

Clinical Trials: Data modeling when there is a large number of Variables

Ton J Cleophas, MD, PhD
European College Pharmaceutical Medicine,
Lyon France

Abstract

Background:

Traditional statistical tests are unable to handle a large number of variables. The simplest method to reduce large numbers of variables is the use of add-up scores. But add-up scores do not account for the relative importance of the separate variables, their interactions and differences in units. Principal components analysis and partial least square analysis account all of that, but are virtually unused in clinical trials.

Objective:

To assess the performance of either of the two methods.

Methods:

A simulated example of 250 patients' gene expression data as predictor and drug efficacy scores as outcome will be used. For principal components analysis SPSS' s Data Dimension Reduction module was used, for partial least square analysis R Partial Least Squares, a free statistics and forecasting software was used.

Results:

Of 27 variables three novel predictor variables were constructed. With principal components analysis the 3 were very significant predictors of the add-up outcome score with t-values of 10.2, 21.6, and 6.7 ($p < 0.000$, $p < 0.000$, $p < 0.000$). Partial least squares included the outcome variables in its program, and was also able to predict the outcome variables although at a lower level of significance with t-values of 6.8, 16.2, and 3.5 ($p < 0.000$, $p < 0.000$, $p < 0.001$). Traditional multiple linear regression with the novel predictors in the form of add-up scores as independent variables produced a consistent further reduction of significance with t-values of 3.4, 11.2 and 2.4 ($p < 0.002$, $p < 0.001$, $p < 0.02$).

Conclusions:

1. Principal components analysis and partial least squares can handle many more variables than the standard covariance methods like MANOVA and MANCOVA can, and is more sensitive than add-up scores.
2. The methods account the relative importance of the separate variables, their interactions and differences in units.
3. They are also very flexible, to the extent that manifest variables can be applied twice, first in the form of clusters for prediction and second unclusteredly as manifest outcome variables.
4. Partial least squares method is parsimonious to principal components analysis, because it can include outcome variables in the model.

Keywords: data modeling, partial least squares, principal components analysis

INTRODUCTION

Current clinical trials tend to include large numbers of variables. For example, a series of gene expressions can be used to predict the efficacy of cytostatic treatment^{1,2}, repeated measurements can be used as endpoint in randomized longitudinal trials^{3,4}, and multi-item personal scores can be used for the evaluation of antidepressants.⁵ Many more examples can be given. Some kind of mathematical modeling of the multiple variables is required for useful information. Often multiple linear models like MANOVA (multivariate analysis of variance) and MANCOVA (multivariate analysis of covariance) are successful and can handle thousands of cases. However, the models have great difficulty to model more than two or three dependent variables. The simplest method to reduce large numbers of variables is the use of add-up scores. But add-up scores do not account for the relative importance of the separate variables, their interactions and differences in units. All of this is accounted for by a technique, developed in the early fifties by the psychometrician Eastman (London UK), called principal components analysis: two or three unmeasured factors, otherwise called components or latent variables (LVs), are identified to explain a much larger number of measured variables, otherwise called

manifest variables (MVs).^{6,7} Partial least squares analysis is an extension of principal components analysis, first described by the econometrist Wold, Stockholm 1966, and is appropriate for estimating more than a single layer of latent variables.⁸

Although both methods have become major research tools in behavioral sciences, social sciences, marketing, operational research, and other applied sciences⁹⁻¹¹, they are rarely applied in clinical trials. When searching the internet we found, except for a few genetic trials¹²⁻¹⁵, virtually no clinical studies. This is a pity given the presence of large numbers of variables in this field of science.

In the current paper we will assess the performance of either of the two methods. A simulated example of 250 patients' gene expression data and drug efficacy scores will be used.

Some theory

Examples of principal components and partial least square models are in Table 1. Both are based on multivariate linear regression. The arrows are the fitted linear correlation coefficients. Compared to the principal components models, the partial least square models are mathematically more complex, because they calculate, in addition to the best fit predictor LVs, the best fit outcome

LVs, and make for that purpose use of both predictor and outcome MVs. The latter models, because they take more into account, will produce smaller test statistics, but are, at the same time, less at risk of biases, e.g., due to differences in importance of manifest outcome variables, their interactions and differences in units.

Principal components analysis

Principal components analysis (Table 1) uses all manifest variables available to linearly model the best fit correlation coefficients between the MVs and the LVs along an x and y-axis. The magnitude of a LV is calculated as the weighted mean of all of the MVs along one axis. Figure 1 gives a simple example with renal and liver function as latent variables. It can be observed that ureum, creatinine and creatinine clearance have a very strong positive correlation with renal function and the other three with liver function. It can be demonstrated that by slightly rotating both x and y-axes the model can be fitted even better. If a third latent variable existed within a data file, it could be represented by a third axis, a z-axis creating a 3-d graph. Also additional factors can be added to the model, but they cannot be presented in a 2- or 3-d drawing, but, just like with multiple regression modeling, the software programs have no problem with multidimensional calculations similar to the above 2-d calculations.

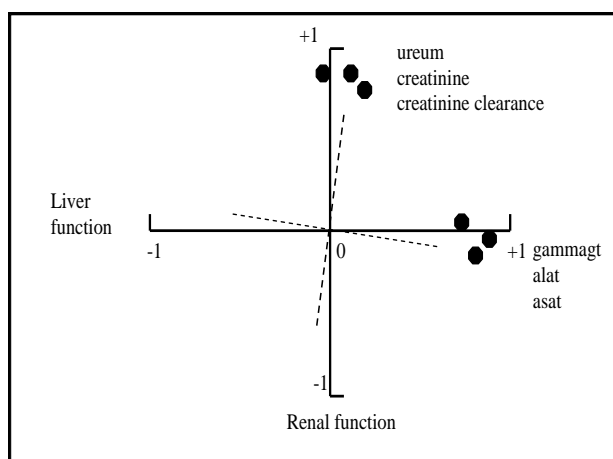


Figure 1. The relationships of 6 manifest variables with two latent variables, the liver and renal function as linearly modeled by principal components analysis. It can be observed that ureum, creatinine and creatinine clearance have a very strong positive correlation with renal function and the other three with liver function. It can be demonstrated that by slightly rotating both x and y-axes the model can be fitted even better.

Partial least squares analysis

Partial least squares analysis (Table 1) does not simultaneously use all of the predictor variables available, but rather uses an a priori clustered set of predictor variables of 4 or 5 to calculate a new latent variable. Unlike principal components analysis which does not consider response variables at all, partial least square analysis does take response variables into account and therefore often leads to a better fit of the response variable. Correlation coefficients are produced from multivariate linear regression rather than fitted correlation coefficients along the x and y-axes.

Example

A 250 patients' data-file was supposed to include 27 variables consistent of both patients' microarray gene expression levels and their drug efficacy scores. All variables were standardized by scoring them on an 11 points linear scale (0-10). The following clusters of genes were highly correlated with one another: the variables 1-5, the variables 16-19, and the variables 24-27. The variables 20-23 were supposed to represent drug efficacy scores and were clustered as the outcome variables. The datafile is given in the appendix.

Principal components analysis

For principal components analysis SPSS' s Data Dimension Reduction module⁷ was used. First the reliability of the model was assessed by testing the test-retest reliability of the original variables. The test-retest reliability of the original variables should be assessed with Cronbach's alphas using the correlation coefficients after deletion of one variable: all of the data files should produce at least by 80% the same result as that of the non-deleted data file (alphas > 80%).

Command:

Analyze....Scale....Reliability Analysis....transfer original variables to Variables box....click Statistics....mark Scale if item deleted....mark Correlations....Continue....OK.

The Table 1 shows, that, indeed, none of the original variables after deletion reduces the test-retest reliability. The data are reliable. We will now perform the principal components analysis with three components, otherwise called latent variables.

Table 1. Data dimension reduction with principal components and partial least square models. Both are based on multivariate linear regression . The arrows are the fit linear correlation coefficients as calculated by the software. Compared to the principal components models, the partial least square models are mathematically more complex, because they calculate, in addition to the best fit predictor LVs, the best fit outcome LVs, and make for that purpose use of both predictor and outcome MVs. The latter models, because they take more into account, will produce smaller test statistics, but are at the same time less at risk of biases, e.g., due to differences in importance of manifest outcome variables, their interactions and differences in units.

| <i>Principal components analysis</i> | | <i>Partial least squares analysis</i> | | | |
|--------------------------------------|---------------|---------------------------------------|---------------|-------------|-------------|
| Predictor MVs | predictor LVs | Predictor MVs | predictor LVs | outcome LVs | outcome MVs |
| 1→ | 1 | 1→ | 1 | | |
| 2→ | | 2→ | | | |
| 3→ | | 3→ | | | |
| 4→ | | 4→ | | → 3 | ← 9 |
| | | | | | ← 10 |
| 5→ | 2 | 5→ | 2 | | ← 11 |
| 6→ | | 6→ | | | |
| 7→ | | 7→ | | | |
| 8→ | | 8→ | | | |

Table 2. Validation of the data: there should be a strong correlation between the scores within the clusters (strong collinearity). The test-retest reliability of the variables is assessed with Cronbach's alphas. All of the reliability assessments should produce at least for 80% the same result (Cronbach's alphas > 80%).

| Item-Total Statistics | | | | | |
|-----------------------|----------------------------|--------------------------------|----------------------------------|------------------------------|----------------------------------|
| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
| VAR00001 | 79,6200 | 277,273 | ,547 | ,486 | ,903 |
| VAR00002 | 79,6200 | 263,980 | ,724 | ,700 | ,896 |
| VAR00003 | 79,5120 | 264,749 | ,743 | ,671 | ,895 |
| VAR00004 | 79,5480 | 284,361 | ,477 | ,385 | ,906 |
| VAR00005 | 81,0720 | 264,501 | ,566 | ,386 | ,903 |
| VAR00016 | 80,3720 | 257,166 | ,714 | ,623 | ,895 |
| VAR00017 | 79,7320 | 268,494 | ,665 | ,582 | ,898 |
| VAR00018 | 80,3080 | 265,869 | ,588 | ,477 | ,902 |
| VAR00019 | 80,9560 | 255,038 | ,719 | ,555 | ,895 |
| VAR00024 | 80,2800 | 245,696 | ,719 | ,611 | ,895 |
| VAR00025 | 80,0200 | 272,702 | ,507 | ,340 | ,905 |
| VAR00026 | 80,6240 | 244,581 | ,714 | ,627 | ,896 |

Command:

Analyze....Dimension Reduction....Factor....enter variables into Variables box....click Extraction....Method: click Principle Components....mark Correlation Matrix, Unrotated factor solution....Fixed number of factors: enter 3....Maximal Iterations plot Convergence: enter 25....Continue....click Rotation....Method: click Varimax....mark Rotated solution....mark Loading Plots....Maximal Iterations: enter 25....Continue....click Scores.... mark Display factor score coefficient matrixOK.

The table 3 shows the best fit coefficients of the original variables constituting 3 components, The component 1 has a very strong correlation with the variables 16-19, the component 2 with the variables 24-27, and the component 3 with the variables 1-4.

These 3 components can, thus, be interpreted as the latent predictor variables. When minimizing the outcome file and returning to the data file, we now observe, that, for each patient, the software program has produced the individual values of these novel predictors.

Table 4. A multiple linear regression was performed with the add-up scores of the outcome variables instead of MANOVA. SPSS statistical software did not execute the command for MANOVA, and explained that too many columns and too many levels were in the data for the purpose of a proper analysis.

| | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|-------|------------------------------------|-----------------------------|------------|---------------------------|---------|------|-------------------------|-------|
| Model | | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 | (Constant) | 27,364 | ,231 | | 118,420 | ,000 | | |
| | REGR factor score 1 for analysis 1 | 4,991 | ,232 | ,735 | 21,558 | ,000 | 1,000 | 1,000 |
| | REGR factor score 2 for analysis 1 | 2,358 | ,232 | ,347 | 10,185 | ,000 | 1,000 | 1,000 |
| | REGR factor score 3 for analysis 1 | 1,552 | ,232 | ,229 | 6,701 | ,000 | 1,000 | 1,000 |

a. Dependent Variable: 20-23

In order to fit these novel predictors with the outcome variables, the drug efficacy scores (variables 20-23), multivariate analysis of variance (MANOVA) should be appropriate given the continuous nature of the 4 outcome variables. However, the large number of columns in the design matrix caused integer overflow, and the large number of columns caused too many levels within some components as well as numerical problems with higher order interactions among components, and the command was not executed. Instead we performed univariate multiple linear regression with the add-up scores of the outcome variables as novel outcome variable. Table 4 gives the results. All of the 3 latent predictors were very significant independent predictors of the add-up outcome variable.

Table 3. Principal components analysis. Component 3 has a strong positive correlation with the MVs 1-4, component 2 with MVs 24-27, and component 1 with MVs 16-19.

Rotated Component Matrix^a

| | Component | | |
|----------|-----------|------|------|
| | 1 | 2 | 3 |
| VAR00001 | ,249 | ,136 | ,797 |
| VAR00002 | ,582 | ,128 | ,652 |
| VAR00003 | ,616 | ,163 | ,586 |
| VAR00004 | ,003 | ,364 | ,770 |
| VAR00005 | ,711 | ,063 | ,211 |
| VAR00016 | ,819 | ,242 | ,127 |
| VAR00017 | ,500 | ,516 | ,217 |
| VAR00018 | ,379 | ,482 | ,306 |
| VAR00019 | ,719 | ,289 | ,235 |
| VAR00024 | ,585 | ,617 | ,079 |
| VAR00025 | ,160 | ,675 | ,228 |
| VAR00026 | ,634 | ,563 | ,050 |
| VAR00027 | ,084 | ,823 | ,172 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Partial least squares analysis

Because partial least is not available in the basic and regression modules of SPSS, we used the software program R Partial Least Squares, a free statistics and forecasting software available on the internet as a free online software calculator.¹⁶ The data-file was imported directly from a Word file. The selected clusters of variables were listed: latent variable 1 (16-19), latent variable 2 (24-27), latent variable 3 (1-4), and latent variable 4 (20-23).

A square boolean matrix was constructed with “0 or 1” if fitted correlation coefficients were to be included in the model “no or yes”. Then the order “compute” was given.

| <i>Latent Variable</i> | <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> |
|------------------------|----------|----------|----------|----------|
| <i>1</i> | 0 | 0 | 0 | 0 |
| <i>2</i> | 0 | 0 | 0 | 0 |
| <i>3</i> | 1 | 1 | 0 | 0 |
| <i>4</i> | 0 | 0 | 1 | 0 |

After 15 seconds of computing the program produced the following results. First, the data were validated using the GoF (goodness of fit) criteria. $GoF = \sqrt{[\text{mean of } r\text{-square values of comparisons in model} * r\text{-square overall model}]}$, where * is the sign of multiplication. A GoF value varies from 0 to 1 and values larger than 0.8 indicate that the data are adequately reliable for modeling. The following results were given:

| | <i>GoF value</i> |
|--|------------------|
| Overall | 0.9459 |
| Outer model (including manifest variables) | 0.9986 |
| Inner model (including latent variables) | 0.9466 |

The data were, thus, adequately reliable. The calculated best fit r-values (correlation coefficients) were estimated directly from the model, and their standard errors would be available from second derivatives. However, the problem with the second derivative procedure is that it requires very large data files in order to be accurate. Instead of an inaccurate estimate of the standard errors, distribution free standard errors were calculated using bootstrap resampling.

| <i>Latent variables</i> | <i>Original r-values</i> | <i>bootstrap r-values</i> | <i>standard error</i> |
|-------------------------|--------------------------|---------------------------|-----------------------|
| 1 versus 3 | 0.57654 | 0.57729 | 0.08466 |
| 2 versus 3 | 0.67322 | 0.6749 | 0.04152 |
| 4 versus 3 | 0.18322 | 0.18896 | 0.05373 |

The table 5 shows that all of the three correlation coefficients were very significant predictors, and that the three predictor latent variables were, thus, very significant predictors of the latent outcome variable.

Comparison of the two models

Table 5 shows the correlation coefficients of predictor latent variables in either of the two models used in this paper. The partial least square method produced somewhat smaller test statistics, but it is less biased because it takes into account the relative importance of

the separate manifest variables, and their interactions. In spite of accounting more, the level of statistical significance of the latter model remained excellent.

When using the add-up scores of the main variables of the 3 components instead of the modeled latent variables, the effects were similarly statistically significant. However, they were so at lower levels of significance (Table 5). Obviously, the principal components and partial least squares analyses provided better fit for the data than did multiple linear regression with add-up variables as both predictor and outcome variables.

Table 5. Comparison of correlation coefficients of predictor latent variables. The partial least square method produced somewhat smaller t-values, but it is less biased and the level of statistical significance is excellent even so. When using the add-up scores of the main variables of the 3 components instead of the modeled latent variables, the effects were similarly statistically significant, but, the magnitudes of the t-values further fell.

| Principal components | | Partial Least Squares | | Add-up scores | |
|---------------------------------|----------------|---------------------------------|----------------|--------------------------------|----------------|
| correlation coefficients | t-value | correlation coefficients | t-value | regression coefficients | t-value |
| 0.74 | 10.2 | 0.58 | 6.8 | 0.15 | 3.4 |
| 0.35 | 21.6 | 0.67 | 16.2 | 0.61 | 11.2 |
| 0.23 | 6.7 | 0.19 | 3.5 | 0.14 | 2.4 |

DISCUSSION

The data dimension reduction methods explained in this paper are wonderful for the analysis of data with many variables, because they can handle many more variables than the standard covariance methods like MANOVA and MANCOVA can. They also have the advantage, compared to models using the composite of multiple variables as endpoint or predictor, that they can account for the relative importance of the separate variables, their interactions and differences in units.

However, the data dimension reduction methods do have a numbers of limitations. They do not comply with all of the requirements of normal distributions in the data, adequate sample sizes to reduce type II errors, adjustments for multiple testing to reduce type I errors etc. Also, they are scientifically less rigorous than traditional methods, because empirical rather than parametric confidence intervals are applied and hypothesis testing is based on re-sampling methods such as jack-knife and bootstrap. However, a reduced scientific rigor is equally true for most traditional multiple variables and multivariate analyses, particularly if they are post hoc and not based on prior hypotheses. We should add that, particularly, with sound underlying clinical arguments, the novel methodologies are helpful for confirm hypotheses, increasing precision of some point estimates, benefit risk analyses, providing relevant arguments for clinical decision making, and other quantitative assessments.

A special advantage of the novel methodologies is their flexibilities, and capacities to handle numerous variables. Forty variables or even more is no problem.¹⁷ If you don't have outcome variables, the two layers of a two-layer partial least squares model can be constructed using the

manifest variables twice, first in the form of clusters for prediction and second unclusteredly as manifest outcome variables.¹⁷

The principal components analysis does not consider response variables, and partial least square analysis does take response variables into account and therefore often leads to a better fit of the response variable. Correlation coefficients are produced from multivariate linear regression rather than fitted correlation coefficients along the x and y-axes. The latter method may be parsimonious to the former, that is, if outcome variables are in your data.

At this time the novel models can not yet be applied for binary data like survival data, but this is a matter of time. Bastien (Aulnay, France) has already proposed a PLS-Cox model for the analysis of the effect of gene expression on survival.¹⁴ Also non linear models do not fit the novel methods. Outlier identification with linear models makes use of tests based on normal distribution and homoscedasticity assumptions like the Durbin Watson test and is not available with the novel methods. And, so, despite the pleasant properties of the novel methods, there is plenty room for improvement.

CONCLUSIONS

Advantages of the novel methods include

- 1/ they can handle many more variables than the standard covariance methods like MANOVA and MANCOVA can, and are more sensitive than add-up scores are,
- 2/ they account the relative importance of the separate variables, their interactions and differences in units,
- 3/ they are very flexible, to the extent that manifest variables can be applied twice, first in the form of clusters for prediction and second unclusteredly as manifest outcome variables.

Limitations of the novel methods include

1. they do not comply with the normal distribution and homoscedasticity,
2. they are at increased risk of type II errors,
3. they are at increased risk of type I errors.

Partial least squares method is parsimonious to principal components analysis because it can include outcome variables in the model.

There is room for improvement of the novel methods, because, to date,

1. binary variables cannot be included,
2. non linear variables cannot be included,
3. no tests for outliers is included.

REFERENCES

1. Tsao DA, Chang HJ, Hsiung SK, Huang SE, Chang MS, Chiu HH, Chen YF, Cheng TL, Shiu-Ru L. Gene expression profiles for predicting the efficacy of the anticancer drug 5 - fluorouracil in breast cancer. *DNA Cell Biol* 2010; 29: 285-93.
2. Latan MS, Laddha NC, Latani J, Imran MJ, Begum R, Misra A. Suppression of cytokine gene expression and improved therapeutic efficacy of microemulsion-based tacrolimus cream for atopic dermatitis. *Drug Deliv Translational Res* 2012; 2: 129-41.
3. Albertin PS, Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 1999; 18: 2863-70.
4. Yang X, Shen Q, Xu H, Shoptaw S. Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures. *Stat Med* 2007; 26: 1552-66.
5. Sverdlov L. The fastclus procedure as an effective way to analyze clinical data. *SUGI Proceedings* 26, paper 224, Long Beach CA, 2001.
6. Barthelemew DJ. Spearman and the origin and development of factor analysis. *Br J Math Stat Psychol* 1995; 48: 211-20.
7. Anonymous. SPSS Statistical Software version 18.0. Module Dimension Reduction, Factor Analysis, Online Help, www.spss.com, 29-04-2012.
8. Wold H. Estimation of principle components and related models by iterative least squares. In: Krishnaiah PR, ed, *Multivariate analysis*. Academic Press, New York 1966; pp 391-420.
9. Anonymous. Factor Analysis. Wikipedia, the free encyclopedia, 25-04-2012.
10. Anonymous. Partial least squares regression. http://en.wikipedia.org/wiki/Partial_least_squares_regression, 13-04-2012.
11. Tenenhaus M, Vinzi VE, Chatelin YM, Lauro C. PLS path modeling. *Comput Statist Data Anal* 2005; 48: 159-205.
12. Meng J. Uncover cooperative gene regulations by microRNAs and transcription factors in glioblastoma using a nonnegative hybrid factor. www.cmsworldwide.com/ICASS2011, 29-04-2012.
13. Hochreiter S, Clevert DA, Obermayer K. A new summarization method for affymetrix probe level data. *Bioinformatics* 2006; 22: 943-9.
14. Bastien T. PLS-Cox model: application to gene expression. *COMPSTAT* 2004; section: Partial Least Squares.
15. Li X, Gill R, Cooper NG, Yoo JK, Datta S. Modeling microRNA-mRNA interactions Using pls regression in human colon cancer. *BMC Medical Genomics* 2011; 4: 44.
16. Anonymous. R Statistical Software. Partial Least Squares, a free statistics and forecasting software, www.wessa.net/rwasp, 25-05-2012.
17. Guinot C, Latreille J, Tenenhaus M. PLS path modeling and multiple table analysis. Application to cosmetic habits of women in Ile de France. *Chem Intellig Lab Syst* 2001; 58: 247-59.

Appendix. Datafile of the example used in the present paper.

| Gene | | | | | | | | | | | | Outcome | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|---------|-------|-------|-------|
| 1 | 2 | 3 | 4 | 16 | 17 | 18 | 19 | 24 | 25 | 26 | 27 | 1 | 2 | 3 | 4 |
| 8,00 | 8,00 | 9,00 | 5,00 | 7,00 | 10,00 | 5,00 | 6,00 | 9,00 | 9,00 | 6,00 | 6,00 | 6,00 | 7,00 | 6,00 | 7,00 |
| 9,00 | 9,00 | 10,00 | 9,00 | 8,00 | 8,00 | 7,00 | 8,00 | 8,00 | 9,00 | 8,00 | 8,00 | 8,00 | 7,00 | 8,00 | 7,00 |
| 9,00 | 8,00 | 8,00 | 8,00 | 8,00 | 9,00 | 7,00 | 8,00 | 9,00 | 8,00 | 9,00 | 9,00 | 9,00 | 8,00 | 8,00 | 8,00 |
| 8,00 | 9,00 | 8,00 | 9,00 | 6,00 | 7,00 | 6,00 | 4,00 | 6,00 | 6,00 | 5,00 | 5,00 | 7,00 | 7,00 | 7,00 | 6,00 |
| 10,00 | 10,00 | 8,00 | 10,00 | 9,00 | 10,00 | 10,00 | 8,00 | 8,00 | 9,00 | 9,00 | 9,00 | 8,00 | 8,00 | 8,00 | 7,00 |
| 7,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 6,00 | 5,00 | 7,00 | 8,00 | 8,00 | 7,00 | 7,00 | 6,00 | 6,00 | 7,00 |
| 5,00 | 5,00 | 5,00 | 5,00 | 5,00 | 6,00 | 4,00 | 5,00 | 5,00 | 6,00 | 6,00 | 5,00 | 6,00 | 5,00 | 6,00 | 4,00 |
| 9,00 | 9,00 | 9,00 | 9,00 | 8,00 | 8,00 | 8,00 | 8,00 | 9,00 | 8,00 | 3,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 |
| 9,00 | 8,00 | 9,00 | 8,00 | 9,00 | 8,00 | 7,00 | 7,00 | 7,00 | 7,00 | 5,00 | 8,00 | 8,00 | 7,00 | 6,00 | 6,00 |
| 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 8,00 | 8,00 | 10,00 | 10,00 | 10,00 | 10,00 |
| 2,00 | 2,00 | 8,00 | 5,00 | 7,00 | 8,00 | 8,00 | 8,00 | 9,00 | 3,00 | 9,00 | 8,00 | 7,00 | 7,00 | 7,00 | 6,00 |
| 7,00 | 8,00 | 8,00 | 7,00 | 8,00 | 6,00 | 6,00 | 7,00 | 8,00 | 8,00 | 8,00 | 7,00 | 8,00 | 7,00 | 8,00 | 8,00 |
| 8,00 | 9,00 | 9,00 | 8,00 | 10,00 | 8,00 | 8,00 | 7,00 | 8,00 | 8,00 | 9,00 | 9,00 | 7,00 | 7,00 | 8,00 | 8,00 |

| Gene | | | | | | | | | | | | | Outcome | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|-------|-------|--|
| 1 | 2 | 3 | 4 | 16 | 17 | 18 | 19 | 24 | 25 | 26 | 27 | 1 | 2 | 3 | 4 | |
| 8,00 | 10,00 | 10,00 | 10,00 | 8,00 | 8,00 | 8,00 | 8,00 | 9,00 | 9,00 | 9,00 | 8,00 | 9,00 | 8,00 | 9,00 | 8,00 | |
| 8,00 | 8,00 | 8,00 | 5,00 | 5,00 | 8,00 | 8,00 | 8,00 | 6,00 | 8,00 | 10,00 | 5,00 | 7,00 | 7,00 | 5,00 | 7,00 | |
| 6,00 | 6,00 | 7,00 | 7,00 | 6,00 | 7,00 | 5,00 | 2,00 | 5,00 | 5,00 | 5,00 | 0,00 | 6,00 | 10,00 | 6,00 | 6,00 | |
| 10,00 | 10,00 | 10,00 | 10,00 | 5,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 5,00 | 10,00 | |
| 8,00 | 7,00 | 8,00 | 8,00 | 7,00 | 9,00 | 9,00 | 7,00 | 6,00 | 8,00 | 8,00 | 8,00 | 7,00 | 7,00 | 5,00 | 6,00 | |
| 8,00 | 7,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 8,00 | 9,00 | 9,00 | 8,00 | 9,00 | 8,00 | 7,00 | 7,00 | 6,00 | |
| 7,00 | 7,00 | 7,00 | 8,00 | 8,00 | 9,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 9,00 | 7,00 | 7,00 | 7,00 | 7,00 | |
| 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | |
| 10,00 | 10,00 | 10,00 | 9,00 | 7,00 | 9,00 | 9,00 | 7,00 | 8,00 | 8,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 8,00 | |
| 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 5,00 | 10,00 | 10,00 | 10,00 | 10,00 | 9,00 | 10,00 | 9,00 | 9,00 | |
| 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 9,00 | 9,00 | |
| 10,00 | 10,00 | 10,00 | 9,00 | 10,00 | 10,00 | 9,00 | 9,00 | 10,00 | 6,00 | 10,00 | 10,00 | 10,00 | 10,00 | 7,00 | 9,00 | |
| 7,00 | 9,00 | 9,00 | 8,00 | 9,00 | 10,00 | 9,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 5,00 | 7,00 | |
| 9,00 | 9,00 | 9,00 | 9,00 | 9,00 | 9,00 | 8,00 | 8,00 | 9,00 | 9,00 | 8,00 | 7,00 | 9,00 | 8,00 | 8,00 | 8,00 | |
| 6,00 | 5,00 | 5,00 | 7,00 | 1,00 | 5,00 | 6,00 | 5,00 | 5,00 | 10,00 | 5,00 | 10,00 | 3,00 | 0,00 | 5,00 | 5,00 | |
| 10,00 | 10,00 | 10,00 | 10,00 | 7,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 10,00 | 5,00 | 10,00 | |
| 8,00 | 9,00 | 10,00 | 9,00 | 9,00 | 10,00 | 9,00 | 9,00 | 9,00 | 10,00 | 10,00 | 9,00 | 10,00 | 9,00 | 10,00 | 9,00 | |
| 6,00 | 8,00 | 8,00 | 9,00 | 3,00 | 8,00 | 5,00 | 5,00 | 5,00 | 5,00 | 7,00 | 6,00 | 5,00 | 5,00 | 6,00 | 6,00 | |
| 9,00 | 9,00 | 9,00 | 9,00 | 5,00 | 8,00 | 5,00 | 6,00 | 9,00 | 9,00 | 8,00 | 10,00 | 8,00 | 8,00 | 8,00 | 8,00 | |
| 8,00 | 9,00 | 9,00 | 8,00 | 5,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 9,00 | 7,00 | 7,00 | 5,00 | 7,00 | |
| 6,00 | 7,00 | 7,00 | 7,00 | 6,00 | 6,00 | 6,00 | 3,00 | 3,00 | 6,00 | 0,00 | 6,00 | 5,00 | 5,00 | 5,00 | 5,00 | |
| 8,00 | 8,00 | 8,00 | 9,00 | 7,00 | 8,00 | 8,00 | 8,00 | 5,00 | 8,00 | 7,00 | 10,00 | 7,00 | 7,00 | 7,00 | 6,00 | |
| 8,00 | 8,00 | 9,00 | 6,00 | 6,00 | 7,00 | 5,00 | 5,00 | 10,00 | 5,00 | 0,00 | 10,00 | 7,00 | 7,00 | 5,00 | 5,00 | |
| 8,00 | 9,00 | 9,00 | 7,00 | 6,00 | 7,00 | 7,00 | 6,00 | 9,00 | 7,00 | 7,00 | 7,00 | 7,00 | 6,00 | 7,00 | 7,00 | |
| 8,00 | 4,00 | 6,00 | 7,00 | 3,00 | 6,00 | 6,00 | 6,00 | 0,00 | 6,00 | 0,00 | 9,00 | 6,00 | 5,00 | 4,00 | 6,00 | |
| 9,00 | 9,00 | 9,00 | 9,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | |
| 6,00 | 7,00 | 7,00 | 6,00 | 6,00 | 6,00 | 4,00 | 4,00 | 5,00 | 6,00 | 8,00 | 5,00 | 2,00 | 3,00 | 3,00 | 4,00 | |
| 6,00 | 7,00 | 7,00 | 7,00 | 4,00 | 6,00 | 4,00 | 4,00 | 4,00 | 8,00 | 4,00 | 5,00 | 6,00 | 7,00 | 7,00 | 5,00 | |
| 8,00 | 7,00 | 7,00 | 9,00 | 7,00 | 10,00 | 5,00 | 6,00 | 8,00 | 8,00 | 6,00 | 9,00 | 6,00 | 7,00 | 6,00 | 7,00 | |
| 10,00 | 10,00 | 10,00 | 9,00 | 8,00 | 7,00 | 8,00 | 7,00 | 8,00 | 9,00 | 9,00 | 8,00 | 5,00 | 5,00 | 5,00 | 4,00 | |
| 8,00 | 7,00 | 8,00 | 10,00 | 8,00 | 9,00 | 6,00 | 7,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 8,00 | |
| 8,00 | 9,00 | 7,00 | 8,00 | 9,00 | 8,00 | 8,00 | 7,00 | 8,00 | 7,00 | 5,00 | 9,00 | 6,00 | 8,00 | 8,00 | 8,00 | |
| 7,00 | 7,00 | 5,00 | 7,00 | 8,00 | 8,00 | 6,00 | 6,00 | 9,00 | 7,00 | 8,00 | 8,00 | 7,00 | 7,00 | 6,00 | 7,00 | |
| 9,00 | 9,00 | 10,00 | 8,00 | 8,00 | 8,00 | 6,00 | 5,00 | 10,00 | 10,00 | 10,00 | 10,00 | 7,00 | 7,00 | 5,00 | 6,00 | |
| 8,00 | 6,00 | 9,00 | 9,00 | 8,00 | 9,00 | 8,00 | 9,00 | 8,00 | 7,00 | 7,00 | 8,00 | 9,00 | 9,00 | 7,00 | 8,00 | |
| 7,00 | 7,00 | 8,00 | 9,00 | 7,00 | 7,00 | 7,00 | 8,00 | 7,00 | 8,00 | 9,00 | 7,00 | 6,00 | 8,00 | 7,00 | 7,00 | |
| 7,00 | 7,00 | 8,00 | 7,00 | 8,00 | 7,00 | 8,00 | 7,00 | 8,00 | 8,00 | 6,00 | 5,00 | 7,00 | 8,00 | 7,00 | 7,00 | |
| 9,00 | 10,00 | 9,00 | 9,00 | 8,00 | 7,00 | 9,00 | 9,00 | 6,00 | 6,00 | 6,00 | 6,00 | 7,00 | 9,00 | 8,00 | 8,00 | |
| 7,00 | 7,00 | 7,00 | 6,00 | 6,00 | 6,00 | 9,00 | 9,00 | 8,00 | 3,00 | 5,00 | 8,00 | 6,00 | 9,00 | 9,00 | 8,00 | |
| 9,00 | 10,00 | 7,00 | 8,00 | 7,00 | 5,00 | 10,00 | 10,00 | 10,00 | 10,00 | 7,00 | 10,00 | 6,00 | 8,00 | 9,00 | 7,00 | |
| 4,00 | 6,00 | 5,00 | 7,00 | 4,00 | 4,00 | 3,00 | 3,00 | 10,00 | 9,00 | 10,00 | 9,00 | 4,00 | 4,00 | 3,00 | 4,00 | |
| 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 6,00 | 10,00 | 8,00 | 10,00 | 10,00 | 10,00 | 8,00 | 8,00 | 7,00 | |
| 8,00 | 6,00 | 3,00 | 8,00 | 6,00 | 8,00 | 5,00 | 1,00 | 10,00 | 7,00 | 10,00 | 10,00 | 7,00 | 7,00 | 8,00 | 6,00 | |
| 7,00 | 7,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 6,00 | 4,00 | 6,00 | 5,00 | 9,00 | 7,00 | 7,00 | 5,00 | 6,00 | |
| 9,00 | 10,00 | 9,00 | 9,00 | 8,00 | 7,00 | 7,00 | 9,00 | 7,00 | 4,00 | 7,00 | 4,00 | 5,00 | 9,00 | 8,00 | 6,00 | |
| 10,00 | 10,00 | 8,00 | 9,00 | 7,00 | 6,00 | 5,00 | 8,00 | 6,00 | 7,00 | 6,00 | 6,00 | 7,00 | 9,00 | 9,00 | 5,00 | |
| 4,00 | 6,00 | 9,00 | 8,00 | 9,00 | 9,00 | 7,00 | 7,00 | 0,00 | 1,00 | 0,00 | 0,00 | 8,00 | 7,00 | 9,00 | 4,00 | |
| 8,00 | 8,00 | 7,00 | 7,00 | 4,00 | 6,00 | 5,00 | 5,00 | 4,00 | 4,00 | 4,00 | 5,00 | 3,00 | 5,00 | 4,00 | 3,00 | |
| 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 7,00 | 8,00 | 6,00 | 8,00 | 7,00 | 8,00 | 9,00 | 6,00 | 7,00 | 7,00 | 9,00 | |
| 8,00 | 8,00 | 3,00 | 4,00 | 9,00 | 8,00 | 7,00 | 6,00 | 6,00 | 8,00 | 6,00 | 8,00 | 7,00 | 8,00 | 8,00 | 5,00 | |
| 6,00 | 8,00 | 8,00 | 7,00 | 5,00 | 7,00 | 7,00 | 7,00 | 8,00 | 7,00 | 8,00 | 8,00 | 7,00 | 7,00 | 7,00 | 7,00 | |
| 9,00 | 10,00 | 9,00 | 8,00 | 7,00 | 8,00 | 8,00 | 8,00 | 6,00 | 7,00 | 6,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | |
| 4,00 | 5,00 | 5,00 | 7,00 | 6,00 | 9,00 | 6,00 | 8,00 | 9,00 | 9,00 | 8,00 | 8,00 | 8,00 | 10,00 | 10,00 | 7,00 | |
| 6,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 6,00 | 5,00 | 4,00 | 6,00 | 6,00 | 9,00 | 9,00 | 9,00 | 8,00 | |
| 9,00 | 9,00 | 7,00 | 7,00 | 4,00 | 6,00 | 5,00 | 3,00 | 3,00 | 5,00 | 3,00 | 6,00 | 4,00 | 3,00 | 7,00 | 5,00 | |
| 8,00 | 9,00 | 5,00 | 5,00 | 7,00 | 8,00 | 7,00 | 6,00 | 6,00 | 7,00 | 6,00 | 6,00 | 7,00 | 6,00 | 7,00 | 4,00 | |
| 10,00 | 10,00 | 8,00 | 7,00 | 7,00 | 8,00 | 6,00 | 5,00 | 6,00 | 8,00 | 5,00 | 7,00 | 5,00 | 6,00 | 8,00 | 8,00 | |
| 7,00 | 9,00 | 8,00 | 7,00 | 7,00 | 8,00 | 10,00 | 8,00 | 9,00 | 8,00 | 7,00 | 8,00 | 9,00 | 8,00 | 8,00 | 7,00 | |
| 9,00 | 9,00 | 9,00 | 9,00 | 6,00 | 3,00 | 4,00 | 4,00 | 7,00 | 8,00 | 6,00 | 8,00 | 6,00 | 3,00 | 4,00 | 4,00 | |
| 6,00 | 5,00 | 5,00 | 7,00 | 6,00 | 7,00 | 6,00 | 4,00 | 9,00 | 8,00 | 8,00 | 9,00 | 7,00 | 7,00 | 7,00 | 6,00 | |
| 9,00 | 9,00 | 9,00 | 9,00 | 9,00 | 10,00 | 10,00 | 8,00 | 7,00 | 6,00 | 8,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | |
| 8,00 | 9,00 | 9,00 | 7,00 | 9,00 | 9,00 | 8,00 | 8,00 | 3,00 | 5,00 | 3,00 | 6,00 | 7,00 | 6,00 | 7,00 | 8,00 | |
| 8,00 | 8,00 | 8,00 | 8,00 | 4,00 | 5,00 | 5,00 | 8,00 | 8,00 | 2,00 | 8,00 | 7,00 | 8,00 | 7,00 | 8,00 | 6,00 | |
| 8,00 | 9,00 | 9,00 | 4,00 | 8,00 | 9,00 | 9,00 | 7,00 | 6,00 | 7,00 | 8,00 | 6,00 | 9,00 | 7,00 | 8,00 | 7,00 | |
| 10,00 | 9,00 | 7,00 | 7,00 | 7,00 | 8,00 | 8,00 | 8,00 | 8,00 | 7,00 | 5,00 | 7,00 | 7,00 | 7,00 | 7,00 | 6,00 | |