The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2615-9821.htm

Gain-probability diagrams as an alternative to significance testing in economics and finance

David Trafimow
Department of Psychology, New Mexico State University, Las Cruces, New Mexico, USA
Ziyuan Wang
University of Lynchburg, Lynchburg, Virginia, USA, and
Tingting Tong and Tonghui Wang
Department of Mathematical Sciences, New Mexico State University, Las Cruces, New Mexico, USA

Received 23 May 2023
Revised 30 May 2023
Accepted 3 June 2023

Abstract

Purpose – The purpose of this article is to show the gains that can be made if researchers were to use gain-probability (G-P) diagrams.

Design/methodology/approach – The authors present relevant mathematical equations, invented examples and real data examples.

Findings – G-P diagrams provide a more nuanced understanding of the data than typical summary statistics, effect sizes or significance tests.

Practical implications – Gain-probability diagrams provided a much better basis for making decisions than typical summary statistics, effect sizes or significance tests.

Originality/value – G-P diagrams provide a completely new way to traverse the distance from data to decision-making implications.

Keywords Mathematical and quantitative methods, Gain-probability diagrams, Skew normal distributions, Delta skew lognormal distributions, Decision-making, Probabilities

Paper type Research Paper

1. Introduction

Consider a medical scenario implying a government financing decision. Suppose researchers invent a medicine to treat a disease. Should the government invest to render the medicine more widely available to its citizens? At present, a widely accepted procedure would be to perform a significance test to determine if the medicine is effective at a statistically significant level. If so, that would constitute a reason to make the investment; if not, the government would not invest. However, significance testing is a poor way to evaluate the effectiveness of the medicine. It is quite possible for a medicine not to be effective, but nevertheless achieve statistical

© David Trafimow, Ziyuan Wang, Tingting Tong and Tonghui Wang. Published in *Asian Journal of Economics and Banking*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

It has come to the attention of the publisher that the article, Trafimow, D., Wang, Z., Tong, T. and Wang, T. "Gain-probability diagrams as an alternative to significance testing in economics and finance" published in Asian Journal of Economics and Banking https://doi.org/10.1108/AJEB-05-2023-0045 gave an incorrect location for Trafimow's affiliation. This error was introduced in the editorial process and has now been corrected in the online version. The publisher sincerely apologises for this error.

Asian Journal of Economics and Banking
Vol. 7 No. 3, 2023
pp. 333-357
Emerald Publishing Limited
e-ISSN: 2633-7991
p-ISSN: 2615-9821
DOI 10.1108/AJEB-05-2023-0045

significance, if the sample of cases is sufficiently large. This is because statistical significance depends on two items: the size of the effect and the sample size. It is possible for a sufficiently large sample size to compensate for even an extremely small effect. The *p*-values upon which significance testing depends confound the effect sizes researchers obtain with the sample sizes they collect. In fact, McQuitty (2004, 2018) famously argued against too large sample sizes on the grounds that they ensure statistical significance even when there is not much of an effect. Of course, a problem with this argument is that statistical significance is not the only matter of concern. There is the much more important issue of the extent to which sample sizes are sufficiently large to engender confidence that the sample statistics provide good estimates of population parameters. The larger the sample size, the better the estimate. Thus, from the point of view of estimation, McQuitty's famous advice is contraindicated. In addition, the fact that the advice was so well received by business and economics researchers and cited over 1,000 times and received an accolade by *Journal of Global Scholars of Marketing Science*, points to one of many problems with significance testing thinking (Trafimow *et al.*, 2021). There have been many criticisms of significance testing, including several reviews (Hubbard, 2016; Ziliak and McCloskey, 2016; Trafimow, 2019) and 43 articles in the 2019 special issue of *The American Statistician*, the highly respected journal of The American Statistical Association. Because the disadvantages of significance testing have been covered so extensively, there is no point in rehashing them here. Rather, the present goal is to present a more useful alternative.

In the medical scenario, there are many factors to consider. These include the seriousness of the disease, the cost of the medicine and many others. In addition, and crucially, there is the issue of how well the medicine works. Put in the form of a question, what is the probability that a randomly selected person who takes the medicine will be better off, or worse off, and by how much, with the medicine than without it?

Nor must we restrict ourselves to economics issues pertaining to medicine. If an economic intervention is proposed to increase people's living space, what is the probability that a randomly selected person will be better off, or worse off, and by how much living space, with the intervention than without it? Or if a government wishes to institute a policy to increase income in a poor area, what is the probability that a person will be better off, or worse off, and by how much income, with the new policy than without it? The issue of probabilities of being better off, or worse off, by varying degrees, is relevant to many potential economics applications.

To estimate the probability of being better off, or worse off, by varying degrees, with the intervention or policy change than without it, it is necessary to look carefully at the data to determine the distribution. We will consider two families of distributions: skew normal distributions, which include normal distributions, and delta log-skew-normal (LSN) distributions, which include lognormal distributions. The subsequent section includes crucial equations, with subsections for skew normal and delta LSN distributions. Following that, we provide examples of how to use the equations to construct gain-probability (G-P) diagrams to draw conclusions far beyond that which significance tests allow (Tong *et al.*, 2022; Trafimow *et al.*, 2022; Wang *et al.*, 2022).

2. Probability of being better off or worse off by how much? Skew normal distributions

Whereas normal distributions have two parameters, mean μ and standard deviation σ; skew normal distributions have three parameters. The location ξ replaces the mean, the scale ω replaces the standard deviation, and there is a shape (or skewness) parameter α. When the shape parameter equals 0 the distribution is normal, and the location equals the mean and the scale equals the standard deviation. But when the shape parameter does not equal 0, the location does not equal the mean and the scale does not equal the standard deviation. In symbols, when $\alpha = 0$, $\xi = \mu$ and $\omega = \sigma$; but when $\alpha \neq 0$, $\xi \neq \mu$ and $\omega \neq \sigma$. The family of normal distributions is a subset of the family of skew normal distributions. Thus, the family of skew normal distributions is more generally applicable than the family of normal distributions.

Let us now derive the probability that a randomly chosen person from one skew normal population has a higher score on the dependent variable than a randomly chosen person from another skew normal population. We will assume dependent populations at any level of correlation between -1 and $+1$, with independent populations (correlation $= 0$) as a special case.

D2.1. Azzalini and Valle (1996) *A random vector* $\mathbf{X} = (X_1, \ldots, X_k)'$ *is said to have an k-dimensional multivariate skew normal distribution with the vector of location parameter* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)' \in \Re^k$, *the scale parameter of the positive definite matrix* Σ, *and the vector of skewness (shape) parameters* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, *denoted as* $\boldsymbol{X} \sim SN_k(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$, *if its density function (pdf) is given by*

$$f_{\mathbf{X}}(\mathbf{x}) = 2\phi_k(\mathbf{x}; \boldsymbol{\mu}, \Sigma)\Phi\left(\boldsymbol{\alpha}'\Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\right), \tag{2.1}$$

where $\phi_k(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ *is the density of the k-dimensional multivariate normal distribution* $N_k(\boldsymbol{\mu}, \Sigma)$ *with mean* $\boldsymbol{\mu}$ *and covariance matrix* Σ, *and* $\Phi(z)$ *is the cumulative distribution function (cdf) of the standard normal random variable* $Z \sim N(0, 1)$. *Here* $\mathbf{x}'$ *is the transpose of the vector* $\mathbf{x} \in \Re^k$.

2.1 Calculation of $\delta_{ab} = P(Z > 0)$ *in independent distributions*

As we mentioned above, Let $X \sim SN(\xi_1, \omega_1, \alpha_1)$, $Y \sim SN(\xi_2, \omega_2, \alpha_2)$ and we assume that X and Y are independent. We want to obtain the distribution of the linear combination, $Z = X + aY + b$, of X and Y for specified $a \in \Re$ and $b \in \Re$ and then calculate $\delta_{ab} = P(Z > 0)$. We need to find the distribution of Z first, which is given below.

There are two advantages of using $Z = X + aY + b$. Both advantages stem from the possibility that a researcher might be interested in more than simply a probabilistic advantage for one group over another. A researcher might be interested in the probabilities of being better off, or worse off, to varying degrees. The equation provides two ways to assess this. One way is by setting $a = -1$ and varying b, to obtain the probability that a randomly selected person from one condition will score higher than a randomly selected person from another condition, by the amount the researcher specifies by setting b to that value. This is the most straightforward advantage.

A second advantage is that the equation provides a way to assess the probability of being better off, or worse off, by varying degrees, in terms of multiples. For example, what is the probability that a randomly selected person from one group will score twice as much, thrice as much and so on, as a randomly selected person from the other group? The equation facilitates such an assessment. The researcher merely sets $b = 0$ and lets a vary at multiples of interest.

Theorem 2.1. *Let* $X \sim SN(\xi_1, \omega_1^2, \alpha_1)$, $Y \sim SN(\xi_2, \omega_2^2, \alpha_2)$. *First we assume that two skew normal populations are independent. Then the probability density function (pdf) of* $Z = X + aY + b$ *is*

$$f_Z(z) = c\phi\left(z; \nu, \tau^2\right)\Phi_2[B(z - \nu); \mathbf{0}_2, \Delta], \tag{2.2}$$

where

$$c = \Phi_2^{-1}\left(\mathbf{0}_2; \mathbf{0}_2, \Delta + \tau^2 BB'\right), \quad \nu = \xi_1 + a\xi_2 + b, \quad \tau^2 = \omega_1^2 + a^2\omega_2^2, \quad B = \mathbf{d}'/\tau^2,$$

$$\mathbf{0}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \Delta = \begin{pmatrix} 1 + \alpha_1^2 & 0 \\ 0 & 1 + \alpha_2^2 \end{pmatrix} - \frac{\mathbf{dd}'}{\tau^2}, \qquad \mathbf{d} = \begin{pmatrix} \alpha_1\omega_1 \\ a\alpha_2\omega_2 \end{pmatrix}.$$

Here $\phi(z;\nu,\tau^2)$ *is the pdf of the normal distribution* $N(\nu,\tau^2)$ *with mean* ν *and variance* τ^2*, and* $\Phi_2(\mathbf{u};\mathbf{0}_2,\Delta)$ *is the cumulative distribution function (cdf) of the bivariate normal distribution with mean vector* $\mathbf{0}_2$ *and covariance* Δ.

The proof of Theorem 2.1 is given in Appendix. Note that the pdf of Z given in Equation (2.2) is the special case of the closed skew normal density. For details of closed skew normal distributions, see Gupta *et al.* and Zhu *et al.*

Now we can compute the probability $P(Z>0)$, which is given by

$$\delta_{ab}=\Pr(Z>0)=\int_0^{\infty}c\phi\left(z;\nu,\tau^2\right)\Phi_2[B(z-\nu);\mathbf{0}_2,\Delta]dz. \tag{2.3}$$

Density curves of Z for different parameters and their corresponding $P(Z>0)$ are given in Figures 1 and 2, respectively. These values can be instantiated into the equations presented earlier, but there is an easier way too. We provide a freely available online calculator at https://probab.shinyapps.io/inde_prob/

2.2 Calculation of $\delta_{pab}=\mathrm{P}(U>0)$ *in dependent distributions*

Now we consider the bivariate skew normal random vector $\mathbf{X}=(X_1,X_2)'\sim SN_2(\boldsymbol{\xi},\Sigma,\boldsymbol{\alpha})$, with location parameter $\boldsymbol{\xi}$, scale parameter Σ and skewness parameter $\boldsymbol{\alpha}$ given by

$$\boldsymbol{\xi}=\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix},\qquad \Sigma=\begin{pmatrix}\omega_1^2 & \rho\omega_1\omega_2\\ \rho\omega_1\omega_2 & \omega_2^2\end{pmatrix},\qquad \boldsymbol{\alpha}=\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}.$$

Then we have the following result and its proof.

Theorem 2.2. Let $\mathbf{X}\sim SN_2(\boldsymbol{\xi},\Sigma,\boldsymbol{\alpha})$ *given above and consider* $U=X_1+aX_2+b$ *with* $a,b\in\Re$*. Then the pdf of* U *is*

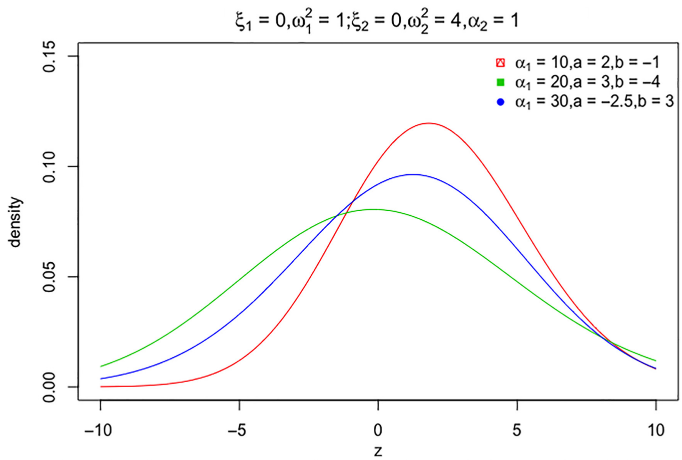
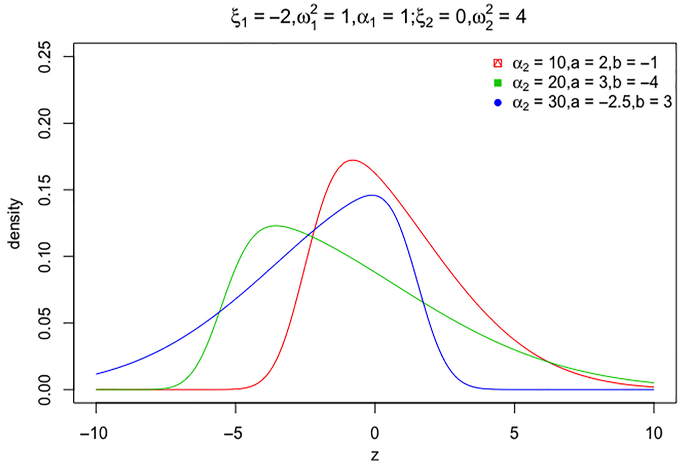


Figure 1.
The density of Z when location parameters of both populations equal 0, the scale parameters for first and second population equal 1 and 2 respectively

Note(s): The shape parameters for second population are 1 and for first population equal 2, 5, 10 with $a=2$, $b=-1$, $a=3$, $b=-4$ and $a=-2.5$, $b=3$. The corresponding probability $P(Z>0)$ is 0.716214, 0.504227 and 0.600769

Source(s): Figure by authors



Note(s): The shape parameters for first population are 1 and for second population equal 2, 5 and 10 with $a = 2$, $b = -1$, $a = 3$, $b = -4$ and $a = -2.5$, $b = 3$. The corresponding probability $P(Z > 0)$ is 0.528823, 0.369435 and 0.245391

Source(s): Figure by authors

Figure 2. The density of Z when location parameters for first and second population equal -2 and 0 respectively, the scale parameters for first and second population equal 1 and 2 respectively

$$f_U(u) = 2\phi\left(u; \xi_1 + a\xi_2 + b, \omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2\right) \times \Phi\left[\alpha_* \frac{(u - \xi_1 - a\xi_2 - b)}{\left(\omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2\right)^{1/2}}\right], \tag{2.4}$$

where

$$\alpha_* = \frac{\alpha_1 d_1 + \alpha_2 d_2}{\sqrt{(1 - \alpha_1^2 - \alpha_2^2)k^2\sigma^2 - (\alpha_1 d_1 + \alpha_2 d_2)^2}}$$

with $d_1 = \omega_1^2 + \omega_1\omega_2(\sqrt{1-\rho^2} + a\rho)$, $d_2 = a\omega_2^2 + \omega_1\omega_2(a\sqrt{1-\rho^2} - \rho)$, $k = \sqrt{\omega_1^2 + \omega_2^2 + 2\sqrt{1-\rho^2}\omega_1\omega_2}$ and $\sigma^2 = \omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2$.

The proof of Theorem 2.2 is given in Appendix.

The probability $\delta_{\rho ab} = P(U > 0)$ is given by

$$\delta_{\rho ab} = P(U > 0) = \int_0^\infty f_U(u)du. \tag{2.5}$$

Density curves of U for different parameters and their corresponding probability $P(U > 0)$ are given in Figures 3 and 4, respectively. For example, in Figure 3, we can see that the density curves are affected by the location parameter $\xi_2 = 0, 2, 4, 2$ with other parameters specified and $\rho = 0.25$, together with their corresponding probabilities $P(U > 0) = 0.433816, 0.566184, 0.226627, 0.10565$, respectively. We provide a different online calculator, still freely available online at https://probab.shinyapps.io/ProbU/

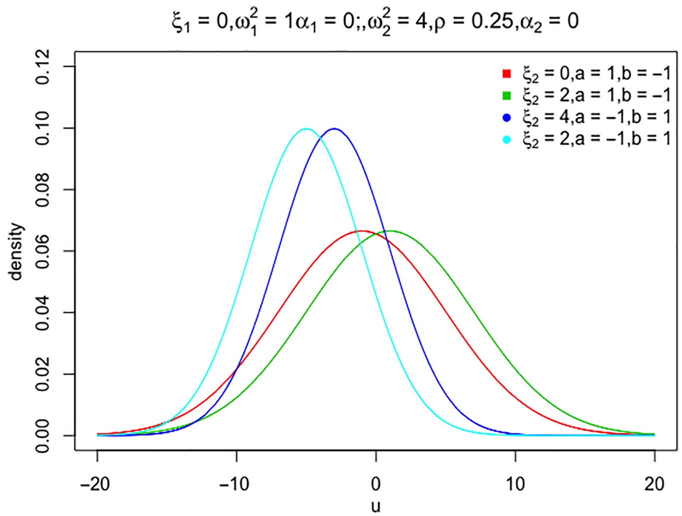


Figure 3.
The density of U when location parameters of the first population equal 0, the scale parameters for first and second population equal 1 and 2 respectively

Note(s): The shape parameters for both populations equal 0 and the location parameters of the second population equal 0, 2, 4, 6 with $a = 1$, $b = -1$ and $a = -1$, $b = 1$. The corresponding probability $P(U > 0)$ is 0.433816, 0.566184, 0.226627 and 0.10565
Source(s): Figure by authors

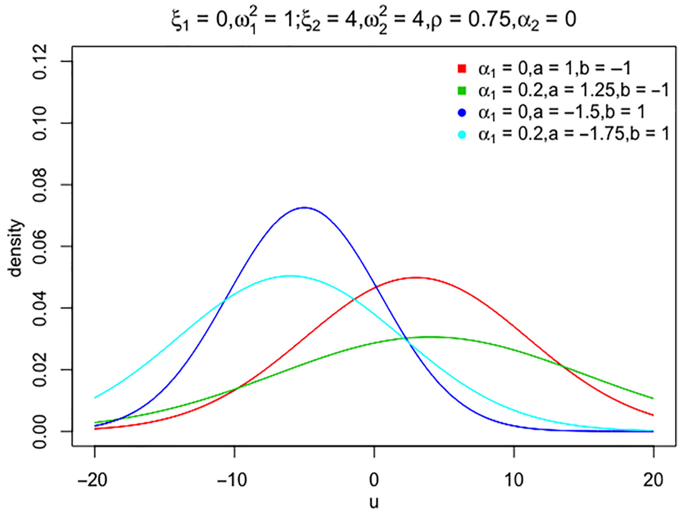


Figure 4.
The density of U when location parameters of the first and second population equal 0 and 4, the scale parameters for first and second population equal 1 and 2 respectively

Note(s): The shape parameter for the second populations equal 0 and for the first population equal 0, 0.2, 0, 0.2 with $a = 1$, $b = -1$, $a = 1.25$, $b = -1$, $a = -1.5$, $b = 1$ and $a = -1.75$, $b = 1$. The corresponding probability P(U > 0) is 0.64617, 0.666848, 0.181651 and 0.224808
Source(s): Figure by authors

2.3 Estimation of $\delta_{ab} = P(Z > 0)$ *and* $\delta_{\rho ab} = P(U > 0)$

The above two theorems imply that if the location, scale and shape parameters are known in two populations, one can calculate the probabilities $\delta_{ab} = P(Z > 0)$ and $\delta_{\rho ab} = P(U > 0)$. But in most situations in real life the population parameters are not known. Thus estimating δ_{ab} and $\delta_{\rho ab}$ is necessary to be studied. The parametric estimation of $P(X > Y)$ for normal distributions in the context of probabilistic environmental risk assignment was discussed by Jacobs *et al.* Here, we consider the method of moment estimator (MME) and maximum likelihood estimator (MLE). The estimators obtained in this two ways are denoted by $\widehat{\delta}_{ab}^{MME}$ and $\widehat{\delta}_{ab}^{MLE}$ respectively.

Method of moment estimator (MME) Equation (2.6), below, relates location to mean and scale to standard deviation.

$$\xi = \mu - \sqrt{\frac{2}{\pi}}\delta\omega \qquad \text{and} \qquad \omega^2 = \sigma^2\left(1 - \frac{2}{\pi}\delta^2\right)^{-1}, \tag{2.6}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. To obtain parameter estimators from samples, it is necessary to obtain an estimate of delta δ, which is the moment estimate $\widehat{\delta}$ in Equation (2.7) below.

$$|\widehat{\delta}| = \sqrt{\frac{\pi}{2}\frac{|\widehat{\gamma}_1|^{\frac{2}{3}}}{|\widehat{\gamma}_1|^{\frac{2}{3}} + ((4-\pi)/2)^{\frac{2}{3}}}} \tag{2.7}$$

where $\widehat{\gamma}_1$ is the sample skewness and the sign of $\widehat{\delta}$ is the same as the sign of $\widehat{\gamma}_1$. In turn, it is easy to obtain an estimate of the shape parameter α using Equation (2.8):

$$\widehat{\alpha} = \frac{\widehat{\delta}}{\sqrt{1-\widehat{\delta}^2}}. \tag{2.8}$$

Rewriting Equation (2.6) in terms of sample estimates, as opposed to population parameters, renders Equation (2.9).

$$\widehat{\xi} = \overline{X} - \sqrt{\frac{2}{\pi}}\widehat{\delta}\widehat{\omega} \qquad \text{and} \qquad \widehat{\omega}^2 = S^2\left(1 - \frac{2}{\pi}\widehat{\delta}^2\right)^{-1}. \tag{2.9}$$

Then we can obtain the $\widehat{\delta}_{ab}^{MME}$ by substituting the MMEs of ξ_1, ξ_2, ω_1, ω_2, α_1 and α_2 in Equation (2.3).

Maximum likelihood estimator (MLE) The log-likelihood function for the skew normal distribution is given by

$$\log(L) = n\log\left(\frac{2}{\omega}\right) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\xi}{\omega}\right)^2 + \sum_{i=1}^{n}\log\left(\Phi\left(\alpha\frac{x_i-\xi}{\omega}\right)\right) \tag{2.10}$$

To derive the MLEs of the parameters in skew normal distribution, we take the partial derivatives of the lnL functions with respect to parameters of interest and equal them to zero. Then the corresponding MLEs are obtained by solving the following equations:

$$
\begin{aligned}
\frac{logL}{\xi} &= \sum_{i=1}^{n}\left(\frac{x_i-\xi}{\omega}\right) - \alpha\sum_{i=1}^{n}\frac{\phi\left(\alpha\frac{x_i-\xi}{\omega}\right)}{\Phi\left(\alpha\frac{x_i-\xi}{\omega}\right)} = 0,\\
\frac{logL}{\omega} &= -n + \sum_{i=1}^{n}\left(\frac{x_i-\xi}{\omega}\right)^2 - \alpha\sum_{i=1}^{n}\frac{\phi\left(\alpha\frac{x_i-\xi}{\omega}\right)}{\Phi\left(\alpha\frac{x_i-\xi}{\omega}\right)}\left(\frac{x_i-\xi}{\omega}\right) = 0, \qquad (2.11)\\
\frac{logL}{\alpha} &= \sum_{i=1}^{n}\frac{\phi\left(\alpha\frac{x_i-\xi}{\omega}\right)}{\Phi\left(\alpha\frac{x_i-\xi}{\omega}\right)}\left(\frac{x_i-\xi}{\omega}\right) = 0.
\end{aligned}
$$

By the invariance property of MLEs, we obtain $\widehat{\delta}_{ab}^{MLE}$ by substituting the MLEs of ξ_1, ξ_2, ω_1, ω_2, α_1 and α_2 in Equation (2.3).

To derive $\widehat{\delta}_{\rho ab}^{MME}$ and $\widehat{\delta}_{\rho ab}^{MLE}$ under matched data setting, instead of estimating ξ_1, ξ_2, ω_1, ω_2, α_1 and α_2 we estimate $\xi_* = \xi_1 + a\xi_2 + b$, $\sigma^2 = \boldsymbol{c}'\Sigma\boldsymbol{c} = \omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2$ and

$$
\begin{aligned}
\alpha_* &= \frac{\delta_*}{\sqrt{1-\delta_*^2}}\\
&= \frac{\alpha_1 d_1 + \alpha_2 d_2}{\sqrt{(1-\alpha_1^2-\alpha_2^2)k^2\sigma^2 - (\alpha_1 d_1 + \alpha_2 d_2)^2}}
\end{aligned} \qquad (2.12)
$$

by using the same methods as in the independent setting. Then by substituting the MMEs and MLEs of ξ_*, σ^2 and α_* in Equation (2.5), we can get $\widehat{\delta}_{\rho ab}^{MME}$ and $\widehat{\delta}_{\rho ab}^{MLE}$.

3. Probability of being better off or worse off by how much? Delta log-skew-normal distributions

The log-normal distributions is very popular in both theory and applications of probability and statistics. However, when modeling complex random phenomena in many applied areas, there are a real need of more flexible models which extend the log-normal. Lin and Stoyanov introduced the LSN distribution which is an extension for modeling positive data using log-normal models. A positive random variable X is LSNly distributed (i.e. $X \sim LSN(\mu, \sigma, \alpha)$) if the logarithm of X is skew normally distributed with location parameter μ, scale parameter σ and slant parameter α. Let $\varphi(.)$ and $\Phi(.)$ be respectively probability density function (pdf) and cumulative distribution function (cdf) of the standard normal distribution, then we have that

$$
f_X(x) = \frac{2}{x\sigma}\,\varphi\left(\log x; \mu, \sigma^2\right)\,\Phi\left(\alpha\frac{\log x - \mu}{\sigma}\right), \qquad x > 0. \qquad (3.1)
$$

Density curves, $f_X(x)$, of X for parameters $\mu = 0$, $\sigma = 0.5$ with different values of $\alpha = -1, 0, 1$ are given in Figure 5. From Figure 5, we can see that the slant parameter α plays an important role in LSN family. The mean and variance of X are respectively given by

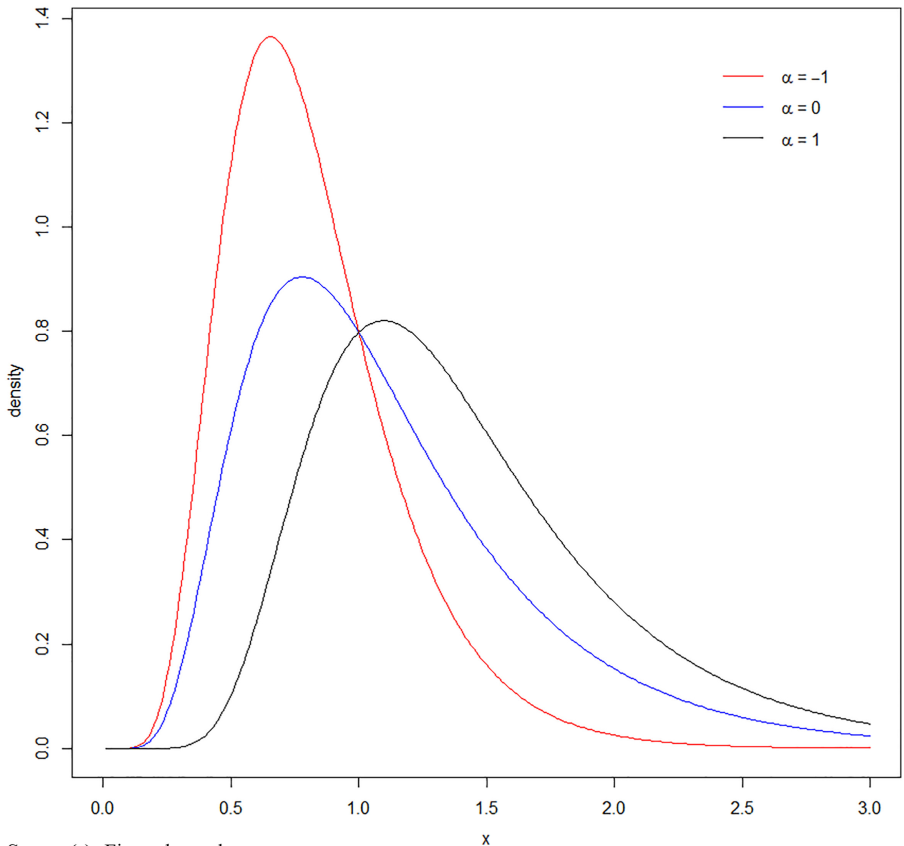
$$\begin{aligned} E(X) &= 2\exp\{\mu + \sigma^2/2\}\Phi(\alpha\sigma), \\ V(X) &= 2\exp\{2\mu + \sigma^2\}\left[\exp(\sigma^2)\Phi(2\alpha\sigma) - 2\Phi^2(\alpha\sigma)\right]. \end{aligned}$$

The following is the definition of the multivariate LSN distribution.

D3.1. *The random vector* $\mathbf{X} = (X_1, \ldots, X_p)'$ *is said to have a p-dimensional LSN distribution if* $\log \mathbf{X} \sim SN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$, *the multivariate skew-normal distribution with location parameter* $\boldsymbol{\mu}$, *scale matrix* Σ, *and the skewness parameter* $\boldsymbol{\alpha}$, *denoted by* $\mathbf{X} \sim LSN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$. *The pdf of* $\mathbf{X}$ *is*

$$h(\mathbf{x}) = \begin{cases} \dfrac{2}{\prod_{i=1}^{p} x_i} |\Sigma|^{-\frac{1}{2}} \varphi_p(\log\mathbf{x}; \boldsymbol{\mu}, \Sigma)\, \Phi\left[\boldsymbol{\alpha}'\Sigma^{-1}(\log\mathbf{x} - \boldsymbol{\mu})\right], & \text{if } \mathbf{x} \in \Re^{p+}, \\ 0 & \text{otherwise}, \end{cases} \tag{3.2}$$

where $\mathbf{x} = (x_1, \ldots, x_p)'$, $\log\ \mathbf{x} = (\log\ x_1, \ldots, \log\ x_p)'$, *and* $\Re^{p+} = \{\mathbf{x} | x_i > 0 \text{ for } i = 1, \ldots, p\}$.



Source(s): Figure by authors

Figure 5.
The density curves of log-skew normal distribution with $\mu = 0$, $\sigma = 0.5$ and values of $\alpha = 1, 0, -1$

Note that, in this paper, we only consider the family of the bivariate distributions, $p = 2$, for both dependent and independent cases.

In life sciences it is common for data to contain a relatively large number of zeros. In many real situations, it is appropriate to model such data using a LSN distribution for the positive values, together with an additional probability mass at all zeros. This type of distribution is commonly referred to as the delta LSN, defined below.

D3.2. The random variable X is said to have a delta-LSN distribution with parameters δ, μ, σ, and α, denoted as $X \sim DLSN(\delta, \mu, \sigma^2, \alpha)$, if the pdf of X is given by

$$g_X(x) = \begin{cases} 0 & if\, x < 0, \\ \delta & if\, x = 0, \\ (1-\delta) f_X(x), & if\, x > 0, \end{cases} \tag{3.3}$$

where $f_X(x)$ is given in Equation (3.1). *Similarly, the random vector $(X_1, X_2)'$ is said to have a bivariate delta LSN distribution with parameters $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2)'$, $\boldsymbol{\mu}$, Σ, and $\boldsymbol{\alpha}$, denoted as $\mathbf{X} \sim DLSN_2(\boldsymbol{\delta}, \boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$ if its cdf of $\mathbf{X}$ is given by*

$$\begin{cases} P(X_1 < 0, X_2 < 0) = 0, & if\, x_1 < 0, x_2 < 0, \\ P(X_1 = 0, X_2 = 0) = \delta_0, & if\, x_1 = 0, x_2 = 0, \\ P(X_1 \le x_1, X_2 = 0) = \delta_1 F(x_1), & if\, x_1 > 0, x_2 = 0, \\ P(X_1 = 0, X_2 \le x_2) = \delta_2 G(x_2), & if\, x_1 = 0, x_2 > 0, \\ P(X_1 \le x_1, X_2 \le x_2) = \left(1 - \sum_{i=0}^{2} \delta_i\right) H(x_1, x_2), & if\, x_1 > 0, x_2 > 0, \end{cases} \tag{3.4}$$

where $0 \le \delta_i < 1$, $i = 0, 1, 2$, and $1 - \sum_{i=0}^{2} \delta_i > 0$, $F(x_1)$ and $G(x_2)$ are the cdf's of $X_1 \sim LSN(\mu_1, \sigma_1^2, \alpha_1)$ and $X_2 \sim LSN(\mu_2, \sigma_2^2, \alpha_2)$, respectively, and $H(x_1, x_2)$ is the joint cdf of $\mathbf{X} \sim LSN_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$.

Note that the requirement of $1 - \sum_{i=0}^{2} \delta_i > 0$ is necessary as $\mathbf{X} \sim LSN_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$ is assumed. Also if X_1 and X_2 are independent, then from Equation (3.4), we can obtain that

$$X_1 \sim DLSN\left(\delta_{*1}, \mu_1, \sigma_1^2, \alpha_1\right) \quad \text{and} \quad X_2 \sim DLSN\left(\delta_{*2}, \mu_2, \sigma_2^2, \alpha_2\right)$$

with $\delta_0 = \delta_{*1}\delta_{*2}$, $\delta_1 = (1 - \delta_{*1})\delta_{*2}$, $\delta_2 = \delta_{*1}(1 - \delta_{*2})$ and $H(x_1, x_2) = F(x_1)G(x_2)$ for all $x_1, x_2 > 0$.

3.1 Linear combination of independent delta log-skew-normal distributions

Let $X \sim DLSN(\delta_1, \mu_1, \sigma_1^2, \alpha_1)$, $Y \sim DLSN(\delta_2, \mu_2, \sigma_2^2, \alpha_2)$ and we assume that X and Y are independent. We want to obtain the distribution of the linear combination, $Z = X + aY + b$, of X and Y for specified constants, $a \neq 0$ and $b \in \Re$; and then find $P(Z > 0)$, the probability of $Z > 0$.

Theorem 3.1. Let $X \sim DLSN(\delta_1, \mu_1, \sigma_1^2, \alpha_1)$, $Y \sim DLSN(\delta_2, \mu_2, \sigma_2^2, \alpha_2)$ and assume that X and Y are independent. Then the pdf of $Z = X + aY + b$ is

(i) for $a < 0$,

$$f_Z(z) = \begin{cases} \int_{\frac{z-b}{a}}^{\infty} f_X(z - ay - b) f_Y(y) dy - \frac{1}{a} f_Y\left(\frac{z-b}{a}\right), & if\, z < b, \\ \delta_1 \delta_2 & if\, z = b, \\ \int_0^{\infty} f_X(z - ay - b) f_Y(y) dy + f_X(z - b), & if\, z > b. \end{cases} \tag{3.5}$$

(ii) for $a > 0$,

$$f_Z(z) = \begin{cases} 0 & if z < b, \\ \delta_1\delta_2 & if z = b, \\ \int_0^{\frac{z-b}{a}} f_X(z-ay-b) f_Y(y)dy + \frac{1}{a} f_Y\left(\frac{z-b}{a}\right) + f_X(z-b), & if z > b. \end{cases} \tag{3.6}$$

where $f_X(x)$, $f_Y(y)$ are pdf's of X when $x > 0$ and Y when $y > 0$, respectively.

The proof of Theorem 3.1 is given in Appendix. Note that if $a = 0$, then $Z = X + b$ so that we only have the univariate case.

Now we can compute the probability $P(Z > 0)$, which is given by.

(*i*) for $a < 0$,

$$P(Z>0) = \begin{cases} \int_0^\infty \int_0^\infty f_X(z-ay-b) f_Y(y)dydz + \int_0^\infty f_X(z-b)dz, & \text{if} b < 0, \\ \int_b^\infty \int_0^\infty f_X(z-ay-b) f_Y(y)dydz + \int_b^\infty f_X(z-b)dz & \\ + \int_0^b \int_{\frac{z-b}{a}}^\infty f_X(z-ay-b) f_Y(y)dydz - \int_0^b \frac{1}{a} f_Y\left(\frac{z-b}{a}\right)dz, & \text{if} b \geqslant 0. \end{cases} \tag{3.7}$$

(*ii*) for $a > 0$,

$$P(Z>0) = \begin{cases} \int_0^\infty \int_0^{\frac{z-b}{a}} f_X(z-ay-b) f_Y(y)dydz + \int_0^\infty \frac{1}{a} f_Y\left(\frac{z-b}{a}\right)dz & \\ + \int_0^\infty f_X(z-b)dz, & \text{if} b < 0, \\ \int_b^\infty \int_0^{\frac{z-b}{a}} f_X(z-ay-b) f_Y(y)dydz + \int_b^\infty \frac{1}{a} f_Y\left(\frac{z-b}{a}\right)dz & \\ + \int_b^\infty f_X(z-b)dz, & \text{if} b \geqslant 0. \end{cases} \tag{3.8}$$

Here is a link of shiny apps and its corresponding R code is available upon the request: https://dlnprobability.shinyapps.io/dlsn_independent/.

3.2 Bivariate delta log-skew-normal distribution

Now we consider the bivariate delta LSN random vector $\mathbf{X} = (X_1, X_2)' \sim DLSN_2(\boldsymbol{\delta}, \boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$, given by (3.4) with parameters given by

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

Then we have the following result and its proof is given in Appendix.

Theorem 3.2. *Let* $\mathbf{X} \sim DLSN_2(\boldsymbol{\delta}, \boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$ *given above and consider* $U = X_1 + aX_2 + b$ *with constants* $a \neq 0$ *and* $b \in \Re$. *Then the pdf of* U *is*

(*i*) *for* $a < 0$

$$f_U(u) = \begin{cases} \int_{\frac{u-b}{a}}^{\infty} (1 - \delta_0 - \delta_1 - \delta_2) h(u - ax_2 - b, x_2) dx_2 - \frac{\delta_2}{a} g\left(\frac{u-b}{a}\right), & if\, u < b, \\ \delta_0 & if\, u = b, \\ \int_0^{\infty} (1 - \delta_0 - \delta_1 - \delta_2) h(u - ax_2 - b, x_2) dx_2 + \delta_1 f(u - b), & if\, u > b. \end{cases} \tag{3.9}$$

(*ii*) *for* $a > 0$

$$f_U(u) = \begin{cases} 0 & if\, u < b, \\ \delta_0 & if\, u = b, \\ \int_0^{\frac{u-b}{a}} (1 - \delta_0 - \delta_1 - \delta_2) h(u - ax_2 - b, x_2) dx_2 & \\ + \frac{\delta_2}{a} g\left(\frac{u-b}{a}\right) + \delta_1 f(u - b), & if\, u > b. \end{cases} \tag{3.10}$$

where $0 \leq \delta_i < 1$, $i = 0, 1, 2$, *and* $1 - \sum_{i=0}^{2} \delta_i > 0$, $f(x_1)$ *and* $g(x_2)$ *are the pdf's of* $X_1 \sim LSN(\mu_1, \sigma_1^2, \alpha_1)$ *and* $X_2 \sim LSN(\mu_2, \sigma_2^2, \alpha_2)$, *respectively, and* $h(x_1, x_2)$ *is the joint pdf of* $\mathbf{X} \sim LSN_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$.

Here is a link of shiny apps and its corresponding code is available upon the request: https://dlnprobability.shinyapps.io/dlsn_dependent/.

4. Applications: invented examples

To explain the possibilities, we invented examples. First, we examine skew normal examples and then lognormal examples.

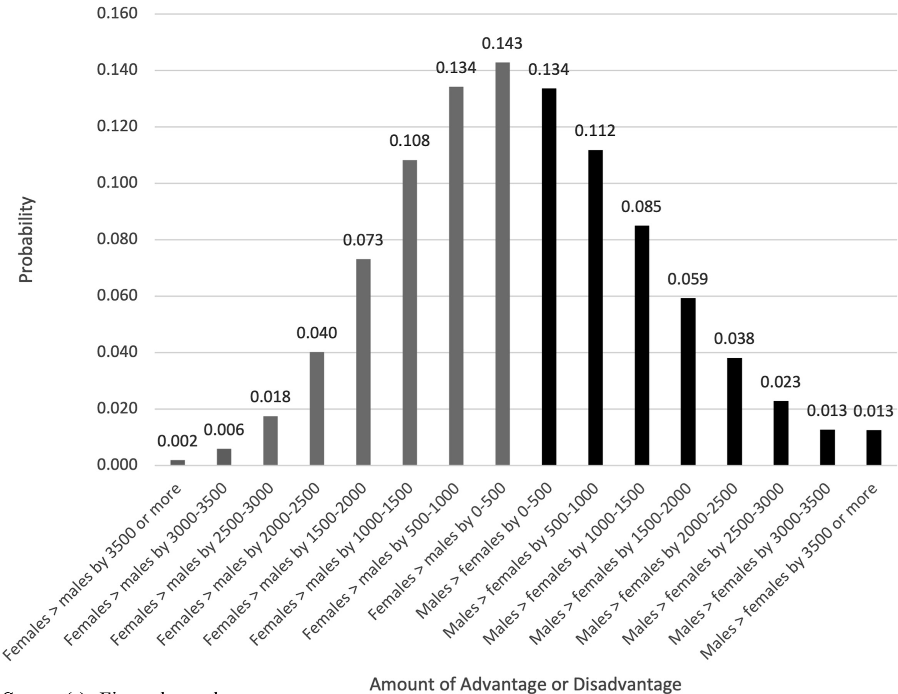
4.1 Skew normal examples

Suppose an economist is interested in comparing salaries of males versus females in a particular geographical area. In this geographical area, based on large samples of randomly selected males and females, suppose the mean salaries are \$45,000 for both males and females, the standard deviations are \$1,000 for each, with skews of 0.5 and −0.5 for males and females, respectively. Because the mean salary is the same for males and females, there seems no reason to promote a policy change to redress inequality in salaries.

However, let us convert the foregoing statistics to estimates of skew normal parameters: location, scale and shape. After making the conversion, the estimated locations are \$43,947.79 and \$46,052.21 for males and females, respectively; the estimated scales are \$1,451.601 for both males and females (the squared value is \$2107144.694); and the estimated distribution shapes are 2.173758 and − 2.173758 for males and females, respectively. We saw earlier that, going by means, there is no relative advantage for males or females. However, we now see that going by locations, there is a relative disadvantage for males relative to females. The difference in means implies a neutral policy whereas the difference in locations implies a policy to increase male salaries to bring them to the level of female salaries. However, neither difference answers the real question, which is: What is the probability of males being advantaged or disadvantaged, relative to females and by how much? Fortunately, the foregoing skew normal mathematical equations allow us to answer this question, especially with the aid of the program that can be accessed via the following link: https://probab.shinyapps.io/inde_prob/.

Figure 6 contains a G-P diagram that provides the probabilities that males will be advantaged (black bars) or disadvantaged (gray bars) relative to females. As will be seen presently, G-P diagrams provide the advantage that researchers can assess probabilistic advantages or disadvantages, at varying degrees of extremity. Note that each gray bar is higher than its corresponding black bar, demonstrating a consistent advantage for females over males. For example, the probability that a randomly selected male would have an advantage of 0–500 is 0.134 whereas the probability that a randomly selected female would have a similar advantage is 0.143. In general, there is asymmetry favoring females over males. Thus, using skew normal statistics to construct a G-P diagram can lead to different conclusions than remaining with normal statistics. Note, too, that if we perform a significance test on the means, the typical t-test would come out consistent with the null hypothesis of no difference whereas Figure 1 shows a very clear difference.

Or consider another example to illustrate the potential importance of seeming trivial skewness. Consider the deductible for doctor visits based on two insurance companies labeled INSURE-A and INSURE-B. For both insurance companies, the location is \$60 and the scale is \$10 (squared scale is \$100). The difference is that the skewness for INSURE-A is 0.10 whereas it is -0.10 for INSURE-B. The small skewness magnitude is so small for both companies that the distributions are very near normal, and typical statisticians and data analysts would treat



Source(s): Figure by authors

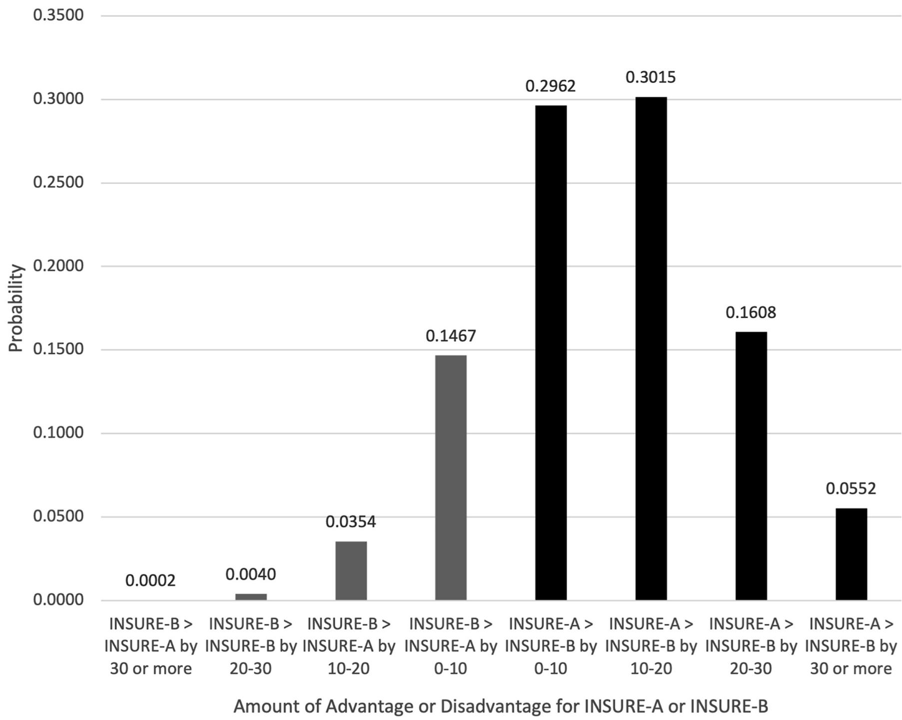
Figure 6. The probability of males or females being better off, or worse off, to varying degrees, with respect to salaries

the distributions as normal. The obvious conclusion is that the two insurance companies are equivalent. However, constructing a G-P diagram suggests a very different conclusion.

Figure 7 shows that despite the two insurance companies having equivalent locations, scales and skews seemingly only trivially different from 0, the G-P diagram shows that a randomly selected doctor's visit is much more likely to be varying degrees more expensive with INSURE-A than with INSURE-B. Each black bar, representing INSURE-A resulting in more expensive payments, is substantially higher than its corresponding gray bar, representing INSURE-B resulting in more expensive payments. The G-P diagram is quite asymmetric; the black bar (INSURE-A) probabilities are higher than corresponding gray bar (INSURE-B) probabilities. Thus, if the goal is to reduce the expected payment amount, the G-P diagram indicates INSURE-B is a substantially statistically better bet than is INSURE-A, despite the seeming equivalence based on similar locations. More generally, Figure 7 shows that seeming trivial skewness differences can have extreme consequences.

4.2 Lognormal examples

Consider time wasted per hour at work in two different industries. Suppose that time wasted is lognormally distributed in companies in Business A and Business B. In Business A, the mean time wasted is 11.373, the standard deviation is 2.888 and the variance is 8.342. In Business B, the mean is 11.078, the standard deviation is 12.376 and the variance is 153.158. The effect size is 0.04, which is trivial, so the obvious conclusion is that type of business has little to do with time wasted per hour. Too, with such a small effect size, a significance test likely would come out insignificant with typical sample sizes.



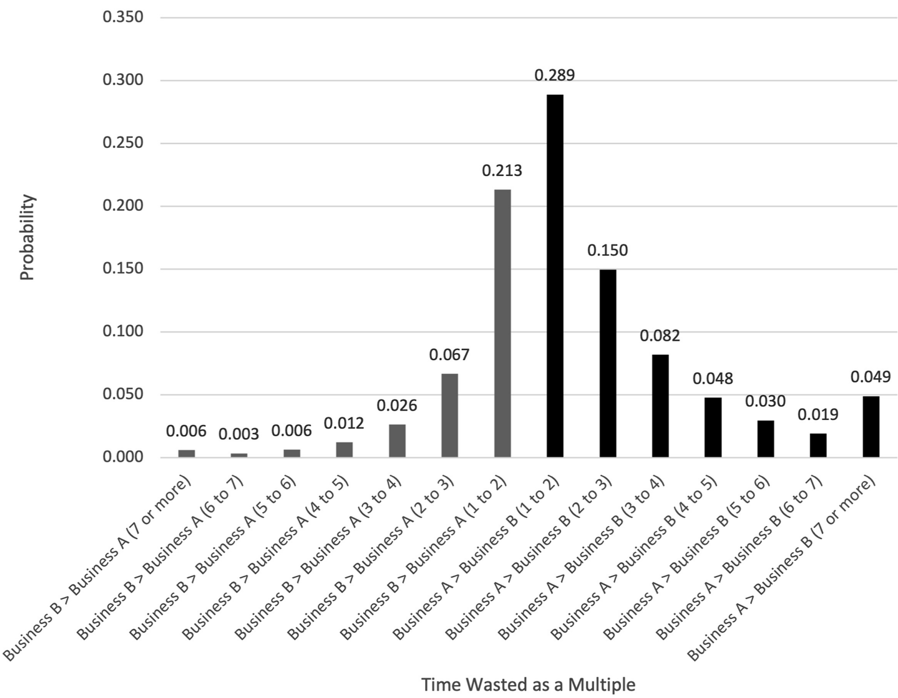
Source(s): Figure by authors

Figure 7.
The probability of being better off, or worse off, by varying degrees, with respect to co-payments for doctor visits

However, let us now consider lognormal statistics. The log transformed mean and standard deviation for Business A is 2.4 and 0.25 (variance is 0.0625), respectively. These values are 2.0 and 0.9 (variance is 0.81) for Business B. Thus, the probability that a randomly selected person in Business A will waste more time than a randomly selected person in Business B is approximately 0.67, despite the miniscule conventional effect size in the previous paragraph. In addition, it is possible to construct a G-P diagram with more nuanced information. One capability of G-P diagrams is that effects can be reported as multiples. For example, it is possible to concern ourselves with the probability that time-wasting is twice, thrice, etc. that in one business relative to the other. The G-P diagram renders clear that which would be obscured otherwise; time-wasting is much more of a problem for Business A than for Business B (see Figure 8).

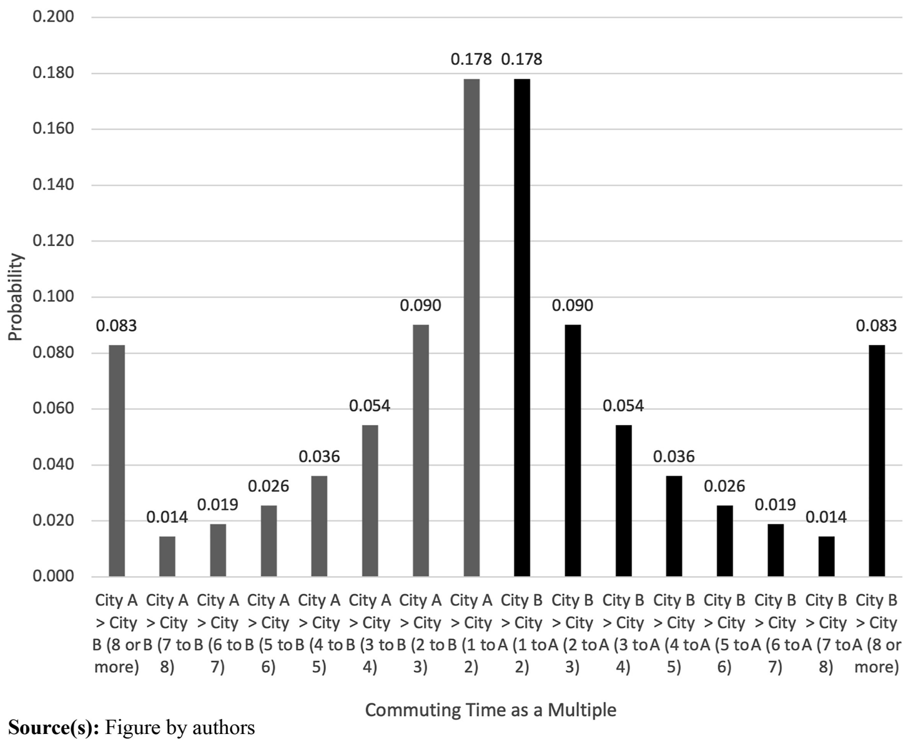
Imagine two large cities, City A and City B and an economist is concerned with comparing commuting times. Suppose commuting times are lognormally distributed and the mean commuting time for City A is 41.264 min, the standard deviation is 74.054 and the variance is 5484.041. For City B, these values are 30.114, 33.641 and 1131.691, respectively. There seems an obvious advantage for City B over City A in that the mean commute is shorter for City B than for City A by 11.15 min.

The mean of the logarithmically transformed distribution for City A is 3.0, the standard deviation is 1.2 and the variance is 1.44. For City B, these values are 3.0, 0.90 and 0.81, respectively. The probability that a randomly selected person from City A would have a shorter commuting time than a randomly selected person from City B is 0.50; despite appearances based on the previous paragraph, there is no probabilistic advantage for City A over City B. Moreover, the G-P diagram is symmetric; as each black bar in Figure 9 is the same size as its corresponding gray bar, there is no probabilistic advantage for either city, no matter the multiple under consideration.



Source(s): Figure by authors

Figure 8. The probability of being better off or worse off, by differing multiples



Source(s): Figure by authors

Figure 9.
The probability of being better off by varying multiples, with respect to commuting times, depending on city

Consider another example, using the same log normal parameter values as the previous example, but with the addition of considering shape. Suppose that the shape parameter for City A is 0.10 and the shape parameter for City B is -0.10. In that case, the probability that a randomly selected person in City A would have a longer commuting tie than a randomly selected person in City B is 0.54, not 0.50. And the probability that a randomly selected person in City B would have a longer commuting time than a randomly selected person in City A is 0.46. Thus, in contrast to the previous example, the slight difference in shape parameters leads to a clear probabilistic disadvantage for City A relative to City B. In addition, in contrast to Figure 9 that is symmetrical, Figure 10 that illustrates the present example is asymmetric.

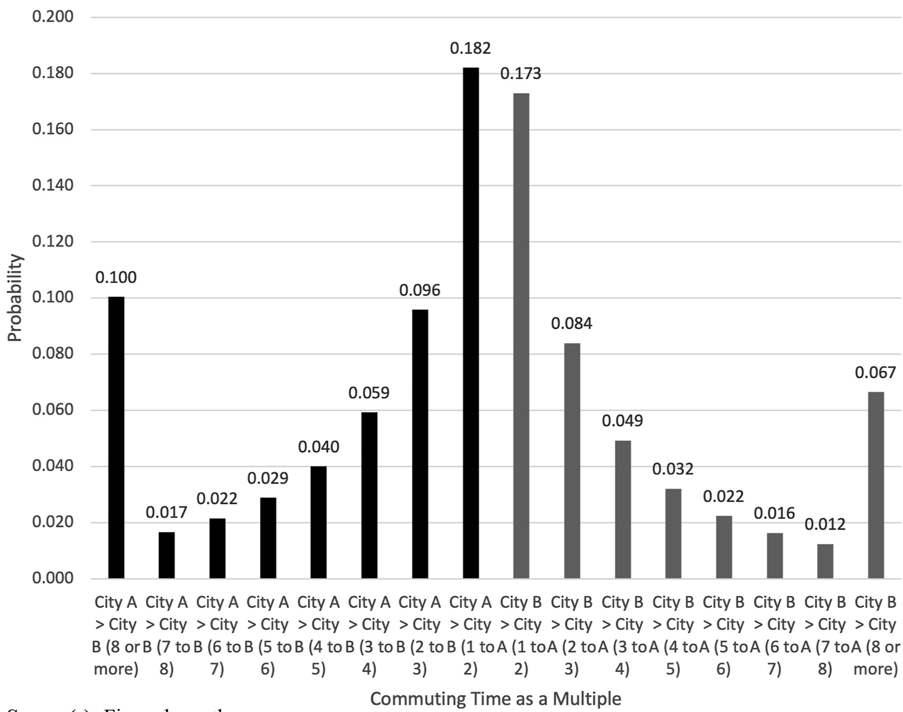
5. Application: real data

In the previous section, we examined invented examples to demonstrate the lessons that can be learned by constructing G-P diagrams. In this section, we present analyses of real data.

5.1 NBA guards (2022–2023 regular season)

We downloaded data from the following website: https://www.espn.com/nba/stats/player/_/season/2023/seasontype/2/table/defensive/sort/avgSteals/dir/desc. Among other types of information, it contains data for minutes played by point guards and shooting guards. Let us compare them using traditional normal statistics versus skew normal statistics.

The means for point guards and shooting guards are 24.0 and 19.8, respectively. The standard deviations are 8.8 and 9.4, respectively. In addition, the skews are -0.3 and 0.2, respectively. Thus, point guards average 4.2 more minutes per game than shooting guards.



Source(s): Figure by authors

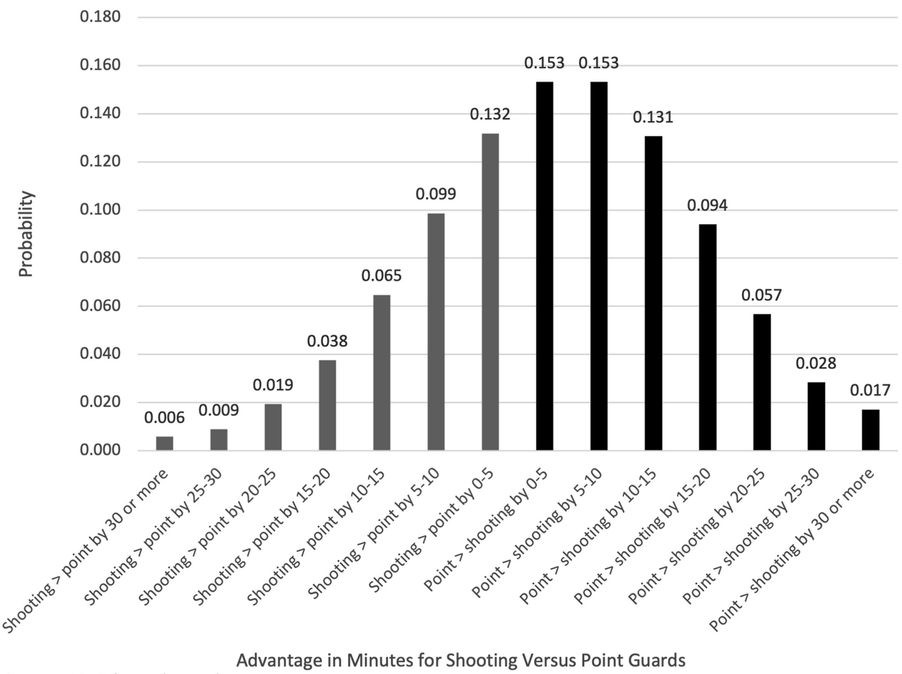
Figure 10. Probability of being better off or worse off, by varying multiples, with respect to commuting times, depending on city and taking shape parameters into account

We might ask about how likely a randomly chosen point guard would average more minutes than a randomly chosen shooting guard. Based on the difference in averages, we might guess that the probability that randomly chosen point guard would average more minutes than a randomly chosen shooting guard would be greater than 0.50, but perhaps not too much more than that. But rather than guess, let us run out the calculations.

As usual, the first step is to find the skew normal values for location, scale and shape. For point guards these are 31.4, 11.5 (scale squared is 131) and −1.4, respectively. For shooting guards, these are 13.1, 11.6 (scale squared is 134) and 1.1, respectively. Based on these values, the probability that a randomly selected point guard would play more minutes per game than a randomly selected shooting guard is 0.63. Figure 11 provides a G-P diagram. An interesting aspect of the figure is that the middle two bars are not very far apart, thereby indicating that at small amounts of difference in minutes played, there is only a slight probabilistic advantage for point guards over shooting guards. However, comparing black against gray bars at greater extremes indicates a strong probabilistic advantage for point guards over shooting guards. The G-P diagram suggests a more subtle substantive story than does a mere difference in means or difference in locations.

5.2 Living space for males versus females

Hyman *et al.* (2002) obtained data comparing living space for their male and female participants, in square feet. The mean amount of living space for men is 2300 and the standard deviation is 705. The mean amount of living space for women is 2186 and the standard deviation is 671. The effect size is miniscule and not statistically significant: Cohen's $d = 0.04$, $t(564) < 1$. However, the data are lognormally distributed and a G-P diagram



Source(s): Figure by authors

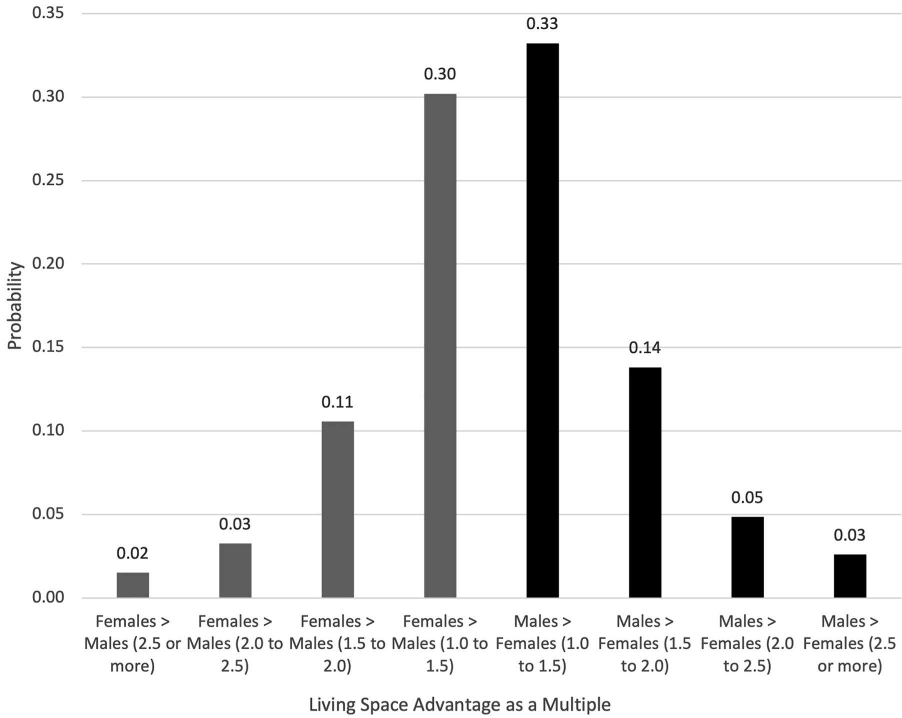
Figure 11. Probability of varying degrees of advantage (or disadvantage) for shooting guards or point guards with respect to minutes played

suggests a very different conclusion. To commence, let us find the means and standard deviations after a logarithmic transformation. The transformed mean and standard deviation for males is 7.69 and 0.31 (variance = 0.0961), respectively. The transformed mean and standard deviation for females is 7.64 and 0.32 (variance = 0.1024), respectively. Thus, the probability that a randomly selected male will have more living space than a randomly selected female is 0.54 whereas the probability that a randomly selected female will have more living space than a randomly selected male is only 0.46, a more impressive difference than is suggested by the Cohen's *d* or the nonsignificant *p*-value.

For a more fine-grained analysis, Figure 12 provides the G-P diagram. Because it is convenient to express gender effects with respect to living space in multiples, Figure 12 is in terms of multiples. That is, the black bars provide the probabilities that a randomly selected male will have 1.0 to 1.5 times as much living space as a randomly selected female, 1.5 to 2.0 times as much living space and so on. The gray bars provide analogous probabilistic advantages for females. Each black bar is larger than its corresponding gray bar thereby indicating that although the probabilistic advantage for males over females is not large, it is not trivial either. Again, we see that a G-P diagram provides both more valid and more nuanced conclusions than typical differences between means, Cohen's *d*, or significance tests.

Consider precipitation in Buffalo for 306 days starting in June in 2019 versus 2020 [https://www.ncdc.noaa.gov/cdo-web/]. Because there are days when there is no precipitation, the data fall under the category of delta skew lognormal distributions. Thus, there are the following bullet-listed summary parameter estimates for each year:

- locations are 2.8712 for 2019 and 2.8327 for 2020,
- squared scales are 3.884841 for 2019 and 3.806401 for 2020,



Source(s): Figure by authors

Figure 12. Probability that males or females are better off, by varying multiples, with respect to living space

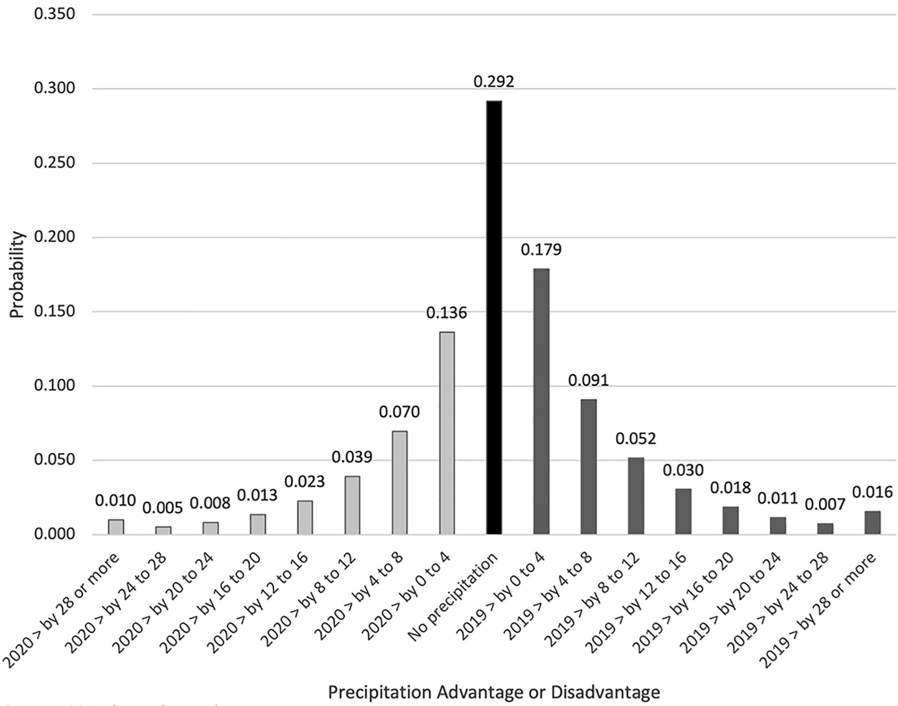
- shapes are −2.929 for 2019 and -3.019 for 2020 and
- probabilities of zeroes (lack of precipitation for a day) are 0.49346 for 2019 and 0.59150 for 2020.

This is an especially interesting example because (a) it engages all the parameters of delta skew lognormal distributions and (b) ties are possible. A tie can occur if there is no precipitation on a randomly selected day in 2019 or 2020. The probability of more precipitation on a randomly selected day in 2019 than on a randomly selected day in 2020 is 0.404. The probability of more precipitation on a randomly selected day in 2020 than on a randomly selected day in 2019 is 0.304 and the probability of a tie is 0.292. Thus, there is a general probabilistic advantage for 2019 over 2020, assuming precipitation is positive. (If precipitation is negative, then there is a probabilistic disadvantage for 2019 relative to 2020. However, we will assume that precipitation is positive.)

Figure 13 shows a G-P diagram. However, the diagram differs from the others because ties are possible. The black bar in the middle shows the probability of ties. The dark gray bars show the probability of more precipitation in 2019 than in 2020, by varying millimeters. And the light gray bars show the probability of more precipitation in 2020 than in 2019, by varying millimeters. The figure shows that the probabilistic advantage for 2019 over 2020 is clear both at small amounts and at large ones.

6. Discussion

The G-P diagrams based on both invented and real data indicate important lessons for economics researchers. One lesson is that both summary statistics and significance tests can



Source(s): Figure by authors

Figure 13.
Probability that there is a precipitation advantage, by varying degrees, for 2019 or 2020, or that there is no advantage for either year (black bar)

be very misleading. It is possible for a difference in means to hide that there is no probabilistic advantage for either group over the other. However, it is possible, too, for a lack of difference in means to hide a very sizable probabilistic advantage for one group over the other. One solution is to use more appropriate summary statistics. For example, if the data are obtained from a skew-normal distribution, locations, scales and shapes can be better than means and standard deviations. However, even distribution-appropriate sample statistics can be misleading. For instance, we have seen an example where the locations and scales of two groups are the same and the difference in skews very small. Nevertheless, the G-P diagram (Figure 7) indicated a very large probabilistic difference.

Although it is dramatic to show that traditional statistics and G-P diagrams can come to opposing conclusions, G-P diagrams have another advantage. Specifically, G-P diagrams can support subtle conclusions that cannot be addressed by either summary statistics or significance tests. Figure 11 provides a nice case in point. It shows that at relatively small differences in minutes played, there is only a miniscule probabilistic advantage for point guards over shooting guards. But at more extreme differences in minutes played, the probabilistic advantage for point guards increases substantially.

In conclusion, there is no need for economics researchers to constrain themselves by depending solely on summary statistics and null hypothesis significance tests. As we have seen, G-P diagrams provide opportunities to contradict the seeming implications of summary statistics or null hypothesis significance tests. In addition, even in cases where there is no contradiction, such as the data pertaining to minutes played by National Basketball Association (NBA) point guards and shooting guards, G-P diagrams provide much more nuanced information than do summary statistics and null hypothesis significance tests. We hope and expect that future economics researchers will avail themselves of the potential advantages to be enjoyed by exploiting the capabilities of G-P diagrams for providing subtle probabilistic information about comparisons of interest.

References

Azzalini, A. and Valle, A.D. (1996), "The multivariate skew normal distribution", *Biometrika*, Vol. 83 No. 4, pp. 715-726, doi: 10.1093/biomet/83.4.715.

Hubbard, R. (2016), *Corrupt Research: the Case for Reconceptualizing Empirical Management and Social Science*, Sage Publications, Los Angeles, CA.

Hyman, M.R., Ganesh, G. and McQuitty, S. (2002), "Augmenting the household affluence construct", *Journal of Marketing Theory and Practice*, Vol. 10 No. 3, pp. 13-31, doi: 10.1080/10696679.2002.11501917.

McQuitty, S. (2004), "Statistical power and structural equation models in business research", *Journal of Business Research*, Vol. 57 No. 2, pp. 175-183, doi: 10.1016/S0148-2963(01)00301-0.

McQuitty, S. (2018), "Reflections on 'Statistical power and structural equation models in business research'", *Journal of Global Scholars of Marketing Science*, Vol. 28 No. 3, pp. 272-277, doi: 10.1080/21639159.2018.1434806.

Tong, T., Wang, T., Trafimow, D. and Wang, C. (2022), "The probability of being better or worse off, and by how much, depending on experimental conditions with skew normal populations", in SriboonchittaKreinovich, W.Y.E.S.V. (Ed.), *Credible Asset allocation, Optimal Transport Methods, and Related Topics (141-149)*, Springer-Verlag.

Trafimow, D. (2019), "A frequentist alternative to significance testing, p-values, and confidence intervals", *Econometrics*, Vol. 7 No. 2, pp. 1-14, available at: https://www.mdpi.com/2225-1146/7/2/26

Trafimow, D., Hyman, M.R., Kostyk, A., Wang, C. and Wang, T. (2021), "The harmful effect of null hypothesis significance testing on marketing research: an example", *Journal of Business Research*, Vol. 125, pp. 39-44, doi: 10.1016/j.jbusres.2020.11.069.

Trafimow, D., Hyman, M.R., Kostyk, A., Wang, Z., Tong, T., Wang, T. and Wang, C. (2022), "Gain-probability diagrams in consumer research", *International Journal of Market Research*, Vol. 64 No. 4, pp. 470-483, doi: 10.1177/14707853221085509.

Wang, Z., Wang, T., Trafimow, D. and Xu, Z. (2022), "A different kind of effect size based on samples from two populations with delta log-skew-normal distributions", in Ngoc Thach, N., Ha, D.T., Trung, N.D. and Kreinovich, V. (Eds), *Prediction and Causality in Econometrics and Related Topics. ECONVN 2021. Studies in Computational Intelligence*, Springer, Cham, Vol. 983, doi: 10.1007/978-3-030-77094-5_10.

Ziliak, S.T. and McCloskey, D.N. (2016), *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, The University of Michigan Press, Ann Arbor, *MI*.

Appendix

Source(s): Appendix by authors

In this section, we provide detailed proofs of Theorem 2.1, 2.2, 3.1, Theorem 3.2.

A.1 . Proof of Theorem 2.1

The joint pdf of $X \sim SN(\xi_1, \omega_1, \alpha_1)$ and $Y \sim SN(\xi_2, \omega_2, \alpha_2)$ is

$$f(x,y) = 4\phi\left(x; \xi_1, \omega_1^2\right)\phi\left(y; \xi_2, \omega_2^2\right)\Phi\left(\alpha_1\frac{x-\xi_1}{\omega_1}\right)\Phi\left(\alpha_2\frac{y-\xi_2}{\omega_2}\right).$$

Let $Z = X + aY + b$, $Y = Y$, then $X = Z - aY - b$, $Y = Y$ and it is easy to see that the Jacobian J of this transformation is 1. Thus, we obtain the pdf of Z as

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - ay - b) f_Y(y) dy,$$

which is

$$\begin{aligned}
f_Z(z) &= 4\int_{-\infty}^{\infty} \phi\left(z - ay - b; \xi_1, \omega_1^2\right)\phi\left(y; \xi_2, \omega_2^2\right)\Phi\left(\alpha_1 \frac{z - ay - b - \xi_1}{\omega_1}\right)\Phi\left(\alpha_2 \frac{y - \xi_2}{\omega_2}\right) \\
&= \frac{2}{\pi\omega_1\omega_2}\int_{-\infty}^{\infty} e^{-\left[\frac{(z-ay-b-\xi_1)^2}{2\omega_1^2} + \frac{(y-\xi_2)^2}{2\omega_2^2}\right]}\Phi_2\left[\left(\alpha_1 \frac{z - ay - b - \xi_1}{\omega_1}, \alpha_2 \frac{y - \xi_2}{\omega_2}\right)'; \mathbf{0}, I_2\right] dy \\
&= \frac{4}{\omega_*}\phi\left(z; \xi_1 + a\xi_2 + b, \omega_1^2 + a^2\omega_2^2\right)\int_{-\infty}^{\infty} \phi\left(y; c, \omega_*^2\right) \\
&\times \Phi_2\left[\left(\alpha_1 \frac{z - ay - b - \xi_1}{\omega_1}, \alpha_2 \frac{y - \xi_2}{\omega_2}\right)'; \mathbf{0}, I_2\right] dy,
\end{aligned} \tag{A.1}$$

where

$$\omega_* = \frac{\omega_1\omega_2}{\sqrt{\omega_1^2 + a^2\omega_2^2}} \quad \text{and} \quad c = \frac{a\omega_2^2(z - b - \xi_1) + \omega_1^2\xi_2^2}{\omega_1^2 + a^2\omega_2^2}.$$

Let $T = \frac{y-c}{\omega_*}$. Then Equation(A.1) will be

$$\begin{aligned}
f_Z(z) &= 4\phi\left(z; \xi_1 + a\xi_2 + b, \omega_1^2 + a^2\omega_2^2\right) \times \\
&E_T\left\{\Phi_2\left[\left(\alpha_1 \frac{z - a(t\omega_* + c) - \xi_1 - b}{\omega_1}, \alpha_2 \frac{t\omega_* + a + c - \xi_2}{\omega_2}\right)'; \mathbf{0}, I_2\right]\right\}
\end{aligned} \tag{A.2}$$

with $T \sim N(0, 1)$. If we denote $X_1 = \alpha_1 \frac{z - a(t\omega_* + c) - \xi_1 - b}{\omega_1}$ and $X_2 = \alpha_2 \frac{t\omega_* + a + c - \xi_2}{\omega_2}$, then E_T in Equation (A.2) can be simplified as

$$\begin{aligned}
E_T\left[\Phi_2(X_1,\ X_2)', \mathbf{0}, I_2\right] &= E_T[P(U_1 \le X_1,\ U_2 \le X_2 | X_1, X_2)] \\
&= P(U_1 \le X_1,\ U_2 \le X_2) \\
&= P(U_1 - X_1 \le 0,\ U_2 - X_2 \le 0),
\end{aligned} \tag{A.3}$$

where U_1, U_2 and T are independent standard normal random variables. Note that $E(U_i - X_i) = -E(X_i)$ and $\mathrm{Var}(U_i - X_i) = 1 + \mathrm{Var}(X_i)$ for $i = 1, 2$. Additionally,

$$Cov(U_1 - X_1,\ U_2 - X_2) = \mathrm{Cov}(X_1,\ X_2) = \frac{-a\alpha_1\alpha_2\omega_*^2}{\omega_1\omega_2}.$$

Therefore, the joint distribution of $(U_1 - X_1,\ U_2 - X_2)'$ is

$$(U_1 - X_1,\ U_2 - X_2)' \sim N_2(\boldsymbol{\mu},\ \Sigma_J),$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} -\alpha_1 \dfrac{z - ac - \xi_1 - b}{\omega_1} \\ -\alpha_2 \dfrac{c - \xi_2}{\omega_2} \end{pmatrix} \quad \text{and} \quad \Sigma_J = \begin{pmatrix} 1 + \alpha_1^2 \dfrac{a^2 \omega_*^2}{\omega_1^2} & -\dfrac{a\alpha_1\alpha_2\omega_*^2}{\omega_1\omega_2} \\ -\dfrac{a\alpha_1\alpha_2\omega_*^2}{\omega_1\omega_2} & 1 + \alpha_2^2 \dfrac{\omega_*^2}{\omega_2^2} \end{pmatrix}.$$

Then Equation(A.3) can be written to be

$$E_T\left[\Phi_2(X_1,\ X_2)', \mathbf{0}, I_2\right] = \Phi_2(-\boldsymbol{\mu}; \mathbf{0}, \Sigma_J),$$

and therefore, Equation (A.1) will be

$$f_{1,2}(z) = 4\phi\left(z; \xi_1 + a\xi_2 + b, \omega_1^2 + a^2\omega_2^2\right)\Phi_2[\boldsymbol{B}(z - (\xi_1 + a\xi_2 + b)); \mathbf{0}_2, \Sigma_J], \tag{A.4}$$

where $\boldsymbol{B} = \left(\frac{\alpha_1\omega_1}{\omega_1^2 + a^2\omega_2^2}, \ \frac{a\alpha_2\omega_2}{\omega_1^2 + a^2\omega_2^2}\right)'$.

A.2 . Proof of Theorem 2.2

Let $\boldsymbol{c} = (1,a)'$, we are trying to find the distribution of $U = X_1 + aX_2 + b = \boldsymbol{c}'\boldsymbol{X} + b$. First we derive the moment generating function (mgf) of U:

$$\begin{aligned} M_U(t) &= E[\exp(t\boldsymbol{c}'\boldsymbol{X} + tb)] \\ &= 2\exp\left\{t(\boldsymbol{c}'\boldsymbol{\xi} + b) + \frac{1}{2}t^2\boldsymbol{c}'\Sigma\boldsymbol{c}\right\}\Phi\left(\boldsymbol{\delta}'\Sigma^{1/2}\boldsymbol{c}t\right). \end{aligned} \tag{A.5}$$

It is easy to obtain that $\boldsymbol{c}'\boldsymbol{\xi} + b = \xi_1 + a\xi_2 + b$ and $\sigma^2 = \boldsymbol{c}'\Sigma\boldsymbol{c} = \omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2$. From Wang *et al.*,

$$\Sigma^{1/2} = \begin{bmatrix} \dfrac{\omega_1^2 + \sqrt{1-\rho^2}\,\omega_1\omega_2}{k} & \dfrac{\rho\omega_1\omega_2}{k} \\ \dfrac{\rho\omega_1\omega_2}{k} & \dfrac{\omega_2^2 + \sqrt{1-\rho^2}\,\omega_1\omega_2}{k} \end{bmatrix},$$

where

$$k = \sqrt{\omega_1^2 + \omega_2^2 + 2\sqrt{1-\rho^2}\omega_1\omega_2}.$$

Thus, $\delta_* = \boldsymbol{\delta}'\Sigma^{1/2}\boldsymbol{c}(\boldsymbol{c}'\Sigma\boldsymbol{c})^{-1/2}$. By equation (2.1), we obtain that $U \sim SN(\xi_1 + a\xi_2 + b, \sigma^2, \alpha_*)$ with

$$\begin{aligned} \alpha_* &= \frac{\delta_*}{\sqrt{1-\delta_*^2}} \\ &= \frac{\alpha_1 d_1 + \alpha_2 d_2}{\sqrt{(1-\alpha_1^2-\alpha_2^2)k^2\sigma^2 - (\alpha_1 d_1 + \alpha_2 d_2)^2}}, \end{aligned} \tag{A.6}$$

where $d_1 = \omega_1^2 + \omega_1\omega_2(\sqrt{1-\rho^2} + a\rho)$ and $d_2 = a\omega_2^2 + \omega_1\omega_2(a\sqrt{1-\rho^2} + \rho)$. Then the density of U is given by

$$f_U(u) = 2\phi(u; \boldsymbol{c}'\boldsymbol{\xi} + b, \sigma^2)\Phi\left[\alpha_* \frac{(u - \boldsymbol{b}'\boldsymbol{\xi} - b)}{\sigma}\right] \quad \text{(A.7)}$$

$$= 2\phi\left(u; \xi_1 + a\xi_2 + b, \omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2\right)$$

$$\times \quad \Phi\left[\alpha_* \frac{(u - \xi_1 - a\xi_2 - b)}{\left(\omega_1^2 + a^2\omega_2^2 + 2a\rho\omega_1\omega_2\right)^{1/2}}\right]. \quad \text{(A.8)}$$

A.3 . Proof of Theorem 3.1

The joint pdf of $X \sim \Delta(\delta_1, \mu_1, \sigma_1^2)$ and $Y \sim \Delta(\delta_2, \mu_2, \sigma_2^2)$ is

$$g(x,y) = \begin{cases} \delta_1\delta_2 & \text{if} x = 0, y = 0 \\ (1-\delta_1)(1-\delta_2)f_X(x)f_Y(y) & \text{if} x > 0, y > 0 \\ \delta_1(1-\delta_2)f_Y(y) & \text{if} x = 0, y > 0 \\ \delta_2(1-\delta_1)f_X(x) & \text{if} x > 0, y = 0 \end{cases} \quad \text{(A.9)}$$

where $f_X(x)$, $f_Y(y)$ are the probability distribution functions of X when $x > 0$ and Y when $y > 0$, respectively.

(1) For $x > 0, y > 0$, let $Z = X + aY + b$, $Y = Y$ so that $X = Z - aY - b$, $Y = Y$. It is easy to see that the Jacobian J of this transformation is 1. Thus, the pdf of Z as

$$f_Z(z) = \int f_X(z - ay - b)f_Y(y)dy,$$

(2) For $x = 0, y > 0$, we have $x = 0$, $Z = aY + b$ and the pdf of Z is

$$f_Z(z) = \delta_1(1-\delta_2)f_Y\left(\frac{z-b}{a}\right)\frac{1}{|a|}.$$

(3) Similarly, for $x > 0, y = 0$, the pdf of Z is

$$f_Z(z) = \delta_2(1-\delta_1)f_X(z-b).$$

Note that the sign of a determines the range of Z. If $a < 0$, then the pdf of $Z = X + aY + b$ is

$$f_Z(z) = \begin{cases} \displaystyle\int_{\frac{z-b}{a}}^{\infty} f_X(z-ay-b)f_Y(y)dy - \frac{1}{a}f_Y\left(\frac{z-b}{a}\right) & \text{if} z < b \\ \delta_1\delta_2 & \text{if} z = b \\ \displaystyle\int_0^{\infty} f_X(z-ay-b)f_Y(y)dy + f_X(z-b) & \text{if} z > b. \end{cases} \quad \text{(A.10)}$$

Also if $a > 0$, the pdf of Z is given by

$$f_Z(z) = \begin{cases} \delta_1\delta_2 & \text{if} z = b \\ \displaystyle\int_0^{\frac{z-b}{a}} f_X(z-ay-b)f_Y(y)dy + \frac{1}{a}f_Y\left(\frac{z-b}{a}\right) + f_X(z-b) & \text{if} z > b \end{cases} \quad \text{(A.11)}$$

so that desired result follows.

A.4 . Proof of Theorem 3.2

Note that from Equation (3.4), we know that the joint pdf of $(X_1, X_2)' \sim \Delta_2(\boldsymbol{\delta}, \boldsymbol{\mu}, \Sigma)$ is given by

$$g_{\mathbf{X}}(x_1, x_2) = \begin{cases} \delta_0 & \text{if} x_1 = 0, x_2 = 0 \\ \delta_1 f(x_1) & \text{if} x_1 > 0, x_2 = 0 \\ \delta_2 g(x_2) & \text{if} x_1 = 0, x_2 > 0 \\ (1-\delta_0-\delta_1-\delta_2)h(x_1,x_2) & \text{if} x_1 > 0, x_2 > 0, \end{cases} \tag{A.12}$$

where $f(x_1)$, $g(x_2)$ are pdf's of log-normal distributions $LN(\mu_1, \sigma_1^2)$ and $LN(\mu_2, \sigma_2^2)$, respectively, and $h(x_1, x_2)$ is the joint pdf of $(X_1, X_2)' \sim LN_2(\boldsymbol{\mu}, \Sigma)$.

(1) For $x_1 > 0, x_2 > 0$, let $U = X_1 + aX_2 + b, X_2 = X_2$, then $X_1 = U - aX_2 - b, X_2 = X_2$. It is easy to see the Jacobian J of this transformation is 1. Thus, we obtain the marginal pdf of U as

$$f_U(u) = \int (1-\delta_0-\delta_1-\delta_2)h(u-ax_2-b, x_2)dx_2,$$

(2) For $x_1 = 0, x_2 > 0$, let $x_1 = 0, U = aX_2 + b$ and the Jacobian of this transformation is $1/a$ so that the pdf of U is

$$f_U(u) = \delta_2 g\left(\frac{u-b}{a}\right)\frac{1}{|a|}.$$

(3) Similarly for $x_1 > 0, x_2 = 0$, the pdf of U is

$$f_U(u) = \delta_1 f(u-b).$$

Therefore the pdf of $U = X_1 + aX_2 + b$ is, for $a < 0$,

$$f_U(u) = \begin{cases} \displaystyle\int_{\frac{u-b}{a}}^{\infty} (1-\delta_0-\delta_1-\delta_2)h(u-ax_2-b,x_2)dx_2 - \frac{\delta_2}{a} g\left(\frac{u-b}{a}\right) & \text{if} u < b \\ \delta_0 & \text{if} u = b \\ \displaystyle\int_0^{\infty} (1-\delta_0-\delta_1-\delta_2)h(u-ax_2-b,x_2)dx_2 + \delta_1 f(u-b) & \text{if} u > b \end{cases} \tag{A.13}$$

and for $a > 0$, the pdf of U is

$$f_U(u) = \begin{cases} \delta_0 & \text{if} u = b \\ \displaystyle\int_0^{\frac{u-b}{a}} (1-\delta_0-\delta_1-\delta_2)h(u-ax_2-b,x_2)dx_2 + \frac{\delta_2}{a} g\left(\frac{u-b}{a}\right) + \delta_1 f(u-b) & \text{if} u > b. \end{cases} \tag{A.14}$$

Thus the proof of Theorem 3.2 is completed.

Corresponding author
David Trafimow can be contacted at: dtrafimo@nmsu.edu

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**