***Aging Wisely Project* Consulting Analysis and Results**

**By Christopher Clark, M.S. Statistics**

**For MATH 6330: Statistical Consulting at the University of Colorado, Denver Spring 2022**

1. **Introduction**

The *Aging Wisely Project* has a simple goal: “How does one age wisely?” Put in more concrete terms, what are the psychological and psycho-analytical markers of one who has aged wisely? To determine this, Dr. Ben Green and Scott Fisher conducted a two-hour interview with fifty-two elderly individuals, ranging in age from 68 to 95 years old, and ranked them on a scale of 1-to-5 for twenty-eight different psychological, psycho-analytical, and presentation markers or factors (with one factor being on a 1-to-7 scale). These markers can be categorized as Meta-Factors, Elder Developmental Tasks, Controllable Protective Factors (scaled from 1-to-7), Elder Identity Revision, Interview Presentation Variables, and Others. As a final metric, each interviewee was given a Global Clinical Impression score, an overall grade on how well they had aged wisely. The Meta-Factors were also combined to create a Composite Meta-Factor Score as a secondary matric for aging wisely. Due to the ranked nature of the scoring, this data qualifies as *Likert* data. Several demographic factors, Gender, Marital Status, Ethnicity, Socio-Economic Status, Residential Status, and Religious Upbringing (as well as Age, mentioned above), were also recorded for each participant. The participants reside mostly in Colorado and Massachusetts, and some are known personally to the investigators, while others are references or residents of the Allied Jewish Home.

The following section will run through our results, starting with whether the investigators became more consistent with their grading as the project progressed. Then, we move onto correlations between factors and the Global Clinical Impression and the Composite Meta-Factor Score. We note here that the actual title of this final score is Meta-Factor Total, but, as we scale this down to the same scale as all the rest, 1-to-5, we renamed it. After this, we explore the relationships, in the form of correlations, between the factors, Global Clinical Impression, and Composite Meta-Factor Score, themselves. We then work towards determining which factors are most important for the two main scores mentioned above. Finally, we determine if there is a relationship between the demographic variable Residential Status and the Elder Identity Revision factor Competence.

The overall goal, which we hope will be answered by the end of our analysis, is: “Which factors most strongly contribute to aging wisely?”

1. **Analysis and Results**

We will proceed by breaking our analysis into distinct questions to be answered.

1. Did the interviewers become more consistent in their grading as the project progressed?

In none of the twenty-eight factors or the Global Clinical Impression did the interviewers become more consistent in their grading toward the end versus the beginning of the project. In fact, the opposite is more strongly supported. In the Global Clinical Impression, Narrative Coherence, Open System, Candor, Energy, Mood, Healthy Grieving, Reliable Attachment Figures, Ego Integrity, and Self-Acceptance/Self-Esteem, the interviewers had more differing grades near the middle and end than in the beginning, as determined by a plot of the difference of the scores per case for each factor and Global Clinical Impression. However, in no cases were the interviewers’ grades more than two different. This can be seen in the size of the “spikes” in the graphs (Appendix D has the graphs for each factor and Global Clinical Impression). Also, the Global Clinical Impression only ever had a difference of one point, showing at least that the investigators were overall close in their assessments for this metric (Appendix D, vii). We also have the average difference per case across all factors and Global Clinical Impression, with a positive value indicating that Scott Fisher on average scored that case higher, while a negative value indicates that Dr. Ben Green on average scored that case higher. The number of cases in which Scoff Fisher scored overall higher were 15, whereas the number of cases in which Dr. Ben Green scored overall higher were 31. The number of cases with an average difference of 0 are 6. The case with the highest average difference was case 39, with an average of -0.8161, meaning that Dr. Ben Green scored this case nearly an average of one whole point higher throughout all factors and Global Clinical Impression. The overall average discrepancy across all cases is -0.09, and the standard deviation of the discrepancies is 0.26

1. What is the correlation between the factors and the Global Clinical Impression and Composite Meta-Factor Score?

For this, several *heatmaps* based on *Spearman’s Correlation Coefficient* (pronounced “rho”) were created to visualize the correlation between the factors and the Global Clinical Impression.

Each of the correlations *is* *statistically significant* at the 0.05 level or better, meaning that we can treat their correlations as very likely representing a genuine association, as opposed to a statistical anomaly.

A heatmap of *all* factors, Global Clinical Impression, and the Composite Meta-Factor Score (a sum of all five Meta-Factors then scaled from one to five) shows us that the lower correlations are in the low 0.30 range, with the *lowest* of 0.31 between Reliable Attachment Figures and Cognition (p-value ). The higher correlations are in the high 0.80 and low 0.90 range, with the *highest* correlation of 0.92 being between Ego Integrity and Global Clinical Impression (p-value ). Between Global Clinical Impression and Composite Meta-Factor Score, we have (p-value ). In short, *all* factors, Global Clinical Impression, and Composite Meta-Factor Score are at least partially positively correlated (Appendix C, i).

Between just the Meta-Factors, we find that the highest correlation is between Resilience and Purpose, (p-value ), with the lowest being between Active Habits and Self-Acceptance/Self-Esteem, , (p-value ) (Appendix C, ii). For Elder Identity Revision factors the largest correlation is between Consciousness and Connection, (p-value ), and the lowest is between Competence and Consciousness, (p-value ) (Appendix C, iii). For Elder Developmental Tasks the highest correlation is between Ego Integrity and Generativity, (p-value ), and the lowest is between Keeper of Meaning and Ego Integrity, (p-value ) (Appendix C, iv). Finally, for Elder Developmental Tasks and Elder Identity Revision together, the highest correlations, looking between the two groups of factors only, which can be seen in the top right block that is five blocks deep and three wide, is between Ego Integrity and Connection, (p-value ), and the lowest is between Connection and Keeper of Meaning, (p-value ) (Appendix C, v).

1. Which factors are most predictive of the Global Clinical Impression and the Composite Meta-Factor Score? Does Age affect the relationship?

For this, several *Multiple Linear Regressions* were run for various sets of predictors (the factors), or *regressors* in our linear regression models, and the outcome of Global Clinical Impression or Composite Meta-Factor Score. As for the question of age, we added Age as a *main effect* and AgeFactor for each factor as the *interaction terms* to the set of regressors that appeared in the *final* models and reran our full model selection (to be described in the Models and Methods section) to determine how, if at all, Age affected the models. We also ran a clustering algorithm, *K-Means Clustering,* with the same sets of regressors for each regression (without Age) to compare the two results.

From the regression with *all* factors as regressors and Global Clinical Impression as the outcome and running a regression *per regressor*, we found that Ego Integrity had the highest value (a measurement indicating model fit and the *percentage of the variability in the outcome explained by the model*) of 0.984, and therefore explained more of the variance as a standalone regressor than any other standalone regressor. Considering all possible regressors, we found the best three-factor model had Ego Integrity, Gratitude, and Connection, in that order of contribution to Global Clinical Impression with (Appendix E, i). The best four-factor model had Ego Integrity, Gratitude, Connection, and Resilience with (Appendix E, i). Why we have two models here is explained in section **4. Models and Methods**. The clustering algorithm found, in order of most important (to be described in the same section), Engagement, Cognitive, and Gratitude. In the three-factor model, Ego Integrity add or explained 36.74% of the variability of Global Clinical Impression *not explained by a model with only the other two factors*, Gratitude added 44.28%, and Connection added 29.41%. In the four-factor model, we have that, in the same manner described above, Ego Integrity added 27.80%, Gratitude added 12.86%, Connection added 33.86%, and Resilience added 16.62%.

From the regression with all non-Meta-Factors as regressors and Composite Meta-Factor Score as the outcome and running a regression per regressor, we found that Mood had the highest of 0.973 as a single regressor. Considering all possible regressors we found, in the order described above, Mood, Generativity, Narrative Coherence, and Tragic Optimism with (Appendix E, ii). We also note that we needed to remove case number 38 to achieve a well-fitting model. The clustering algorithm found, using the same dataset with case number 38 removed, in the order as described above, Engagement, Cognitive, and Continuity. In the final model, Mood added 12.11%, Generativity added 4.79%, Narrative Coherence added 11.32%, and Tragic Optimism added 8.09% of the variability in Composite Meta-Factor Analysis.

From the regression with all Meta-Factors as regressors and the Global Clinical Impression as the outcome and running a regression per regressor, we found that Gratitude had the highest of 0.980 as a single regressor. Considering all regressors, we found Gratitude, Self-Acceptance/Self-Esteem, and Resilience with (Appendix E, iii). From the clustering algorithm, we have Active Habits, Resilience, and Self-Acceptance/Self-Esteem. In the final model, Resilience added 8.37%, Gratitude added 15.18%, and Self-Acceptance/Self-Esteem added 14.17% of the variability of Global Clinical Impression.

From the regression with Elder Identity Revision as regressors and the Global Clinical Impression as the outcome and running a regression per regressor, we found that Control had the highest of 0.969 as a single regressor. Considering all regressors, we found Connection and Control with (Appendix E, iv). We note that we had to remove case 43 to achieve a well-fitting model. The clustering algorithm, using the same dataset with case 43 removed, found Continuity, Control, and Connection, the last two having the same “score” as described above. In the final model, Connection added 34.79% and Control added 44.47% of the variability in Global Clinical Impression.

For the regression with Elder Developmental Tasks as regressors and Global Clinical Impression as the outcome and running a regression per regressor, we found that Ego Integrity had the highest of 0.984 as a single regressor. Considering all regressors, we found Ego Integrity and Generativity with (Appendix E, v). The clustering algorithm found Ego Integrity and Generativity. In the final model, Generativity added 16.83% and Ego Integrity added 64.26% of the variability of Global Clinical Impression.

For the regression with Elder Identity Revision as regressors and Composite Meta-Factor Score as the outcome and running a regression per regressor, we found that Control had the highest of 0.963 as a single regressor. Considering all regressors, we found Control, Continuity, and Connection with (Appendix E, vi). We note that we had to remove cases 38 and 43 to achieve a good model fit. The clustering algorithm found, using the same dataset with cases 38 and 43 removed, Continuity, Connection, and Control. In the final model, Continuity added 11.34%, Connection added 4.86%, and Control added 9.04% of variability in Composite Meta-Factor Score.

For the regression with Elder Developmental Tasks as regressors and Composite Meta-Factor Score as the outcome and running a regression per regressor, we found that Ego Integrity had the highest of 0.959 as a single regressor. Considering all regressors, we found Ego Integrity and Generativity with (Appendix E, vii). The clustering algorithm found Ego Integrity and Generativity (just choosing the top two). In the final model, Ego Integrity added 30.00%, Generativity added 20.36% of the variability in Composite Meta-Factor Score. This was also the only regression in which Age contributed significantly to the outcome (Appendix E, viii).

Finally, for this section, it was requested to fit a model for Global Clinical Impression with *each possible set of the following factors*: Ego Integrity, Gratitude, Connection, Purpose, and Resilience. The resulting values can be found the Appendix E, ix. However, we note here that this method uncovered a three-factor model our forward stepwise regression skipped over (to be explained in **4. Models and Methods**). It is one with Ego Integrity, Connection, and Resilience with . Going out to four decimal places to compare the model with Ego Integrity, Gratitude, and Connection with this one, we have 0.9936 versus 0.9938, showing a *slight* improvement.

1. What is the relationship between demographic variable Residential Status and the Elder Identity Revision factor Competence?

We found that, after grouping the participants into groups of “H” and “Not-H”, the “H” status individuals had an average Competence score of 4.1 and the “Not-H” had an average Competence score of 2.94. We also ran a *Fisher’s Exact Test* and found strong statistical evidence that there *is* an association between Residential Status and Competence such that higher Residential Status often results in higher Competency Scores.

1. **Discussion**

From the Analysis and Results section, we can answer the questions which motivated this analysis.

First, and probably the most important, is: “Which factors are the most indicative of a high Global Clinical Impression Score?”

From the simple correlation heatmaps, we find that Ego Integrity has the highest correlation with Global Clinical Impression. We can now look to the multiple linear regressions with Global Clinical Impression as the dependent variable. Ego Integrity shows up as the largest coefficient when *all* factors are included in the model and has the highest value when compared to models of only single regressors. We encountered the *same* results regressing only the Elder Developmental Tasks with Global Clinical Impression, which should not be a surprise. From the regression with just the Elder Identity Revision factors as regressors and the Global Clinical Impression as the dependent variable, we found that Control and Connection provided the best fit by our criteria. Further, they both explained similar amount of the variance of the full model and had almost identical coefficients. Given the previous statement, the additional fact that the score of the models fit with only each regressor are nearly identical as well, *and* a strong correlation between them of 0.79 (p-value ), we can confidently say that the two are essentially interchangeable. Thus, this researcher would say that the final answer to which factors are most indicative of a high Global Clinical Impression is Ego Integrity overall, and among the Elder Identity Revision factors, Connection and/or Control, and from the Meta-Factors (to be elaborated on in the next paragraphs), is Gratitude and/or Resilience.

Second question: “What is the relationship between the Meta-Factors/Composite Meta-Factor Score and the Global Clinical Impression?”

The correlation between Composite Meta-Factor Score and Global Clinical Impression is 0.83, a strong, positive correlation (p-value ). We see that Gratitude is among the most contributory factors in the regression with *all* factors as regressors and Global Clinical Impression as the outcome. Gratitude also appears as a high-ranking factor in the clustering analysis with the same factors. In the regression with only Meta-Factors as regressors and the Global Clinical Impression as the dependent variable, we that Gratitude is the most contributary of the Meta-Factors. Among the Meta-Factors themselves, we see an overall high level of correlation, with the weakest being between Active Habits and Self-Acceptance/Self-Esteem and the highest being between Purpose, Meaning and Resilience. So, this researcher can say with confidence that there is a strong relationship between the Meta-Factors and the Global Clinical Impression, with the Meta-Factor Gratitude being the most significant.

Finally, Age does not, except in the regression with Elder Developmental Tasks as regressors and the Composite Meta-Factor Score, seem to have any impact on the models.

We can also note that, though we have some reservations with the clustering algorithms (to be explained in the next section), they, except for the models with non-Meta-Factors as regressors and Composite Meta-Factor as the dependent variable, generally support the outcomes of the multiple linear regressions, further giving us some confidence in the veracity of our findings.

From here, the next step would be to try and generate a larger dataset. As mentioned in the introduction, most of these individuals are from two states and many are from the same assisted living home. Further, though the investigators attempted to create a dataset representative of the overall racial percentages in America, the small number of participants, even accounting for White versus Non-White, make analysis of the effect of race difficult. So, expanding the investigation to many different areas across the country and maintaining the same racial breakdown while doing so would allow for better generalization of these results.

While doing so, it might be beneficial to reduce the number of factors. Given the strong correlation between many of the factors, perhaps focusing on just a few select ones would be a better use of the investigators’ resources.

Finally, given the investigators’ sometimes differing grading, perhaps introducing more quantitative methods which yield consistent results between graders would help the analysis.

1. **Model and Methods**

This section will be a non-technical (in as much as it can be) review of the methods used here. Mathematics will be kept to a minimum, and the objective is to provide the clients with a rough understanding of how the analysis was performed for explication to an interested party or for continued analysis.

To begin, we will discuss the data itself. Likert data is generally on a 1-to-5 (sometimes 1-to-7) scale and there is an assumption that the variable being measured is *linear*, meaning a proportional increase in the attitude/datum should result in a rise in ranking that is roughly the same for each unit increase in the attitude/datum. However, because of the nature of qualitative Likert data (what is collected here), an increase of, say, 2 to 3, might *not* carry the *exact* same weight as one from 1 to 2. So, certain common statistics, such as the mean and standard deviation, are not appropriate [1]. This does not present a problem, but we must tailor our analysis to what we *do* know about this data type. Likert data is *ordinal*, meaning there is a clear order [2]. It is referred to as *non-parametric* data because we cannot assume the data was collected from some distribution (with *parameters* we might discover, hence the name “*non-*parametric”) [3]. We will see how deal with this data shortly. However, we note here that *everything*, including the single factor recorded on a 1-to-7 scale, Age, and the interaction terms, are *all* scaled to 1-to-5 for consistency.

First, and easiest to describe, is our method of determining if the two investigators converged on their grading as the project progressed. For this, we simply chose to take the value for each case for each factor and Global Clinical Impression (we did not include the Composite Meta-Factor Score, as that is just a composition of the five Meta-Factors) determined by Scott Fisher and subtracted the corresponding value determined by Dr. Ben Green. If the two investigators agreed on the value for that case for the factor or Global Clinical Impression, this would return a value of 0, for example. We then plotted them sequentially. A perfect sequence of grading would be a line straight through zero. However, jumps, positive or negative, indicate a grading difference for that case.

An example is shown below:

Chart, line chart

Description automatically generated

In the above graphic, we see roughly similar numbers of disagreements in the beginning and end of the project, but a marked increase in the middle. This is common in the graphs. We can also plot each type of factor together.

Here is the Meta-Factors plot:

Chart, histogram

Description automatically generated

Though some spikes might not be visible if they were graphed before subsequent spikes overlaying them, we can see that *all* Meta-Factors oscillate around the zero, with Purpose, Resilience, and Self-Acceptance/Self-Esteem reaching differences of two. All the graphs can be found in Appendix D.

Turning to the concept of correlation, we can define this word, *correlation*, roughly as “statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate).” [4] For this, we will use *Spearman’s Correlation Coefficient*, or *Spearman’s* (pronounced “rho”) [5]. We calculate this *for each pair of factors/outcomes/variables*, and we will receive a value between and 1. A value of signifies that an increase of one member of the pair will result in a decrease in the other by the same amount, and vice versa for a value of . A value near indicates that the two sets of numbers have no real relationship to each other.

So, we began by performing a p-value test for each possible pair. Given that there are 30 factors ( factors, Global Clinical Impression, and Composite Meta-Factor Score), we find 870 possible non-repeating pairs (, not because the correlation between two equal sets of data will be 1, so we can skip them). We find a p-value lower than for each. We interpret this as meaning that *there is a less than 5% chance of getting the given correlation between them by chance if there is no correlation*. Given that this is quite low, we can reject the hypothesis that there is no correlation between them, meaning we can proceed quite confident that there *is*. We can visualize these correlations with heatmaps, where each square corresponds to the correlation of the intersecting row and column. Each square has a color, ranging from black, representing a low correlation, to bright orange/white, representing a strong one. We see that the diagonals are all white and have a value of 1. This is to be expected, as the diagonals are where a row meets the *same* column, which, from our description above, will result in a perfect correlation of 1. The heatmaps can be found in Appendix C.

Now, to determine which of the factors are most important for the Global Clinical Impression and Composite Meta-Factor Score, as mentioned above, we performed several multiple linear regressions. However, we need to plot each of the factors versus the Global Clinical Impression and Composite Meta-Factor Score to confirm there is a rough linear relationship to be discovered.

Here is Gratitude and Global Clinical Impression:

Chart, scatter chart

Description automatically generated

We see it generally looks linear. Here’s another:

Chart, scatter chart

Description automatically generated

This is Generativity versus Composite Meta-Factor Score. We see a point that seems to buck our linear trend. Remembering that we outliers for regressions with Composite Meta-Factor Score as a dependent variable, we shouldn’t be surprised. We run this for each factor and Global Clinical Impression and Composite Meta-Factor Score, but have left them in the Python script, as they are all like the above two.

In general, our data seems to be linear, and we are justified in using multiple linear regressions.

Now, multiple linear regressions take a set of factors and attempt to fit a model where each receives a coefficient, which we can interpret as a weight (how much each factor contributes to the dependent variable), and then adds them to arrive at the value of the outcome. These models are easily interpretable, as a positive coefficient implies that the factor contributes positively to the outcome and vice versa. Further, smaller coefficients contribute less than larger ones. So, we can easily read off which factors add or subtract from the outcome, and which are more and less important for the final model.

For example, the model that fit Elder Developmental Tasks as regressors and Global Clinical Impression as the outcome returned (after our various checks, which will be explained shortly) the equation:

We interpret this by saying that, if two cases had the *same* Generativity score, but a difference of 1 for the Ego Integrity score, there should be a difference in the Global Clinical Impression score, and vice versa with the Generativity factor. We see that both have positive coefficients, so both *add* to Global Clinical Impression, but Ego Integrity adds much more, and is thus more important.

However, one might notice that there is no Keeper of Meaning in the above equation. For this, we will now discuss exactly how we arrived at our models.

To begin, fitting the 28 factors as regressors with Global Clinical Impression as the outcome would result in possible combinations of factors, and therefore, models, a number that we simply cannot fit in a reasonable timeframe. Thus, we needed to do a *Forward* *Stepwise Search*, wherein we fit the models from simple (low numbers of factors) to more complex (large numbers of factors) by some criteria. We note that the models fitting the non-Meta-Factor factors with Composite Meta-Factor Score would have models, which is similarly prohibitive. So, we will use the same criteria for our stepwise search for both.

The criteria we choose to use is the *Akaike Information Criterion* (AIC). This measurement weighs the fit of the model with the complexity. Generally, a low AIC is preferred, and as the fit becomes better, the AIC will decrease. *However*, adding factors will *increase* the AIC. So, we seek a model with a good fit which is as simple as can be, which will result in the lowest AIC. We can set up our stepwise search by a simple algorithm. First, take each factor individually and fit a model with only that factor, recording the AIC. Take the model with the lowest AIC. Now, begin with that same model and fit a second factor, running through the non-used ones. Record the AIC. Find the model with the lowest AIC. If this AIC is *not* lower than the lowest AIC from the first round of model fitting, the original model with the lowest AIC from the first round of fitting is our model and we are done. If it is smaller than the model fit with just one factor, this is the model we continue with. Now, take that model and begin fitting for a third factor. Continue doing this until you find that *no* AIC values from the current round of model fitting are lower than the lowest from the previous. As in, we have reached a minimum AIC at the previous step. This is the model we go with, the lowest from the previous round of fitting (which will be the lowest we have thus encountered).

Now, we can print out the coefficients and various other measures of fit to see our final model.

Here is the output after a forward stepwise search by AIC with all 28 factors as regressors and the Global Clinical Impression as the outcome:

A screenshot of a computer

Description automatically generated with medium confidence

There is quite a lot of information, but we will only focus on the *coef* and columns and the score. First, the score is quite high (with 1 being perfect), meaning this is a good model. Looking at the *coef* column, we see we have many factors included and some have a very low value coefficient. This isn’t a problem, but when we now look at the column, we see that factors with low coefficients *also* have higher p-values. We interpret this as meaning that they are not contributing much to the model and are not statistically significant.

Before continuing, we want to do another check of the model fit by showing a *Histogram* and *QQ-Plot*:

Chart, histogram

Description automatically generated

What we hope to see is a nice bell-shaped (or nearly) graph on the left and all the blue dots within the red, dotted lines on the right. We see this, so we believe our model is fit well.

What would a *bad* fit look like? Let us look at the first fitting of the 23 non-Meta-Factor factors as regressors and the Composite Meta-Factor Score as the dependent variable:

Chart, histogram

Description automatically generated

We see that the left graph doesn’t look too bell-shaped, and the right has a point *far* outside of the red, dotted lines. This is an outlier, and we need to remove it and refit the model. We mention above which cases in each regression we had to remove. Once we have a model that fits well, we can continue.

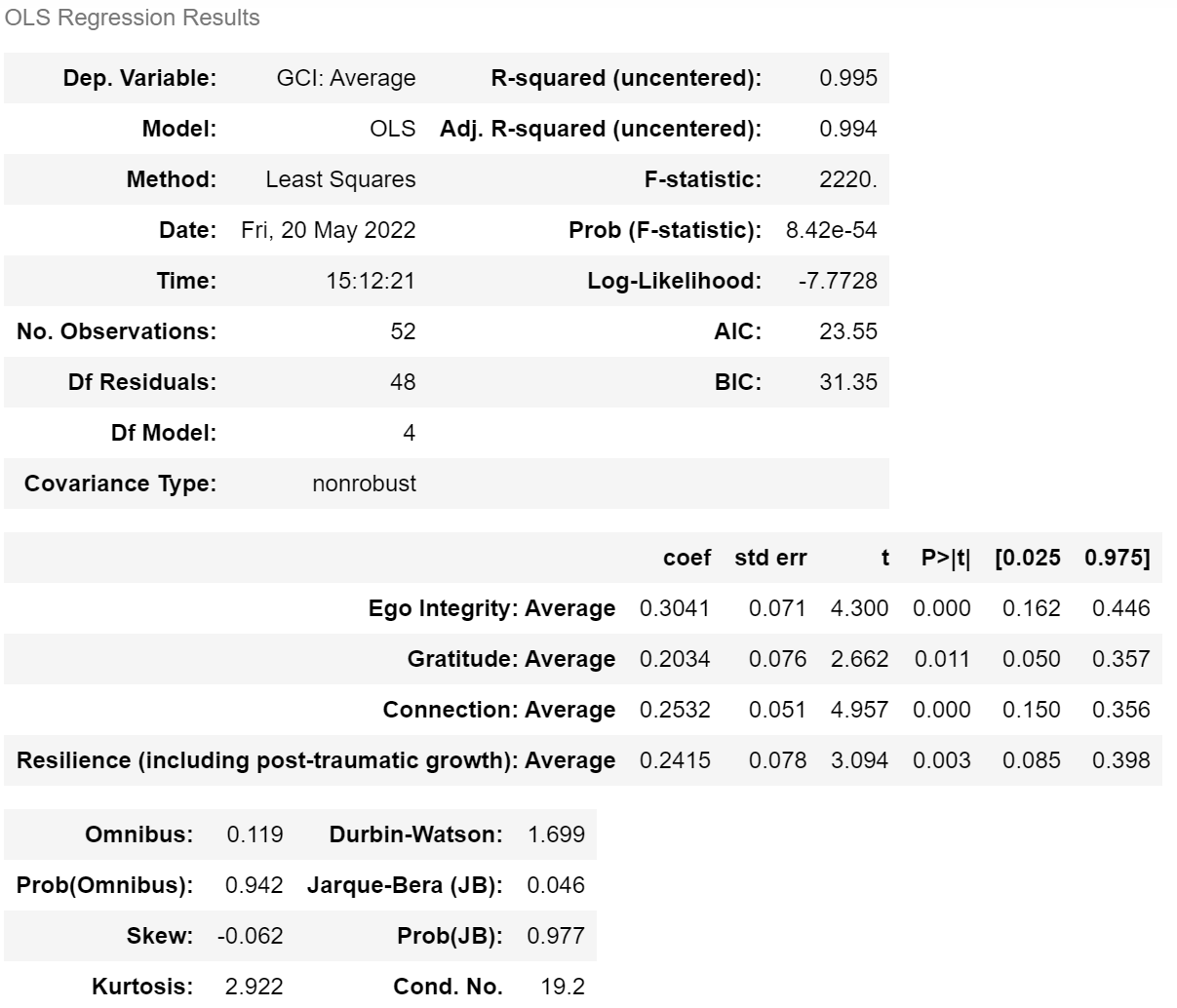
Returning to the first regression, with the 28 factors as regressors and the Global Clinical Impression as the dependent variable, we have a good fit, so we will continue parsing down the model. We will look at the AIC score versus steps in the stepwise regression to see if there is a point where we encounter diminishing returns from adding factors.

Here it is:

Chart, line chart

Description automatically generated

We see that, though the AIC *does* continue to decrease from step 3 (index 2, as Python starts at index 0) to step 4 (index 3), it is less than the previous step. Further from step 5 (index 4) to 6 (index 5), the decrease is becoming *slight*. We also notice that at index 7, or the step where the stepwise algorithm adds the *eighth* factor, we are now increasing our AIC, meaning our model is becoming too complex for how little we are gaining from adding factors. We see this in the original results. When the forward stepwise algorithm attempted to find an eighth factor to add, all possible models increased our AIC, so it stopped at just 7 factors. We now invoke the principle of parsimony to rerun our forward stepwise search but stop at just 3 or 4 factors. Stopping at 4 factors, we have the following model output (on the next page):



We see a high score and all the factors have a low p-value. Continuing, what happens if we stop at just 3 factors? We get:



We have lost a small amount of , but we now have only a single Meta-Factor, Elder Developmental Task, and Elder Identity Revision. We also note that the correlation between Gratitude and Resilience is 0.84, or quite high, hence the small gain from adding the fourth factor. This is the reason we include both in section **2.** **Analysis and Results**, so we can decide which we prefer, though *both are well-fit, have acceptable p-values, and have high explanatory power*.

For the other five regressions, because we have far fewer factors, we have correspondingly fewer possible groupings for models. So, we can perform an *Exhaustive Search*, where we fit a model with *each possible combination* of factors. For example, if we have factors A, B, and C, we can fit a model with A, B, C, A and B, A and C, C and B, and finally, A and B and C. This is far more reasonable computationally. We again choose our preferred set of factors by AIC. After we have the set of factors with the lowest AIC, we run our checks, plotting the histogram and QQ-plot, removing any outliers, re-fitting, checking for outliers, and so on until we have a good fit. Then, like above, we can visualize AIC versus step to see if we can simplify our models further until we have a well-fit model with low p-values for each of the included factors.

Now, why did our exhaustive search with the factors Ego Integrity, Gratitude, Resilience, Purpose, and Connection return a better model than Ego Integrity, Gratitude, and Connection with Global Clinical Impression as the outcome?

Well, the forward stepwise found the best one-factor model to be Ego Integrity, then started with Ego Integrity and began checking all other 27 possible combinations. It found the best second-factor to add to the model with Ego Integrity is Gratitude. Then, with Ego Integrity and Gratitude, it would have fit the other 26 possible models until it found Connection. So, we see that it never would have even checked Ego Integrity, Connection, and Resilience. This is an unfortunate tradeoff between computational speed/ability and accuracy.

To determine the percent of the variance explained by each factor in the final model of each regression, we fit a model with the other factors, excluding the one we are measuring. Then, we find the *Sum of Squared Errors* (SSE), a measurement of the total incorrectness of our model for the full model (with all the factors) and the reduced model (without the factor we are interested in). We can then find the proportion of the SSE *reduction* we get from adding the removed factor to the model with the other factors. An example should help:

Consider the model with the 28 factors as regressors with Global Clinical Impression as the dependent variable. We found the best model has Ego Integrity, Generativity, and Connection. So, to determine the percent of the variance explained by Ego Integrity, for example, we would find the SSE for the full model (with all factors). This is about 4.92. Now, we remove Ego Integrity and fit a model with only Connection and Generativity. This has an SSE of 7.78. So, to find the proportion, we do:

So, in comparison with a model fit with only Generativity and Connection, adding Ego Integrity explains of the variance of Global Clinical Impression not accounted for by just Generativity and Connection.

To determine the impact of Age, we do two things, one is to add Age (scaled to 1 to 5, like everything else), to the final model of each regression and then multiply each factor in the final model by Age (to create interaction terms) and add those into the model (also scaling them down). So, in the previous example where our final model had Ego Integrity, Generativity, and Connection, we now would have Age, Ego Integrity, Ego IntegrityAge, Generativity, GenerativityAge, Connection, and ConnectionAge. Then, we run our forward stepwise regression with all checks and procedures, same as before, until we arrive at a model with a good fit and all low p-values for the included factors. If Age or any of the interaction terms appear, we can say that Age is influential in the model.

In two cases, the regression between all non-Meta-Factors as regressors and Composite Meta-Factor Score as the outcome and when determining the role of Age in the regression with Elder Developmental Tasks as regressors and Composite Meta-Factor Score as the outcome, we ended up with a model with negative coefficients but with *all* regressors with a low enough p-value. At this point, we then ran a regression with the factor with a negative coefficient as the outcome and the other factors as regressors. The goal was to see if the other regressors could account for much of the value added to our models by the regressor given a negative coefficient. In both cases, we found this, allowing us to eliminate the regressor with a negative factor.

Finally, for our clustering algorithms, we mentioned above that we would not put too much weight into them, especially when they disagree with our regressions. Why is this? Well, to determine if clustering is a good idea, we first need to figure out of our data seems to be clustered at all. With so many factors, visualizing them directly is not possible, but we can perform *Principal Component Analysis* (PCA) to extract some information from them. This will return a new data set whose columns are *mixtures of the original ones, are uncorrelated, and contained decreasing amounts of the variance of the data set*.

We are using PCA *only* to help us visualize our higher-dimensional datasets. Using the example we have been using, with 28 factors as regressors with Global Clinical Impression as the dependent variable, we would take the 28 factors, perform our PCA, and plot the results:

Chart

Description automatically generated

Now, what this says is that *just the first component of our new PCA matrix* accounts for nearly 65% of the variance of the dataset. Then, just two components account for over 70% together, and so on. We see that at 28 components (all of them), we have 100% of the variance, which is to be expected. Now, we can plot the first two components of each case against each other.

When we do this, we get:

Chart, scatter chart

Description automatically generated

We see some clustering mostly in the middle/lower-left, and then a bit of scattering otherwise. Note that the axes are *not* the original, so we do not want to read too much into the values. What we are looking for is simply points close together.

Here are the first three components, accounting for just over 75% of the variance:

Chart, scatter chart

Description automatically generated

Again, in three dimensions, we can see some clustering. So, clustering algorithms might yield some information, and they often agree with our regressions, but because the data is not in neat clusters, we should be cautious about them. Each regression has a corresponding PCA decomposition and plotting in the Python script, which we have not included in the appendices, as they are all quite similar, but can be found in the Python script for the curious.

Now, how can we determine how many clusters we should attempt to build? Well, for this, we can use a *Silhouette Score*. We plot the silhouette score versus number of clusters and will choose the number of clusters with the highest score.

For the 28 factors together, the silhouette score versus number of clusters looks like:

Chart, line chart

Description automatically generated

We clearly see that two clusters are preferred by this metric. Now, we run the clustering algorithms and compare their results.

Here are a few of the first lines/cases:

GCI Score: 4.0 Group: 0

GCI Score: 4.0 Group: 0

GCI Score: 3.0 Group: 1

GCI Score: 4.5 Group: 0

GCI Score: 2.0 Group: 1

It appears that our clustering is grouping higher values into group 0 and lower ones into group 1. Note that the values/names of the groups mean nothing, and if we rerun the algorithm, it might, due to randomizations in where it starts looking, give different names to the clusters, but we always get the same members in the same cluster, which is what we’re interested in. Now, we can check our assumption about how the algorithm is dividing the cases. If we divide our data right at a Global Clinical Impression value of 3, then we find that our cluster has correctly put the “high” values together and “low” value together with 92.31% accuracy.

Now, to determine which factors the algorithm deems most important for each grouping, we can find where it put the center of the cluster. This represents the *prototypical member of each group*. So, we can see what factors the clustering algorithm has decided best represent a “high” value group member by how far along the “axis” of each factor it placed the cluster center, which, as stated above, are Engagement, Cognitive, and Gratitude. So, a member of the “high” value group is expected to have a high score for these three. Running this analysis for each set of factors used in the regressions gave us our results reported earlier. We found that when the number of factors was much higher, there was a bit more disagreement in our clustering versus regression algorithms, but with lower numbers of factors, they often agreed on which factors were most important and in order of importance. Again, this is interesting, but as we see in the plots above, the data isn’t quite as clustered as we’d like, so we should be careful ascribing too much importance to the clustering results, but they can be informative, nonetheless.

Finally, to see if any relationship exists between the demographic variable Residential Status and the Elder Identity Revision factor Competence, we simply relabel all non-H labels as “Not-H” to provide us with two more numerous groups of Residential Status. We then run a Fisher’s Exact Test to determine if the two variables (Residential Status and Competence), are independent.

The test is determined by creating a table such as:

|  |  |  |
| --- | --- | --- |
| Residential Status vs Competence | Low Competence ( | High Competence |
| Not H | 10 | 7 |
| H | 1 | 34 |

We then use calculate Fisher’s Exact p-value, which gives . With such a small value, we *reject* the notion that they’re independent and state that they *are*. So, there is a very likely a genuine relationship between Residential Status and Competence.

**References**

1. “Likert Scale Definition, Examples and Analysis | Simply Psychology.” *Www.simplypsychology.org*, [www.simplypsychology.org/likert-scale.html](http://www.simplypsychology.org/likert-scale.html).
2. Formplus Blog. “What Is Ordinal Data? Examples, Variables & Analysis.” *Formpl.us*, Formplus, 10 Oct. 2019, [www.formpl.us/blog/ordinal-data](http://www.formpl.us/blog/ordinal-data).
3. “Parametric and Nonparametric Statistics.” *Www.ibm.com*, [www.ibm.com/docs/en/db2woc?topic=procedures-statistics-parametric-nonparametric](http://www.ibm.com/docs/en/db2woc?topic=procedures-statistics-parametric-nonparametric).
4. “Correlation.” *Www.jmp.com*, [www.jmp.com/en\_ca/statistics-knowledge-portal/what-is-correlation.html](http://www.jmp.com/en_ca/statistics-knowledge-portal/what-is-correlation.html).
5. Frost, Jim. “Spearman’s Correlation Explained.” *Statistics by Jim*, 29 Mar. 2021, [www.statisticsbyjim.com/basics/spearmans-correlation/](http://www.statisticsbyjim.com/basics/spearmans-correlation/)
6. **Appendices**
7. Advanced Methods Section

This section will essentially be a re-write of the Models and Methods section, but with a more technical/mathematical focus.

We begin with the Spearman’s, defined as [1]:

Here just means that has been converted into a rank variable. Luckily, our variables are already rank variables, so we needn’t worry about this step, and our equation simplifies to:

From here, we can use the SciPy addition to the Python package NumPy to calculate the correlation and p-values of our pairs.

Now that we know that each of our correlations is significant at the or greater level, we can proceed to reject the null hypothesis that there is no correlation for each pair. Then, we visualize each correlation using heatmaps.

To determine whether the investigators scored more consistently as the project progressed, we simply looked at the scores given by each investigator for each factor or Global Clinical Impression, call them and , and plotted versus ( ranging over the 52 cases) to give us a nice visualization of when they agreed on the grading of case and when not. We interpreted the results in the manner described above.

For our multiple linear regressions, we first state the general form of a model fit with regressors:

In the above, is interpreted as the intercept and each , is the coefficient of fit for the model for regressor , and is the error or residual term. We are looking for a model with the lowest sum of squared residuals/errors, which we define as [2]:

However, as mentioned above, two of our regressions have very large sets of factors, and the number of models possible for a set of factors is , or the size of the powerset of minus 1 (we note here that the comes from that fact that the powerset returns *all* subsets, including the empty set, which we do not want to use, so we remove it). When is large, as we have and , we cannot, in reasonable time with the resources available to us, check *every* subset, as has size . So, we must choose some criteria by which we might perform a stepwise search.

Before we determine which of the various criteria we will use, we need to think about the number of data instances and the number of regressors. We have only 52 cases, so 52 data points. Attempting a *Backwards Stepwise Search* method with a model fit with 23 or 28 regressors will yield an *underfit* model in the beginning, and there isn’t much we gain from this. So, we will begin the other way, with a single factor and move from there, giving us a forward stepwise search.

Our criteria will be Akaike Information Criterion(AIC), defined as [3]:

In the above,  is the maximum value of the likelihood function for the model, roughly how well the model reproduces the data [4], and is the number of factors/regressors. Thus, we see that a better fit reduces AIC while increasing the number of regressors increases it. We can thus use this metric to build models that reproduce the data well while being simpler.

Now, for our two regressions with large , we run a forward stepwise search, beginning by fitting each model with only a single regressor, keeping only the model with the lowest AIC. We also mention this model as the best single-regressor fit, as this is interesting information itself. Then, we begin with that model, fit another model with that regressor and each of the non-used ones, and check if any of our two-regressors models have lower AIC. If so, we take the model with the lowest AIC, then begin fitting for a third regressor, and so on until arrive at a situation where all the models from the current round of fitting have higher AIC than the lowest AIC model from the previous round of fitting. When this occurs, we go with the last model with the lowest AIC.

After fitting, we run a histogram and QQ-plot, hoping to see a normal distribution in the histogram and all the points falling within a confidence bounds in the QQ-plot.

If we have a good fit, we can move on. If not, we find the offending case, remove it, refit, replot, and so on until we arrive at a good fit.

Looking at the output of our well-fit model, we hope to find that all the regressors included have a p-value near 0, with 0.05 being our cutoff for statistical significance. If we have regressors that fail to meet this, we can run an AIC versus stepwise step to determine when we begin to encounter diminishing returns for AIC. As in, though we are getting slightly lower AIC values by continually adding regressors, we run the risk of those factors not being statistically significant. So, we trade simple models with regressors all statistically significant for slightly higher AIC. After visualizing AIC versus stepwise step, we see an “elbow”, where the gain from adding a regressor to the model stops providing much benefit for AIC. We can then choose to stop our forward stepwise search at that number of regressors. When we do so, we find models that are well-fit, have statistically significant regressors, *and* are easily interpretable.

This last point is important, as with large numbers of regressors which are strongly correlated, we begin to find regressors with *very* small or *negative* coefficients, even when we expect them to be positive. Thus, we must keep interpretability in mind and how strongly correlated factors can muddle what the model is telling us.

For the smaller sets of factors, we can perform an exhaustive search, wherein we iterate over all possible subsets of (except the empty set) and choose the one with the lowest AIC. From there, we check for fit, remove outliers, and so on again until we arrive at a well-fit model. We then see if we can reduce the number of regressors with our AIC versus stepwise step plot. To determine the effect of Age, we add Age as a main effect and the interaction terms, AgeFactor for each factor *to the final models*, remembering to keep everything scaled to 1-to-5. We then run the exact same procedure described above with forward stepwise searching, *not exhaustive search*, and see if Age or an interaction term makes into our final model after all our model simplification procedures.

Of note, in two cases, the regression between all non-Meta-Factors as regressors and Composite Meta-Factor Score as the outcome and when determining the role of Age in the regression with Elder Developmental Tasks as regressors and Composite Meta-Factor Score as the outcome, we ended up with models with negative coefficients but with *all* regressors with a low enough p-value. At this point, we then ran a regression with the factor with a negative coefficient as the outcome and the other factors as regressors. The goal was to see if the other regressors could account for much of the value the regressors in question added to our original models. We determined this by a high . In both cases, we found this, allowing us to eliminate the regressor with a negative coefficient and to feel confident that we weren’t losing much in removing the regressor from final model, either stopping the forward stepwise search before hitting that regressor (in the first case mentioned in paragraph) or just removing it from out final model’s regressor set (in the question of Age’s interaction in the second case).

Now, as another method to determine which of the factors are most important for Global Clinical Impression and Composite Meta-Factor Score, we will see what our clustering algorithm has to say. To determine if clustering algorithms are appropriate, we will attempt to visualize our data to look for clustering there, which would give us reason to believe a clustering algorithm might uncover interesting relationships in the data. However, with any data matrix with more than two or three dimensions (two or three factors, in this case), this isn’t feasible. So, we will run Principal Component Analysis and plot the results from that.

To do this, we use *Singular Value Decomposition* to decompose the matrix of *all* variables/factors into:

Here, is a diagonal matrix of the square root of the eigenvalues of or . We call these and order them in descending order. is a matrix whose columns give the coefficients for a linear combination of the original variables to create *uncorrelated* variables. is a matrix whose columns are the values of the uncorrelated random variables created by linear combinations defined by the columns of divided by .

We then truncate our original to . Here, we choose the first eigenvalues to reduce the complexity of our , as now only has diagonal entries up to , and elsewhere.

This will be the best approximation to the original and if we choose or , we can plot this to look for clustering.

We find some clustering for each set of factors using PCA or not, but some significant scattering, as well. Thus, we will say that clustering will provide some information, but we caution putting too much weight into the outputs. In fact, we will find that when the number of factors is small, our clustering agrees *very* well with our regressions, but there are marked differences when is large.

We now provide a brief overview of K-Means Clustering, our algorithm of choice. We chose this because it is *unsupervised*, meaning that, though *we* know what good choices for clusters might be, the algorithm doesn’t. We hope that the algorithm’s results will match our intuition and the regressions’ results.

To begin, we will describe the algorithm. Set the number of clusters to , which we will describe how to choose momentarily. Now, assign a random vector of dimension , with being the number of dimensions of the data, to be the center of each cluster. These will be called and we collect them into , a set of vectors. From this, our goal is to find an assignment of data points to clusters, as well as the set of centers, , such that the sum of the square distances between the points and their closest centers is a minimum.

We now define by if data point is assigned to cluster and 0 if that data point isn’t in cluster . However, how do we choose our initial assignments?

We perform the following:

In words, the initial assignment of a datum is simply whatever center is closest by chance. Now, with our held fixed, we can optimize the set by minimizing the cost function of :

Note that is just the Euclidean distance between and . Now, we can take the derivative of each with respect to to get:

We set this equal to and solve to get:

Now that we have a *new* set of cluster centers, we reassign our data points, determining new values of , re-solve for the optimal vectors in , and continue until we are no longer seeing any data point assignment changes or until we reach a pre-determined number of iterations.

However, now that we have our algorithm in place, how do we determine to begin with?

We choose the silhouette score, a mean of the silhouette coefficients. Each datum recieves a silhouette coefficient, defined by [5]:

In the above equations, is the cluster data instance belongs to and is any other cluster it is not in. So, is the mean distance between data instance and *all other instances in the same cluster*, and is the *minimum extra-cluster distance*, or distance between and all the other data instances in the *closest* cluster.

Now, we take the mean of all the silhouette coefficients, , and *that* is the silhouette score for that number of clusters. We can then fit several different values for , calculate a silhouette score for each, and choose the highest one.

From here, we attempt to see if *we* can see any pattern in the clusters and we can check the accuracy of the clusters by this. In the example given in the Models and Methods section, we saw that it appeared as if the algorithm with only two clusters in 28 dimensions was clustering “high” and “low” Global Clinical Impression scores together. Running a check, it appeared to so more than 90% of the time. We then look at the vectors to find the centers of the clusters. From there, we take the of the “high” cluster and find which dimensions of have the highest value, corresponding to which axes the algorithm placed the center furthest, which we interpret as being the most important for that cluster.

Finally, to determine whether there is a relationship between Residential Status and Competence, we set up our table as above and us SciPy’s in-built function to our score. We begin with the following table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Variable 1, option 1 | Variable 1, option 2 |  |
| Variable 2, option 1 | A | B | A+B |
| Variable 2, option 2 | C | D | C+D |
|  | A+C | B+D | A+B+C+D=n |

We then calculate [6]:

If this is smaller than our level of significant, we can reject , which is that the two variables are independent, and support the idea that there is an association with them, which is what we found for Residential Status and Competence.

1. References for Appendix A
2. Wikipedia Contributors. “Spearman’s Rank Correlation Coefficient.” *Wikipedia*, Wikimedia Foundation, 7 Sept. 2019, www. en.wikipedia.org/wiki/Spearman%27s\_rank\_correlation\_coefficient.
3. “Sum of Squares - Definition, Formulas, Regression Analysis.” *Corporate Finance Institute*, corporatefinanceinstitute.com/resources/knowledge/other/sum-of-squares/.
4. Wikipedia Contributors. “Akaike Information Criterion.” *Wikipedia*, Wikimedia Foundation, 4 Sept. 2019, en.wikipedia.org/wiki/Akaike\_information\_criterion.
5. Bevans, Rebecca. “Akaike Information Criterion | When & How to Use It.” *Scribbr*, 26 Mar. 2020, [www.scribbr.com/statistics/akaike-information-criterion/](http://www.scribbr.com/statistics/akaike-information-criterion/).
6. Wikipedia Contributors. “Silhouette (Clustering).” *Wikipedia*, Wikimedia Foundation, 9 Aug. 2019, en.wikipedia.org/wiki/Silhouette\_(clustering).
7. Wikipedia Contributors. “Fisher’s Exact Test.” *Wikipedia*, Wikimedia Foundation, 5 Sept. 2019, en.wikipedia.org/wiki/Fisher%27s\_exact\_test.
8. Heatmaps
9. All factors, Global Clinical Impression, and Composite Meta-Factor

Background pattern

Description automatically generated with medium confidence

This map is quite small to accommodate everything, and further, the names have been shrunk to just the first four letters to fit.

Most can be easily determined, but we provide a list (in the same order as they appear in the heatmap) to help pick them out:

['Gratitude: Average', 'Active Habits: Physical, Mental, Social: Average', 'Purpose, Meaning: Average', 'Resilience (including post-traumatic growth): Average', 'Self-Acceptance, Self-Esteem: Average', 'Generativity: Average', 'Keeper of Meaning: Average', 'Ego Integrity: Average', 'Controllable Protective Factors: Only factor that goes up to 7: Average', 'Reliable Attachment Figure(s): Average', 'Healthy Grieving: Average', 'Sense of Humor: Average', 'Tragic Optimism (realistic but hopeful): Average', 'Mood: Average', 'Anxiety: Average', 'Energy: Average', 'Cognitive: Average', 'Engagement: Average', 'Candor: Average', 'Open System: Average', 'Narrative Coherence: Average', 'Transparency: Average', 'Youthfulness: Average', 'Control: Average', 'Competence: Average', 'Continuity: Average', 'Consciousness: Average', 'Connection: Average', 'GCI: Average', 'Meta-Factor Total']

1. Figure 2: Meta-Factors

Chart, treemap chart

Description automatically generated

1. Figure 3: Elder Identity Revision

Chart, treemap chart

Description automatically generated

1. Figure 4: Elder Developmental Tasks

Graphical user interface

Description automatically generated

1. Figure 5: Elder Developmental Tasks and Elder Identity Revision

Chart

Description automatically generated

1. Grading Difference Plots
2. Meta-Factors

Chart, histogram

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generated

1. Elder Developmental Tasks

Chart, histogram

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generated

1. Controllable Protective Factors

Graphical user interface, chart

Description automatically generated

1. Other

Graphical user interface, chart

Description automatically generatedChart, line chart

Description automatically generatedChart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generated

1. Interview Presentation Variables

Graphical user interface, chart, application

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generated

1. Elder Identity Revision

A picture containing text, athletic game, screenshot

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generated

1. Global Clinical Impression

Chart

Description automatically generated

1. Multiple Linear Regression
2. All Factors as Regressors and Global Clinical Impression as Dependent Variable

Individual Regressors (Top 3):

Graphical user interface, text, application

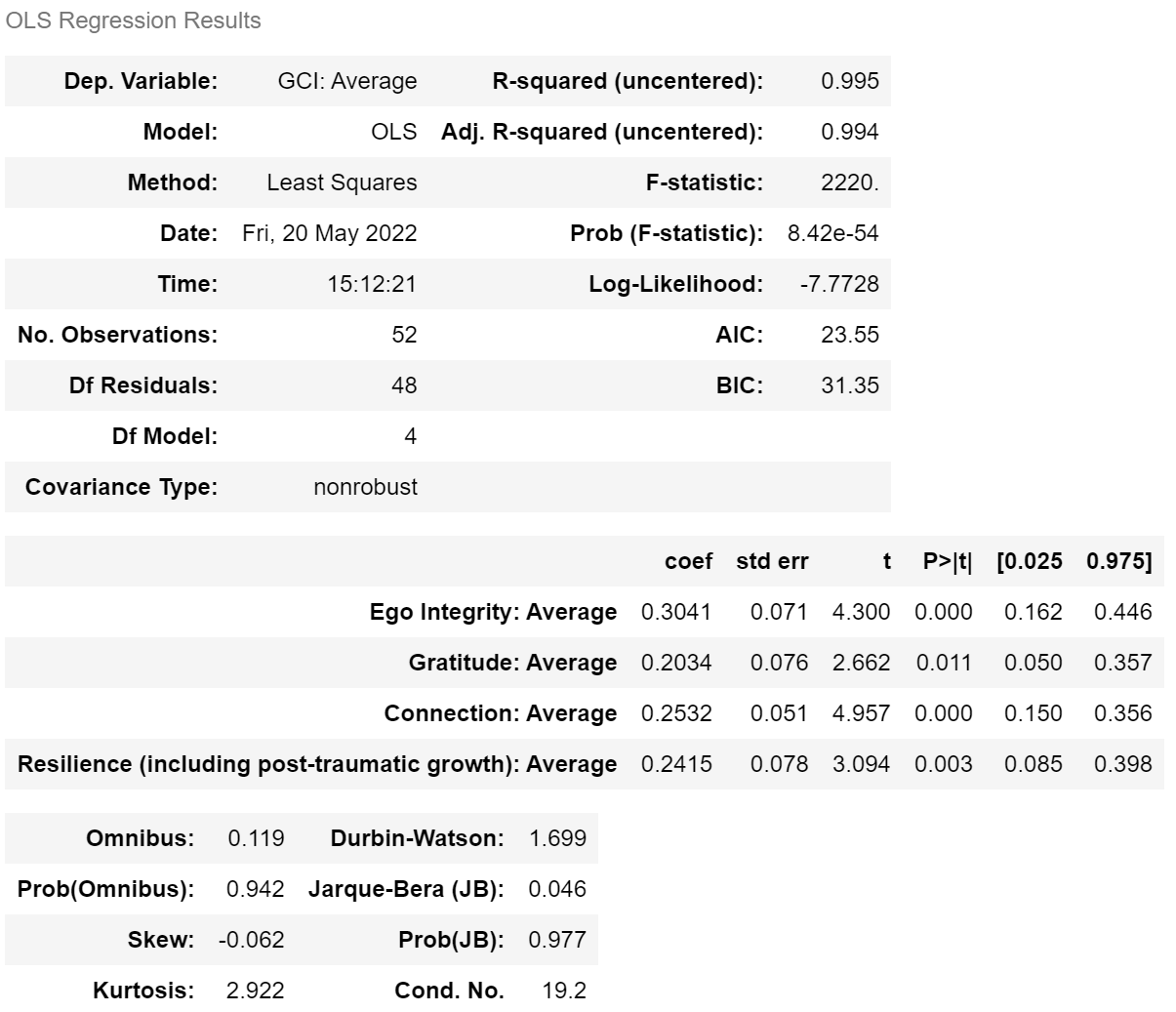
Description automatically generated

Final Three-Factor Model:

Table

Description automatically generated

Final Four-Factor Model:



1. All Non-Meta-Factors as Regressors and Composite Meta-Factor Score as Dependent Variable

Individual Regressors (Top 3):

Graphical user interface, application

Description automatically generated

Final Model:

Table

Description automatically generated

1. Meta-Factors as Regressors and Global Clinical Impression as Dependent Variable

Individual Regressors:

Table

Description automatically generated

Final Model:

Table

Description automatically generated

1. Elder Identity Revision as Regressors and Global Clinical Impression as Dependent Variable

Individual Regressors:

Table

Description automatically generated

Final Model:

Table

Description automatically generated

1. Elder Developmental Tasks as Regressors and Global Clinical Impression as Dependent Variable

Individual Regressors:

Table

Description automatically generated

Final Model:

Table

Description automatically generated

1. Elder Identity Revision as Regressors and Composite Meta-Factor Score as Dependent Variable

Individual Regressors:

Table

Description automatically generated

Final Model:

Table

Description automatically generated

1. Elder Developmental Tasks as Regressors and Composite Meta-Factor Score as Dependent Variable

Individual Regressors:

Table

Description automatically generated with low confidence

Final Model:

Table

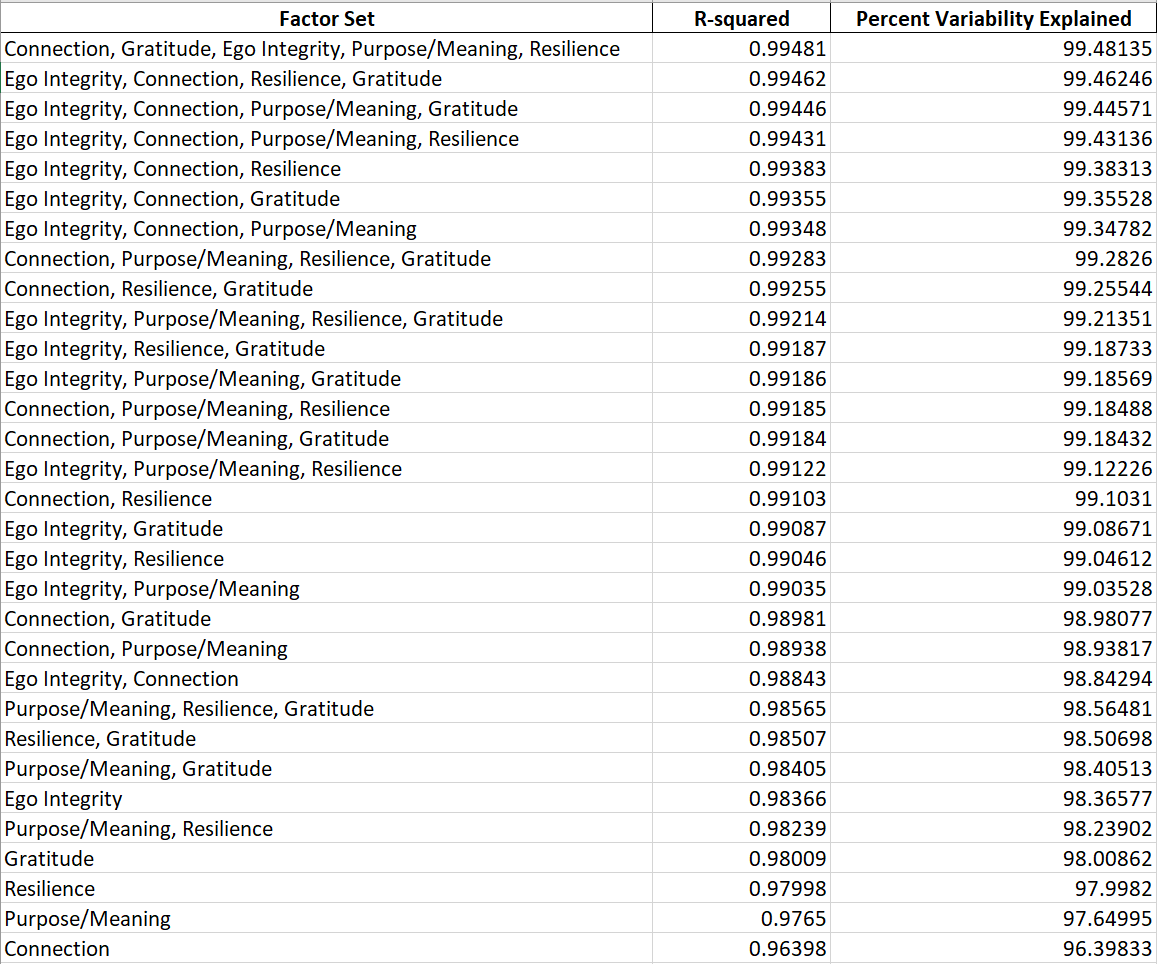
Description automatically generated

1. Elder Developmental Tasks factors from previous final model *and* Age as Regressors and Composite Meta-Factor Score as Dependent Variable to determine the effect of Age:

Table

Description automatically generated

1. Multiple Linear Regressions with All Combinations of Ego Integrity, Connection, Gratitude, Resilience, and Purpose and their values (also included as Excel sheet):



1. Online Repository

The full code and can be found at:

https://github.com/ClarkUCD/Aging-Wisely-Project