

Prediction of time spent on the research administrative support requests

Amy Song

Client: Leslie Heal (UNC Research Finance Service Center)

December 05, 2022

Abstract

The Research Administration Service Center manages grants and supports research administration. The project aims to see if there are any statistically significant effects of attributes on the actual time spent at each level of work by using multiple linear regression analysis. Multiple linear regression models are built to predict the actual time spent on various tasks and then variable selections are done to find the best subset of variables to predict the time. The actual time spent on efforts completed for the pre-award and post-award stages is best predicted by the subsets of data collected from employees. For the pre-award and post-award tasks, or the service request tickets, attributes of both efforts and tasks show significance. Lastly, only the variables from effort show a significant effect on proposals while attributes of both efforts and projects show significance in predicting the actual time spent on projects. In conclusion, it is recommended to collect data from employees over a longer period for a larger number of tasks and projects for future analyses.

1 Introduction

The Research Administration Service Center at the University of North Carolina at Chapel Hill provides administrative support to the research teams and the University. The Center handles numerous administrative support requests for projects, including grant proposals, in both *pre-award* and *post-award*. The term *task* is defined as the service request ticket in the Research Administration Service Request (RASR) system. When a research team submits a proposal for a grant, their project is in the pre-award stage and the tasks performed for this project are correspondingly labeled as pre-award tasks. Similarly, the tasks related to projects in the post-award stage are labeled as post-award tasks. An *effort* is defined as a weekly task completed by an employee, and each effort is considered an experimental unit in the analyses. Efforts and the details, including the estimated time and work time spent, are reported by employees from the week of October 11, 2021, to December 21, 2021. Each effort is completed for a task, which may or may not be on the RASR system.

This project aims to quantify the effects of the different attributes of work on predicting the actual time spent at each level of work using multiple linear regression analysis. Using variable selection, this report presents the set of predictor variables at each level that results in the highest adjusted R-squared value and statistical significance of each predictor variable.

In each model considered in this project, the variables collected from employees on efforts, including estimated time, complexity score, and preference score, show statistical

significance on predicting actual work time. Some detailed attributes of tasks, including types and the total number of days, have shown significant effects on time spent for efforts and tasks. For projects in the post-award stage, only the status of account among other attributes of projects, along with variables from efforts, showed statistical significance in predicting the actual time spent on projects. In conclusion, it is recommended to collect similar data on efforts in a longer period of time for a larger number of projects and tasks in both pre-award and post-award for stronger results and future analyses.

The report is organized as follows. The data used in each part of the analyses are described in section 2. Section 3 discusses the results of the analyses, and section 4 explains the methods used in the analyses. The appendix has supplemental tables from the analyses.

2 Data

2.1 Efforts data

Efforts are reported by the employees who completed tasks in weekly basis. The data on efforts are collected from the week of October 11, 2021, to week of December 21, 2021. During this period, total of 5613 efforts are reported by 12 unique employees. Among these efforts, 517 are efforts related to pre-award tasks and 5096 are related to post-award tasks. The longer definition of each attribute of efforts is in Table 2.1 with the type of the variable.

Table 2.1: Descriptions of variables in efforts data

Variable Name	Variable Type	Description
data_type	Factor	Type of tasks: Pre-Award or Post-Award
task_id	Factor	Task ID in RASR System
employee	Factor	ID of unique employee
project_ipf	Factor	Project or proposal unique ID for the task
work_week	Date	Week the effort was worked on
dept_support	Factor	Department's support on the effort
estimated_time	Numeric	Estimated time of the effort
actual_time	Numeric	Actual time employee took to complete the effort
complexity_score	Numeric	Complexity score by the employee working on the effort: 1 to 10, easy to hard
preference_score	Numeric	Preference score by the employee working on the effort: 1 to 10, least to most preferred
work_week_iso	Integer	ISO week number of the week the effort was completed

Each effort has following attributes: estimated time in minutes, actual time taken in minutes, complexity score, preference score, the work week, and details of the task, including the task ID, presence of department support, and project ID, if exists. The variable `work_week` is replaced by `work_week_iso`, ISO week number in year, in the analyses to consider the work week as numeric value instead of factor value. This transformation of work week is done to consider how early or late in the year the task was completed impacts the actual time spent to finish the task.

2.2 Tasks data

Tasks can either be on the RASR system or not, but the tasks on the RASR system are used in the analysis for the client's interest to see if there is any linear relationship between the attributes from the RASR and actual time spent on tasks. A total of 384 tasks are used in the analysis, and these tasks have efforts that fall within their start and end date on the RASR system. In the task dataset, there are 16 pre-award tasks and 368 post-award tasks. Table 2.2 shows the longer definition of each variable of the tasks.

Table 2.2: Descriptions of variables in tasks data

Variable Name	Variable Type	Description
data_type	Character	Type of tasks: Pre-Award or Post-Award
task_id	Factor	Task ID in RASR System
rasr_start	Date	Started (created) date of task in RASR
rasr_end	Date	Ended (closed) date of task in RASR
RASR task time (days)	Numeric	Number of days the task was open for
# of RASR task touches	Numeric	Number of touches made when task open
Total estimated time (mins)	Numeric	Sum of estimated time reported in efforts for task
Total actual time (mins)	Numeric	Sum of actual time reported in efforts for task
Total # of efforts	Numeric	Total number of efforts made for the task
Avg. complexity score	Numeric	Average complexity score by the employee working on the effort: 1 to 10, easy to hard
Avg. preference score	Numeric	Average preference score by the employee working on the effort: 1 to 10, least to most preferred
Type	Factor	Type of task: Eform, Journal Budget Transfer, Journal ID, Proposal Number, Purchase Order, Voucher, Web Traffic, or None

The actual time, estimated time, average complexity score, and average preference score are calculated based on the efforts related to each task. Each task has the following attributes from the RASR system: the RASR ticket start and end dates, the number of RASR ticket touches, the total number of days the RASR ticket was open, and the type of task, including *Eform*, *Journal Budget Transfer*, *Journal ID*, *Order ID*, *Proposal Number*, *Purchase Order*, *Voucher*, and *Web Travel*. These attributes of the tasks from the RASR system are merged into the efforts dataset. However, 3018 out of 5096 efforts do not have those RASR-related values, indicating that these efforts are related to tasks not on the RASR system. The number of efforts associated RASR is discussed in results, section 3.1.

2.3 Proposals and Projects data

There are total of 69 proposals and 477 projects, e.g. grants, that efforts or tasks are completed on. Proposals have 10 unique attributes described in detail in Table 2.3, and Table 2.4 shows the longer description of attributes of projects.

Table 2.3: Descriptions of variables in proposal data

Variable Name	Variable Type	Description
proposal_number	Factor	Unique ID of proposal
status	Factor	Status of proposal: Proposal Approved (19), Not Funded (17), Unsubmitted (2), Funded (28), Award Received (2), PS Review (1)
proposal_type	Factor	Type of proposal: New (58), Non-Competing Continuation/Progress Report (1), Recurring Contract (2), Resubmission or Amendment (7), Supplement (1)
project_amount	Numeric	Amount of project grant in dollar value
award_admin_dept_no	Factor	Award administration department unique ID
award_type	Factor	Type of award: Contract (17), Cooperative_Agreement (2), Grant (22), Incoming_Subcontract (20), NIH_Grant (8)
prime_sponsor	Character	Name of prime sponsor
sponsor	Character	Name of sponsor
project_start_date	Date	Start date of the proposal
project_end_date	Date	End date of the proposal

Table 2.4: Descriptions of variables in proposal data

Variable Name	Variable Type	Description
project_ipf	Factor	Unique ID of project
account_status	Factor	Status of account: Active (348), Frozen / Inactive (127), Preliminary (2)
account_begin_date	Date	Start date of the project account
account_end_date	Date	End date of the the project account
final_rpt_date	Date	Final date reported on RPT reporting system
dept_no	Factor	Unique ID of department
pi_dept_no	Factor	Unique ID of principal investigator department

Total actual time, total estimated time, total number of tasks, count of work weeks, count of employees, average complexity score, and average preference score are calculated based on the efforts and tasks worked for each proposal and project.

3 Results

3.1 Efforts Analysis

A multiple linear regression model is fit to predict the actual time spent in both pre-award and post-award using these variables from efforts: the estimated time, complexity score, preference score, work week in ISO, and these variables from tasks: type of tasks, number of days on RASR system, and number of touches on RASR. However, as described in the data section, efforts can either be worked for tasks that are on RASR or not on RASR. To avoid the elimination of observations by fitting the model on the entire data, each stage of data is further divided into RASR efforts and non-RASR efforts.

First, the model is fitted to the pre-award efforts associated to RASR tasks. As shown in Table 3.1, the attributes from the RASR do not show any statistical significance in predicting the actual time spent.

Table 3.1: Summary of linear regression model of pre-award efforts associated to RASR tasks

	Estimate	Std. Error	Pr(> t)
Intercept	24.631	42.394	0.562
Department Support	26.644	8.126	0.001
Estimated Time	0.813	0.030	0.000
Complexity Score	-0.818	1.352	0.546
Preference Score	4.740	2.472	0.056
Work Week (ISO)	-0.633	0.816	0.438
Type: Proposal Number	0.396	22.120	0.986
Type: Voucher	15.096	43.551	0.729
Number of RASR Task Touches	-0.480	0.317	0.131
RASR Task Days	0.034	0.026	0.181

Multiple R-squared: 0.743, Adjusted R-squared: 0.735

With this result, the model is fitted to all of pre-award efforts using the attributes from efforts and the summary of this model is shown in Table 3.2.

Table 3.2: Summary of linear regression model of pre-award efforts

	Estimate	Std. Error	Pr(> t)
Intercept	29.380	28.140	0.297
Department Support	17.493	4.221	0.000
Estimated Time	0.882	0.022	0.000
Complexity Score	-0.230	0.950	0.808
Preference Score	1.974	1.680	0.241
Work Week (ISO)	-0.715	0.639	0.264

Multiple R-squared: 0.778, Adjusted R-squared: 0.776

From the summary of this first model, the complexity score, preference score, and work week in ISO have p-values higher than 0.05, failing to reject the null hypothesis of no significance in the parameter estimates. On the other hand, the estimated time and department's support resulted in p-values lower than 0.05, rejecting the null hypothesis of insignificance in the parameter estimates.

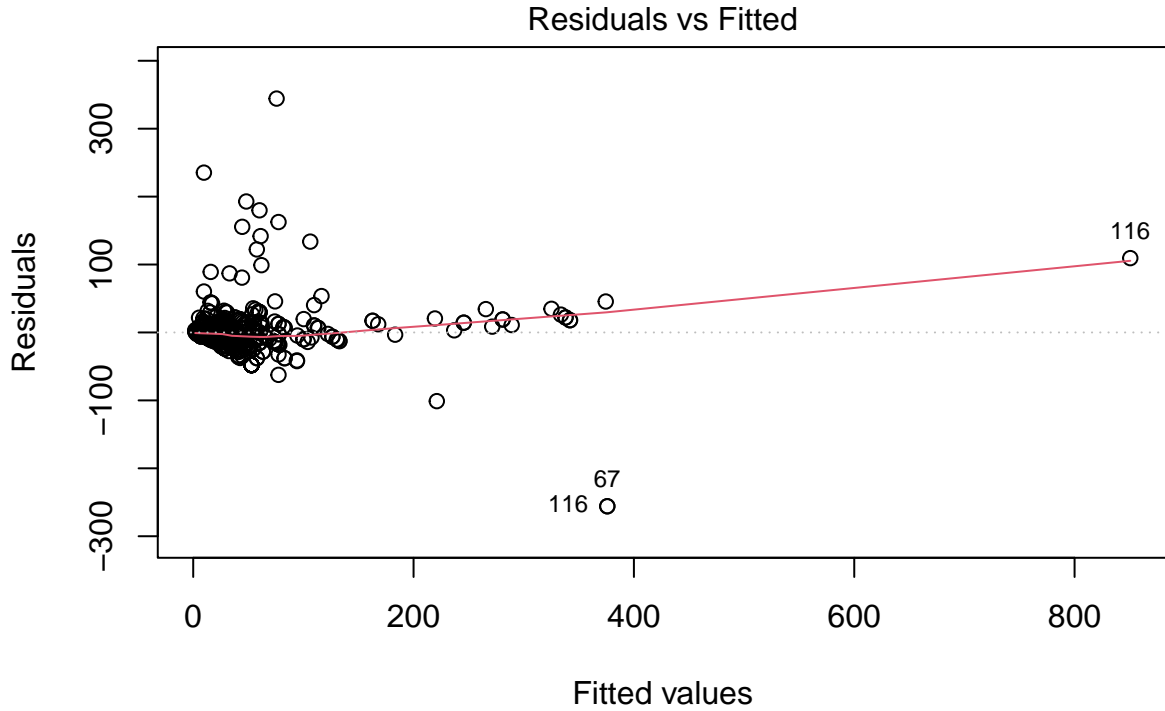


Figure 1: Residuals versus fitted plot of the model in Table 3.2

Furthermore, Figure 3.1 shows the residuals, the difference between the predicted value from the model and the observed value in the data, versus the fitted values. This plot identifies

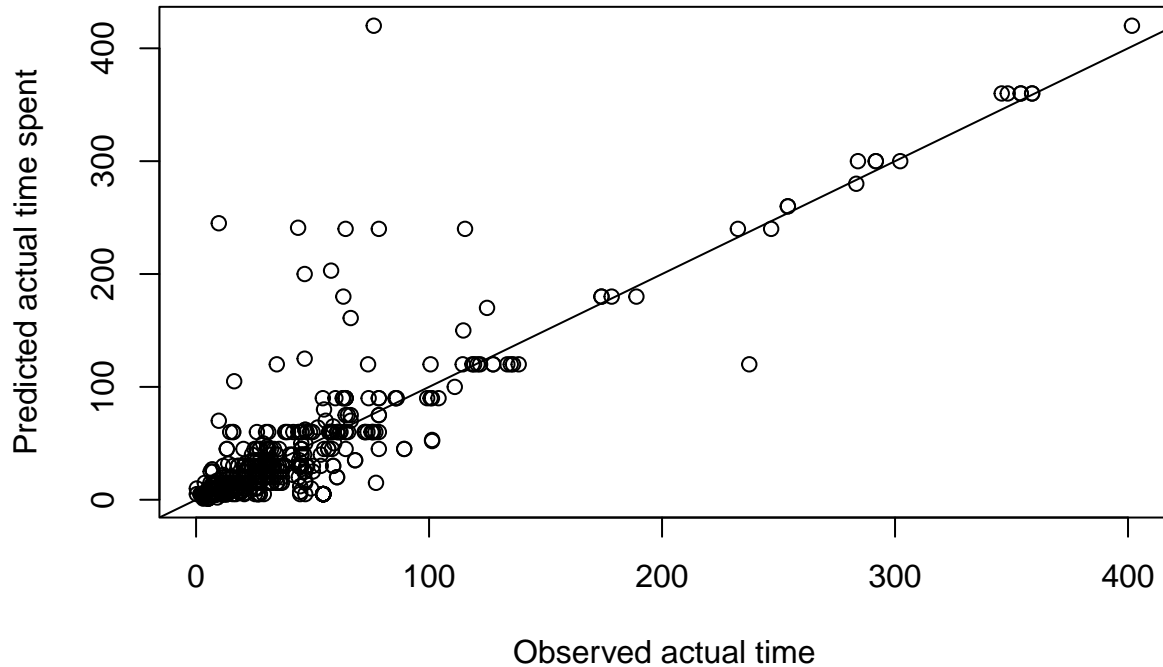
these observations as the outliers: point 67, 107 and 116. The next model is fitted after these observations are excluded from the data.

Table 3.3: Summary of linear regression model of pre-award efforts

	Estimate	Std. Error	Pr(> t)
Intercept	48.809	23.300	0.037
Department Support	14.273	3.812	0.000
Estimated Time	0.944	0.026	0.000
Preference Score	2.189	1.312	0.096
Work Week (ISO)	-1.220	0.562	0.030

Multiple R-squared: 0.761, Adjusted R-squared: 0.759

As shown in Table 3.3, looking at individual predictors, the departmental support, estimated time, and preference score increases the actual time spent in efforts. The work week of the effort is estimated decrease the actual time spent. Also, Figure 3.2 is the plot of observed actual time spent versus the fitted values. The points on the plot follow the line pretty closely, showing that the regression model has significance in predicting the actual time spent on pre-award efforts.



For efforts associated to post-award RASR tasks, similar to the results from pre-award RASR tasks, the attributes from RASR do not show any statistical significance in predicting the actual time. Initially, the attributes from efforts are used to build the regression model

on the post-award efforts. As shown in the summary of this initial model, included in the appendix, preference score is the only variable with insignificance as the p-value is higher than 0.05.

Table 3.4: Summary of linear regression model of post-award efforts

	Estimate	Std. Error	P-value
Intercept	-20.229	3.363	1.932e-09
Department Support: Yes	-3.419	0.615	2.900e-08
Estimated Time	0.706	0.010	0.000e+00
Complexity Score	1.040	0.218	1.884e-06
Work Week (ISO)	0.467	0.077	1.509e-09

Multiple R-squared: 0.489, Adjusted R-squared: 0.488

The final model is built on the data excluding the preference score, and the summary of this model is presented in Table 3.4. The estimated time, complexity score, and work week in ISO are predicted to increase the actual time spent, and the presence of departmental support is predicted to decrease the actual time spent in these post-award efforts.

3.2 Tasks Analysis

For the tasks in the pre-award stage, the first model is fit using the following variables: the number of days on RASR, the total number of touches on RASR, the total estimated time, total number of efforts, average complexity score, and average preference score. From the summary of this model, included in the appendix, the total number of efforts, average complexity score, and type of tasks are insignificant predictor variables for the p-values are higher than 0.05. All predictor variables but the number of touches on the RASR system positively predict the actual total time spent on the tasks.

Table 3.5: Summary of linear regression model of pre-award tasks

	Estimate	Std. Error	P-value
Intercept	-204.509	163.038	0.238
RASR Task Time (Days)	0.626	0.295	0.059
# of RASR Task Touches	-36.729	18.509	0.075
Total Estimated Time	0.628	0.113	0.000
Avg. Complexity Score	26.360	11.134	0.039
Avg. Preference Score	104.072	23.802	0.001

Multiple R-squared: 0.84, Adjusted R-squared: 0.76

Using the backward selection, Table 3.3 shows the summary of the model which is fit excluding the type of tasks. The number of touches on RASR still shows a p-value higher than 0.05, but all the other predictor variables have p-values lower than 0.05, appeared to be useful to predict the time spent in task. Also, all but the number of touches on RASR predict to increase the actual time spent on the task. It also should be noted that only 16 pre-award tasks are used to fit the model and it may not be significant to predict the actual time spent in larger set of pre-award tasks.

The same set of variables is used to fit the saturated model for the tasks in the post-award stage. From the summary of the model, included in the appendix, all but the total estimated time, the total number of efforts, the average complexity score, and a few types of tasks have p-values higher than 0.05, showing insignificance.

Table 3.6: Summary of linear regression model of post-award tasks

	Estimate	Std. Error	P-value
Intercept	-3.535	2.328	0.1298
Total Estimated Time	0.123	0.005	3.035e-72
Total # of Efforts	0.290	0.152	0.0565
Avg. Complexity Score	1.592	0.584	0.0067
Avg. Preference Score	1.004	0.600	0.0951
Type: Journal Budget Transfer	108.517	10.819	5.222e-21
Type: Journal ID	7.702	2.057	2e-04
Type: None	5.817	1.587	3e-04
Type: Order ID	9.842	4.925	0.0464
Type: Purchase Order	12.826	3.222	1e-04
Type: Voucher	10.874	1.856	1.068e-08
Type: Web Travel	27.117	6.249	1.864e-05

Multiple R-squared: 0.699, Adjusted R-squared: 0.69

Using stepwise variable selection, the total number of RASR days on RASR and the total number of touches on RASR are excluded from the predictor variables for the final model. Table 3.4 shows the summary of the model. As shown in the table, all the predictor variables, except the average preference score, show significant as predictor variables with p-values less than 0.05. The average preference score and total number of efforts are kept as the predictor variables because excluding these variables resulted in a more predictor variables to be insignificant in predicting the actual time. Also, the variable for proposal award type of task is excluded from the model because there was only one post-award task with that type.

3.3 Proposals and Projects Analysis

The actual time spent on proposals, or the projects that are in the pre-award stage, is predicted with the model where the following variables are first used as predictors: total number of tasks, number of employees worked on efforts, number of work weeks, presence of department support, total estimated time, average complexity score, proposal type, award administration department, and award type. The project amount of the proposal is not included in the full model because more than half of the proposals are missing in this variable.

Using the stepwise variable selection, the final model for proposals uses the number of tasks, count of work weeks, and total estimated time to predict the actual time spent on the proposals. The summary of this model is shown in Table 3.5. All three predictor variables

have a p-value higher than 0.05, showing significance as predictor variables, and predict to increase the actual time spent on proposals.

Table 3.7: Summary of linear regression model of proposals

	Estimate	Std. Error	P-value
Intercept	-81.642	28.672	0.0059
Number of Tasks	31.145	7.685	1.377e-04
Count of Work Weeks	41.355	15.872	0.0114
Total Estimated Time (mins)	0.827	0.050	4.579e-25

Multiple R-squared: 0.929, Adjusted R-squared: 0.926

Similar to the proposals, the following variables are used to fit the saturated model: number of work weeks, presence of department support, total estimated time, average complexity score, account status, and department number. From this saturated model, stepwise variable selection is used, which identified the department number, the total number of tasks, and the average complexity score to be statistically insignificant in predicting the actual time spent on a project.

Table 3.8: Summary of linear regression model of projects

	Estimate	Std. Error	P-value	NA
Intercept	15.827	13.115	1.207	0.2281
Count of Employees	-30.806	4.778	-6.448	2.822e-10
Count of Work Weeks	19.182	2.268	8.458	3.465e-16
Total Etimated Time (mins)	0.788	0.028	28.325	1.066e-103
Avg. Preference Score	-7.687	3.685	-2.086	0.0375
ACCOUNT_STATUS: Frozen / Inactive	18.344	7.650	2.398	0.0169
ACCOUNT_STATUS: Preliminary	19.818	51.456	0.385	0.7003

Multiple R-squared: 0.829, Adjusted R-squared: 0.827

The model is fit again excluding these variables, and the summary of this model is shown in Table 3.6. From the table, it can be noticed that the factor variable of account status being preliminary does not have significance with high p-value. This is because there are only two projects whose account statuses are preliminary. The count of employees who worked on the projects and average preference score predict to decrease the actual time spent, and the

count of work weeks, total estimated time, and account status predict to increase the actual time spent on projects.

4 Methods

4.1 Multiple Linear Regression

Multiple linear regression is used to predict a response variable with multiple predictor variables. In this project, the response variable is the actual time spent at each level of work and predictor variables are the attributes given at each level. The `lm` function in R provides the estimates of the parameters in predicting the response variable. In each part of the analysis, all variables of interest are selected to fit the model first and then variable selections are performed to find the best subset of variables to predict the actual time. The summary of the linear regression model shows the coefficients estimates, standard error, t-value for the T-test, and the p-value for the test. The null hypothesis, H_0 for this test is that the estimated coefficient is 0, thus insignificant in predicting the response variable, and the alternative hypothesis, H_1 , is estimated coefficient is not 0. The T-test is performed under the assumption of the null hypothesis, so if the p-value is lower than 0.05, it is concluded to reject the null hypothesis of insignificance, i.e. the variable is significant.

4.2 Variable Selection

The variable selection chooses predictor variables from the data have statistical significance in predicting the variable of interest. In this project, stepwise selection is used, and the criterion called AIC (Akaike Information Criterion) is assessed in this process. The `step` function in the `stats` package in R successively builds a model in steps and predictors are added or deleted at each step [1]. At each step, this procedure aims to minimize the AIC.

4.3 Diagnostics

Once the model is fit, the outliers and points of high influence are identified. By identifying these points, further decisions are made as to whether these points should be eliminated when fitting the model or should be kept in the data. The point of high influence impacts the model by leveraging the slope of the fitted line. Figure 4.1, from Chatterjee and Hadi [2], in which the y value is a child's score in an aptitude test and x value denotes the age of child, illustrates examples of outliers and point of influence. Point labeled as "O" is considered as an outlier because if linear regression line is fitted to the data, the estimated

slope of the fitted line “will hardly change if this data point is deleted” [2]. On the other hand, the point labeled as “I” in Figure 4.1 illustrates a data point that is influential to the estimated slope of the fitted line if it is removed from the data.

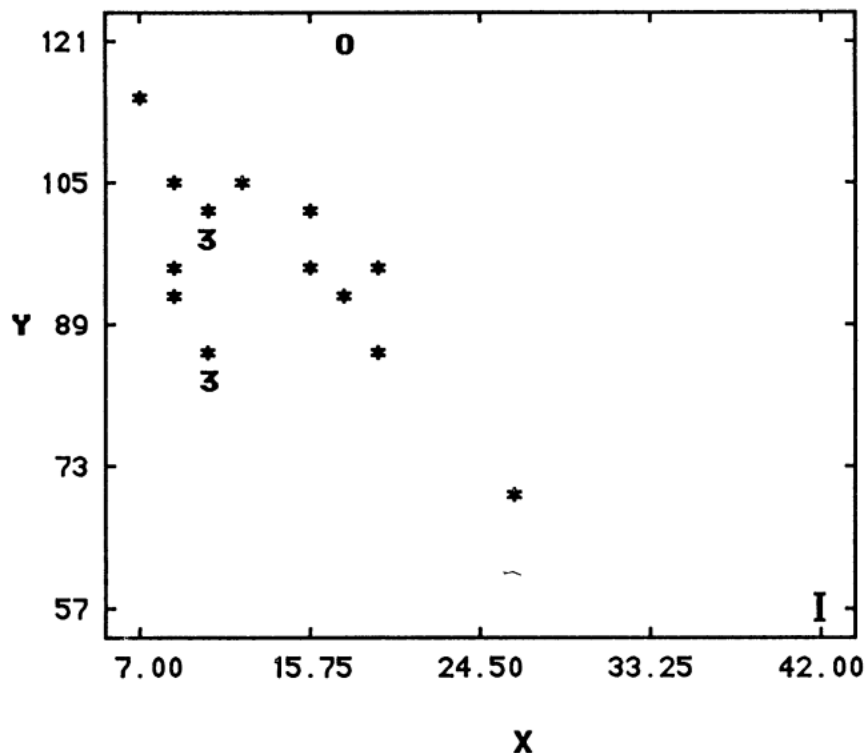


Figure 2: Plot of Mickey, Dunn, Clark (1967) data from Chatterjee and Hadi (2)

Although the figure is displaying a case of simple linear regression case, it can be easily seen that these points can influence the parameter estimates and the fit of the model. For some models built in this project, these points of high influence are removed to fit the data, and some have been kept. As explained in more detail in Section 3 of the results, if these points are removed, some models will have an insufficient amount of data to confidently conclude the significance of the variables or the model to predict the response variables.

5 Conclusion

The report presents the results from multiple linear regression analyses performed at each level of work at the Research Administration Service Center. The actual time taken at each level of work and both stages of the award is best predicted with a unique subset of attributes of work and data collected from employees. At all levels of work, attributes collected from

employees on the efforts show a significant effect on predicting the actual time, which is also reported by the employees. At the tasks level, the client expected certain variables associated with tasks to show significance in predicting the actual time spent on tasks; however, not all of the expected variables showed significance and the data on efforts from employees showed high significance in the model. Similar results are presented for projects and proposals, where data from employees showed high significance in predicting the actual time spent on either proposals or projects. For future analysis, it is recommended to collect data from employees over a longer period for a larger number of tasks and projects.

Reference

1. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
2. Samprit Chatterjee, Ali S. Hadi. “Influential Observations, High Leverage Points, and Outliers in Linear Regression.” Statistical Science, 1(3) 379-393 August, 1986. <https://doi.org/10.1214/ss/1177013622>

Appendix

Table 5.1: Summary of linear regression saturated model of pre-award efforts

	Estimate	Std. Error	Pr(> t)
Intercept	29.380	28.140	0.2969
Department Support	17.493	4.221	0
Estimated Time	0.882	0.022	9.634e-160
Complexity Score	-0.230	0.950	0.8084
Preference Score	1.974	1.680	0.2406
Work Week (ISO)	-0.715	0.639	0.264

Multiple R-squared: 0.778, Adjusted R-squared: 0.776

Table 5.2: Summary of linear regression saturated model of post-award efforts

	Estimate	Std. Error	Pr(> t)
Intercept	-18.163	3.607	4.9e-07
Department Support	-3.226	0.627	2.8e-07
Estimated Time	0.708	0.010	0.0e+00
Complexity Score	1.091	0.220	7.6e-07
Preference Score	-0.424	0.267	0.11
Work Week (ISO)	0.447	0.078	1.1e-08

Multiple R-squared: 0.489, Adjusted R-squared: 0.488

Table 5.3: Summary of linear regression saturated model of pre-award tasks

	Estimate	Std. Error	Pr(> t)
Intercept	71.696	250.856	0.782
RASR Task Time (Days)	0.754	0.319	0.046
# of RASR Task Touches	-81.948	35.452	0.050
Total Estimated Time	0.486	0.149	0.011
Total Number of Efforts	26.820	18.682	0.189
Avg. Complexity Score	23.406	11.190	0.070
Avg. Preference Score	99.266	24.203	0.003
Type: Proposal Number	31.380	72.093	0.675

Multiple R-squared: 0.875, Adjusted R-squared: 0.765

Table 5.4: Summary of linear regression saturated model of post-award tasks

	Estimate	Std. Error	Pr(> t)
Intercept	-3.759	2.362	0.1124
RASR Task Time (Days)	-0.008	0.019	0.694
# of RASR Task Touches	0.031	0.052	0.5451
Total Estimated Time	0.123	0.005	1.2e-71
Total Number of Efforts	0.287	0.152	0.06
Avg. Complexity Score	1.594	0.588	0.0071
Avg. Preference Score	1.008	0.601	0.0945
Type: Journal Budget Transfer	108.461	10.845	6.7e-21
Type: Journal ID	7.820	2.085	2.1e-04
Type: None	5.836	1.595	3e-04
Type: Order ID	10.887	5.684	0.0562
Type: Proposal Number	-0.690	10.671	0.9485
Type: Purchase Order	13.424	3.644	3e-04
Type: Voucher	11.068	1.935	2.3e-08
Type: Web Travel	27.167	6.266	1.9e-05

Multiple R-squared: 0.7, Adjusted R-squared: 0.688

Table 5.5: Summary of linear regression saturated model of proposals

	Estimate	Std. Error	Pr(> t)
Intercept	33.338	211.811	0.8757
Number of Tasks	33.889	15.533	0.0348
Count of Employees	-25.797	45.557	0.5742
Count of Work Weeks	21.181	25.108	0.4037
Department Support	-40.024	80.118	0.62
Total Estimated Time (mins)	0.830	0.075	4.169e-14
Avg. Complexity Score	17.666	17.039	0.3058
Avg. Preference Score	57.521	36.861	0.1261
PROPOSAL_TYPE: Non-Competing Continuation/Progress Report	-131.495	164.021	0.4272
PROPOSAL_TYPE: Recurring Contract	65.428	111.608	0.5609
PROPOSAL_TYPE: Resubmission or Amendment	126.626	78.755	0.1154
PROPOSAL_TYPE:Supplement	124.536	194.032	0.5245
AWARD_ADMIN_DEPT_NO:201430	-235.956	207.698	0.2624
AWARD_ADMIN_DEPT_NO:221010	26.619	329.676	0.936
AWARD_ADMIN_DEPT_NO:318000	-221.732	200.619	0.2753
AWARD_ADMIN_DEPT_NO:318400	-238.882	161.555	0.1467
AWARD_ADMIN_DEPT_NO:462001	-213.933	221.261	0.3391
AWARD_ADMIN_DEPT_NO:631300	68.635	210.758	0.7463
AWARD_ADMIN_DEPT_NO:633100	-346.557	171.676	0.0499
AWARD_ADMIN_DEPT_NO:633200	-362.262	159.659	0.0285
AWARD_ADMIN_DEPT_NO:633500	-372.416	150.362	0.0174
AWARD_ADMIN_DEPT_NO:633600	-371.205	170.146	0.0348
AWARD_ADMIN_DEPT_NO:635200	-287.023	144.403	0.0534
AWARD_TYPE:Cooperative_Agreement	44.232	106.095	0.6789
AWARD_TYPE:Grant	-39.687	56.515	0.4864
AWARD_TYPE:Incoming_Subcontract	13.261	54.220	0.808
AWARD_TYPE:NIH_Grant	-72.968	82.710	0.3827

Multiple R-squared: 0.962, Adjusted R-squared: 0.939

Table 5.6: Summary of linear regression saturated model of projects

	Estimate	Std. Error	Pr(> t)
Intercept	-4.831	74.345	0.948
Number of Tasks	1.137	0.621	0.068

Table 5.6: Summary of linear regression saturated model
of projects (*continued*)

	Estimate	Std. Error	Pr(> t)
Count of Employees	-28.270	5.545	0.000
Count of Work Weeks	12.406	2.906	0.000
Department Support	-10.934	10.717	0.308
Total Estimated Time (mins)	0.777	0.041	0.000
Avg. Complexity Score	-4.050	5.096	0.427
Avg. Preference Score	7.406	5.271	0.161
ACCOUNT_STATUS: Frozen / Inactive	15.331	8.148	0.061
ACCOUNT_STATUS: Preliminary	-37.196	72.881	0.610
DEPT_NO:315600	-5.622	88.474	0.949
DEPT_NO:318200	17.538	102.288	0.864
DEPT_NO:318300	10.109	102.508	0.921
DEPT_NO:318400	23.290	73.528	0.752
DEPT_NO:318600	-1.496	102.410	0.988
DEPT_NO:318900	-5.273	102.042	0.959
DEPT_NO:319100	1.985	102.608	0.985
DEPT_NO:382000	14.632	101.858	0.886
DEPT_NO:400101	-2.500	102.028	0.980
DEPT_NO:411601	-1.111	88.520	0.990
DEPT_NO:412009	0.004	88.988	1.000
DEPT_NO:412019	-15.767	102.765	0.878
DEPT_NO:412250	-4.727	102.014	0.963
DEPT_NO:412802	-15.995	102.758	0.876
DEPT_NO:414401	10.535	101.842	0.918
DEPT_NO:423501	2.645	101.738	0.979
DEPT_NO:424903	55.664	125.061	0.656
DEPT_NO:426401	10.605	101.774	0.917
DEPT_NO:426801	3.181	79.136	0.968
DEPT_NO:427601	14.785	102.970	0.886
DEPT_NO:427801	-8.984	84.025	0.915
DEPT_NO:429001	-8.777	102.203	0.932
DEPT_NO:429201	30.447	102.050	0.766
DEPT_NO:429601	37.616	102.226	0.713
DEPT_NO:440100	18.430	102.313	0.857
DEPT_NO:452100	-0.525	89.206	0.995
DEPT_NO:452300	10.605	101.774	0.917
DEPT_NO:460502	11.109	102.508	0.914
DEPT_NO:462002	-1.617	88.219	0.985

Table 5.6: Summary of linear regression saturated model
of projects (*continued*)

	Estimate	Std. Error	Pr(> t)
DEPT_NO:463001	22.700	88.339	0.797
DEPT_NO:463501	-16.214	88.614	0.855
DEPT_NO:464501	18.430	102.313	0.857
DEPT_NO:466001	1.985	102.608	0.985
DEPT_NO:631100	-7.455	88.324	0.933
DEPT_NO:631200	-6.957	88.385	0.937
DEPT_NO:631300	78.888	81.079	0.331
DEPT_NO:633100	6.389	72.875	0.930
DEPT_NO:633200	20.815	72.429	0.774
DEPT_NO:633500	-8.820	73.438	0.904
DEPT_NO:633600	-55.301	73.047	0.449
DEPT_NO:635200	-7.292	72.531	0.920

Multiple R-squared: 0.848, Adjusted R-squared: 0.83