Intelligence 64 (2017) 67–70

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Predicting group differences from the correlation of vectors

Kimmo Sorjonen*, Jon Aurell, Bo Melin

Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, 171 65 Solna, Sweden

ARTICLE INFO

Keywords:
Factor loadings
Intelligence
Jensen effect
Sigmoid association
Simulation

ABSTRACT

It has been proposed that a correlation between a vector of factor loadings of intelligence tests and a vector of group differences on these same tests (= correlation of vectors) indicate that the group difference is mainly in *g*. In the present simulation, we show that there is an inverse sigmoid association between the difference between population means on a latent variable and the correlation of vectors and that the appearance and precision of this association is moderated by sample size and the standard deviation of factor loadings. In high powered studies, a weak correlation of vectors would falsify, while a strong correlation would not be able to verify, a hypothesis about a sizeable difference between population means.

1. Introduction

According to Jensen (1985, 1998), a correlation between a vector of factor loadings of intelligence tests and a vector of group differences on these same tests (Table 1) indicates that the group difference is mainly in *g* rather than in more specific abilities and that the cause of this difference is the same as the cause of variation within populations. Such a correlation between vectors has been found in various comparisons (see te Nijenhuis et al., 2016, for a review).

It has been argued that some of the findings taken to support the claim that group differences are due to the same cause as variation within populations could be due to violation of the assumption of measurement invariance across groups, something that can lead to a positive correlation between factor loadings and the degree of group differences on the tests even when there is no difference in *g* between the groups (Dolan, Roorda, & Wicherts, 2004). The use of Multigroup Confirmatory Factor Analysis (MGCFA), rather than Jensen's method of correlated vectors, has been recommended with the argument that if variance between groups is due to the same factor as variance within groups, e.g. *g*, factorial invariance across groups should be possible to demonstrate (e.g. Dolan, 2000; Lubke, Dolan, & Kelderman, 2001; Lubke, Dolan, Kelderman, & Mellenbergh, 2003). Re-analyzes using MGCFA have sometimes failed to demonstrate measurement invariance (Dolan et al., 2004). On the other hand, Ashton and Lee (2005) demonstrate that the method of correlated vectors also can fail to reveal a correlation between the vector of *g*-loadings and the vector of correlations between subtests and a variable V even when V has a strong correlation with *g*. Ashton and Lee also show that *g*-loadings depend on the included subtests and this can affect the results obtained through the method of correlated vectors.

Schönemann (1989, 1997) performed simulations and argued that a correlation between the vectors of tests' *g*-loadings and the size of differences between populations is a tautological consequence that will arise if (1) these tests are positively correlated, and (2) people in one of the populations tend to score higher on the tests than people in the other population. However, Schönemann's simulations have, in their turn, been criticized to produce trivial results due to using the scores on the first factor as a selection variable, and it has been argued that a correlation between *g*-loadings and the size of the group differences is *not* a mathematical necessity (Dolan, 1997; Dolan & Lubke, 2001).

The objective of the present simulation was to look for a function that can be used to predict the difference between population means on a latent variable from the correlation between the vectors of factor loadings and group differences on items.

2. Method

Using R 3.2.2 statistical software (R Core Team, 2015), a dataset was simulated through the following steps (code and dataset available as supplementary material): (1) Two samples with between 50 and 12,800 (= eight doublings) persons each were created (the same number in each sample); (2) for both samples, values were randomly drawn from a normally distributed variable *T* with *SD* = 1 and with a defined difference between population means varying between − 1.5 and + 1.5; (3) fifteen normally distributed items with varying correlations with the variable *T* were created. The range of the correlations was decreased in steps from 0.1–0.9 to 0.475–0.525 (evenly spaced) in order to vary the standard deviation of the factor loadings. The average

* Corresponding author.
E-mail address: kimmo.sorjonen@ki.se (K. Sorjonen).

http://dx.doi.org/10.1016/j.intell.2017.07.008
Received 13 February 2017; Received in revised form 12 July 2017; Accepted 26 July 2017
Available online 02 August 2017
0160-2896/

Table 1
An example where the correlation between the vector of factor loadings and the vector of group differences equals 0.79 and the standard deviation of loadings equals 0.22.

Test	Loading	Group 1[a]	Group 2[a]	Diff[b]
1	0.2	85 (19)	88 (19)	− 0.16
2	0.3	90 (18)	81 (21)	0.46
3	0.4	95 (12)	94 (17)	0.07
4	0.5	110 (13)	99 (12)	0.88
5	0.6	115 (14)	107 (14)	0.57
6	0.7	120 (15)	106 (12)	1.04
7	0.8	125 (19)	111 (17)	0.78

[a] Group mean (*SD*).
[b] Difference between group means in pooled standard deviations.

correlation was thus kept at approximately 0.5 in all of the simulations; (4) using the psych package (Revelle, 2015) the loadings of these fifteen items on one factor were calculated. As a group difference on the latent factor may inflate factor loadings when using the pooled group (Jensen, 1998), the calculation was conducted in one of the two samples; (5) the correlation of the vectors of factor loadings and group differences on items and the standard deviation of factor loadings were calculated and saved, together with the defined sample size and difference between population means, in a data frame. These five steps were run a thousand times for nine sample sizes, 31 defined differences between population means on the variable *T*, and for 16 different ranges of factor loadings, resulting in 4,464,000 simulated datasets.

3. Results

The difference between population means was found to be an inverted sigmoid function of the correlation of vectors (Fig. 1). A sigmoid curve is given by the following formula:

$$y = \frac{d}{1 + e^{-s \cdot x}} - c$$

In this formula, d = the distance between the floor and the ceiling of the curve; s = the steepness of the curve; c = the ceiling of the curve. In the present case, the function is symmetrical on both sides of zero, giving that $d = 2c$, and the function can be simplified. After inversion we get that the difference between population means (D_P) can be predicted by:

$$D_P = -\frac{1}{s} \cdot ln\left[\frac{2}{\frac{R_V}{c} + 1} - 1\right]$$

In this formula, R_V stands for the correlation between vectors. An analysis found the values of the s and c parameters in this function to be moderated by the standard deviation of the factor loadings (SD_L) and the sample size (N, in each sample) according to the following:

$$s = 0.883 + 0.000249 \cdot N + 23.8 \cdot SD_L + 0.00853 \cdot N \cdot SD_L$$

$$c = 2.38 - 0.158 \cdot \ln(N) + 1.22 \cdot \ln(SD_L) - 0.142 \cdot \ln(N) \cdot \ln(SD_L)$$

It is also apparent in Fig. 1 that the standard error of the prediction is affected by the standard deviation of the factor loadings, sample size, and the correlation of vectors. In order to avoid negative predictions, the standard error was logarithmized. This $\ln(SE(D_P))$ was found to be a curvilinear function of the correlation of vectors (R_V) and after exponentiation:

$$SE(D_P) = e^{b_0 + b_1 \cdot R_V + b_2 \cdot R_V^2}$$

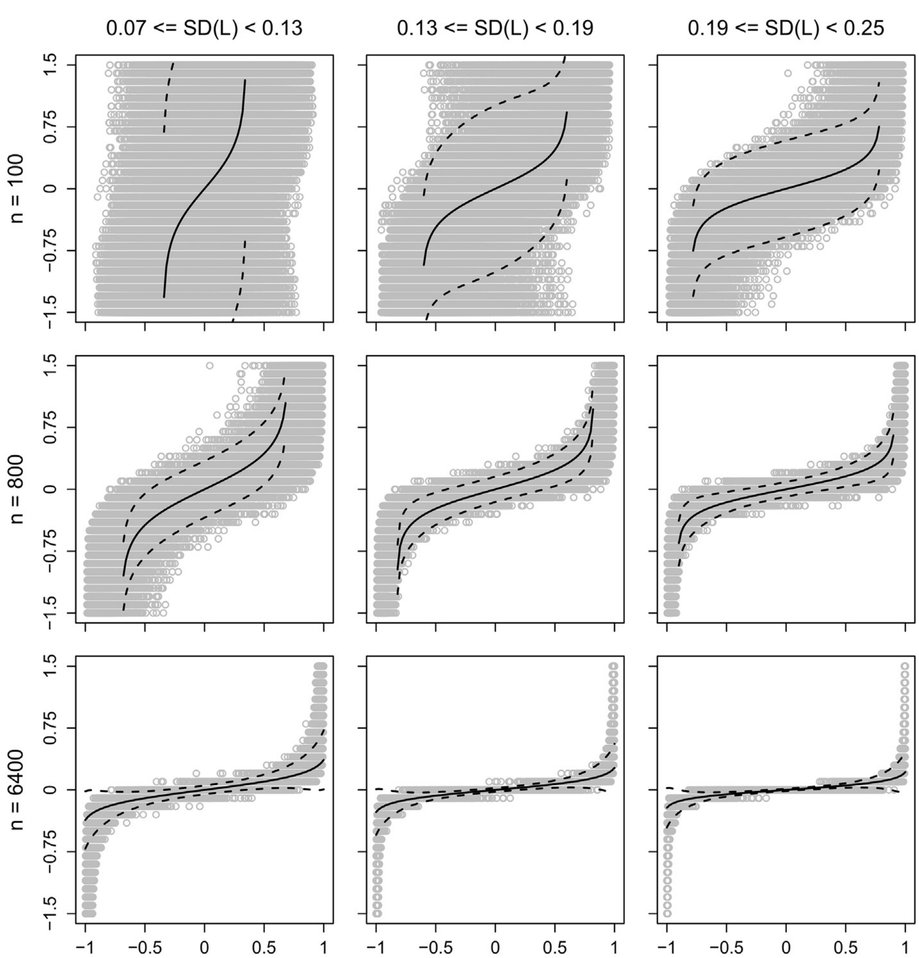

Fig. 1. The association between the difference between population means on a latent variable (y-axis) and the correlation between the vectors of factor loadings and group differences on items (x-axis) for the combinations of three different sample sizes (n in each sample) and three different ranges of standard deviations of factor loadings (*SD*(*L*)). The predicted differences between population means (solid lines) with 95% CI (dotted lines) have been calculated using the formulas presented in the text.

Table 2
Predicted differences between population means (with 95% CI) for the studies presented in Jensen's (1985) Table 3.

Study	N_1	N_2	R_V [a]	Pred. diff. (95% CI)
Jensen & Reynolds (1982)	1868	305	0.73	1.050 (0.681; 1.419)
Reynolds & Gutkin (1981)	285	285	0.51	0.510 (− 0.039; 1.059)
Sandoval (1982)	332	314	0.36	0.283 (− 0.217; 0.783)
Mercer (1984)	668	619	0.66	0.590 (0.250; 0.930)
National Longitudinal Study	12,275	1938	0.78	0.268 (0.091; 0.446)
Nichols (1972)	1940	1460	0.75	0.424 (0.194; 0.653)
Dept. of Defence (1982)	5533	2298	0.39	0.104 (0.023; 0.185)
Dept. of Labor (1970)	4001	2416	0.71	0.243 (0.087; 0.399)
Kaufman & Kaufman (1983)	813	486	0.56	0.419 (0.089; 0.749)
Veroff et al. (1971)	179	186	0.36	0.348 (− 0.444; 1.140)
Hennessy & Merrifield (1976)	1818	431	0.66	0.558 (0.233; 0.882)

Note: Standard deviation of factor loadings assumed to be 0.13 except for Jensen & Reynolds (1982) where it, for computational reasons, is assumed to be 0.14.
[a] Correlation of vectors.

The coefficients in their turn are given by:

$$b_0 = 0.346 - 0.904 \cdot \ln(N) - 1.710 \cdot \ln(SD_L)$$

$$b_1 = -0.0464 + 0.0116 \cdot \ln(N) + 0.0180 \cdot \ln(SD_L)$$

$$b_2 = -1.587 + 0.726 \cdot \ln(N) + 1.247 \cdot \ln(SD_L)$$

The 95% CI for the predicted difference between population means has been calculated by adding and subtracting two standard errors to the predicted difference (Fig. 1).

4. Discussion

The present simulation found an inverse sigmoid association between the difference between population means on a latent variable and the correlation between the vectors of factor loadings and group differences on items. The appearance of this association, as well as standard errors of the prediction, was moderated by sample size and the standard deviation of factor loadings. As an example (Table 2), we have used our function (available in the supplementary script) to predict differences between population means in *g*, or whatever the used tests are indicators of, for eleven correlations of vectors presented in Jensen (1985). In these calculations, we have, with one exception, used 0.13 as the standard deviation of factor loadings, as Jensen (1998) noted this to be the case for 149 tests. It is apparent in Table 2 that, due to large predicted standard errors, the predicted difference between population means is small and non-significant for the studies with small sample sizes. It is interesting to note that although the sample size is smaller and the correlation of vectors weaker in the study by Jensen and Reynolds compared to the National Longitudinal Study and the study by Nichols, the predicted difference between population means is bigger. The reason for this is that with a high powered study, with a large sample size and standard deviation of factor loadings, and a nontrivial difference between population means, we expect a very strong correlation of vectors, and lack thereof indicates a small difference between population means. For a given correlation of vectors ($\neq 0$), the predicted difference between population means will be smaller but the confidence interval narrower for a study with higher compared to a study with lower power. For studies with very high power (see bottom row in Fig. 1), we expect a very strong correlation of vectors already with a modest difference between population means, and the predicted difference will never be large. If using the correlation of vectors as a method to estimate the difference between population means on a latent variable, such high powered studies will, maybe to Popper's (1959) appreciation, be good at falsifying a hypothesis of a sizable difference while lacking the capacity to verify it. It has been concluded before that a repeated demonstration of a strong correlation of vectors is necessary, but not sufficient, to infer a difference between population means on a latent variable and that the method lacks in specificity (Dolan & Hamaker, 2001, see also Wicherts, 2017). With less power, a strong correlation of vectors would predict a larger, but also a more uncertain, difference between population means and it would, again, be difficult to assume a large difference between population means. In low powered studies, the range of predicted possible correlations between the difference between population means and the correlation of vectors ($-c$ to c in the formulas above) is also more restricted. For example, given a standard deviation of factor loadings of 0.1, the range of possible correlations is predicted to be between − 0.35 and 0.35, between − 0.70 and 0.70, and between − 1 and 1 for sample sizes 100, 800, and 6400, respectively (leftmost column in Fig. 1).

4.1. Limitations

In this paper, we present approximate formulas for predicting the difference between population means on a latent variable from the correlation between the vectors of factor loadings and group differences on items. There might be exact irrational numbers lurking behind the scenes, but they have failed to reveal themselves for us. We cannot be completely sure that our formulas do not behave erratically in some situations and problematization of using the correlation of vectors as a predictor of group differences on latent variables, rather than presenting formulas that can be used for accurate predictions, should be regarded as the main point of the present paper.

It should be noted that the present formulas have been calculated with data that fulfills the assumption of measurement invariance across groups, and probably cannot be trusted to give accurate predictions if this assumption is violated. The assumption of measurement invariance is central to group comparisons and testable (see e.g. Dolan et al., 2004; Lubke et al., 2003 for a description of how). With the lavaan R-package (Rosseel, 2012), for example, it is possible to conduct multigroup factor analyses while constricting factor loadings, residuals, and intercepts to be the same across groups. If these constrictions do not result in a reduction of model fit, compared with an unconstricted model, it indicates measurement invariance and that within- and between-group differences are due to the same factor (Lubke et al., 2003). However, there are findings indicating that IQ test batteries often fail to be measurement invariant across groups based on ethnicity, gender, educational background, cohort, or age, which means that these tests cannot always be trusted to give a valid and unbiased measurement of group differences in IQ (Wicherts, 2016).

Our simulations, and therefore also the applicability of the formulas, is based on the idealized scenario in which only *g* is responsible for the group differences on items. This means that if *g* would be regressed out of item scores, the remaining group difference would be zero on all of these scores. This would mean complete fulfillment of the "equal intercept" part of the assumption of measurement invariance, and is probably nothing that ever apply to genuine data.

In the present simulation, we have varied sample size and the standard deviation of factor loadings. However, it is possible that other factors, such as number of items and the mean of factor loadings, also moderate the association between the difference between population means and the correlation of vectors. It should also be stressed that our simulations are based on normally distributed continuous item scores and the presented formulas cannot be used with dichotomously scored items. Wicherts (2017) demonstrated complexly non-linear associations between vectors of item-total correlations (often considered to be the item's *g* loading for dichotomous items) and phi coefficients (measure of group difference on dichotomous items).

References

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, *33*, 431–444.
Dolan, C. V. (1997). A note on Schönemann's refutation of Spearman's hypothesis.

Multivariate Behavioral Research, *32*, 319–325.
Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*, 21–50.
Dolan, C. V., & Hamaker, E. L. (2001). Investigating black–white differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC, and a critique of the method of correlated vectors. *Advances in Psychology Research*, *6*, 31–60.
Dolan, C. V., & Lubke, G. H. (2001). Viewing Spearman's hypothesis from the perspective of multigroup PCA: A comment on Schönemann's criticism. *Intelligence*, *29*, 231–245.
Dolan, C. V., Roorda, W., & Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: The GATB in Holland and the JAT in South Africa. *Intelligence*, *32*, 155–173.
Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences*, *8*, 193–263.
Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences on cognitive tests using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research*, *36*, 299–324.
Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*, 543–566.
te Nijenhuis, J., Bakhiet, S. F., van den Hoek, M., Repko, J., Allik, J., Žebec, M. S., ... Abduljabbar, A. S. (2016). Spearman's hypothesis tested comparing Sudanese children and adolescents with various other groups of children and adolescents on the items of the Standard Progressive Matrices. *Intelligence*, *56*, 46–57.
Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
R Core Team (2015). R: A language and environment for statistical computing. *R foundation for statistical computing*. Retrieved from https://www.R-project.org/ (Vienna, Austria) .
Revelle, W. (2015). *Psych: Procedures for personality and psychological research*. Evanston, Illinois, USA: Northwestern University. Retrieved from http://CRAN.R-project.org/package=psychVersion=1.5.8.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/.
Schönemann, P. H. (1989). Some new results on the Spearman hypothesis artifact. *Bulletin of the Psychonomic Society*, *27*, 462–464.
Schönemann, P. H. (1997). Famous artefacts: Spearman's hypothesis. *Cahiers de Psychologie Cognitive*, *16*, 665–694.
Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, *30*, 1006–1016.
Wicherts, J. M. (2017). Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence*, *60*, 26–38.