



Munich Personal RePEc Archive

**An overview of the elementary statistics
of correlation, R-squared, cosine, sine,
and regression through the origin, with
application to votes and seats for
Parliament**

Colignatus, Thomas

Thomas Cool Consultancy Econometrics

20 February 2018

Online at <https://mpra.ub.uni-muenchen.de/84722/>

MPRA Paper No. 84722, posted 21 Feb 2018 05:26 UTC

An overview of the elementary statistics of correlation, R-squared, cosine, sine, and regression through the origin, with application to votes and seats for Parliament

Thomas Colignatus
February 20 2018

MSC2010
62J20 Statistics. Diagnostics
00A69 General applied mathematics
28A75 Measure and integration. Length, area, volume, other geometric measure theory
97M70 Mathematics education. Behavioral and social sciences

JEL
A100 General Economics: General
D710 Social Choice; Clubs; Committees; Associations,
D720 Political Processes: Rent-seeking, Lobbying, Elections, Legislatures, and Voting Behavior
D630 Equity, Justice, Inequality, and Other Normative Criteria and Measurement

Abstract

The correlation between two vectors is the cosine of the angle between the centered data. While the cosine is a measure of association, the literature has spent little attention to the use of the sine as a measure of distance. A key application of the sine is a new “sine-diagonal inequality / disproportionality” (SDID) measure for votes and their assigned seats for parties for Parliament. This application has nonnegative data and uses regression through the origin (RTO) with non-centered data. Textbooks are advised to discuss this case because the geometry will improve the understanding of both regression and the distinction between descriptive statistics and statistical decision theory. Regression may better be introduced and explained by looking at the angles relevant for a vector and its estimate rather than looking at the Euclidean distance and the sum of squared errors. The paper provides an overview of the issues involved. Also a new relation between the sine and the Euclidean distance is derived.

1. Introduction	2
1.1. The subject of discussion	2
1.2. Votes and seats	3
1.3. Structure of the paper	5
2. Notation and basics	6
2.1. Well-known basics	6
2.2. Regression through the origin (RTO), for nonnegative vectors	7
2.3. A view from didactics	9
3. Evolving statistics	9
3.1. The statistical triad of Design, Description and Decision	9
3.2. A possible reason why RTO may be less prominent in the textbooks	10
3.3. Statistical significance	11
3.4. Causality	11
3.5. Specification search	11
4. Application to votes and seats	12
4.1. Descriptive statistics and decisive apportionment	12
4.2. Different worlds for votes and seats: DR and EPR	12
4.3. Apportionment in EPR	13
4.4. Different models and errors	13
4.5. Disproportionality, dispersion and education	14
4.6. True variables v^* and s^* and particular observations v and s	15

4.7.	Cos, slope and concentrated numbers of parties (CNP).....	16
4.8.	Analyses of squares for the direct error	17
4.9.	Symmetry.....	18
5.	More on interpretation.....	18
5.1.	Statistics and heuristics on the slope	18
5.2.	Electoral justice and inequality	19
5.3.	From the Humanities to Science	20
5.4.	The example of Brexit.....	22
6.	Conclusions	23
7.	References.....	24

1. Introduction

1.1. The subject of discussion

Karl Pearson (1857-1936) designed correlation between two vectors with the deliberate focus on centered data, namely to capture as much variation as possible. This arrangement also applied to the coefficient of determination, R-squared, between a vector and its estimate. We now have reason to look at the original and non-centered data, and in particular at nonnegative data. This application arises for votes and the assigned seats for parties for Parliament. When we want equal proportions of votes and seats, or to maximise the association, or to minimise the inequality / disproportionality (ID), then it would not make sense to center the data. Having a vector of seat shares at a constant distance of a vector of vote shares would be rather curious. For such non-centered data we apply regression through the origin (RTO), see Kozak & Kozak (1995) and Eisenhauer (2003). Textbooks tend to warn against dropping the constant in the regression, since it is better to test whether modeling errors show up in an estimated constant, yet in this case RTO appears to be the proper approach. Obviously, both Pearson's centered data and RTO share the property that the constant is zero, yet remarkably there are some differences for nonnegative data. This article will use this framework of nonnegative vectors, following Colignatus (2017ab) (2018abcd), but potentially some properties might be generalised for other data.

Pearson also concentrated on the cosine as a measure of similarity. The literature has spent little attention to the use of the sine as a measure of distance. For shares of votes and seats one might say that they are 97% close but the literature had focussed on developing measures of inequality / disproportionality, as people might be more sensitive to the 3% dissimilarity. Colignatus (2018d) discusses measures for votes and seats, and develops the new *sine-diagonal inequality / disproportionality* (SDID) measure. SDID uses not only the sine but also the square root on the sine, as a magnifying glass for small values, like the logarithm in the Richter scale. Colignatus (2018a) gives a more general perspective on distance and norm, and rejects the Aitchison geometry for compositional data for this comparison of votes and seats. ¹ There are particular aspects in voting, like majority switches, that are not relevant for other applications. The sine measure is more sensitive than the angular distance, and also interesting since it links up to regression, since R-squared finds its translation in RTO as the squared cosine of the angle between the vectors. A new finding in **Section 4.8** below is a direct relation between the sine and the Euclidean distance between two vectors.

Our objective here is to provide an overview of the elementary statistics of correlation, R-squared, cosine, sine, and regression through the origin (RTO), with application to votes and seats for Parliament. While our focus is on the relevance for statistics, with an eye on education in statistics, ² we cannot avoid highlighting aspects of votes and seats, since this

¹ Comparing votes and seats is a topic of itself. The log or logit transformations could still be used for explaining votes by say the economy, though Gelman & King (1994) stopped using the logit, since the vote shares in their data tended to be between 0.2 and 0.8.

² Mood & Graybill (1963:1) describe statistics as “the technology of the scientific method.”

content determines the analysis. For earthquakes there exists the Richter scale, but for votes and seats there isn't yet a similar "change" measure, while the advice now is to use SDID.

We use lower case x for a real vector and $\|x\| = \sqrt{x'x}$ for the norm of x , and $\|y - x\|$ for the Euclidean distance between x and y . There is also θ for the angle between x and y . Linear algebra provides an expression for the cosine of this angle, as $\text{Cos}[x, y] = x'y / \sqrt{(x'x y'y)}$, and thus $\theta = \text{ArcCos}[\text{Cos}[x, y]]$. The cosine is the improduct of the normalised vectors $x / \|x\|$ and $y / \|y\|$ on the unit circle. It is essentially the projection divided by the radius 1. ³ The formula for the cosine is scale invariant, as the angle does not change for positive scalars λ and μ , with $\text{Cos}[\lambda x, \mu y] = \text{Cos}[x, y]$.

1.2. Votes and seats

Let v be a vector of votes for parties and s a vector of their seats gained in the House of Commons or the House of Representatives. We discard zeros in v and use a single zero in s for the lumped category of "Other", of the wasted vote, for parties that got votes but no seats. Let $V = 1'v$ be total turnout and $S = 1's$ the total number of seats, and $w = v / V$ and $z = s / S$ the perunages. For presentation we will use $10 w$ and $10 z$ and the range $[0, 10]$ in general. For votes and seats, percentages $[0, 100]$ generate too much an illusion of precision, while $[0, 1]$ generates too many leading zeros. Distance measures on $[0, 10]$ read as an inverted (Bart Simpson) report card (with much appreciation for a low score). ⁴

In political science, the main current inequality / disproportionality (ID) measures are (conventionally for percentages but now with 10): ⁵

- Absolute difference / Loosemore-Hanby (ALHID): $10 \text{Sum}[\text{Abs}[z - w] / 2]$. The division by 2 corrects for double counting. An outcome of 1 means that one seat in a House of 10 seats is relocated from equality / proportionality.
- Euclid / Gallagher (EGID): $10 \sqrt{\text{Sum}[(z - w)^2 / 2]}$. For two parties this equals ALHID.
- χ^2 / Webster / Sainte-Laguë (CWSID): $10 \text{Sum}[w (z / w - 1)^2] = 10 \text{Sum}[(z - w)^2 / w]$. The Chi-Square expression has nonzero w . One can compare CWSID with $\text{ALHID} = 10 \text{Sum}[w \text{Abs}[z / w - 1] / 2]$ and $\text{EGID} = 10 \sqrt{\text{Sum}[w^2 (z / w - 1)^2 / 2]}$.
- The difference in shares for the "largest" party, i.e. with the most seats: $10 (z_L - w_L)$. This is an easy, rough and ready indicator with some history in the literature, and Shugart & Taagepera (2017:143) show a remarkable approximation $\text{EGID} \approx 10 (z_L - w_L)$.

The proposed new *sine-diagonal inequality / disproportionality* (SDID) measure has the formula $\text{SDID}[v, s] = \text{sign} 10 \sqrt{\text{Sin}[v, s]}$. The sine is invariant to scale: $\text{Sin}[v, s] = \text{Sin}[w, z]$. With $k = \text{Cos}[v, s]$ given by linear algebra, we might use $\theta = \text{ArcCos}[k]$ and then find $\text{Sin}[v, s] = \text{Sin}[\theta]$ but we can also use $\text{Sin}[\theta] = \sqrt{(1 - k^2)}$ directly. The additional square root on Sin works as a magnifying glass for inequalities / disproportionalities. The *sign* indicates majority switches, and is 1 for zero or positive covariance and -1 for negative covariance. ⁶

To clarify the distinction between the new proposal and the conventional measures in political science, **Figure 1** gives ALHID (*blue*) and SDID (*red*), and the intermediate steps given by the angle (*yellow*) and sine itself (*green*). ⁷ For two parties normalised to $[0, 10]$ we plot with *seats*

³ For projection of y onto x , the projection matrix $P = x x' / x'x$, so that $P y = x (x'y / x'x) = b x$. Normalising x and y onto the unit circle gives $b = \text{Cos}$.

⁴ $\text{D} = 10$ (Word: ctrl ' + Shift D) can be used to abbreviate "5.4 per 10" into 5.4 / D or 5.4 D^H , with $H = -1$. The official name of the letter is Eth, but for $\text{D} = 10$ the pronunciation "deka" is more appropriate, and D^H would be "decim" = "per 10". Thus $10\% = 1 \text{D}^H = 1 \text{decim}$.

⁵ We might leave out the factor 10 in these definitions, and use 10 only for presentation, but for SDID the factor 10 is a key design feature, and thus the others are best at the same scale. Other presentations introduced the scale via the inputs, like $10 w$ and $10 z$, but it is better to keep w and z unambiguously on the unit simplex and introduce the scale at the level of the measure.

⁶ In speech, one might not hear the difference between "sine" and "sign": then use "sinus" and perhaps "signum".

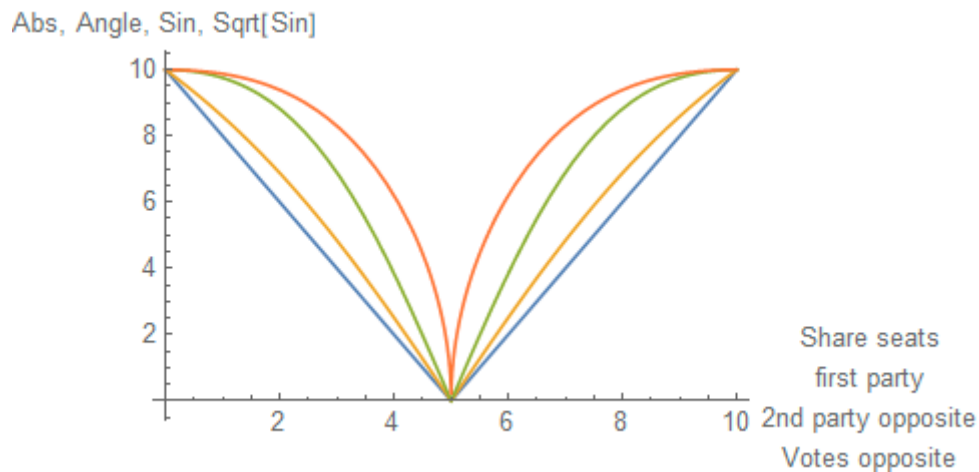
⁷ For CWSID, see Colignatus (2018ad). This present paper will not discuss it.

= $\{t, 10 - t\}$, with t the seats for the first party. We also consider opposite values $votes = \{10 - t, t\} = 10 - seats$. Since this implies negative correlation, the SDID becomes negative, but the plot gives the absolute value. To wit:

- (1) The angular measure (AID) is $10 \theta / (90 \text{ degrees})$ (yellow), Since we look at nonnegative vectors, the maximum angle is 90 degrees.
- (2) The sine plotted is $10 \text{ Sin}[\theta]$ (green). For small angles there is little difference between $\text{Sin}[\theta]$ and θ in radians. There is a large difference between $\text{Sin}[\theta]$ and $\theta / (90 \text{ degrees})$.

The key point of **Figure 1** is that the SDID indeed works like a magnifying glass to determine inequalities / disproportionalities in votes and seats. This relates to the Weber-Fechner law on psychological sensitivity.⁸ When a frog is put into a pan with water at room temperature and subsequently is slowly boiled it will not jump out. When a frog is put into a pan with hot water it will jump out immediately. People may notice big differences between vote shares and seat shares, but they may be less sensitive to small differences, while these differences actually can still be quite relevant for the decision to jump out. For this reason, the SDID uses a sensitivity transform. Like with the Richter scale, it will now be easier to relate the smaller values to the larger values. At the values $t = 4.5$ or 5.5 , when the absolute distance ALHID registers a 1 on a scale of 10, SDID generates a staggering 4.4 on a scale of 10, which outcome better relays the message that this difference is alarming.

Figure 1. Plot of $d[votes, seats]$ for $votes = 10 - seats$ and $seats = \{t, 10 - t\}$, for $d = \text{Abs}/2$, AngularID, Sine, and |SDID| (eliminating the latter's negative sign)



As another example: When votes $\{4.9, 5.1\}$ are translated into seats then the absolute difference (ALHID) and the Euclidean distance (EGID) regard outcomes $\{4.8, 5.2\}$ or $\{5.0, 5.0\}$ as at the same distance, namely 0.1 seat difference (correcting for double counting), while common sense and the sine would hold that the seats $\{4.8, 5.2\}$ are closer to the votes and less disruptive than the seats $\{5.0, 5.0\}$ that suggests that there is equality. The values are: $10 \text{ Sin}[\{4.9, 5.1\}, \{4.8, 5.2\}] = 0.1998 < 0.19996 = 10 \text{ Sin}[\{4.9, 5.1\}, \{5.0, 5.0\}]$.⁹ These values are so close together, though, that also the magnifying glass SDID hardly sees a difference: $1.41351 < 1.41407$. However, the latter values are still at the high level or 1.4 on a scale of 10, rather than at the low value of 0.1 on a scale of 10 for ALHID.¹⁰

⁸ Wikipedia is a portal and no source:

https://en.wikipedia.org/wiki/Weber%E2%80%93Fechner_law

⁹ We didn't divide the sine by 2 to correct for double counting. If we would do so then $0.1998 / 2 \approx 0.1$ or the ALHID score. But then we would have to multiply by 20 instead of 10 to get to the $[0, 10]$ range again. For small values $\text{Sin}[w, z] = \text{Sin}[\theta] \approx \theta \approx \|z\| / \|z\| - w / \|w\|$ and for the unit simplex we might consider $\|z - w\| / \sqrt{(\|z\| \|w\|)}$ using the geometric mean.

¹⁰ Readers familiar with percentages might compare with the case of $\{49, 51\}$, where 1 seat in a House of 100 is relocated. The ALHID recovers the 1% but SDID magnifies to a score of 14 on a scale of 100. This 1% of the US House is 4.35 seats, and of the UK House 6.5 seats, or,

Table 1 contains the real world example of the US House of Representatives of 2016 (435 seats) and the UK House of Commons 2017 (650 seats). SDID properly conveys the insight that there is shocking inequality / disproportionality.

Table 1. Votes and seats in the USA 2016 and UK 2017 ¹¹

<i>USA, House, November 8 2016, S = 435</i>			<i>UK, House, June 8 2017, S = 650</i>		
<i>Party</i>	<i>Votes</i>	<i>Seats</i>	<i>Party</i>	<i>Votes</i>	<i>Seats</i>
Republicans	4.91	5.54	Conservatives	4.22	4.88
Democrats	4.80	4.46	Labour	3.99	4.03
Other	0.29	0	Other	1.79	1.09
<i>100%</i>	10	10	<i>100%</i>	10	10
10 ($z_L - w_L$)	0.63		10 ($z_L - w_L$)	0.66	
ALHID	0.63		ALHID	0.7	
AID	0.67		AID	0.92	
SDID	3.2		SDID	3.8	

1.3. Structure of the paper

The square root within SDID is psychologically important and only a presentation feature of descriptive statistics. This present discussion collects the key steps in Colignatus (2018d) for the content of statistics and targets at an overview. Part of the present text has already been used on my weblog Colignatus (2017b).

The next section provides notation and basics. The subsequent section places our topic within the perspective of the statistical triad of experimental design, description and decision. Subsequently we apply the cosine and sine for nonnegative data, and state the relevant formulas. We close with a summary of the findings.

The reader might be advised to peek at **Table 2** for a glance at the different models, to see what this discussion is about specifically. I have considered putting this table up front, but it remains better to first rekindle awareness about the basics before delving into the models.

PM 1. See Colignatus (2007) for the approach with determinants rather than angles - as area and volume might generalise easier to more dimensions than the angle, that might remain stuck to the 2D plane created by the two vectors. PM 2. There is also the notion of "distance correlation" ¹² but the Pearson correlation remains relevant here precisely because of the linearity contained in the notion of equality / proportionality. PM 3. There is "least angle regression" ¹³ but such is different, and we remain in the realm of "simple regression". ¹⁴

with double effect 8.7 and 13 seats. This 1% might make quite a difference. The value of 1.4 on a scale of 10 would seem to be acceptable as the indication that something is wrong.

¹¹ The interpretation of this table requires **Section 5.3**.

¹² Wikipedia is a portal and no source. https://en.wikipedia.org/wiki/Distance_correlation

¹³ https://en.wikipedia.org/wiki/Least-angle_regression

¹⁴ https://en.wikipedia.org/wiki/Simple_linear_regression

2. Notation and basics

2.1. Well-known basics

We use the underlined upper case \underline{X} for the centered value $\underline{X} = x - \text{Mean}[x]$. The angle θ^* is between the centered values \underline{X} and \underline{Y} . The Pearson correlation coefficient is $r[x, y] = \text{Cos}[x - \text{Mean}[x], y - \text{Mean}[y]] = \text{Cos}[\underline{X}, \underline{Y}] = \text{Cos}[\theta^*]$ so that $\theta^* = \text{ArcCos}[r[x, y]]$. The covariance of x and y is the improduct of the centered values, divided by the number of observations n , or $\text{cov}[x, y] = \underline{X}' \underline{Y} / n$. Using covariance, the correlation coefficient $r[x, y] = \text{cov}[x, y] / \sqrt{(\text{cov}[y, y] \text{cov}[x, x])}$. See e.g. Egghe & Leydesdorff (2009) for a visualisation of the shift towards centered data, and Theil (1971:165) for a discussion of the geometric meaning that $r = \text{Cos}[\theta^*]$.

Along (1) θ and (2) θ^* , we also consider the linear cases of (3) the “regression through the origin” (RTO) for y given x , without a constant, (4) the “regression” in general, with a standard constant, hence for \underline{Y} given \underline{X} . For linearity, the standard case with a constant may also be formulated in terms of y and x , but some formulas then require the mention of the means. The symbol \hat{y} will be the estimate of y , but e and \hat{e} are just different kinds of error.

For (4) with centered $\underline{Y} = \underline{X} b + \hat{e}$ for matrix \underline{X} , then $\underline{Y}' \underline{Y} = b' \underline{X}' \underline{X} b + \hat{e}' \hat{e}$ using $\underline{X}' \hat{e} = 0$. This is commonly expressed as $\text{SST} = \text{SSX} + \text{SSE}$. In this, $\underline{Y}' \underline{Y} = \text{SST} = \text{sum of squares total}$, $\hat{e}' \hat{e} = \text{SSE} = \text{sum of squared errors}$, and $\text{SSX} = \text{sum of squares of the explanation} = \text{SST} - \text{SSE}$.¹⁵

The coefficient of determination is $R^2 = \text{SSX} / \text{SST}$. Thus $1 - R^2 = \text{SSE} / \text{SST}$. For the calculation of SST and SSE we must use centered data, though the regression itself might also be formulated as $y = X b + \hat{e}$ with a column of 1 in X for the constant. With n observations and m explanatory variables in X , the root mean squared error (RMSE) adjusted for the degrees of freedom is $\text{RMSE} = \sqrt{(\text{SSE} / (n - m))}$. Colignatus (2006) discusses the sample distribution of (adjusted) R-squared. When the (explanatory) variables are given without measurement errors then there is not a “population” but a “space”, and the relevant parameter for R is best denoted as $\rho[X]$ to express the conditionality on the data.

While the coefficient of determination, R-squared, in this setup seemingly has an independent definition as SSX / SST , it appears that it is actually the square of the correlation r between y and its estimate \hat{y} . It might be possible to present this identity as a great insight and wonder, but it is better to infer that such independence of definition actually wasn't possible. It is better to start with $r[y, \hat{y}]$ and then show that steps in its calculation can be abbreviated as SST, SSX and SSE. The R-squared for two vectors thus is the squared cosine of the angle θ^* between the centered values. The root mean squared error (RMSE) then relates to $\text{Sin}[\theta^*] = \sqrt{(1 - R^2)} = \sqrt{(\text{SSE} / \text{SST})}$, as $\text{RMSE} = \text{Sin}[\theta^*] \sqrt{(\text{SST} / (n - m))}$.

The angle itself is a measure of distance. The angle divided by 360 degrees gives a measure in $[0, 1]$. See Colignatus (2009, 2015) for a suggestion to measure angles on $[0, 1]$ anyway, using the plane itself as the unit of account, and to speak about turns. When it doesn't matter whether the angle is positive or negative, then 180 degrees would be a relevant maximum. For nonnegative data the relevant maximum is 90 degrees. Subsequently 1 minus such value is an angular measure of association.

The cosine is a measure of association too. Some have suggested to take $1 - \text{Cos}$ as a distance, calling it “cosine distance”, but this actually is not a metric, see Van Dongen & Enright (2012). The latter authors clarify that the sine and its root are a metric. Using $\text{Sin}[\theta^*] = \sqrt{(1 - R^2)}$ as a measure of distance might be dubious. The square root causes a management of the signs, and when the angle is larger than 90 degrees, then the same values of the sine can only be distinguished by looking at the sign of the cosine again. It might well be that this present discussion remains relevant for nonnegative vectors only.

¹⁵ Often the abbreviation SSR is used, $\text{SSR} = \text{sum of squares of the regression}$, but then confusingly with $\text{SSR} = \text{sum of squares of residuals} = \text{SSE} = \text{sum of squared errors}$. The use of SSX has X , of both “explanation” and the variable x in $y = b x + e$.

2.2. Regression through the origin (RTO), for nonnegative vectors

At stake now is the use of θ of the original and non-centered vectors, which leads us to regression through the origin (RTO). Also, we consider nonnegative vectors. In this case, the angle between the vectors is between 0 and 90 degrees.

With $k = \text{Cos}[\theta] = \text{Cos}[x, y]$, and faced with a choice of a distance measure, there is no fundamental difference between using $\theta = \text{ArcCos}[k]$ or $\text{Sin}[\theta] = \text{Sqrt}[1 - k^2]$. The advantage of using the sine instead of the angle is that we have some interpretations in terms of regression because of the cosine. Also, $\theta / (90 \text{ Degree})$ has a lower slope than the sine at small values of the angle, whence the sine is more sensitive, which fits our purposes. While we can add lengths and angles, we cannot add values of the sine though.

There are slopes b and p from the regressions through the origin (RTO) $z = b w + e$ and $w = p z + \epsilon$. Then $k = \text{Cos}[v, s] = \text{Cos}[w, z] = \text{Sqrt}[b p]$. The geometric mean slope is a symmetric measure of similarity of the two vectors. Also $\text{Sin}[v, s] = \text{Sin}[w, z] = \text{Sin}[\theta] = \text{Sqrt}[1 - b p]$ is a metric and a measure of distance or inequality or disproportionality in general.

All this is straightforward but there are some reasons to call attention to it.

- Political scientists have been looking for a sound inequality / disproportionality measure without finding one, i.e. not finding the sine. They have been setting for the less adequate Euclidean distance EGID discussed (not proposed) by Gallagher (1991), though with an awareness that it wasn't perfect, see Taagepera & Grofman (2003), Karpov (2008) and Koppel & Diskin (2009). Colignatus (2018d) provides an overall evaluation.
- The use of angle, cosine and sine gives a perspective on compositional data (i.e. nonnegative values on the unit simplex), see Barceló-Vidal and Martín-Fernández (2016). This is discussed in Colignatus (2018a). Votes and seats are somewhat special data however. Often seats are apportioned given the votes and the available seats in the House, as $\hat{s} = Ap[v, S]$. This means that the influence of S cannot be neglected as perhaps may be done in "pure" compositional data.
- Compositional data generally would use a log transform (with the geometric average as the mean) and drop one equation because of the addition condition. For the present paper, however, we don't employ statistical decision theory, with regression as one of its applications, but we employ statistical description (without distinction between true coefficient b and its estimate). For determination of the angle between the vectors all elements are relevant, though still with the scalar invariance. A contribution to RTO by (seeming) "compositional data" like votes and seats is that we consider two errors, not only $z = b w + e$ (used by SDID) but also $z = w + \tilde{e}$ (used by ALHID, EGID and relatively by CWSID).

The inoptimal situation of discussion in these different though related areas might have to do something with that, remarkably, the angular distance and the sine tend not to be in the textbooks.¹⁶ Currently, errors are related to the *normal distribution*. From there, the Euclidean distance $\|y - \hat{y}\| = \|\hat{e}\| = \sqrt{\text{SSE}}$, is a measure of variation and estimator for true variation, whence regression is explained, and proceeds by minimising SSE to find the normal equations and solve for coefficients. Instead it might make more sense to explain that the angle can be used as a distance measure, and minimizing the angle means maximising $\text{Cos}[y, \hat{y}]$. For regression with a constant this involves $\text{Cos}[y, c + b x]$.

However, a binary regression in RTO has $y = b x + e$ and $\hat{y} = b x$, so that $\text{Cos}[y, \hat{y}] = \text{Cos}[y, x]$ because of scale invariance, so that the angle is fixed. We still require another criterion than the angle on y to find the parameter b .

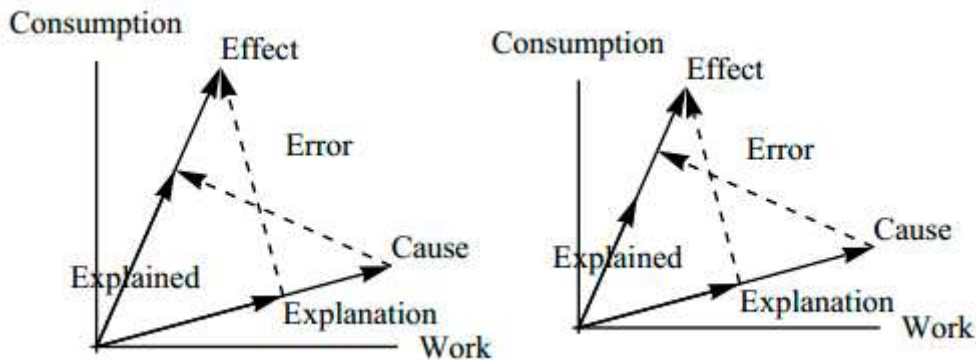
- For RTO, $\text{SSE} = e'e = (y - b x)'(y - b x) = y'y - 2 b y'x + b^2 x'x$ gives a minimum for the estimate $b = y'x / x'x$. On occasion $b = \sqrt{(y'y / x'x)} \text{Cos}[y, x]$ may be useful.

¹⁶ My sample though contains Mood & Graybill (1963) and Johnston (1972) and isn't very much larger.

- From $y = b x + e$ the alternative approach is to $x'y = b x'x + x'e$, take $x'e = 0$, so that $b = y'x / x'x$. This is shorter but requires geometry instead of calculus. This namely projects y onto x and imposes perpendicularity between x and e , so that $\text{Cos}[x, e] = \text{Cos}[x, y - b x] = 0$ which gives $x'e = 0$. The projection of y on x is given by $b x$, with $b = x'y / x'x$ (using projection matrix $P = x x' / x'x$). **Figure 2**, taken from Colignatus (2011:143), shows the geometry how y is projected onto x , which determines the size of the Explanation ($b x$), so that y follows from addition of perpendicular $b x$ and e . In the figure, the LHS is normalised with Cause and Effect on the unit circle, so that the coefficient is Cos or R at 63%. The RHS is not normalised with coefficient $b = 0.546$.

Thus there is a geometric approach that is at least as intuitive as the minimisation of SSE with reference to the normal distribution. The conceptual link between perpendicular x and e and minimal SSE need not be intuitive however, and can only be proven exactly by looking at the normal equation.

Figure 2. Projection of Effect {4, 9} on Cause (11, 3) ¹⁷



Both approaches generate the same solution, and there is only the difference in presentation, either via the angles between the vectors or the Euclidean norm of the error. Both relate to the assumption of i.i.d. normal errors, but the latter can also be seen as step 2, when one agrees that it is more informative to start with the use of the angle as step 1, rather than derive the angle as a corollary or be actually silent on it.

My suggestion is that the world of statistics develops a greater awareness of the angle and sine as distance metrics in relation to R-squared, at least for applications of RTO for such nonnegative data. For textbooks in statistics, this particular combination might be regarded as a missing link. The geometry would contribute to a better understanding of students of both regression and the distinction between descriptive statistics (no distinction between b and an estimate) and statistical decision theory (a true b and its estimate).

Shalizi (2015:19) states that “I have never found a situation where it [R-squared] helped at all.” – see also Ford (2015). This statement is somewhat curious where minimising the sum of squared errors is equivalent to maximising R-squared, so that R-squared is crucial. Obviously one should be careful in interpreting the actual outcome of a R-squared calculation, e.g. in specification search.

¹⁷ The RHS: The projection of Effect y on Cause x generates the Explanation ($b x$). Addition of $b x$ and perpendicular Error e (from Explanation to Effect) then generates y again. The LHS: Cause and Effect are normalised into the unit circle. Then $y = b x + e$ gives $y^* = b^* x^* + e^*$, using $y^* = y / \|y\|$, $x^* = x / \|x\|$, $e^* = e / \|y\|$, and $b^* = b \|x\| / \|y\|$. Obviously $b^* = \text{Cos}[y, x]$. Thus the Explanation ($b^* x^*$) on the LHS gives the Cos, in this example 63%. For the Explained part of the Effect we use $y^* y^* = b^* x^* x^* b^* + e^* e^*$ or $1 = b^{*2} + e^* e^*$. We can say that b^* of y is explained too. Thus on the LHS Cause and Effect are on the unit circle (radius 1) and the smaller arrows from the origin are on the circle with radius 0.63.

2.3. A view from didactics

In education it happens far too often that a textbook starts a new section with “Now something completely different” while it appears that one essentially has the same topic though only from another perspective. Consider for example $\{x, y\} = x + i y = r (\text{Cos}[\varphi] + i \text{Sin}[\varphi]) = \text{Exp}[\rho + i \varphi]$. Obviously such perspectives exist and each perspective has something to say for it, and obviously it is a result in itself when one can show the equality. But one also feels rather exasperated, finding out that one only learns different languages for the same.

In the same way for correlation. The “explanation” on correlation, that “correlation between two vectors is the cosine of the angle between the centered data”, is only required because the word “correlation” has been introduced without explicit reference to the basic angularity of the notion. Most students, who first are introduced to “covariance” and who then are presented with correlation by a formula that uses covariance, will miss out on the notion that correlation refers to an angle. Even when this is derived it tends to remain a mystery because students have built up associations on “correlation” that tend to be quite different from angles.

Perhaps it was a deliberate decision by Pearson to keep trigonometry outside of the realm of statistics, for fear that the subject might appear more dreadful than needed. Instead, it would be better to make trigonometry more acceptable, see “Trig rerigged” in Colignatus (2009, 2015). When α is the size of the turn on the angular circle with circumference 1, then functions $X = Xur[\alpha]$ and $Y = Yur[\alpha]$ give the values of $\{X, Y\}$ on the unit radius circle, such that $X^2 + Y^2 = 1$. The upper case variables X and Y , as standard co-ordinates on the unit circle, have the same role as the lower case variables x and y as the standard co-ordinates for the plane. This approach avoids the mysterious names and uninformative labels “sine” and “cosine”, and avoids the needless calculation that 90 degrees is a quarter of 360 degrees. Useful is also $\Theta = 2\pi$, pronounced as “archi” from Archimede, and written with capital theta.¹⁸ The only reason to stick to the use of sine and cosine defined on the unit radius circle itself is that their derivatives translate into each other. Once students have learned trigonometry with Xur and Yur while avoiding the current opacity, there would be easier acceptance of sine and cosine for who can deal with calculus anyhow.

With Xur defined (derived) for vectors as $Xur[\alpha] = Xur[x, y] = (x / ||x||)' (y / ||y||)$, and a new function for centered data as $XurCD[x, y] = Xur[x - \text{Mean}[x], y - \text{Mean}[y]]$, then the meaning of this $XurCD$ should be clear. With each value of $X = Xur$ there is a $Y = Yur$ on the unit radius circle, and the angle α that gives the size of the turn in $[0, 1]$. Once the meaning is clear, there is obviously no objection to calling $XurCD$ “correlation”, though.

It would be preferable to first set up the basic structure of correlation and regression in this clean manner, before entering upon the error distribution, such that “mean” is replaced by “expectation”, and with such use of covariance. Present textbooks (curriculum) however shy away from re-engineering trigonometry, and work around corners by introducing correlation as if it were something really new. In practice they block the understanding by many.

3. Evolving statistics

3.1. The statistical triad of Design, Description and Decision

Statistics has the triad of Design, Description and Decision. Up to fairly recent, statistics relied much upon the paradigm by R.A. Fisher that focused on population and sample distributions. With the dictum “correlation is not causation”, statistics assumed that causation was given by the scientific model, and then concentrated on correlation for cases with clear causality. Since Pearl (2000), the issue of causality is more in focus again, though this doesn't change the triad.

¹⁸ <https://boycottholland.wordpress.com/2014/07/14/an-archi-gif-compliments-to-lucas-v-barbosa/>

- Design is especially relevant for the experimental sciences. Design is much less applicable for observational sciences, like macro-economics and national elections when the researcher cannot experiment with nations.
- Descriptive statistics has measures for the center of location – like mean or median – and measures of dispersion – like range or standard deviation. Important are also the graphical methods like the histogram or the frequency polygon. Measures like the Richter scale for earthquakes belong in this category too. Description relates to decisions on content (e.g. in medicine or economics). Description becomes more important because of Big Data.
- Statistical decision making involves the formulation of hypotheses and the use of loss functions to choose alpha and beta values to evaluate hypotheses. A hypothesis on the distribution of the population provides an indication for choosing the sample size. A typical example is the decision error of the first kind, i.e. that a hypothesis is true but still rejected. The probability of that error, the alpha, is called the level of statistical significance. This notion of statistical significance differs from causality and decisions on content. (See e.g. Varian (2015).)

Historically, statisticians have been working on all three areas, but the most difficult was the formulation of decision methods, since this involved both the calculus of reasoning and the more involved mathematics on normal, t, chi-square, and other frequency distributions. In practical work, the divide between the experimental and the non-experimental (observational) sciences appeared insurmountable. The experimental sciences have the advantages of design and decisions based upon samples, and the observational sciences basically rely on descriptive statistics. When the observational sciences do regressions, there may be an ephemeral application of statistical significance that invokes the Law of Large Numbers, that all error is approximated by the normal distribution.

This statistical tradition is being challenged by Big Data including the ease of computing – see also Wilcox (2017). When the relevant data are available, and when you actually have the *space* or *population data*, then the idea of using a sample may evaporate, and you would not need to develop hypotheses on those distributions anymore. In that case descriptive statistics tends to become the most important aspect of statistics. Decisions on content then are less compounded by statistical decision making on statistical phenomena. It comes more into focus how descriptive statistics relate to decisions on content. Such questions already existed for the observational sciences like for macro-economics and national elections, in which the researcher only had descriptive statistics, and lacked the opportunity to experiment and base decisions upon samples. The disadvantaged areas may now provide insights for the earlier advantaged areas of research.

The suggestion is: to transform the loss function into a descriptive statistic itself. An example is the Richter scale for the magnitude of earthquakes. A measurement on that scale is both a descriptive statistic and a factor in the loss function. A community making a cost-benefit analysis has on the one hand the status quo with the current risk on human lives and on the other hand the cost and benefit of investments in new building and construction including the risk of losing the investments and a different estimate on human lives. In the evaluation, the descriptive statistic helps to clarify the content of the issue itself. For the amount of destruction it would not matter how earthquakes are measured, but for human judgement it would, as the human mind need not be sensitive to relevant differences. The key issue is no longer a decision within statistical hypothesis testing, but the key issue is the adequate description of the data and the formulation of the decision problem in terms for better human understanding of what is involved.

3.2. A possible reason why RTO may be less prominent in the textbooks

Statistics and specifically textbooks apparently found relatively little use for original (non-centered) data and RTO. A possible explanation is that statistical theorists fairly soon regarded descriptive statistics as less challenging, and focused on statistical decision making. Textbooks prefer the inclusion of a constant in the regression, so that one can test whether it

differs from zero with statistical significance. The constant is essentially used as an indicator for possible errors in modeling. The use of RTO or the imposition of a zero constant would block that kind of application. This (traditional, academic) focus on statistical decision making apparently caused the neglect of a relevant part of the analysis, that now comes into focus again.

I am not familiar with the history of statistics – see Stigler (2008), Lehmann (2008) and John Aldrich's website¹⁹ – and it is unknown to me what Pearson (1857-1936), Gosset (1876-1937), Fisher (1890-1962) and other founding and early authors wrote about the application of the cosine or sine, other than what transpires from current textbooks. The choice to apply the cosine to centered data to create correlation and R -squared is deliberate. Pearson would have been aware that the cosine might also be applied to original (non-centered) data, but he rejected this for his purposes on variation. RTO is available in the mantra, though, and not obliterated.²⁰ This history is interesting yet history is not my focus. Quite likely the theoretical challenge was determined by the lack of Big Data. Thus we can understand that these founders focused on statistical decision making and hypotheses on distributions rather than on description.

3.3. Statistical significance

Part of the history is that R.A. Fisher with his attention for mathematics emphasized precision for statistical purposes while W.S. Gosset with his attention to practical application on content emphasized the effect size of the coefficients found by regression. Somehow, precision in terms of statistical significance became more important in textbooks than content significance, and empirical research has rather followed Fisher than the practical relevance of Gosset. This history and its meaning is discussed by Ziliak & McCloskey (2007), see also the discussion by Gelman (2007) referring to Gelman & Stern (2006).

3.4. Causality

Since the cosine is symmetric, the R -squared is the same for regressing y given x , or x given y . Shalizi (2015, p18) infers from the symmetry: "This in itself should be enough to show that a high R^2 says nothing about explaining one variable by another." This is too quick, with too much reliance on "in itself". When theory shows that x is a causal factor for y then it makes little sense to argue that y explains x conversely. Thus, for research the percentage of explained variation can be informative. Obviously it matters how one actually uses this information. For standardised variables y and x (difference from mean, divided by standard deviation),²¹ $y = R x$, so that the regression coefficient is the R , and then the R -squared can also be understood with attention for the effect size. For some applications a low R -squared would still be relevant for the particular field. Researchers do not tend to work with standardised variables, and don't have to when the R is available by itself.

For standardisation, let s_x and s_y be the standard deviations, and $\underline{y} = Y / s_y$ and $\underline{x} = X / s_x$ the standardised variables. Then $Y = b X + e$ gives $\underline{y} = (b s_x / s_y) \underline{x} + e / s_y$ or $\underline{y} = r \underline{x} + \underline{e}$, using $r = b s_x / s_y$ and defining $\underline{e} = e / s_y$. We may also standardise $\underline{Y} = b \underline{X}$. Its standard deviation is $b s_x$, and thus $\underline{\hat{y}} = \underline{\hat{Y}} / (b s_x) = \underline{\hat{x}}$. Thus we may also write $\underline{y} = r \underline{\hat{y}} + \underline{e}$. This could be non-informative on details for more variables, though.

3.5. Specification search

When it is reported that a regression has an R -squared of say 70% then this means that 70% of the variance of y is explained by the variance of \hat{y} . In itself such a report does not say much, for it is not clear whether 70% is a little or a lot for the particular explanation. For

¹⁹ <http://www.economics.soton.ac.uk/staff/aldrich/kpreader.htm> and more general <http://www.economics.soton.ac.uk/staff/aldrich/figures.htm>

²⁰ https://en.wikipedia.org/wiki/Simple_linear_regression

²¹ http://andrewgelman.com/2009/07/11/when_to_standard/

evaluation we obviously also look at the issue on content (and the regression coefficients). The use of R-squared is primarily for specification search.

One can always increase R-squared by including other and even nonsensical variables. For a proper use of R-squared, we would use the adjusted R-squared. R-adj finds its use in model specification searches - see Giles (2013). For an increase of R-adj, coefficients of new variables must have an absolute t-value larger than 1. A proper report would show how R-adj increases by the inclusion of particular variables. A researcher would compare to studies by others on the same topic. Comparison on other topics obviously would be rather meaningless. Shalizi (2015) also rejects R-adj and suggests to work directly with the mean squared error (MSE), also corrected for the degrees of freedom. Since R-squared is the squared cosine, then the MSE relates to the sine, and these are basically different sides of the same coin, so that this discussion is much a-do about little. As said, for standardised variables, the R-squared also generates the regression coefficient, and then it is relevant for the effect size.

Giles (2013) restates the "uselessness" of R-squared: "My students are often horrified when I tell them, truthfully, that one of the last pieces of information that I look at when evaluating the results of an OLS regression, is the coefficient of determination (R²), or its "adjusted" counterpart. Fortunately, it doesn't take long to change their perspective!" Such a statement should not be read as providing the full clarification on cosine or sine in general, or as rejection of the relevance of the effect size also for $y = R \hat{y} + u$.

R-squared is not devoid of meaning. For a satisfactory regression it sets the level that must be surpassed by the next satisfactory regression. Reporting on it is important for future researchers, though they would have to use the same dataset.

4. Application to votes and seats

4.1. Descriptive statistics and decisive apportionment

Vectors s and $z = s / s^{\prime 1}$ have been created by human design upon v , and not by some natural process as in common statistics. A statistical test on $s | v$ would require to assume that seats have been allocated with some probability, and this doesn't seem to be so fruitful when there was an underlying system of rules. We can use the same linear algebra however, now for descriptive statistics.

The ID measures are used to compare outcomes of electoral systems across countries, though such comparisons have limited value when countries have different designs. Taagepera & Grofman (2003) mention also some other reasons for an ID measure: (i) comparison on President, Senate, House, or regional elections (what they call "vote splitting" but is better called: votes for different purposes), (ii) comparison on years in similar settings (both votes and seats) (what they call "volatility" but what is better called: votes on different occasions).

Above measures ALHID, EGID and CWSID have drawbacks and are inoptimal. There appears to be some distance between the voting literature on inequality / disproportionality and the statistics literature on association, correlation and concordance. A main point is that voting uses $\tilde{e} = z - w$ (conventionally) and now $z = b w + e$ (for SDID) as descriptive, while statistical theory tends to think in terms of hypotheses tests on general relationships like $s = c + B v + u$ and then require stochastics.

4.2. Different worlds for votes and seats: DR and EPR

A general distinction is between District Representation (DR) and Equal or Proportional Representation (EPR). "Elections" in systems of EPR differ from those in DR, and we should actually avoid the single term "election" for both cases when the meanings are fundamentally different, see Colignatus (2018c).

EPR recognises that elections for Parliament concern *multiple seats*, such that there are conditions for overall optimality. In DR, each district may have a *single seat* election, with neglect of conditions for overall optimality.

In DR, votes for candidates other than the district winner(s) are not translated into seats, and the system discards those votes. DR-elections have much strategic voting for fear that the vote is lost. The true first preferences thus are masked. Comparing votes and seats is actually comparing masked votes and seats, with unknown discarded votes. Only geography might cause a semblance of balancing at the national level. Also the median voter theorem might cause that voters concentrate around the middle, but this should not deceive us in thinking that we could achieve a proper "comparison" of votes and seats in DR.

4.3. Apportionment in EPR

Only in EPR there is a deliberate *apportionment of the seats* given the votes.

- In general the apportionment will not be perfect, since the distribution over perhaps millions of votes must be approximated by perhaps a few hundred seats (with integer values). The apportionment involves some political philosophies that have been adopted by the national parliaments.
- There need not be a real distance between the voting literature and statistics, at roots, because (i) the Chi-Square / Webster / Sainte-Laguë (CWSID) apportionment philosophy obviously compares with the Chi square, and (ii) the apportionment according to Hamilton / Largest Remainder (HLR) minimises the absolute difference, or the Loosemore-Hanby index (ALHID), but also minimises the sum of squared differences, or the Euclidean distance (or the EGID index). Perhaps this early historical linkage also caused the presumption that voting theory already "had enough" of what was available or relevant in the theory of statistics.
- Researchers on voting may have a tendency to remain with these philosophies when they *measure* the outcomes from such apportionments too. Apportionment (deciding) and measuring (describing) have different purposes and methods, though there may be a family resemblance.
- When comparing results from different countries, however, it would make sense, to use a common best measure, rather than reporting that each country applies its own method.

There are more aspects, yet this present article does not focus on voting theory but on giving an overview of statistics, that is also applicable for this setting. The present discussion however highlights where comparing votes and seats differs from other purposes in statistics: (i) First the requirement on the diagonal of the scatter plot of w and z . (ii) Secondly, comparing votes and seats cannot rely on stochastic assumptions for testing, and thus wants to describe & measure. We use the same linear algebra but for a different purpose. The discussion helps to see that the choice of an inequality / disproportionality measure apparently is not self-evident, at least with the current literature and textbooks so dispersed over the topics that come together here.

4.4. Different models and errors

We do not want to *explain* s by v , in which case we would be very careful w.r.t. the exclusion of the constant. Instead, we want to design a measure. This still uses the same linear algebra. The relevant distinctions are (i) between true values versus estimates, and (ii) between level variables versus unitised variables.

$Q = V / S$ is the natural quota, or number of votes to cover a seat. There may be a threshold to get a seat, or just the natural quota. Voters may vote for parties that do not pass the threshold and that thus get no seats. This sums to the "wasted vote" W . Standardly the wasted vote and zero seats are collected in one category "Other", so that v and s still have the same length. The votes that cause a seat are $V_e = V - W$. For regression it is conventional to write $s = T v$, so that s is explained by v . We call this vector-proportionality because of the lack of a constant. Any such relation also holds for its sum totals, and we can

usefully define $T = S / V = 1 / Q$. In reality we have $s = B v + u$ or $z = b w + e$ with proportionality parameter b and error e . There is only unit proportionality or equality if $z = w$ or $b = 1$ and $e = 0$. Let $a = S v / V = T v = v / Q = S w$ be the proportionally accurate average of seats that a party might claim. A common error term is $s - a = S (z - w)$. The unit simplex is the natural environment to look at this, though we should not forget about the role of S .

The major distinctions are in **Table 2**.

Table 2. Basic models and their errors

$T = S / V = 1 / Q, B = b T$	Original	Unitised
Without parameter ²²⁾	$s = T v + \tilde{u} = S w + \tilde{u} = a + \tilde{u}$	$z = w + \tilde{e}$ with $\tilde{e} = \tilde{u} / S$
Regression through origin	$s = B v + u$	$z = b w + e$
		$w = \rho z + \varepsilon$
With constant (centered)	$s = c + B v + \hat{u}$	$z = \gamma + \beta w + \hat{e}$

The analysis better uses regression through the origin (RTO) and not regression with a constant (Pearson). The value $a = S w$ will be the average, and apportionment of s will tend to be for $\text{Floor}[a] \leq s \leq \text{Ceiling}[a]$. There are three different error measures:

- \hat{e} from the standard regression with a constant, using centered data (Pearson).
- e from regression through the origin (RTO), with coefficient $b = z'w / w'w$. (SDID uses e .)
- $\tilde{e} = z - w$ or "direct error" for the plain difference, using $b = 1$. (ALHID and EGID use \tilde{e} .)

Some useful mnemonics directly are:

- $\hat{e}'\hat{e} \leq e'e \leq \tilde{e}'\tilde{e}$ because regression parameters allow the reduction of error.
- e picks up a potential source for proportionality that \tilde{e} does not allow for.
- $1' \tilde{e} = 0$ and $b = 1 - 1'e$ because $1'z = 1'w = 1$.
- $b = z'w / w'w$ because we multiply $z = b w + e$ with w' while $w'e = 0$. The regression selects the b with perpendicular w and e , or with $w'e = 0$ and minimal $e'e$.
- Taking the plain differences $\tilde{e} = z - w$ and weighing them by the vote shares and normalising on their squares, gives $w' \tilde{e} / w'w = b - 1 = -1'e$. This might perhaps be seen as an "implicit estimate".

For centered data, we had a simple direct relation between on the one hand the Euclidean norm of the error, via SSE and RMSE, and on the other hand the sine $\sqrt{(1 - R^2)} = \sqrt{(SSE / SST)}$. We now wonder about RTO. Observe that the Euclidean norm of the error is $SSE = \sum e'e = \|e\|$ while the Euclidean distance between the vectors is $\sqrt{\tilde{e}' \tilde{e}} = \|\tilde{e}\| = \|z - w\|$.

4.5. Disproportionality, dispersion and education

For education we want to maintain that we want to explain to students that any line through the origin in 2D represents a proportional relationship. For vectors and their scatter plot we would speak about "vector-proportionality". Thus, also for z and w , a relation $z = \alpha w$ is a vector-proportional relationship, and $z = w$ is only *unit* or *diagonal* vector-proportional. It so happens that the unit simplex is defined such that $1'z = \alpha 1'w$, thus any pure proportional relation in that space requires $\alpha = 1$ of necessity. I would phrase this as that the space is defined such that those other *pure* vector-proportions *do not exist*. I would *not* phrase it as saying that (quote) " $z = \alpha w$ for $\alpha \neq 1$ would not be vector-proportional and that only $\alpha = 1$ is vector-proportional" (unquote). It is better to say: we only have $z = b w + e$ with vector-proportionality parameter b and scattered e . Thus, overall, it would be didactically preferable to speak about "unit or diagonal proportionality" for voting rather than "proportionality". Only

²² This is not unreasonable for official data, when a law declares some $\hat{s} = Ap[S, v]$ to be optimal. Regression then would misstate the error, though b would still be a measure.

when $e = 0$ then we also have equality. *Equal Representation* is better anyway. It may be difficult to change a convention, but it would help for mathematics education.

This calls attention to the relation with dispersion. **Table 3** reviews the relations. The key relationship in RTO is that $b = 1 - 1'e$. Thus $b = 1 \leftrightarrow 1'e = 0$. The upper right cell is impossible: $\text{Not}[e = 0 \ \& \ 1'e \neq 0]$. The lower right case of disproportionality implies dispersion, but dispersion ($e \neq 0$) need not imply such disproportionality. The left column with $b = 1$ is in opposition to the right column, but we must distinguish between *unit* or *diagonal* vector-proportionality without dispersion (equality) and *average unit* or *average diagonal* vector-proportionality with dispersion.

Table 3. Disproportionality and dispersion in Regression Trough the Origin (RTO)

RTO: $b = 1 - 1'e$	$1'e = 0 \ \& \ b = 1$	$1'e \neq 0 \ \& \ b \neq 1$
<i>No dispersion</i> $e = 0$	<ul style="list-style-type: none"> • <i>Unit or diagonal</i> vector proportionality • Equality 	<ul style="list-style-type: none"> • Logically impossible
<i>Dispersion</i> $e \neq 0$	<ul style="list-style-type: none"> • <i>Average unit</i> vector proportionality can come with dispersion. E.g. $w = \{3, 3, 2\}/8$ and $z = \{2, 1, 1\}/4$. 	<ul style="list-style-type: none"> • Disproportionality by slope and dispersion

With $b = 1$ then $e = z - b w = z - w = \check{e}$. We already had $1' \check{e} = 0$ but $b = 1$ causes also $1'e = 0$. With $b \neq 1$ then $w'(z - w) = 0$, of which $z - w = 0$ is only a special case.

4.6. True variables v^* and s^* and particular observations v and s

A proportional relationship for 1D variables is best described by the 2D line $\lambda y + \mu x = 0$, which coefficients may be normalised on the unit circle. For nonzero λ this reduces to $y = T x$ with $T = -\mu / \lambda$, where slope T also is the tangent of the angle of the line with the (horizontal) x -axis. For vectors this generalises into vector-proportionality, with now a plane $u = \lambda y + \mu x$ and then choosing $u = 0$ so that $y = T x$ again. For example, $y = \{1, 2, 3\}$ and $x = \{2, 4, 6\}$, then $T = 1/2$, and we would see a line without dispersion in the scatter plot.

- The notion of unit proportionality as in the line $y = 1 x + 0$ is a mathematical concept, while in statistics with dimensions we can rebase the variables, so that there need not be a natural base for 1.
- For voting, there are natural bases in the individuals and seats. Larger parliaments may have more scope for a better fit. Still, normalisation onto the unit simplex makes sense.
- At first sight it is not clear where voting differs from other applications, say the ratio of 1 car per 2 persons. Any vector-proportional relation $s = T v$ also holds for the totals. Thus $S = 1's = T 1'v = T V$. In $\{w, z\}$ space it reduces to a scatter with diagonal $z = w$ because division gives $z = s / S = T v / S = w$. When any vector-proportional relationship (also non-unity) is transformed onto the unit simplex, then they become unit or diagonal proportional in the scatter, and we lose the original information.
- There also is dispersion. For the sake of understanding this issue of proportionality, but not for the sake of estimation and hypothesis testing, we now distinguish true elements and their observations. The basic solution namely is to distinguish the true vector-proportionality $s^* = T^* v^*$ that holds for all *true elements* on one hand, and on the other hand observations (e.g. errors in variables) $s = B v + u$, for which we only have the definition for the *sum totals* as $T = S / V$. This also generates \check{u} from $s = T v + \check{u}$.
- For theory we have $z^* = w^*$ but for the data $z = w b + e$. We divided by S and took $b = B / T$ and $e = u / S$. Thus we should not focus only on parameters B and b but also be aware of the hidden $Q = V / S$ or $T = 1 / Q$, and perhaps consider T as an estimate on T^* .

4.7. Cos, slope and concentrated numbers of parties (CNP)

We have $b = z'w / w'w = \sqrt{(z'z / w'w)} \text{Cos}$. The relation between Cos and b is by means of $w'w$ and $z'z$. They take the place of the covariances that are not used in RTO, and are also known as the Hirschman-Herfindahl concentration indices. They are known in the voting literature by the inverse "effective number of parties" $Nv = 1 / w'w$ and $Ns = 1 / z'z$. Since it hasn't been clarified what "effectiveness" would be, a better term is "concentrated number of parties" (CNP).

For theory we have $s^* = T^* v^*$ for the vectors, but for the data we have $s = B v + u$ and thus only $T = S / V$ for the totals. Thus there are not only parameters B and b but also an "estimate" T on T^* (or perhaps institutionally given $T = T^*$). **Table 4** reviews the relations. For readability we drop the stars in the theory column.

Table 4. Norm and estimation, in levels and unitised, regression through origin (RTO)

Norm or true situation ^{23 24)}	RTO in levels	RTO on unit simplex
Theoretical s and v	Observed s and v	$w = v / V$ and $z = s / S$
$s = T v$	$s = B v + u$	$z = b w + e, e = u / S$
$1' s = T 1' v$ or $T = S / V$	$1' s = S = B V + 1' u$ $T = S / V$ and $s = T v + \tilde{u}$	$1' z = 1 = b + 1' e$ $z = w + \tilde{e}$ with $\tilde{e} = \tilde{u} / S$
$s = S / V v$, thus $z = w$	RTO: $1' u$ need not be 0 $B = S / V - 1' u / V$	RTO: $1' e$ need not be 0 $b = 1 - 1' e = B V / S$
$s's = SST = SSX + SSE$	Compare B with $T = S / V$	Compare b with 1
$\text{Cos} = s'v / \sqrt{(v'v s's)} = 1$	$\text{Cos} = s'v / \sqrt{(v'v s's)}$	$\text{Cos} = z'w / \sqrt{(w'w z'z)}$
$B = s'v / v'v = T \text{Cos}$	$B = s'v / v'v = \sqrt{(s's / v'v)} \text{Cos}$	$b = z'w / w'w = \sqrt{(z'z / w'w)} \text{Cos}$
$R^2 = T^2 v'v / s's$ (SSX / SST)	$R^2 = B^2 v'v / s's$ (SSX / SST)	$R^2 = b^2 w'w / z'z$ (SSX / SST)
$R = T \sqrt{(v'v / s's)} = \text{Cos}$	$R = B \sqrt{(v'v / s's)} = \text{Cos}$	$R = b \sqrt{(w'w / z'z)} = \text{Cos}$
$T = s'v / v'v$ (alike covar)	$1 - R^2 = 1 - \text{Cos}^2 = u'u / s's$	$1 - R^2 = 1 - \text{Cos}^2 = e'e / z'z$
$T = s's / s'v$ (alike covar)	$\text{Cos} = 1 \leftrightarrow u = 0$	$\text{Cos} = 1 \leftrightarrow e = 0$
$T = \sqrt{(s's / v'v)}$ (no covar)	$\text{Sin}^2 = u'u / s's = \text{SSE} / \text{SST}$	$\text{Sin}^2 = e'e / z'z = \text{SSE} / \text{SST}$

Key points of RTO on the unit simplex are, using mostly the last column:

- The sum of errors $1' u$ or $1' e$ need not be 0, but $1' \tilde{u} = 0$ and $1' \tilde{e} = 0$.
- It is a contribution to RTO by compositional data that we now also look at $T = S / V$ and $s = T v + \tilde{u}$, or $z = w + \tilde{e}$. This features in voting theory but might be inoptimal.
- Perhaps useful to be aware of: $b w'w = p z'z = w'z = b / Nv = p / Ns$ while $b / p = Ns / Nv$ is also the square of another geometric average of slope, $\text{Sqrt}[b / p]$.
- The coefficient of determination R^2 applied to non-centered data gives the ratio of SSX / SST, with: SST = sum of squares total, SSX = sum of squares of the explanation = SST - SSE.
- For RTO, Cos takes the role of the coefficient of determination R^2 in OLS with a constant. We might simply use Cos^2 . However, it is clearest to write $R = \text{Cos}$, that is, within the confines of RTO. When we compare RTO with standard regression, then it we better distinguish the two again.
- $v' u = w' e = 0$, following the estimates for B and b (or they are chosen such).

²³ Thus actually $s^* = T^* v^*$, with perfection $\tilde{u}^* = 0$. This column is not to be confused with observations $s = T v + \tilde{u}$, see **Table 2**.

²⁴ This column drops the stars for readability.

- The "analysis of squares" (no deviations and thus no "analysis of variance") on SST gives:
 $z'z = (b w' + e')(b w + e) = b^2 w'w + 2bw'e + e'e = b^2 w'w + e'e$ with $w'e = 0$. Then:
 $1 = b^2 w'w / z'z + e'e / z'z = R^2 + e'e / z'z$. We also have $R = \text{Cos}$.
- $\text{Sin}^2 = 1 - \text{Cos}^2 = e'e / z'z = e'e / (b^2 w'w + e'e)$
- Thus there is a perfect fit $e = 0$ if and only if $\text{Cos} = 1$, when $b = 1$ and $z = w$.
- Observe that $b = 1 - 1' e = B V / S$ so that B has a relevant dimensional factor.
- With $z = s / S$ and $e = u / S$, $\text{Sin}^2 = e'e / z'z = u'u / s's$, and thus it doesn't matter whether we regress on the levels or the unitised variables to find Cos and Sin .
- We may regard $\text{Cos} = b \sqrt{(w'w / z'z)} = b \sqrt{(Ns / Nv)}$ as the explanation of Cos . We may also see this as an explanation of the slope as $b = \text{Cos} \sqrt{(Nv / Ns)}$.
- The footnote in the table is relevant unless you already saw the distinction between the columns, or the distinction between theoretical s^* and v^* and observed values s and v .

4.8. Analyses of squares for the direct error

Table 4 decomposes SST into SSX and SSE for the model $z = b w + e$, and we deduce that $\text{Sin}^2 = e'e / z'z = \text{SSE} / \text{SST}$. There is also the model $z = w + \tilde{e}$. The square of the Euclidean distance between the vectors is $\tilde{e}'\tilde{e} = (z - w)'(z - w)$. We would be interested in a relation between this Euclidean distance and the sine for variables in RTO.

We find this relationship: $\text{Sin}^2 = (\tilde{e}'\tilde{e} / z'z - h) / (1 - h)$ for $h = (1 / b - 1)^2$. While my historical knowledge about this area is limited, my impression is that this could well be a new finding.

Together: $\text{SSE} / \text{SST} = \text{Sin}^2 = e'e / z'z = (\tilde{e}'\tilde{e} / z'z - h) / (1 - h)$.

The interpretation of the RHS is: The sine makes the Euclidean distance $\tilde{e}'\tilde{e}$ relative to $z'z$ and subtracts a value h , because the cosine uses the slope to see proportionality that the Euclidean distance doesn't pick up, and then rebases with this h to remain in $[0, 1]$.

The deduction is:

- (1) A direct result from $z = z$ is that $e = \tilde{e} + (1 - b) w$ or $\tilde{e} = e - (1 - b) w$
- (2) The LHS of (1) with e gives $e'e = \tilde{e}'\tilde{e}$ because $w'e = 0$.
- (3) The LHS of (1) with \tilde{e} gives $\tilde{e}'\tilde{e} = \tilde{e}'\tilde{e} + (1 - b) w'\tilde{e} = e'e$.
- (4) Using w on (1) gives: $e'w = \tilde{e}'w + (1 - b) w'w$, and with $w'e = 0$ also $\tilde{e}'w = - (1 - b) w'w$. Also $b = 1 + \tilde{e}'w / w'w$. We already had this "implicit regression" above.
- (5) Eliminating $w'\tilde{e}$ from (3) and (4) gives $\tilde{e}'\tilde{e} = e'e + (1 - b)^2 w'w$. We already had $\tilde{e}'\tilde{e} \leq e'e \leq \tilde{e}'\tilde{e}$ and see this confirmed by the sum of squares
- (6) From (5):

$$\tilde{e}'\tilde{e} = e'e + (1 - b)^2 w'w$$

$$\tilde{e}'\tilde{e} / z'z = e'e / z'z + (1 - b)^2 w'w / z'z$$

$$\tilde{e}'\tilde{e} / z'z = \text{Sin}^2 + (1 - b)^2 \text{Cos}^2 / b^2$$

$$\tilde{e}'\tilde{e} / z'z = \text{Sin}^2 + h (1 - \text{Sin}^2) \text{ using } \text{Cos}^2 = 1 - \text{Sin}^2 \text{ and } h = (1 - b)^2 / b^2 = (1 / b - 1)^2$$

$$\tilde{e}'\tilde{e} / z'z = h + (1 - h) \text{Sin}^2$$

$$\text{Sin}^2 = (\tilde{e}'\tilde{e} / z'z - h) / (1 - h)$$

For voting theory this can be translated into a relation between $\text{EGID} = \sqrt{(\tilde{e}'\tilde{e} / 2)}$ and SDID . Colignatus (2018d) contains plots and tables that clarify the different behaviours.

4.9. Symmetry

From symmetry, $w = p z + \varepsilon$ generates the same insights. Cos and Sin are symmetric, and thus we have $\text{Sin}^2 = \varepsilon' \varepsilon / w'w$. We can write down the result straight away, yet in the repeat deduction, the negative sign in $w = z - \tilde{\varepsilon}$ causes a distraction, and thus it is better to repeat the steps, also writing $w = z + f$ so that $f' f = \tilde{\varepsilon}' \tilde{\varepsilon}$. Then

- (1) From $w = w$ or $z + f = p z + \varepsilon$ we find $\varepsilon = f + (1 - p) z$
- (2) With ε we have $\varepsilon' \varepsilon = \varepsilon' f$ since $z' \varepsilon = 0$
- (3) With f we have $f' \varepsilon = f' f + (1 - p) f' z = \varepsilon' \varepsilon$ because of (2)
- (4) With z we get $z' \varepsilon = z' f + (1 - p) z' z$ and with $z' \varepsilon = 0$ also $z' f = - (1 - p) z' z$
- (5) Eliminating $z' f$ from (3) and (4) gives $f' f = \varepsilon' \varepsilon + (1 - p)^2 z' z = \tilde{\varepsilon}' \tilde{\varepsilon}$ from $f = - \tilde{\varepsilon}$
- (6) From (5) and also using $p = z'w / z'z = \sqrt{(w'w / z'z)}$ Cos:

$$\tilde{\varepsilon}' \tilde{\varepsilon} = \varepsilon' \varepsilon + (1 - p)^2 z' z$$

$$\tilde{\varepsilon}' \tilde{\varepsilon} / w'w = \varepsilon' \varepsilon / w'w + (1 - p)^2 z' z / w'w$$

$$\tilde{\varepsilon}' \tilde{\varepsilon} / w'w = \text{Sin}^2 + (1 - p)^2 \text{Cos}^2 / p^2$$

$$\tilde{\varepsilon}' \tilde{\varepsilon} / w'w = g + (1 - g) \text{Sin}^2 \quad \text{with } g = (1 / p - 1)^2$$

$$\text{Sin}^2 = (\tilde{\varepsilon}' \tilde{\varepsilon} / w'w - g) / (1 - g)$$

Thus, both SDID and EGID are symmetric measures. The statement that SDID picks up disproportionality from an imputed angle b does not imply asymmetry (because of the asymmetry in the regression $z = b w + e$). The same sensitivity can also be formulated from the inverted regression.

The latter form can be rewritten using $w^* = w / |w|$. Then the relative sum of squared differences can also be seen as another weighted sum of relative differences.

$$\tilde{\varepsilon}' \tilde{\varepsilon} / w'w = \text{Sum}[(z - w)^2 / w'w] = \text{Sum}[w^{*2} (z / w - 1)^2]$$

5. More on interpretation

5.1. Statistics and heuristics on the slope

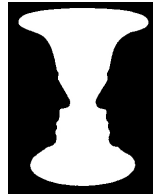
This paper uses descriptive statistics, and we are not in hypothesis testing. The linear algebra in this paper should not be confused with statistical decision methods. The latter methods use the same linear algebra but also involve assumptions on distributions: and we will make no such assumptions. However, when the linear algebra results into a new measure, then this measure can be used for new statistics again.

A heuristic is this: the voting literature has the Chi-Square / Webster / Sainte-Laguë (CWSID) measure $10 \text{Sum}[w (z / w - 1)^2]$, as the weighed sum of the squared deviations of the ratios from the ideal of unity provided by the diagonal. CWSID is not symmetric. If we would try to make CWSID symmetric then this causes division by zero, since some parties meet with zero seats (the "wasted vote"). An insight is: the idea to compare each ratio to unity is overdone, because the criterion or proportionality rather applies to the whole situation, and we should not be distracted by single cases. If b is the slope of the regression of $z = b w + e$, then $(1 - b)^2$ is a measure on the regression coefficient. This shifts the focus from individual parties to the *relation* between z and w . Thus we now consider the *slope-diagonal deviation*. This gives scope for symmetry by also looking at the regression $w = p z + \varepsilon$, so that we can compare b and $1 / p$.

The latter was the heuristic that started the Colignatus (2018d) paper. Taking the geometric average $\text{Sqrt}[b p]$ gave the recognition that this gave the same mathematical expression of the cosine as well. In other words $\text{Cos}[v, s] = \text{Sqrt}[b p]$. At some point it appeared that the role

of the cosine was more important by itself, and thus not regarded as a slope, as it generates the inequality / disproportionality measure Sin. This again was first seen as a *slope-diagonal deviation* measure but eventually the name *sine-diagonal* seems more accurate. This double nature of cosine and sine may be illustrated by Rubin's Vase, see **Figure 3**.

Figure 3. Rubin's Vase ²⁵



Alongside CWSID other common measurea are the absolute difference / Loosemore-Hanby index (ALHID) and Euclid / Gallagher index (EGID). CWSID, ALHID and EGID are parameter-free, and clearly fall under descriptive statistics. My original view was that SDID is parameter-dependent, since it refers to b and p . Paradoxially, however, while we started out looking for a slope measure, Cos is actually parameter-free, since its value as a similarity measure can be found without thinking about slopes at all. It is just a matter of perspective, see **Figure 3**. Thus the bonus of above heuristic is only that it helps to better understand what the measure does. The awareness of this double nature is important for the comparison of ALHID and SDID. For ALHID one might argue that it takes error $z - w = \tilde{\epsilon}$ in **Table 2** under the assumption that official regulations have chosen $s = \hat{s} = Ap[S, v]$ with some optimality. Thus, the use of $z = b w + e$ can be seen as changing the error. However, when Cos is regarded as a parameter-free measure of similarity, then no errors have been changed. Thus we can compare ALHID and SDID with this argument out of the way. Interpreting Cos as the outcome of RTO is only a bonus, that helps in the development of more perspectives. The useful perspective is that SDID may pick up more inequality / disproportionality than ALHID in the official $s = \hat{s} = Ap[S, v]$ for example by an *imputed* b that systematically differs from 1. And the cosine is a slope again in the equation with the standardised variables.

5.2. Electoral justice and inequality

Balinski & Young (1976:2) quote Daniel Webster:

"To apportion is to distribute by right measure, to set off in just parts, to assign in due and proper proportion."

Webster's emphasis was on "due and proper" and not on "proportion", but the world adopted "proportionality". Arend Lijphart has written about proportionality as "electoral justice". I have considered adopting the term "justice" too but settle for the notion that $z = w$ means equal proportions, or *equality*. A 2D line $y = \frac{1}{2} x$ is proportional too. Thus, rather than speaking about "proportionality" in comparing votes and seats, it is better to speak about *unit or diagonal vector-proportionality* and *electoral equality*. Similarly, while the standard expression is "Proportional representation" (PR) it would be better to speak about *Equal or Proportional Representation* (EPR) or even *Equal Proportional Representation* (EPR). I don't think that there is much chance that the world will rephrase this for the sake of education, but at least it is important to explain the vocabulary as a possible source for confusion. ²⁶

²⁵ https://en.wikipedia.org/wiki/Rubin_vase

<https://commons.wikimedia.org/wiki/File:Rubin2.jpg>

²⁶ France has the slogan "*Liberté, Egalité, Fraternité*", and thus it seems important to explain to the French that their system isn't *equal*. Using the word *disproportional* wouldn't ring their bell.

5.3. From the Humanities to Science

The issue of votes and seats contains an awkward element. The issue merges a topic of content with methods of statistics, and it appears that political science can use an infusion from statistical ethics as well. For description we know that displaying squares by their heights is misleading, since comparing 5 and 10 is quite else than comparing 25 and 100. This principle also causes the need to magnify what might be overlooked though. The underlying issue however is much more involved.

Colignatus (2018c) evaluates the “political science of electoral systems” and concludes that this branch of study still remains in the *Humanities*, without the methods of clear definitions, modeling and measuring as is required for proper *Science*. A telltale is the use of the word “election”, while EPR-elections are quite different from DR-elections. The invitation to empirical scientists is to help re-engineer the “political science of electoral systems” to become a real science.

Remarkably, political scientists Shugart & Taagepera (2017) themselves, in a recommendable book that indeed is a major advance, tend to agree that “political science of electoral systems” in general isn’t a science yet indeed, and they present their own book as a rare exception: “Thus the book is a rare *scientific* book about politics, and should set a methodological standard for all social sciences.” (p:320) They point to the trap of testing for statistical significance and the direction of the coefficient, while neglecting the effect size that would be relevant on content. “This can produce valuable insights, but these so-called “empirical models” are not really models at all. (...) Every peasant in Galileo’s time knew the direction in which things fall – but Galileo felt the need to predict more than direction.” (p324).

Unfortunately, while S&T mention different properties of DR and EPR, they apparently are blind to the key distinction, and thus in this key respect they are widely off-track.²⁷ They employ the words “vote” and “seat” as if these would be the same phenomenon everywhere while this is only superficially so. While we would expect that political scientists would understand what they are studying, apparently they don’t. And apparently it is necessary to belabour the point.

Let us focus on what “vote” and “seat” are about, and thus on what the term “representative” might mean. A common dictionary²⁸ emphasizes the legal meaning of “representative”. “2 a : standing or acting for another especially through delegated authority, b : of, based on, or constituting a government in which the many are represented by persons chosen from among them usually by election”. The dictionary also assumes that DR-elections and EPR-elections are all “elections” in a legal sense. Since direct democracy might only be feasible in town halls, we need representative democracy indeed, but the dictionary focuses on the legal setting and hides the key distinction between DR and EPR. As opposed to this common dictionary and legalistic thinking, in science we need strict definitions that fit the observations. We might make the assumption that the Moon is made of green cheese, but obviously we don’t. In a *single seat* case like for the President, one might argue that the President would represent the whole nation, but the House of Representatives has *multiple seats*, and this changes the logic of the case. Thus we get the distinction:

(1) In EPR, “representation” for Parliament means “standing for the people who have voted for you, by marking your name or party”. In EPR, a candidate gets a seat when the natural quota $Q = V / S$, the national average votes per seat, is covered, while this criterion is only lowered for the remainder seats. Thus there tends to be full backing by those like-minded, and this avoids the green cheese moon of “assuming” that the winner of a district seat

²⁷ When the article Colignatus (2018c) was updated, I had seen only a few pages of this new book by S&T. I have now studied it as required, with a curious mixture of great pleasure (science) and great worry (no science).

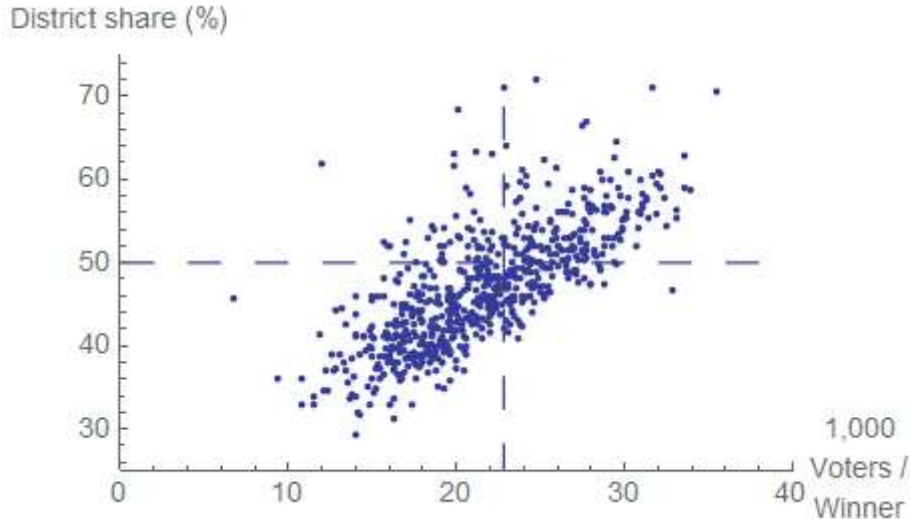
²⁸ <https://www.merriam-webster.com/dictionary/representative> (2018-02-18)

“represents” all conflicting interests of both who voted for him or her and who explicitly didn’t.
29

(2) In DR, there is the confusion between the single seat election and the multiple seats election. The *de iure* House of Representatives is *de facto* a House of District Winners, and *de facto* not a House of Representatives in the sense of EPR. DR systems can only be understood from history as proto-democracy with its confusions. Researchers who do not call attention to this, focus on legal notions and words in a dictionary, like might be common in the Humanities, rather than on what actually happens in empirical reality.

Let us reconsider **Table 1**, with outcomes of such DR-elections. The USA and UK use both single seat districts (SSD) and the criterion of Plurality per district, meaning that the district seat is given to the candidate with the most votes - also called "first past the post" (FPTP). The values in the table do not indicate: (1) that the votes are masked and do not give the first preferences, so that we rather don't know what voters would really prefer, (2) that the votes for candidates other than the district winners are obliterated. The latter would show by comparing the votes per winning candidate to the natural quota Q , which would highlight that the winning candidates are below this quota. **Figure 4** plots the UK "elections" of 2010, reproduced from Colignatus (2010:10). Only the plotted points are translated into seats and the other votes are obliterated. The 2010 quota was 45.6 thousand votes per seat. With 650 districts and seats, a MP would tend to need 100% of the district to get the quota. None of the MPs gained the electoral quota of 45.6 thousand, and more than half do worse than 50% of the quota, and 66% is even below 50% of their own district. Such information is not provided by the UK Electoral Commission (2017abc), even though they claim to do an "analysis".³⁰ We can diagnose a situation of incomplete information to the general public, and actual disinformation, biased towards proto-democratic DR and against proper democratic EPR.

Figure 4. MPs of UK 2010: Winning % (District share) per votes won per seat



²⁹ The empirical observation is that it takes complicated bargaining to find workable compromises for such conflicting interests, and we should not assume that a district winner would be able to do so all by himself or herself, even by invoking magic.

³⁰ Website and texts: “The Electoral Commission is the independent body which oversees elections and regulates political finance in the UK. We work to promote public confidence in the democratic process and ensure its integrity.” and “We carry out policy work across a range of areas. We want to ensure that the rules around all aspects of elections are as clear and simple as possible and that the interests of voters are always put first.” Yet the Electoral Commission silently accepts that most votes are discarded and not translated into seats. Most problematic is that they do not provide the statistics that show this discarding of most votes.

Basically the majority of the USA still miss out on “No taxation without representation”. A legal approach, like in current “political science on electoral systems”, would adopt the green cheese moon of the legality of the 1773 UK Parliament and the Royal Governour Thomas Hutchison as “representing” the American people, so that the Boston Tea Party was criminal indeed. Let us see what it means in terms of the EPR notion of representation. In a schematic setup, assume that Republicans and Democrats have the majority in the House each 50% of the time. Assume {Rep, Dem, Other} vote shares {4.6, 4.4, 1} for half of the districts, thus with Rep winners. Assume vote shares {4.4, 4.6, 1} for the other half of the districts, thus with Dem winners. Let minor fluctuations (say on Other) cause the majority switches. Then in the first half of districts only 4.6 out of 10 are represented all of the time and 5.4 never since their votes are discarded. In the second half the same. Thus overall only 4.6 gets represented and 5.4 never. If we make this more random then a voter would be represented 46% of his or her life and 54% not. We have a US House of District Winners and not a House of Representatives. The switches of a majority around the 5-5 middle in the House are a product of the system, that allows that “representatives” get less than the natural quota, and that hides what is really happening on masking and discarding. Also, these majority switches should not deceive us, because we are looking at representation in the House and not at the participation of parties in the majority coalition. One might argue that a Republican voter who loses out in his or her district then is “represented” by a Republican winner in another district, but if this would be the true meaning of “representation”, then even more is wrong with the dictionary, and even lawyers might protest.

Shugart & Taagepera (2017) use the term “wasted vote” but they do not clarify to their readers that this term has quite different meanings in DR and EPR. In EPR small parties may simply be too tiny, whence those votes go lost. In Holland this is 0.2 out of 10. DR however has systematic obliteration of the votes for other candidates than the district winners. Above schematic example has 5.4 of 10, not far from reality in **Figure 4**.

S&T ingeniously and remarkably succeed in modeling key statistical variables on votes and seats, notably by using the assembly size S (US 435, UK 650) and the district magnitude M (in SSD $M = 1$), and by deriving constraints that drive the outcomes. They suggest that these findings could be used for better electoral design. However, they overlook (1) that M is only relevant for DR, since EPR has $M = S$, and (2) that the scientifically correct advice is to switch from DR to EPR, since DR can only be understood from history and confusion about single seat and multiple seats elections. The shocking but accurate analogy is that S&T are skilled surgeons, but they are still blind to bacteria and do not wash their hands. The problem, like with the UK Electoral Commission that emphasizes integrity, isn't quite with integrity but with hygiene, and the willingness to see what you are looking at. Fish might not know the water they are swimming in, but here we are speaking about basic notions of empirical science. It would however become an issue of research integrity when researchers in political science are alerted to the bias and when they would not correct it.

There as well appears to exist an argument that DR would be more “accountable” than EPR, but this appears to be based upon faulty logic, recently “supported” by invalid regression analysis. Colignatus (2018c) provides the overall deconstruction.

The legalistic and essentially unscientific bend in the “political science on electoral systems” is a key factor that keeps the USA and UK locked in proto-democracy since 1917, while Holland switched to EPR in 1917. Thus let me repeat that the scientific world is advised to look into this area of “political science on electoral systems”, see Colignatus (2018c). There might be a special role for the ethics in statistics, when statisticians deal with statistics on votes and seats. Statisticians should be clear on what the figures and charts mean, and should not deceive the public by playing into common confusions and the use of words only.

5.4. The example of Brexit

I came to this subject because of the 2016 Brexit referendum and the reactions in the UK. Oxford Chancellor Chris Patten (2017) expresses a similar amazement: “A parliamentary democracy should never turn to such populist devices. Even so, May could have reacted to

the 52 per cent vote to quit Europe by saying that she would hand the negotiations to a group of ministers who believed in this outcome and then put the result of the talks in due course to parliament and the people. Instead, she turned the whole of her government into a Brexit machine, even though she had always wished to remain in the EU. (...) But, as with any divorce, we can be fairly confident that it is the children who will suffer the most.” Eventually Colignatus (2017e) concludes that these events can be explained by the UK system of DR. The UK Electoral Commission went along with a Brexit referendum question that is binary only in a *legal* sense (Remain or Leave), while it does not satisfy the standards for a statistical questionnaire, that would list the *policy options* of how to Remain or Leave, see Colignatus (2017c). When the company YouGov.com made the proper statistical questionnaire, it appeared that 17% of those polled put the option of Remain between different options of Leave, see Colignatus (2017d). Basically it still is unclear what the UK electorate would prefer. Colignatus (2017e) concludes that the UK can only find this out when the UK switches from DR to EPR and has new elections, so that a properly elected House of Commons can clarify the situation, and perhaps find a compromise on their views. Perhaps this case might be an eye-opener for other nations with DR. (And chancellor Patten is advised to look into the political science department at his university.)

6. Conclusions

This overview generates various insights for our topic of discussion.

- (1) Regression of $y = b x + e$ likely is better introduced by the geometric approach using angles, with projection of y on x , such that $b x$ and e are perpendicular and $x'e = 0$. One would start with regression through the origin (RTO), and only subsequently introduce the constant, the notion of correlation (with centered values), the sum of squared errors and normal equations, the normal distribution for the error, and then multiple regression.
- (2) The comparison of votes and seats would be a key example for such an approach. This is descriptive statistics and uses the same linear algebra as regression but without an assumption on a distribution of the errors. In this setting, the sine-diagonal inequality / disproportionality (SDID) finds a quick understanding as a measure of the distance between votes and seats.
- (3) For descriptive statistics, there is an important role for psychology. We must avoid deceiving presentations like portraying squares by their height only. This principle also causes the need to magnify what might be overlooked. We must also be clear on the different meanings of votes and seats under systems of DR and EPR.
- (4) Big Data might cause a shift from a focus on statistical decision theory to more attention for descriptive statistics. Elements within the loss functions for statistical decision can be used to construct measures for statistical description. Examples are the Richter scale for earthquakes and SDID for votes and seats.
- (5) R-squared has a key role, first since it is a cosine, and secondly since minimising the sum of squared errors is tantamount to maximising R-squared. The particular value of R-squared would be informative for specification search. The sum of squared errors can also be related to the sine, and for votes and seats in RTO to the SDID.
- (6) For compositional data z and w such that $1'z = 1$ and $1'w = 1$ there is the contribution to regression analysis that $z = b w + e$ is supplemented with a meaningful $\tilde{e} = z - w$. This would not hold for all compositional data, of course, and it might be inoptimal for votes and seats.
- (7) For the latter there is the relation $SSE / SST = \text{Sin}^2 = e'e / z'z = (\tilde{e}' \tilde{e} / z'z - h) / (1 - h)$, with $h = (1 / b - 1)^2$. This helps to understand that SDID picks up disproportionality caused by an imputed slope that the Euclidean norm $\|\tilde{e}\|$ doesn't see.
- (8) For the latter case we also have $w = p z + \varepsilon$ and $\text{Cos}[w, z] = \text{Sqrt}[b p]$. The parameter-free measure of similarity is also a geometric average of slope. This dual face appears to be an advantage, allowing us to switch between these perspectives.
- (9) On content, we have highlighted the key difference between DR-elections and EPR-elections, so that outcomes about inequality / disproportionality are only comparable for EPR, and rather incomparable even within DR situations (with masked data and obliterated votes). Empirical researchers are invited to help re-engineer the “political science on electoral systems”, see Colignatus (2018c).

The new measure SDID provides a better description of the inequality / disproportionality of votes and seats compared to existing measures ALHID, EGID and CWSID. The new measure has been tailored to votes and seats, by means of (1) greater sensitivity to small inequalities, and (2) a sign, since a small change in inequality may have a crucial impact on the (political) majority. For different fields, one could tailor measures in similar fashion.

The proposed measure SDID provides an enlightening bridge between descriptive statistics and statistical decision making. This comes with a better understanding of what kind of information the cosine or R-squared provides, in relation to regressions with and without a constant. Statistics textbooks would do well by providing their students with this new topic for both theory and practical application. If textbooks would not change then they run the risk of perpetuating the misunderstanding that they already created and which have been highlighted here too.

7. References

Colignatus is the name in science of Thomas Cool, econometrician and teacher of mathematics, Scheveningen, Holland, <http://econpapers.repec.org/RAS/pco170.htm>

References in the text to Wikipedia refer to it as a portal and no source.

Balinski, M. and H.P. Young (1976), "Criteria for proportional representation", IIASA Research Report December, RR-76-020, <http://pure.iiasa.ac.at/525/1/RR-76-020.pdf>

Barceló-Vidal, C. and Martín-Fernández, J.A. (2016), "The Mathematics of Compositional Analysis", *Austrian Journal of Statistics*, 45(4), 57-71. DOI: <http://dx.doi.org/10.17713/ajs.v45i4.142>.

Colignatus, Th. (2006), "On the sample distribution of the adjusted coefficient of determination (R2Adj) in OLS", <http://library.wolfram.com/infocenter/MathSource/6269/>

Colignatus, Th. (2007), "Correlation and regression in contingency tables. A measure of association or correlation in nominal data (contingency tables), using determinants", <https://mpra.ub.uni-muenchen.de/3660/>

Colignatus, Th. (2009, 2015), "Elegance with substance", <https://zenodo.org/record/291974>

Colignatus, Th. (2010), "Single vote multiple seatselections. Didactics of district versus proportional representation, using the examples of the United Kingdom and The Netherlands", <https://mpra.ub.uni-muenchen.de/22782/>

Colignatus, Th. (2011), "Conquest of the Plane", <https://zenodo.org/record/291972>

Colignatus, Th. (2017a), "Two conditions for the application of Lorenz curve and Gini coefficient to voting and allocated seats", <https://mpra.ub.uni-muenchen.de/80297/>

Colignatus, Th. (2017b), "Statistics, slope, cosine, sine, sign, significance and R-squared", <https://boycottholland.wordpress.com/2017/10/21/statistics-slope-cosine-sine-sign-significance-and-r-squared/>

Colignatus, Th. (2017c), "Voting theory and the Brexit referendum question", *RES Newsletter* 177, <http://www.res.org.uk/view/art4Apr17Features.html>

Colignatus, Th. (2017d), "Great Britain's June 2017 preferences on Brexit options", *RES Newsletter*, October, 6-9, <http://www.res.org.uk/view/art2Oct17Features.html>

Colignatus, Th. (2017e), "Dealing with denial: Cause and cure of Brexit", <https://boycottholland.wordpress.com/2017/12/01/dealing-with-denial-cause-and-cure-ofbrexit/>

Colignatus, Th. (2018a), "Comparing the Aitchison distance and the angular distance for use as inequality or disproportionality measures for votes and seats", <https://mpira.ub.uni-muenchen.de/84387> also at <https://www.wolframcloud.com/objects/thomas-cool/Voting/2018-01-18-Aitchison.nb>

Colignatus, Th. (2018b), "Measures of policy distance and inequality / disproportionality of votes and seats", <https://mpira.ub.uni-muenchen.de/84324/> also at <https://www.wolframcloud.com/objects/thomas-cool/Voting/2018-02-02-PolicyDistance.nb>

Colignatus, Th. (2018c), "One woman, one vote. Though not in the USA, UK and France", first version in 2017 but updated in 2018, <https://mpira.ub.uni-muenchen.de/84482/>

Colignatus, Th. (2018d), "Comparing votes and seats with cosine, sine and sign, with attention for the slope and enhanced sensitivity to inequality / disproportionality", first version in 2017 but updated in 2018, <https://mpira.ub.unimuenchen.de/84469>

Dongen, S. van & A.J. Enright (2012), "Metric distances derived from cosine similarity and Pearson and Spearman correlations", <https://arxiv.org/abs/1208.3145>

Egghe, L. and L. Leydesdorff (2009), "The relation between Pearson's correlation coefficient r and Salton's cosine measure", *J. of the Am. Soc. for Information Science and Technology*, 60(5), p 1027-1036, <http://hdl.handle.net/1942/8494>

Eisenhauer, J.G. (2003), "Regression through the origin", *Teaching Statistics*. Volume 25, Number 3, Autumn 2003, p76-80

Ford, C. (2015), "Is R-squared Useless?", <http://data.library.virginia.edu/is-r-squared-useless/>

Gallagher, M. (1991), "Proportionality, disproportionality and electoral systems", *Electoral Studies*, 10:1, 33-51, https://www.tcd.ie/Political_Science/draft/staff/michael_gallagher/ElectoralStudies1991.pdf

Gelman, A. and G. King (1994), "A Unified Method of Evaluating Electoral Systems and Redistricting Plans", *American Journal of Political Science*, 38, Pp. 514–554, <https://gking.harvard.edu/files/abs/writeit-abs.shtml>

Gelman, A. (2007), "Significance testing in economics: McCloskey, Ziliak, Hoover, and Siegler", http://andrewgelman.com/2007/10/05/significance_te/

Gelman, A. and H. Stern (2006), "The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant", *The American Statistician*, November 2006, Vol. 60, No. 4, p328-331

Giles, D. (2013), "Unbiased Model Selection Using the Adjusted R-Squared", <http://davegiles.blogspot.nl/2013/08/unbiased-model-selection-using-adjusted.html> (actually try <http://davegiles.blogspot.nl/search?q=R-squared>)

Johnston, J. (1972), "Econometric methods", McGraw-Hill

Karpov, A. (2008), "Measurement of disproportionality in proportional representation systems", *Mathematical and Computer Modelling*, Volume 48, Issues 9–10, November 2008, Pages 1421-1438, <http://www.sciencedirect.com/science/article/pii/S0895717708001933>
Politics

Koppel, M, and A. Diskin (2009), "Measuring disproportionality, volatility and malapportionment: axiomatization and solutions", *Social Choice and Welfare*, August, 33:281, https://www.researchgate.net/publication/225444815_Measuring_disproportionality_volatility_and_malapportionment_Axiomatization_and_solutions

- Kozak, A. and R. A. Kozak (1995), "Notes on regression through the origin", *The Forestry Chronicle*, May/June, Vol 71 no 3, p326-330
- Lehmann, E.L. (2008), "On the history and use of some standard statistical models", *IMS Collections. Probability and Statistics: Essays in Honor of David A. Freedman*, <https://arxiv.org/abs/0805.2838>
- Mood, A. and F. Graybill (1963), "Introduction to the theory of statistics", McGraw-Hill
- Patten, C. (2017), "What Brexit means for Britain's future, according to Oxford University's Chancellor", <https://blog.politics.ox.ac.uk/brexit-means-britains-future-according-oxford-universitys-chancellor/>
- Pearl, J. (2000), "Causality", CUP
- Shalizi, C. (2015), "Lecture 10: F-Tests, R2, and Other Distractions", (October 16), <http://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/10/lecture-10.pdf>
- Shugart, M.S. & R. Taagepera (2017), "Votes from seats. Logical models of electoral systems", CUP, <https://doi.org/10.1017/9781108261128>
- Stigler, S. (2008), "Karl Pearson's Theoretical Errors and the Advances They Inspired", *Statistical Science*, Vol. 23, No. 2, 261–271, arXiv:0808.4032v1
- Taagepera, R. and B. Grofman (2003), "Mapping the indices of seats-votes disproportionality and inter-election volatility", *Party Politics*, 9(6), p659-677, <http://escholarship.org/uc/item/0m9912ff#page-1>
- Theil, H. (1971), "Principles of Econometrics", Wiley
- UK Electoral Commission (2017a), "Voting in 2017. Understanding public attitudes towards elections and voting", https://www.electoralcommission.org.uk/__data/assets/pdf_file/0011/234893/Voting-in-2017-Final.pdf
- UK Electoral Commission (2017b), "2017 UK general election results", <https://www.electoralcommission.org.uk/find-information-by-subject/elections-and-referendums/past-elections-and-referendums/uk-general-elections/2017-uk-general-election-results>
- UK Electoral Commission (2017c), "UK Parliamentary General Election, June 2017", https://www.electoralcommission.org.uk/__data/assets/pdf_file/0004/234976/UKPGE-2017-electoral-data-report.pdf see also the data files <https://www.electoralcommission.org.uk/our-work/our-research/electoral-data/electoral-data-files-and-reports>
- Varian, H. (2016), "Causal inference in economics and marketing", July 5, 2016 vol. 113 no. 27, 7310-7315, <http://www.pnas.org/content/113/27/7310.abstract>
- Wilcox, R. (2017), "New statistical methods would let researchers deal with data in better, more robust ways", *Significance*, May 9, <https://www.significancemagazine.com/science/548-new-statistical-methods-would-let-researchers-deal-with-data-in-better-more-robust-ways>
- Ziliak, S.T. and D.N. McCloskey (2007), "The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives", Univ. of Michigan Press