36-617 Background Survey – Fall 2021

 Never used it
 -0 (0%)

 Used it in a class
 -23 (92%)

 Used it for a project outside of class
 -11 (44%)

 0
 5
 10
 15
 20
 25

Your experience with linear regression (check all that apply) ²⁵ responses

8 Project Descriptions...

Learned it as a small part of a machine learning class.

Performed data analysis on products and gave data-driven suggestions on targeting clients

Auto regression model for homeruns prediction

I have used regression in an independent research project.

I have used linear regression models on a few projects outside of class within summer programs. The first one was on basketball data with CMU's Sports Analytics Camp and the other was a biostatistics/epidemiology project with Columbia University's Mailman School of Public Health.

I used linear regression in R in a faculty guided research project during my undergrad. Among other things, we modeled the role that different serve types had on ace percentage in college volleyball. I also used R in an internship last summer with a hockey team. We built a GAM to predict protection lists in anticipation of the NHL expansion draft.

Used R for work. Mostly tidyverse but I have also used datatable and some glm.

I used R and regression analysis to do identification on which variables are the most important one for a specific consulting-related question

Multiple linear regression model and checked model assumptions and did some back or step elimination to fit it.

Your Experience with R (check all that apply) 25 responses



7 Project descriptions...

I have used R for several research projects that involved randomization tests and t-tests for surveys

I used linear regression in R in a faculty guided research project during my undergrad. Among other things, we modeled the role that different serve types had on ace percentage in college volleyball. I also used R in an internship last summer with a hockey team. We built a GAM to predict protection lists in anticipation of the NHL expansion draft.

Used R for work. Mostly tidyverse but I have also used datatable and some glm.

I have used R for several research projects that involved randomization tests and t-tests for surveys. I used Python for two large modeling projects at my internship this previous summer. Lastly, I have used regression in an independent research project.

I used R while working as a statistics research assistant for the Carnegie Mellon Cognitive Development Lab. I used k-means clustering in R to summarize children's performance on a task measuring their attention spans. I also assessed the psychometric properties of tasks using correlations and t-tests in R in order to demonstrate the validity of the task as a measure of semantic structure in children.

Created various Shiny apps to analyze colors in images

I used R and regression analysis to do identification on which variables are the most important one for a specific consulting-related question

Your Experience with Python (check all that apply) 24 responses



7 Project Descriptions...

Data wrangling for survey data.

I used Python for two large modeling projects at my internship this previous summer.

Used python for an internship; basic data processing work

Use Python for machine learning. Using packages like scikit-learn. Use linear regression mainly for simple predictions.

I used Python to create a decision tree model to detect a fraud user during my last internship.

Use python with some package like sklearn. Use linear regression for some simple predictions.

I took a Coursera course, and finished the corresponding assignment and tests.

Have you ever heard of a "hat matrix"? If so, tell me what it is, as best you can remember (don't look anything up in a book or online etc., just tell me what you remember).

Correct (2)

It is a symmetric idempotent, Y^hat=HY.

project the actual outcome to predicted values

On the Right Track (6)

Has something to do with pseudo-inverse matrices and the solution to the regression equation.

A matrix that is formed by multiplying the matrix of the explanatory variables with itself and it's inverse multiple times (don't remember the exact order).

It's a matrix that has something to do with second derivatives.

I know that I learned about this in a regression course and there was some kind of formula with inverse and transposed matrices.

The matrix that transforms the the response variable to the estimated coefficients, $H^{Y} = \mathbb{B} (I \text{ think})$

A hat matrix is a matrix of estimated values which are calculated using observations

In the Ball Park (3)

yes - xtx-1y??

Beta = (X^T X)^-1 X^T beta

I believe it refers to the center matrix (the one containing the eigenvalues) of the eigen-decomposition of the model matrix in the context of linear regression.

Not So Much... (3)

I've never heard that term before but I'm betting that I've seen it

I'm not sure exactly what a hat matrix is but "matrix" leads me to linear algebra which I know is the foundation of linear regression.

The scalar value in matrix form

And 5 "no's"...

Have you ever heard of the "level of a test"? If so, tell me what it is, as best you can remember (don't look anything up in a book or online etc., just tell me what you remember).

Correct (9)

Significance of a test is the probability of type I error i.e., rejecting the null hypothesis when it is true. Type I error results in the conclusion that there is a difference between the population and sample value when there is no actual difference. And the significance level (alpha) gives us the risk of this false positive conclusion.

I'm guessing this has to do with the significance level of a test - which indicates the threshold for the p-value.

The probability of a type 1 error (I think)

Does this mean level of significance/alpha? If not, then no.

It's the alpha value (?)

The "level" of a test refers to the threshold we set for how unlikely an event must be under the null for us to reject the hypothesis.

I'm not sure. Does it mean the significance level of a test? So a significance level of a test means that when you're doing a hypothesis test, you need to decide the probability of making mistakes if the null hypothesis is true.

significance level

Significant level alpha, the probability of the type one error.

In the Ball Park (7)

The level of the test would determine if the alternative hypothesis is significant or not, depending on the p-value.

Hypothesis testing done on dummy (categorical) variables with at least two possible categories. The significance of the regression estimator determines whether there is a linear relationship between the explanatory and response variables.

The level of significance is how easy or difficult would be to reproduce the result with a different sample.

Yes, I remember it is related to hypothesis testing and something about type I and II errors.

pvalue

T value, P value

level of the test refers to significant level, which is alpha, and we usually have the threshold of 0.05, meaning that if the value of z/t is smaller than 0.05, we reject the hypothesis; otherwise, we fail to reject the hypothesis.

Not So Much... (2)

The level of a test corresponds to the width of the confidence interval, so it tells you how confident you can be in the conclusions you draw from you analyses.

Significance level of a test: the min probability of rejecting an event when doing a hypothesis test

And 3 "no's"...

Is there anything you'd like me to know about your background or your hopes for this class?

Learn about regression and its generalizations (7)

Hope for a more in-depth understanding of regression and it's applications as well as how to better use statistical computing software like R and Python.

Hope to learn a decent amount of theory, I took 36-402 as an undergrad so I'm hoping this extends beyond that

The statistic courses I mainly took are about probability and stochastic process.

I want learn more about how to apply regressions in our future work. I only have some basic knowledge of Python.

want to have more practice on the way to use linear regressions maybe in our future career

The statistic knowledge I mostly learn in college is about probability and stochastic process.

Hope to learn more about linear models and understand how they could be utilized to solve real life problems.

Learn about R and/or Python (5)

Hope for a more in-depth understanding of regression and it's applications as well as how to better use statistical computing software like R and Python.

I have little to no experience with R and Python - never used it in a class or for a project. I've learnt the basics via online resources. Hoping to get some extra help outside of class (this one as well as Statistical Computing) until I'm all caught up.

I want learn more about how to apply regressions in our future work. I only have some basic knowledge of Python.

I hope to learn more uses of R and Python (mostly Python)

I hope I can practice more in python since have learned a lot in R during my undergraduate study.

...and 4 responses just looking forward to learning in the class...