Determining the Effect of Happiness on Job Performance

Stephen Merrill
Nathan Bean

Brigham Young University

Febrruary 18, 2016

Background

- Relationship between employee performance and company success, including profitability
- 480 observations from employees of a large university
- 6 measurements: Age, Tenure, IQ, Job Performance, Job Satisfaction, Overall Well-Being

Data

- Missing values - 73% of employees are missing data
- Missing at random structure

	Job Satisfaction	Well Being	Job Perf
Missing Data Points	160	160	64
Percentage Missing	33.33%	33.33%	13.33%

Table 1: There are significant amounts of the employee score data missing from the 480 total observations.

Data

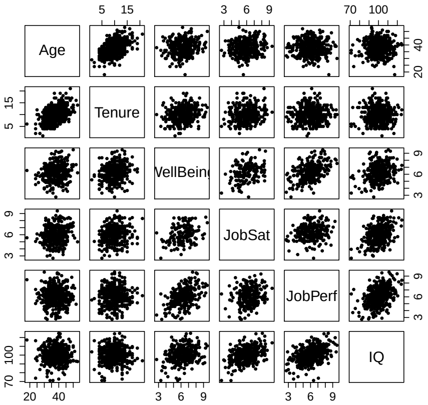

Figure 1: Matrix scatterplot of the six quantitative variables.

Goals

- Make inference on the effect these covariates have on Job Performance
- Report to the university the effect of employee happiness on performance

MLR Model

$$y_i = \beta_0 + \sum_{i=1}^{N} X_i \beta_i + \epsilon_i$$

$\epsilon_i \sim N(0, \sigma^2 I)$

- β_0: Intercept
- β_1: Effect of Age
- β_2: Effect of Tenure
- β_3: Effect of Well-Being
- β_4: Effect of Job Satisfaction
- β_5: Effect of IQ

β interpretation: For every one unit increase in ..., job performance increases by β. (Thus fulfilling goals of the study)

Multiple Imputation

- Iterative algorithm that takes advantage of the properties of the Multivariate Normal Distribution to "fill-in" the missing data points
- Handles any number of missing observations on any one employee
- Calculate and store β values and errors each iteration - pool these values later to get final results
- No model selection process - effect sizes and errors are calculated for each covariate

Trace Plots

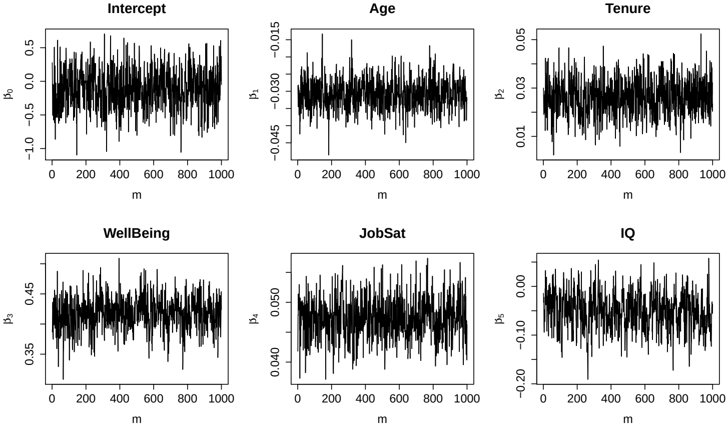

Figure 2: The trace plots show the convergence of the six β estimates.

Model Assumptions

Standard MLR assumptions still need to hold:

- Linearity
- Independence
- Normality
- Equal Variance

In addition, multiple imputation requires that the observed values be distributed Multivariate Normal ($MVN(\boldsymbol{\mu}, \Sigma)$)

Linearity

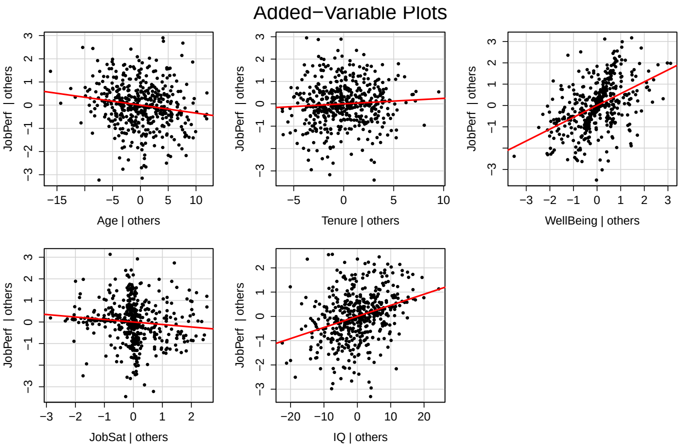

Figure 3: Added variable plots indicate that the linearity assumption is satisfied.

Bivariate Normality Assumption

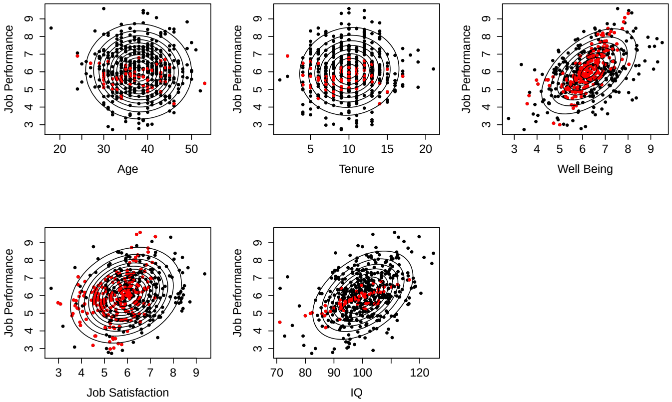

Figure 4: Each of the five predictors plotted against the response variable, Job Satisfaction. Red points indicate imputed values.

Normality of Residuals

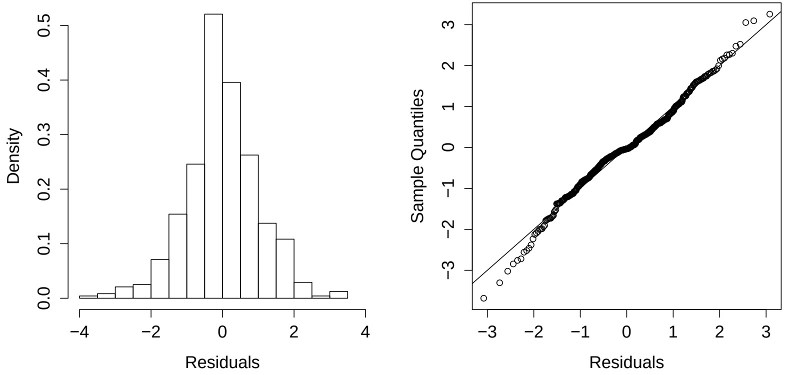

Figure 5: The histogram of residuals and Q-Q plot show no major concerns.

Equal Variance of Residuals

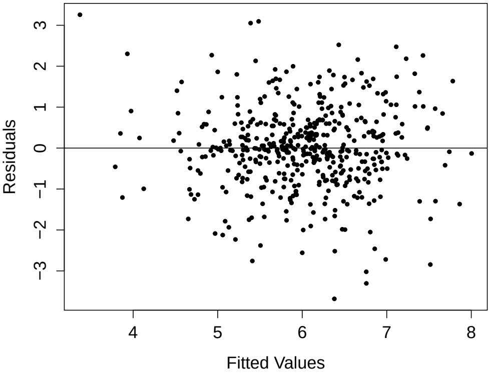

Figure 6: The residual plot shows no signs of heteroscedasticity.

Results

$R^2 = 0.3662$: 36.62% of the variation in job performance can be explained by age, tenure, well-being, job satisfaction, and IQ.

Variable	Estimate	CI Lower Bound	CI Upper Bound
Intercept	-0.114	-1.558	1.330
Age	-0.031	-0.053	-0.008
Tenure	0.026	-0.012	0.065
Well-Being	0.417	0.315	0.520
Job Satisfaction	-0.052	-0.168	0.064
IQ	0.047	0.033	0.062

Table 2: Estimates and 95% Confidence Intervals for Variables

Conclusion

- Looking at the confidence intervals, only age, well-being, and IQ appear to be significant (i.e., the confidence intervals do not contain 0).
- Since well-being is significant, we conclude that happiness is related to job performance. However, this may be a weak relationship.

Limitations and Further Studies

- Imputed values are predictions and not true values, causing potential bias from measurement error.
- Use of imputed values increases variation beyond sampling variation.
- For future studies, consider data that are not missing at random.
- Explore other variables, such as income and education level, that may relate to job performance.