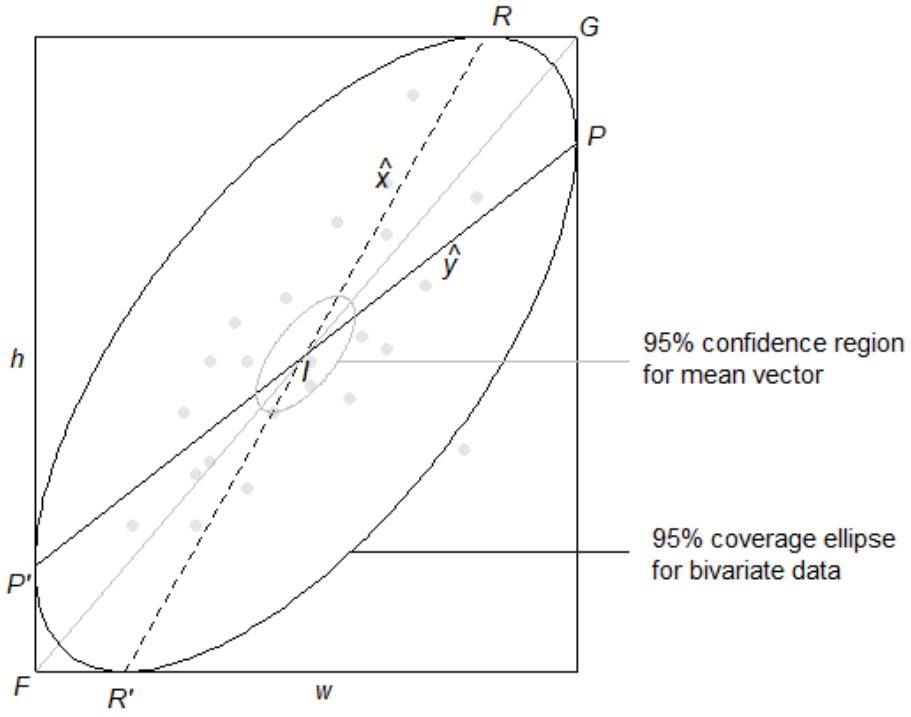


Figure 6 Given a rectangle *FG* of height *h* and width *w* and a point *P* on its right boundary, one and only one ellipse is internally tangential to the rectangle at *P*. The ellipse is also internally tangential to the rectangle at three other points P', *R*, R' such that $PG = P'F$, $GR = FR'$, and $PG : GR = P'F : FR' = h : w$. In particular, PP' and RR' intersect at the center of the rectangle.

Next, consider the rectangle with sides parallel to *x*- and *y*-axes, and with one diagonal *FG*. Accordingly, we shall call this rectangle *FG*. The height and width of rectangle *FG* are $h = 2\sqrt{6}s_y$ and $w = 2\sqrt{6}s_x$, respectively. The $\hat{y}$ line intersects rectangle *FG* at the point $P = (\bar{x} + \sqrt{6}s_x, \bar{y} + r\sqrt{6}s_y)$.

There is one and only one ellipse $\mathcal{E}$ which is internally tangential to rectangle *FG* and which passes through *P*. Clearly, $\mathcal{E}$ is also tangential to rectangle *FG* at another point $P' = (\bar{x} - \sqrt{6}s_x, \bar{y} - r\sqrt{6}s_y)$ on the $\hat{y}$ line such that $PI = IP'$. Moreover, the ellipse $\mathcal{E}$ and the rectangle *FG* are also tangential at two more points $R = (\bar{x} + r\sqrt{6}s_x, \bar{y} + \sqrt{6}s_y)$ and $R = (\bar{x} - r\sqrt{6}s_x, \bar{y} - \sqrt{6}s_y)$, which are on the $\hat{x}$ line and satisfy $RI = IR'$. In particular, $GP = FP'$ is a fraction $(1-r)/2$ of the height of the rectangle and $GR = FR'$ is the same fraction $(1-r)/2$ of the width of the rectangle. We leave the proofs of these claims to the astute reader.

So far, we simply mentioned the existence of a unique ellipse $\mathcal{E}$ tangential to rectangle *FG* and passing through *P*. In Section 8, we will document some additional properties of the ellipse $\mathcal{E}$ which will equip us to construct the ellipse explicitly. Furthermore, we chose $IH = \sqrt{6}s_x$ to ensure that, under the bivariate normal model, the probability that a random vector (*X*, *Y*) falls inside the ellipse is 95%. Of course, if we chose a different multiplier, the coverage probability will differ. We invite the reader to choose the multiplier that achieves a desired coverage. We will reveal the choice at the end of this paper.

8. Recovering Bivariate Statistics from the Coverage Ellipse

Now reversing the logic discussed above, given any one coverage ellipse $\mathcal{E}$, corresponding to any desired coverage, one can recover the bivariate statistics: First, draw the smallest rectangle with horizontal and vertical sides that cover the coverage ellipse $\mathcal{E}$. Join the points of tangency between the ellipse and the rectangle on the left and the right sides of the rectangle to recover $\hat{y}$ regression line; and join the points of tangency on the upper and the lower sides of the rectangle to recover the $\hat{x}$ regression lines. The intersection of these regression lines gives the mean vector. Construct the SD line as the diagonal (with positive slope) of the covering rectangle; recover the coefficient of determination, the correlation coefficient, and the ratio of SD's. Lastly, if the coverage of $\mathcal{E}$ is known to be 95%, then we can recover the two SDs via $s_x = w/(2\sqrt{6})$ and $s_y = h/(2\sqrt{6})$.

Let us document some more properties of the ellipse $\mathcal{E}$ which enable its construction.

(1) Suppose that the ellipse $\mathcal{E}$ has a coverage probability of $(1-\alpha)$ under the fitted bivariate normal distribution. Then we shall call $\mathcal{E}$ the $(1-\alpha)$-coverage ellipse for the random vector (*X*, *Y*) in the sense that (*X*, *Y*) belongs to $\mathcal{E}$ with probability $(1-\alpha)$. Among all regions (of whatever shape) with a coverage probability of $(1-\alpha)$ under the fitted bivariate normal distribution, $\mathcal{E}$ has the smallest Euclidean area. This is a consequence of the fact that contour plots of a bivariate normal density are concentric ellipses with the highest

density at the center (μ_x, μ_y) and the density decreases exponentially as the ellipse expands in size.

(2) If the $(1-\alpha)$-coverage ellipse $\mathcal{E}$ is contracted by a factor $\sqrt{n}$ around center $I = (\bar{x}, \bar{y})$, then the contraction is another ellipse called the $(1-\alpha)$-confidence ellipse for the mean vector (μ_x, μ_y). This is a consequence of the distribution of $(\bar{X}, \bar{Y})$ being bivariate normal with mean vector (μ_x, μ_y) and dispersion matrix Σ / n.

(3) The major axis of the coverage ellipse $\mathcal{E}$ shows the direction onto which the scatter points must be projected so that the dot plot of the univariate projections will attain maximum SD. Likewise, the minor axis is the direction onto which the projected scatter points are most densely packed. Thus, we have answered the question raised in the abstract and at the end of Section 3. The two axes are given by

$$\text{Major axis:}\quad y-\bar{y} = \frac{\sqrt{(s_x^2-s_y^2)^2+(2rs_xs_y)^2}-(s_x^2-s_y^2)^2}{2rs_xs_y}(x-\bar{x});$$

$$\text{Minor axis:}\quad y-\bar{y} = -\frac{\sqrt{(s_x^2-s_y^2)^2+(2rs_xs_y)^2}+(s_x^2-s_y^2)^2}{2rs_xs_y}(x-\bar{x}).$$

It is worth pointing out that only in the special cases when either $s_x = s_y$ or $r = \pm 1$, the major axis of $\mathcal{E}$ agrees with the SD-line.

9. Conclusion

We conclude this paper by drawing the readers' attention to two papers:

1) Sarkar and Rashid [8] proves property (3) above. It also recommends superimposing the 95%-coverage ellipse, corresponding to $IH = \sqrt{6}s_x$, on the scatter plot and shows how the two regression lines can be reconstructed from it. If this multiplier is declared in the scatter plot overlayed with a coverage ellipse, then we can recover the two SD's in addition to their ratio.
2) Sarkar and Rashid [5] depicts all summary statistics associated with a simple linear regression model. An R package called `shutterplot` (see Phuyal *et al.* [2]) allows users to draw such a comprehensive summary for any data set consisting of two quantitative variables.

Dear reader, we raised a question in the title of this paper. Should you ask us back the same question, we would say: "Superimpose on a scatter plot a 95% coverage ellipse." This is a small modification to the scatter plot that, without requiring any additional space or text, summarizes all bivariate statistics in a simple linear regression model with normally distributed error or in a bivariate normal distribution model. Of course, one may change the coverage to 98% or 99% or any other value $100(1-p)\%$, in which case the ratio IH / s_x is the square root of the $100(1-p)$-th percentile of a chi-square distribution on two degrees of

freedom (using the code `sqrt(qchisq(1-p,2))` in R). For example, if $p = .01$, then the ratio IH / s_x is 3.034854; if $p = .02$, then $IH / s_x = 2.79715$, and if $p = .05$, then $IH / s_x = 2.4477 \approx \sqrt{6}$.

The coverage ellipse contains within itself a wealth of information, which we have explained in this paper how to extract. Moreover, any scatter point outside the coverage ellipse is a potential *x*-, *y*-, regression- or bivariate outlier deserving special attention.

Acknowledgements

The authors thank our colleagues and students for playing the game of guessing the coefficient of determination from diagrams depicting the two regression lines. Also, the authors thank the editor and two anonymous reviewers for their constructive comments and suggestions.

References

[1] Nguyen, T., Sarkar, J., and Rashid, M. (2021). IVYplot: produces an IVY plot (similar to dot plot) with/without frequencies, R package version 0.1.0 https://CRAN.R-project.org/web/ packages/IVYplot/index.html.

[2] Phuyal, S., Sarkar, J., and Rashid, M. (2021). Shutterplot: The R Shutter Plot Package. R package version 0.1.0. https://CRAN.R-project.org/package=shutterplot.

[3] Rashid, M. and Sarkar, J. (2018). Cyber Mentoring in an Online Introductory Statistics Course, *Educational Research Quarterly*, **41**(3), 25-38.

[4] Sarkar, J. and Rashid, M. (2019). Portraying standard deviation via revolution, *Journal of Probability and Statistical Science*, **17**(1), 109-119.

[5] Sarkar, J. and Rashid, M. (2020). Shutter plot: a visual display of summary statistics over a scatter plot, *International Journal of Statistical Sciences*, **20**(2), 99-116.

[6] Sarkar, J. and Rashid, M. (2021a). IVY plot and Gaussian interval plot, *Teaching Statistics*, to appear.

[7] Sarkar, J. and Rashid, M. (2021b). Two regression lines suffice to determine r^2 and r, *Educational Research Quarterly*, to appear.

[8] Sarkar, J. and Rashid, M. (2021c). Depicting bivariate relationship with a Gaussian ellipse, *Statistics and Applications*, **19**(2), to appear.

[9] Steele, J. M. (2004). The Cauchy-Schwarz Master Class, Cambridge University Press, Cambridge, http://dx.doi.org/10.1017/CBO9780511817106.

[10] Sturges, H. A. (1926). The choice of a class interval, *Journal of the American Statistical Association*, **21**(153), 65-66.

