Job Performance

Gabe Adams and Shelby Taylor

January 2019

1 Introduction

When employees are happy, they work harder, leading to better job performance and success for the company. The primary goal of this analysis is to help universities understand the impact of employee satisfaction on job performance. We are also curious about the relationship between age and job performance, specifically whether or not older professors care less about their jobs than younger professors. Some students believe the smarter the professor, the less dedicated they are to teaching, thus we want to understand the relationship between IQ and job performance. Lastly, we want to know whether job satisfaction and well-being lead professors to stay at the university longer.

To answer our research questions, we will have to fit two different models, one with job performance as the response, and the other with tenure as the response. We will describe these models in more detail later in the report.

Our data consists of information from 480 current employees of a large university. For each employee, various measurements were to be recorded: a randomly assigned ID number, employee age, the number of years the employee has been employed at the university, a measure of the employee's happiness, a measure of the employee's job satisfaction, a measure of the employee's job performance, and the employee's IQ.

Looking at the data, we see there are a lot of missing values. There are 160 observations missing data on Well Being, 160 missing data on Job Satisfaction, and 64 missing data on Job Performance. In fact, of the 480 total observations, only 131 observations have complete data. This is an issue we will address later in the analysis.

We will disregard the ID variable for the rest of this analysis as it is randomly assigned and should have no bearing on the other variables.

2 Model Selection

For our analysis, we will assume all of our variables together follow a multivariate normal distribution. We will justify this assumption later. Because of this assumption, we will be able to account for our missing data and use a couple of multiple linear regression models to answer our research questions. One model will have job performance as the response variable, with all the other variables serving as explanatory variables. The other model will have tenure as the response variable, with the other variables serving as explanatory variables. We will describe these models in more detail later in this report.

Because we have so much missing data, we will need to use imputation methods to "fill in" our missing data before fitting our models. If we simply threw out any observations with missing data, we would be left with only 131 observations with which to fit our models. This would not be good because it could introduce bias and we would be throwing away a lot of potentially useful information.

Instead of simply throwing out incomplete cases, we will use an imputation method called stochastic regression imputation. This will help us avoid biased estimates and will, overall, serve our purposes well. This method works as follows:

1. We will use a joint multivariate normal distribution to fill in missing data. This distribution will have a mean vector of the means (of data for which we have complete cases) for each of the variables. The covariance matrix for this distribution will be calculated from the complete data as well. Using

this distribution, we will derive conditional distributions (response = missing variables, explanatory = non-missing variables). Using these conditional distributions, we will draw values for each of the missing data points.

2. We will then fit our two models (one for each response) to the new "complete" data set. Using the fitted models, we will store the estimates for the coefficients (β's) and their corresponding standard errors, as well as the R^2 values of each model.

3. We will calculate a new mean vector and covariance matrix for our joint multivariate normal distribution from the new "complete" data set. We will then draw a new set of values for each of the missing data using appropriate conditional distributions.

4. We will then repeat steps 2 and 3 a large number of times (randomly generating new data to fill in the missing data, fitting models, storing values).

5. After several thousand iterations of steps 2 and 3, we will average our estimates, standard errors, and R^2 values for our two models. We will also be able to look at various measures of uncertainty regarding our estimates.

Again, for the goals of our analysis, a couple of multiple linear regression models will be appropriate tools to analyze our data. One model will have job performance as the response, while the other model will have tenure as the response.

When we want to understand how job satisfaction, age, and IQ impact job performance, we use the following model (Model 1):

$$JobPerformance_i = \beta_0 + \beta_1(Age)_i + \beta_2(Tenure)_i + \beta_3(WellBeing)_i + \beta_4(JobSatisfaction)_i + \beta_5(IQ)_i + \epsilon$$

$$\text{where } \epsilon \overset{iid}{\sim} \text{N}(0,\, \sigma^2)$$

In this model, β_0 represents the average job performance measure if all of the explanatory variables (age, tenure, well being, job satisfaction, and IQ) are 0. Every other β represents the average increase in job performance when the corresponding explanatory variable increases by 1. For example, β_1 represents the average increase in job performance when age increases by 1, holding all other variables constant.

This model will help us answer our first three research questions. By looking at the estimates and confidence intervals of the coefficients in this model, we will be able to see the effects of well-being and job satisfaction on job performance. By look at the coefficient for age, we will be able to tell whether or not older professors care less (and by how much) about their jobs than younger professors. We will also be able to determine whether or not the IQ of a professor has an effect on his or her job performance. If there is an effect, we will be able to quantify it.

When we want to understand how job satisfaction and well being affect tenure, however, we use the following model (Model 2):

$$Tenure_i = \beta_0 + \beta_1(Age)_i + \beta_2(JobPerformance)_i + \beta_3(WellBeing)_i + \beta_4(JobSatisfaction)_i + \beta_5(IQ)_i + \epsilon$$

$$\text{where } \epsilon \overset{iid}{\sim} \text{N}(0,\, \sigma^2)$$

In this model, each β_j (j = 1,...,5) represents the average increase in tenure when the j^{th} variable increases by 1. For example, β_2 represents the average increase in tenure when job performance increases by 1, holding all other variables constant. β_0 represents the average tenure if all of the explanatory variables are 0.

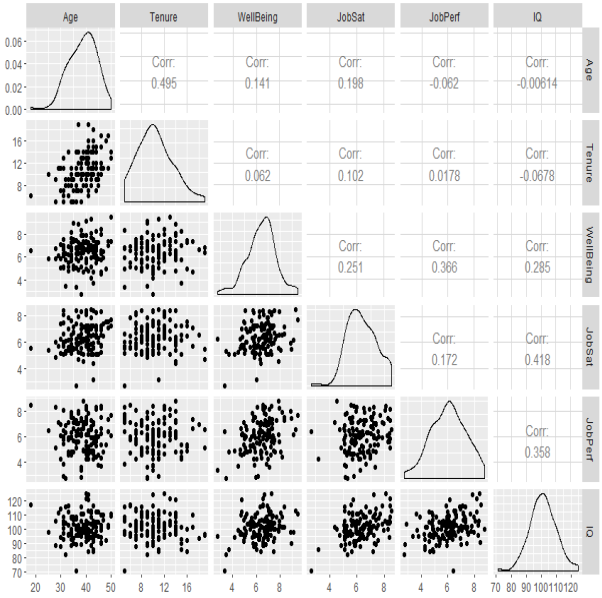
This model will help us answer our fourth research question. We will be able to look at the estimates and confidence intervals for the Job Satisfaction and Well Being effects to see how and if those two variables affect tenure.

In order for us to be able to use these models and the prerequisite imputation method, our data must follow a multivariate normal distribution. We will verify this through the use of various plots in the next section.

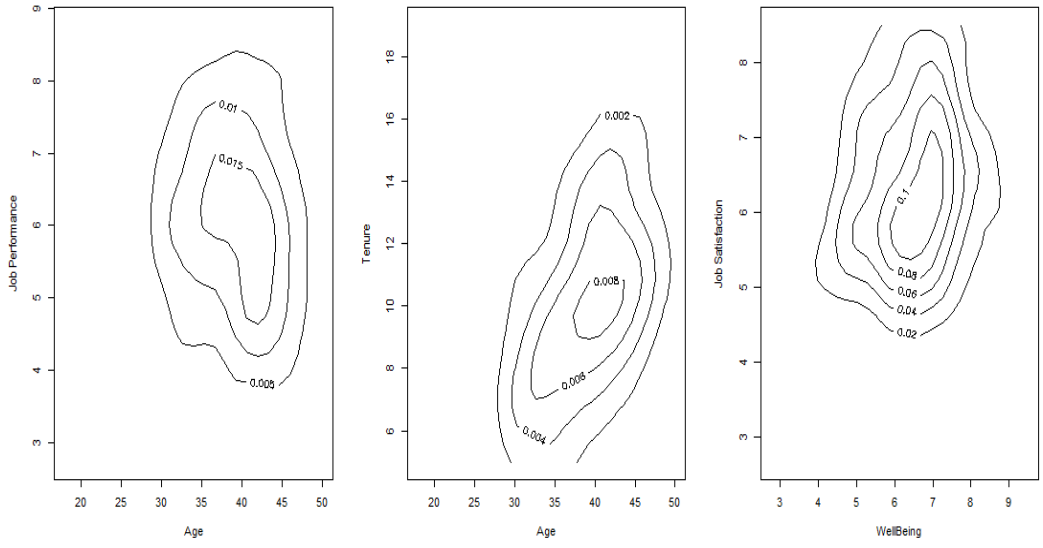
3 Model Justification and Performance Evaluation

Our models described in the previous section are only valid if the underlying data is multivariate normal. To verify this, we will look at various plots of the complete cases of our original data.

The scatterplots in the figure below indicate roughly linear relationships between all the variables in our data. Importantly, the density plots on the diagonal of the figure indicate each variable to be normally distributed.



The contour plots below confirm linear bivariate relationships between a small selection of variables (age by job performance, age by tenure, and well being by job satisfaction). Similar relationships can be found between other combinations of variables, but it is infeasible to show all of the corresponding contour plots here.

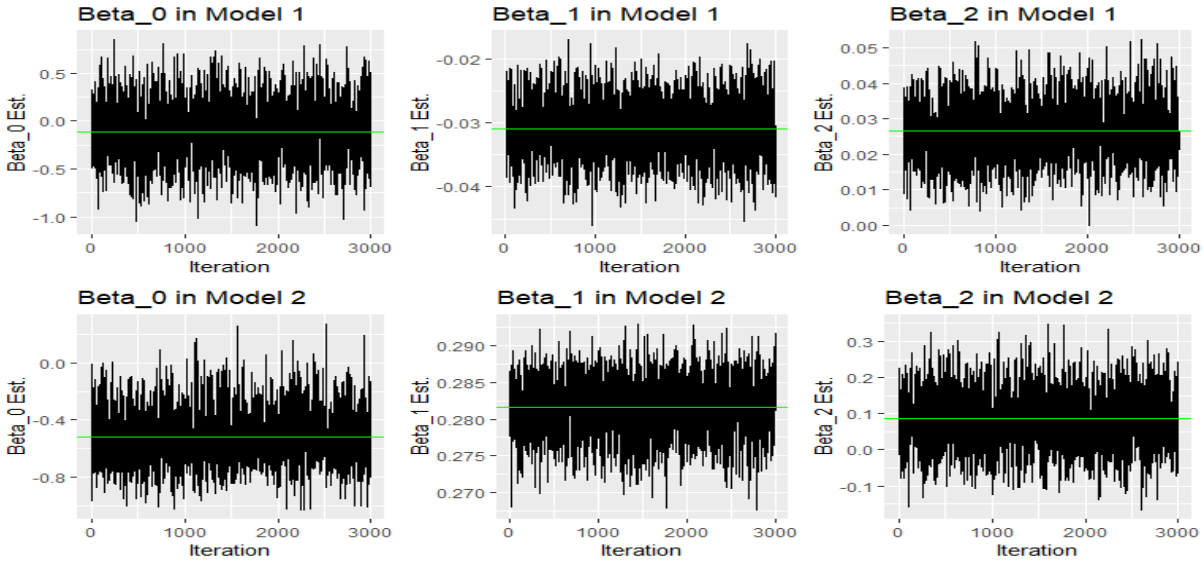


Because individual variables are distributed normally, and linear bivariate relationships are apparent between variables, we can reasonably conclude that the data follows a multivariate normal distribution. This

means that our imputation method is valid. It also implies the normality and equal variance assumptions for multiple linear regression hold. Independence is reasonable for two reasons. First, tenure and job performance of an individual probably does not depend too much on the tenure or job performance of another professor. Second, missing data is generated randomly, so independence is reasonable to assume.

The plots below show how the values of β_0, β_1, and β_2 (for each model) change as we iterate through the algorithm previously mentioned. Similar plots of the other β's are not shown here. We can see that the various values "converge." They mix well. There are no noticeable upward or downward trends. Thus, we can be confident in the estimated values of our β's.

The green lines show the means of the generated β's. We will use these means as our estimates for the corresponding coefficients in our models.



Our data are multivariate normal and the assumption of independence is reasonable. Our estimates for our β's appear to converge well. Thus, we can feel comfortable continuing with our analysis.

The fraction of missing information (FMI) measures the percent variability in our betas due to how we filled in the missing data. The table below gives fractions of missing information for each variable in our two models:

Model 1:

β_0(Intercept)	β_1(Age)	β_2(Tenure)	β_3(WellBeing)	β_4(Job Satisfaction)	β_5(IQ)
0.1604	0.1469	0.1482	0.2788	0.4447	0.2078

From the table above, we see the beta with the most variation due to our imputation method is Job Satisfaction, with an FMI of 0.4447.

Model 2:

β_0(Intercept)	β_1(Age)	β_2(Job Performance)	β_3(WellBeing)	β_4(Job Satisfaction)	β_5(IQ)
0.0120	0.0252	0.2961	0.2921	0.1484	0.0674

From the table above, we see the beta with the most variation dues to our imputation method is Job Performance, with an FMI of 0.2961.

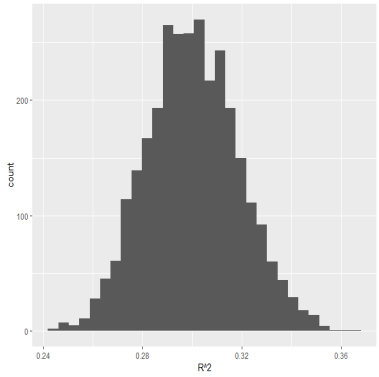
Overall, our imputation method does have a moderate amount of bearing on the variability of our estimates. It is not too much, however, so we can feel confident in our analysis.

4 Results

The table below gives estimates for each effect in Model 1, along with 95% confidence intervals. For example, as Well Being increases by 1, holding all other variables constant, Job Performance increases by 0.416 on average, and we are 95% confident it increases between 0.313 and 0.519 on average.

Effect	**Point Estimate**	**95% Confidence Interval**
Intercept	-0.109	(-1.556,1.337)
Age	-0.031	(-0.054, -0.008)
Tenure	-0.027	(-0.012, 0.065)
Well Being	0.416	(0.313, 0.519)
Job Satisfaction	-0.055	(-0.173, 0.064)
IQ	0.048	(0.033, 0.062)

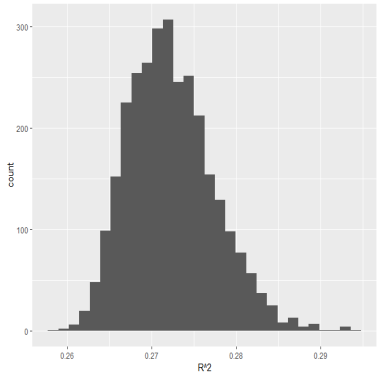
Based on our iterative process, we estimate R^2 for Model 1 to be 0.30, with 95% confidence it is between 0.27 and 0.34, meaning 30% of the variation in Job Performance can be explained by the variables in the model. The histogram below shows the distribution of R^2 for Model 1.



The table below gives estimates of each variable in Model 2, along with 95% confidence intervals. For example, as Well Being increases by 1 holding all other variables constant, Tenure increases by 0.262 on average, and we are 95% confident it increases by between 0.0002 and 0.524 on average.

Effect	**Point Estimate**	**95% Confidence Interval**
Intercept	-0.517	(-3.861, 2.826)
Age	0.282	(0.235, 0.328)
Job Performance	0.086	(-0.198, 0.370)
Well Being	0.262	(0.0002,0.524)
Job Satisfaction	0.167	(-0.077, 0.411)
IQ	-0.032	(-0.067, 0.003)

Based on our iterative process, we estimate R^2 for Model 2 to be 0.27, with 95% confidence it is between 0.26 and 0.28, meaning 27% of the variation in Tenure can be explained by the variables in the model. The histogram below shows the distribution of R^2 for Model 2.



Our first model helps answer the first three questions posed in the analysis, while our second model helps us with the fourth question.

1. Well-being has a positive impact on Job Performance. As Well-being increases by 1, holding all other variables constant, Job Performance increases by 0.416 on average, with 95% confidence it increases between 0.313 and 0.519. Job Satisfaction, however, does not have a significant impact on Job Performance because 0 is contained in its confidence interval.

2. The student's hypothesis that older professors care less about their jobs than younger professors appears to be true, but not by much. As age increases by 1, holding all other variables constant, job performance decreases by 0.031 on average, with 95% confidence it decreases by between 0.008 and 0.054.

3. IQ has a positive impact on job performance. As IQ increases by 1, holding all other variables constant, job performance increases by 0.048 on average, with 95% confidence it increases between 0.033 and 0.062.

4. Well Being has a positive effect on tenure. Holding all other variables constant, as Well Being increases by 1, tenure increases by 0.262 on average, with a 95% confidence interval from 0.0002 to 0.524. Because 0 is contained in its confidence interval, it appears that job satisfaction does not have a significant effect on tenure.

5 Conclusions

Through stochastic regression imputation, we were able to complete the goals of this analysis by filling in missing data, fitting appropriate models, and performing inference. We discovered that well-being has a positive impact on job performance, while job satisfaction does not appear to have a significant effect. Older professors do appear to care slightly less about their jobs than younger professors. IQ has a positive relationship with job performance: as IQ increases, job performance generally increases as well. Finally, well-being has a positive effect on tenure, while job satisfaction appears to have no significant impact on tenure.

One potential shortcoming of this analysis is that, because we used stochastic regression imputation, our standard errors could be too small. This would cause the confidence intervals for our estimates to be too small. Further analysis could try different imputation methods to fill in missing data. We could then compare standard errors. Also, since all our data came from one university, we cannot infer our results to other universities or other workplaces. Future research could involve gathering data from a greater variety of workplaces.

6 Teamwork

Gabe worked a lot on the code for imputing the missing data while Shelby worked on the write-up, although we worked together on the majority of both writing and coding.