Predict Part 2: Terminology in Predict

Back to Basics in Construct and Predict Series

November 15th, 2022

Rapid Insight

Meet Your Presenters

Lily Brennan

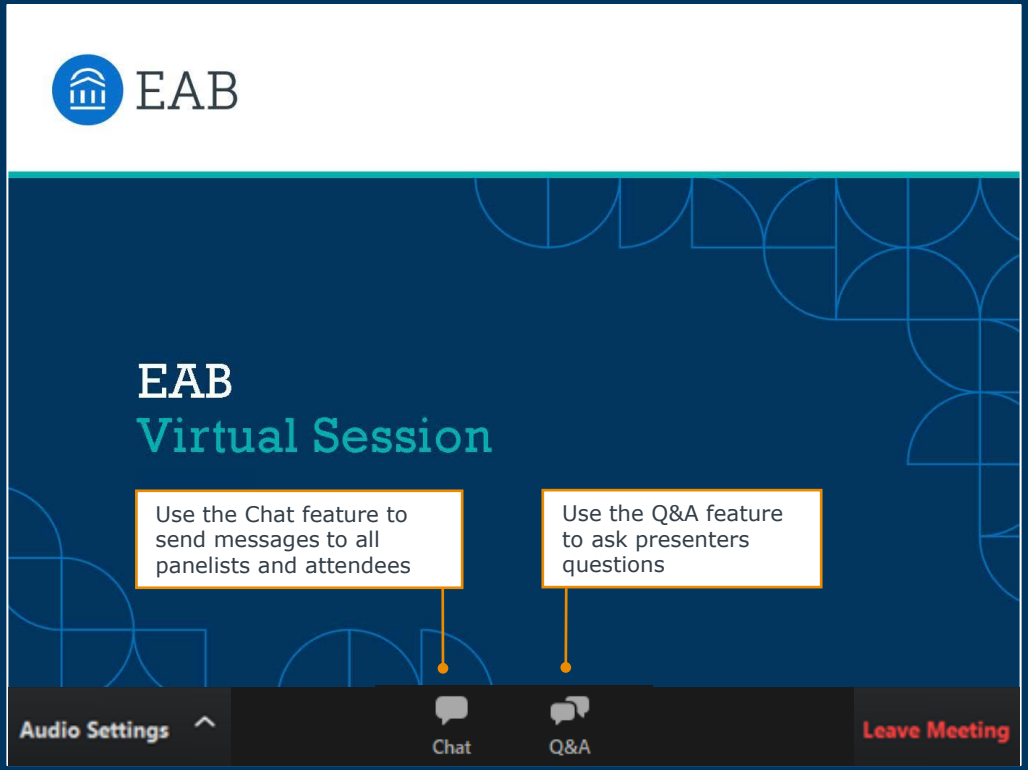
Strategic Leader,
Data and Analytics

Earl Sires

Senior Analyst,
Product Marketing

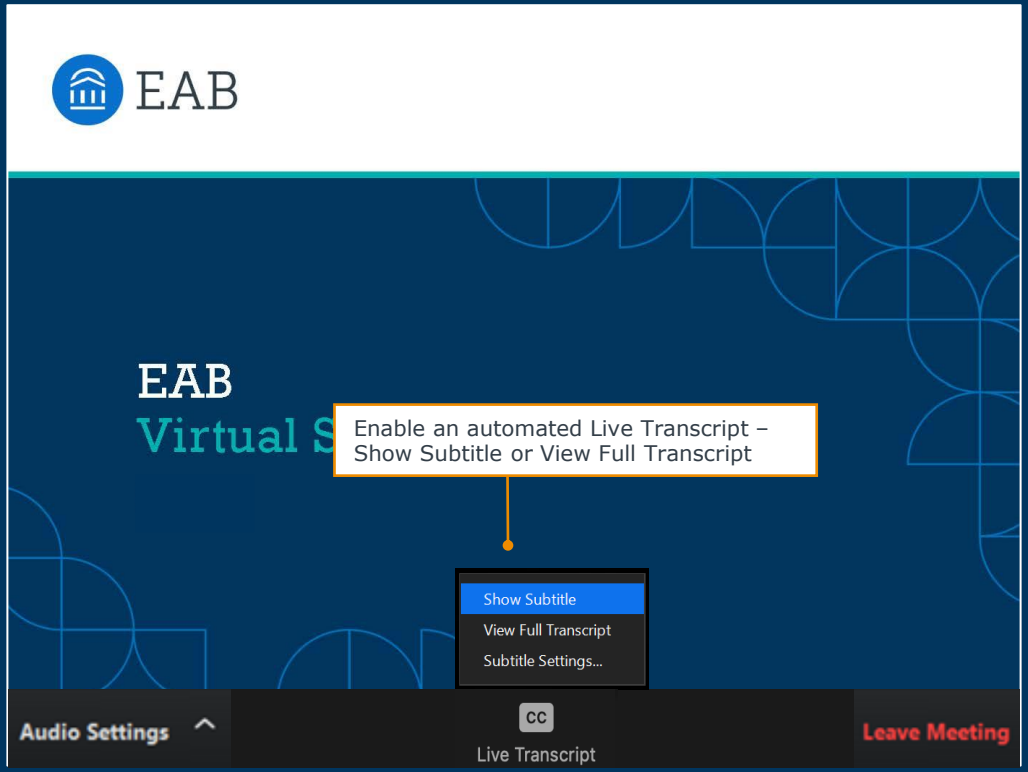
©2022 by EAB. All Rights Reserved. **eab.com**

Submit a Question or Comment



©2022 by EAB. All Rights Reserved. **eab.com**

Turn on Captions



©2022 by EAB. All Rights Reserved. **eab.com**

1. **Back to Basics Series**
2. Terminology– Slides & Live Demo
3. Preview of Next Session
4. Audience Q & A

©2022 by EAB. All Rights Reserved. **eab.com**

Back to Basics in Construct and Predict Series

Catch Up on On-Demand Content and Register for the Last Session

- **Construct Parts 1, 2 & 3**
  Learn tips and tricks in Construct
- **Predict Part 1: Basic How-To's**
  Master the basic flow and features of Predict

Catch up on sessions online

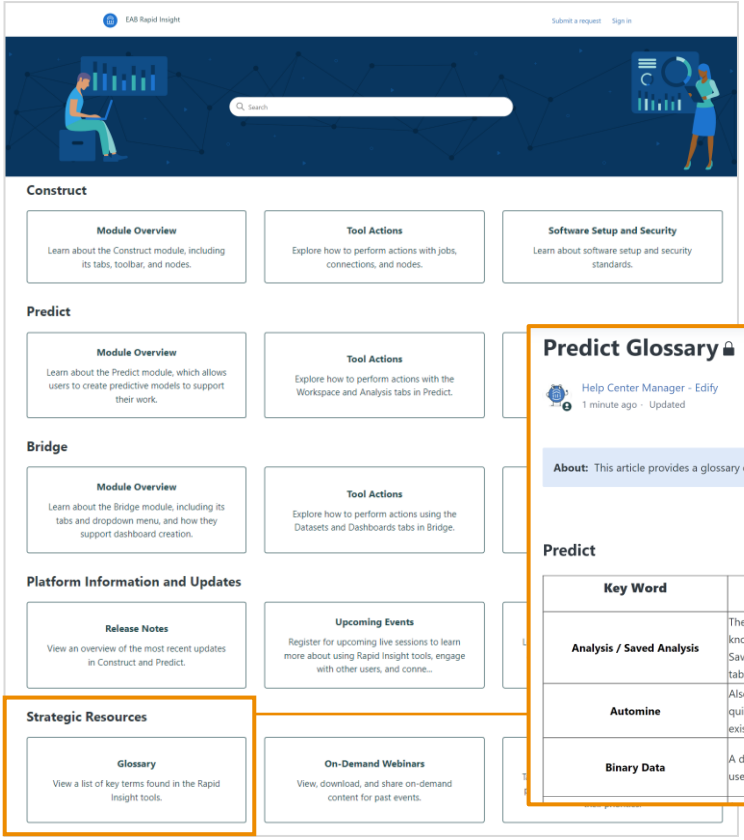
- **Final Session: Navigating Help Resources**
  December 6 | 2:00pm Eastern

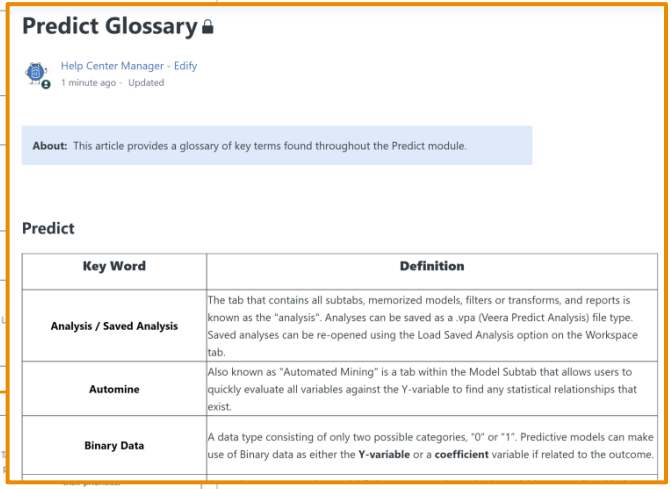
  Learn about our latest support offerings and best practices for getting help!

Register today!

©2022 by EAB. All Rights Reserved. **eab.com**

NEW Predict Glossary Article in the Help Center





Predict Glossary

Help Center Manager - Edify
1 minute ago · Updated

About: This article provides a glossary of key terms found throughout the Predict module.

Predict

Key Word	Definition
Analysis / Saved Analysis	The tab that contains all subtabs, memorized models, filters or transforms, and reports is known as the "analysis". Analyses can be saved as a .vpa (Veera Predict Analysis) file type. Saved analyses can be re-opened using the Load Saved Analysis option on the Workspace tab.
Automine	Also known as "Automated Mining" is a tab within the Model Subtab that allows users to quickly evaluate all variables against the Y-variable to find any statistical relationships that exist.
Binary Data	A data type consisting of only two possible categories, "0" or "1". Predictive models can make use of Binary data as either the **Y-variable** or a **coefficient** variable if related to the outcome.

Visit the Help Center!
https://support.rapidinsight.eab.com/hc/en-us

©2022 by EAB. All Rights Reserved. **eab.com**

Poll

Why are you joining today?

a) Brand new Predict user looking for help

b) Looking for a refresher in Predict

c) Just curious and hoping to learn something new

©2022 by EAB. All Rights Reserved. **eab.com**

1. Back to Basics Series
2. **Terminology– Slides & Live Demo**
3. Preview of Next Session
4. Audience Q & A

©2022 by EAB. All Rights Reserved. **eab.com**

Definitions Categories

Data Definitions

Interface Definitions

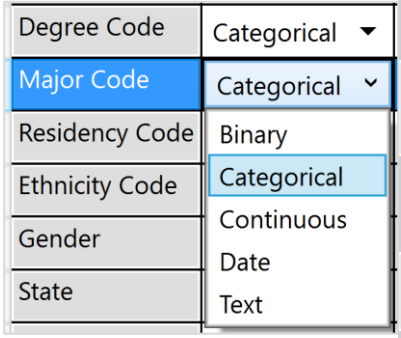
μ

Statistical Definitions

©2022 by EAB. All Rights Reserved. **eab.com**

Data Definitions

Data Types in Construct



Data Type	Definitions	Example
Binary	Values consisting of only two possible categories.	0/1
Categorical	Values consisting of a finite number of possible categories or types	Gender: {male, female}
Continuous	A number that may include decimals as an approximation of a real number	"215.9983"
Date	A calendar date	"11/15/2022"
Text	Any combination of alpha-numeric characters	"Zebra123!"

©2022 by EAB. All Rights Reserved. eab.com

Trivia

Which of the following data types can be used in a predictive model?

1) Text
2) Binary
3) Categorical
4) Date
5) Continuous

Answers:

A) All
B) 1,2,3
C) 2,3,5
D) 1,4,5
E) 2,3,4

©2022 by EAB. All Rights Reserved. **eab.com**

Data Definitions

Data Types Used in Predictive Models

Binary

- Can be used as a Y-variable (Logistic Regression) or a related variable

Categorical

- Can NOT be used as a Y-variable
- Binary transformations of each category are automatically created and can be used as related variables
- Example: Instead of Gender {Male, Female}, this would transform into Male{0,1}, Female{0,1}

Continuous

- Can be used as a Y-variable (OLS Regression) or a related variable

©2022 by EAB. All Rights Reserved. **eab.com**

Definition Categories

Data Definitions

Interface Definitions

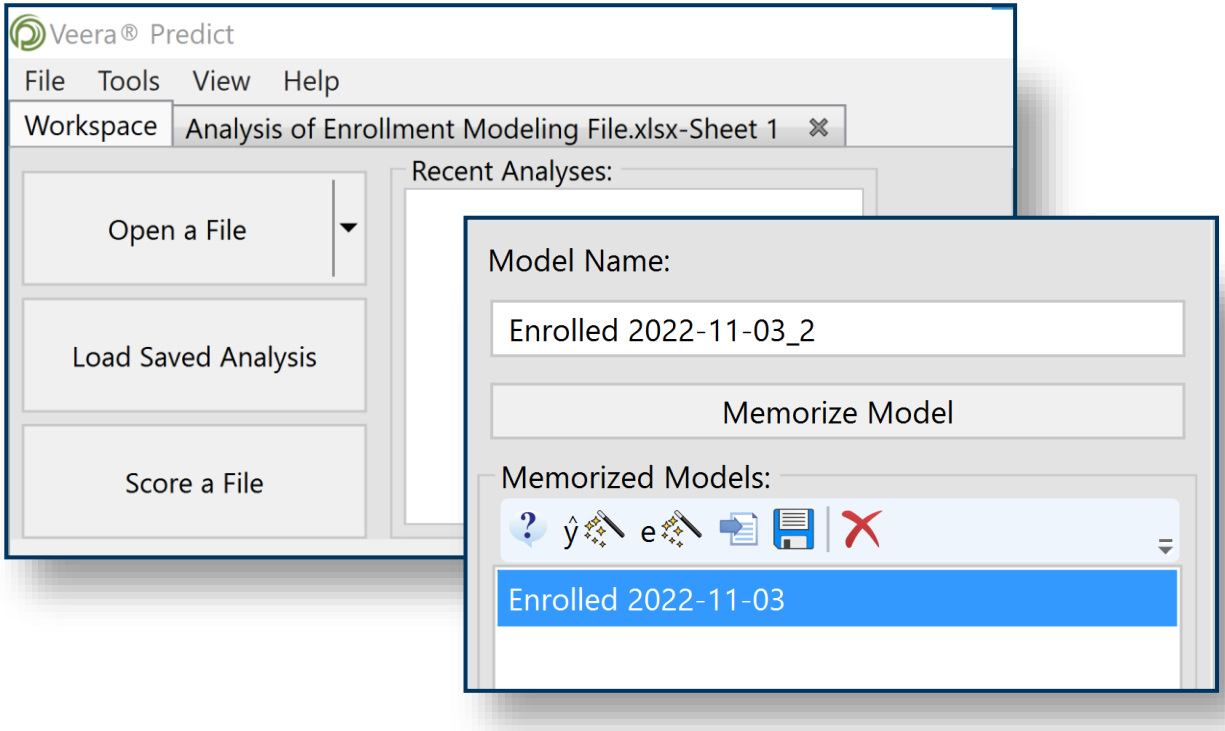
μ

Statistical Definitions

©2022 by EAB. All Rights Reserved. **eab.com**

Interface Definitions

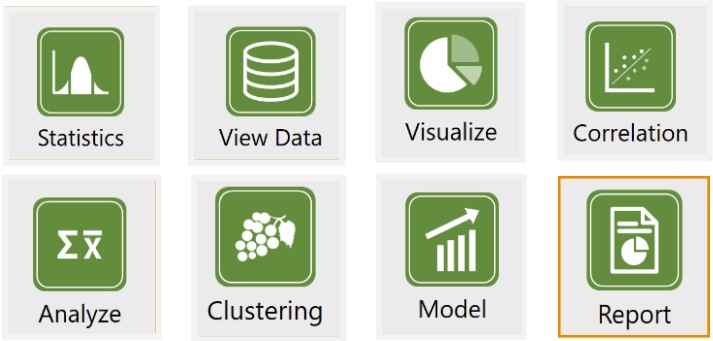
Analysis vs. Memorized Model



©2022 by EAB. All Rights Reserved. eab.com

Interface Definitions

Subtabs



- **Statistics:** Summary Statistics and data types for all variables in the dataset
- **View Data:** Record by record view of the entire dataset
- **Visualize:** Multivariate and univariate graphical representations of data relationships
- **Correlation**: Correlation matrix for all variables in the dataset
- **Analyze**: Means analysis, frequency analysis, profiling analysis
- **Clustering**: K-means clustering tool to partition similar data points into clusters to be included in modeling
- **Model:** Model building and regression analysis
- **Report:** A hub to save visualizations and outputs for reporting

©2022 by EAB. All Rights Reserved. **eab.com**

Interface Definitions

Y-Variable

An outcome that depends on the state of several other factors.

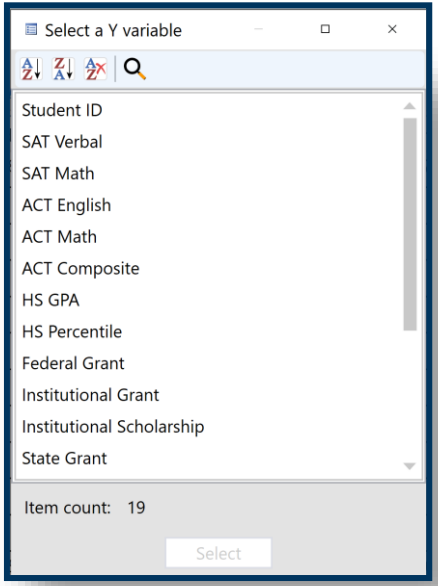
- A variable must be binary or continuous to be used as the Y-variable.

Synonyms

- Dependent variable
- Outcome variable
- Response variable
- Predicted variable

Example

A test score may depend on how much you studied, how much sleep you got, how long ago you ate, etc.

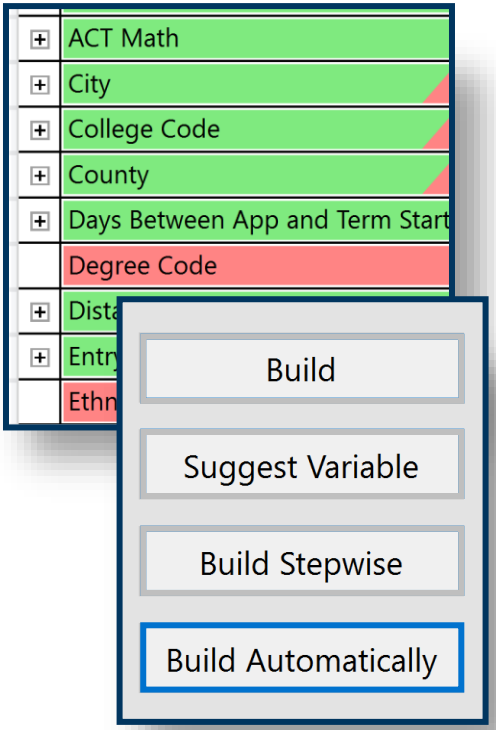


©2022 by EAB. All Rights Reserved. eab.com

Interface Definitions

Subtabs - Model

- **Automine –** "Automated Mining" of all variables against the y-variable for statistical relationships
- **Build –** Fits a model with only the variables manually added to the Included Variables section
- **Suggest Variable –** Adds a single variable at a time to the model from the available Variables pool that has been identified as the best variable based on what has already been included.
- **Build Stepwise –** Performs Stepwise Regression (adding and removing independent variables iteratively and testing for significance after each iteration) only using the variables that have been added to the Included Variables section
- **Build Automatically** – Also uses Stepwise Regression but considers every variable according to its significance in the Automine Tab, not just the Included Variables. No manual selection is required.



©2022 by EAB. All Rights Reserved. eab.com

Definition Categories

Data Definitions

Interface Data Definitions

μ

Statistical Definitions

©2022 by EAB. All Rights Reserved. **eab.com**

Trivia

When the Y-variable is a binary outcome, what type of regression is used?

1) Linear Regression

2) Ordinary Least Squares Regression

3) Logistic Regression

©2022 by EAB. All Rights Reserved. **eab.com**

Statistical Definitions

Regression Options in Predict

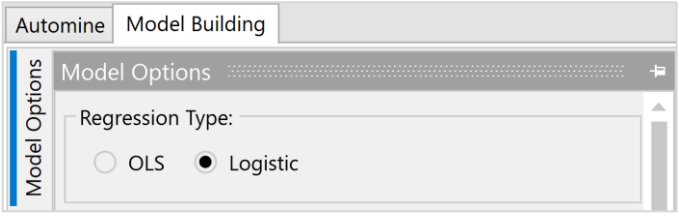
Ordinary Least Squares Regression/Linear Regression

- A type of regression analysis used when the Y-variable is a continuous variable
- Estimates the coefficients of a linear regression equation
- Models the relationship between independent variables and the dependent variable (Y-variable)
- The outcome is a point estimate of the value

Logistic Regression

- A type of regression analysis used when the Y-variable is a binary variable
- The outcome is a probability value (0-1.0) of the positive outcome (Y-variable = 1) to occur

Regression Type can be found in the Model Options!



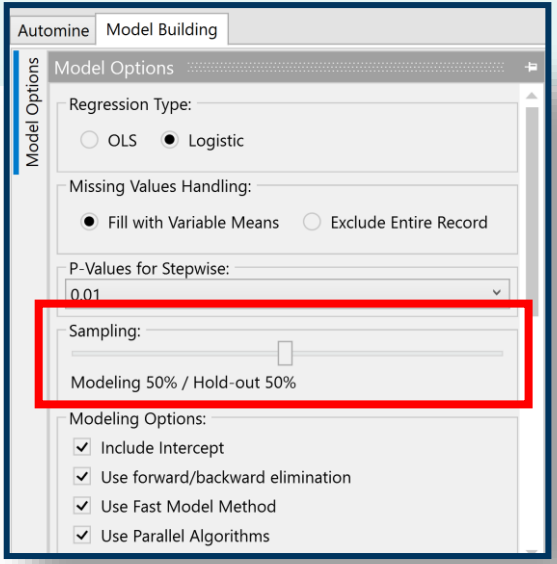
©2022 by EAB. All Rights Reserved. eab.com

Hold Out Sample

Hold Out Sample

A random sample from the data set that is withheld and not included in the modeling process.

After a model is built using the non-hold-out data (the "training" data), it is then applied to the hold-out sample to test and validate the accuracy of the model.



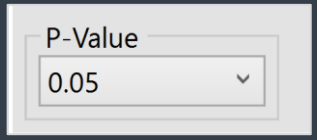
©2022 by EAB. All Rights Reserved. eab.com

Statistical Definitions

P-Value (Probability Value)

P-Value

The probability that the null hypothesis is true



- In regression the null hypothesis is that each independent variable **is not related** to the y-variable
- Low P-Value = reject the null hypothesis = independent variable is related to the y-variable
- The P-Value of a coefficient indicates whether the relationship to the outcome (y-variable) is statistically significant
- Precision in the app means that "0.000000" actually indicates "something lower than 0.000000"

Final Regression Model

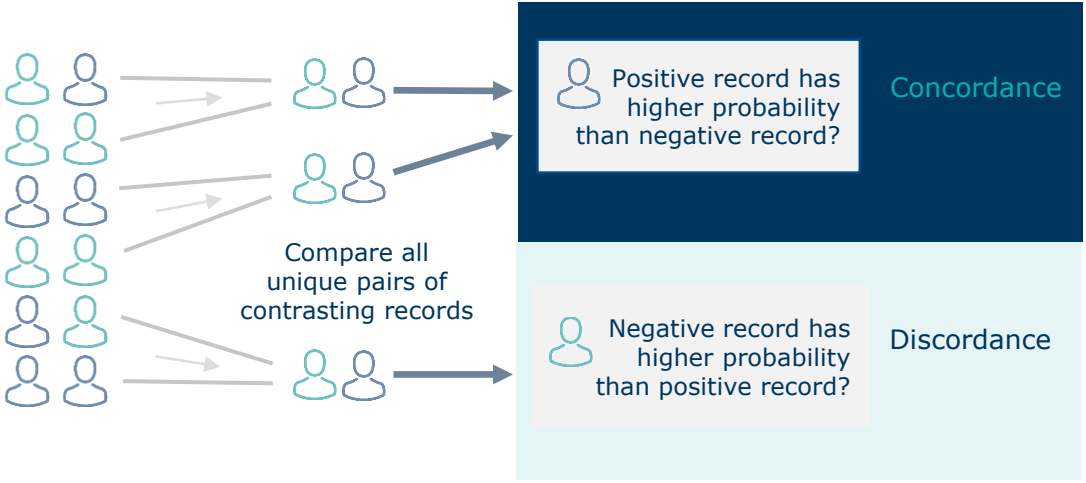
Predicting: Enrolled

Variable	Coef	S.E.	Wald chi-sq	p-value
Intercept	29.16	3.729	61.14	0.000000
Binary(County,Carroll)	1.158	0.1444	64.34	0.000000
Binary(Residency Code,I)	1.097	0.1020	115.61	0.000000
LOGe(SAT Composite or ACT Equivalent)	-2.020	0.3028	44.50	0.000000
Total Offered	0.000109	0.000011	95.65	0.000000
HS GPA	-0.3317	0.06936	22.87	0.000002

©2022 by EAB. All Rights Reserved. eab.com

Statistical Definitions

% Concordant and % Discordant



©2022 by EAB. All Rights Reserved. eab.com

Statistical Definitions

% Concordant and % Discordant

% Concordant:

The frequency at which the model accurately assigned higher probabilities to those who did observe the outcome than it assigned to those who did not (i.e. "how often the model was right")

% Discordant:

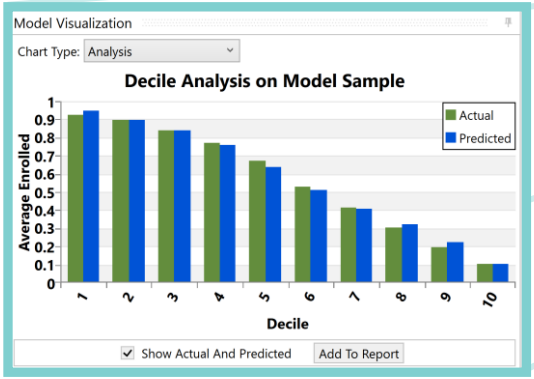
The frequency at which the model incorrectly assigned higher probabilities to those who did not observe the outcome than it assigned to those who did (i.e. "how often the model was incorrect")

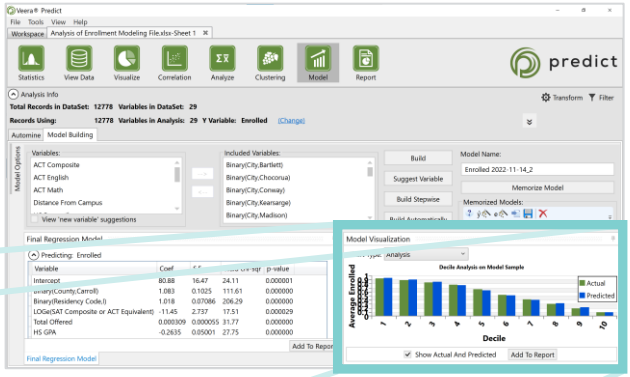
©2022 by EAB. All Rights Reserved. **eab.com**

Statistical Definitions

Decile Analysis

- Assign probabilities to each record
- Sort records according to probability
- Divide records into 10 equal groups
- Take the average outcome for each group
- Plot



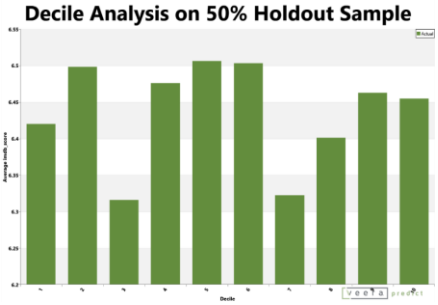


©2022 by EAB. All Rights Reserved. eab.com

Statistical Definitions

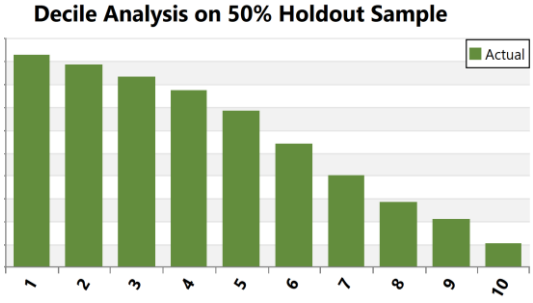
Decile Analysis Archetypes

Less Than Ideal Analysis



- Model is not doing a good job of predicting actual outcomes
- Model is not performing any better than random guessing
- Model should be improved before moving forward

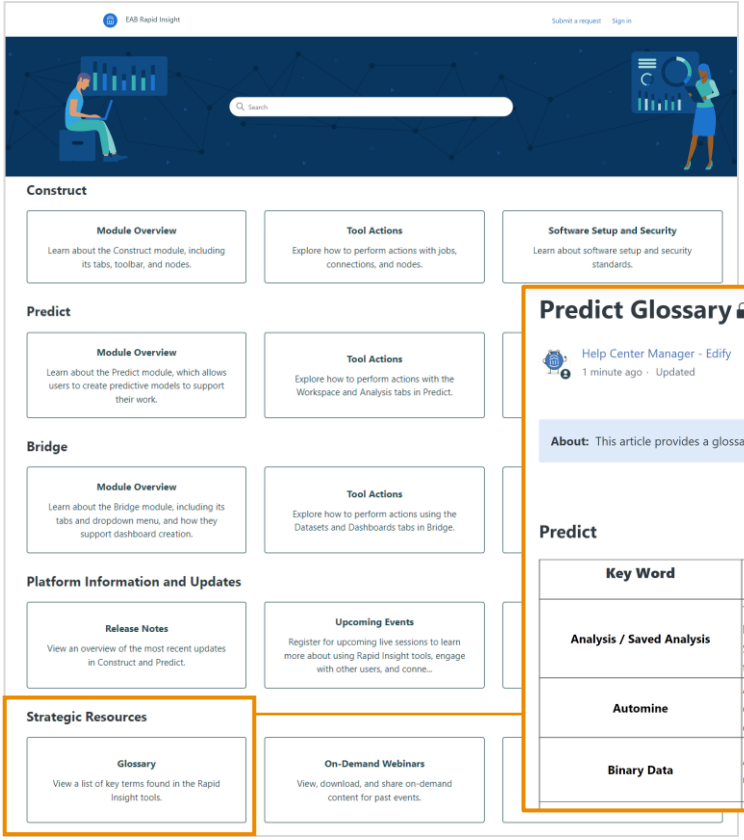
Ideal Analysis



- Consistent negative correlation to decile – "Staircase effect"
- Tells you that the model "binned" your constituents correctly from most likely to least likely
- A model exhibiting a good staircase decile analysis is one you can consider moving forward with

©2022 by EAB. All Rights Reserved. eab.com

Don’t Forget to Check Out the Predict Glossary



Visit the Help Center!
https://support.rapidinsight.eab.com/hc/en-us

Predict Glossary

Help Center Manager - Edify
1 minute ago · Updated

About: This article provides a glossary of key terms found throughout the Predict module.

Predict

Key Word	Definition
Analysis / Saved Analysis	The tab that contains all subtabs, memorized models, filters or transforms, and reports is known as the "analysis". Analyses can be saved as a .vpa (Veera Predict Analysis) file type. Saved analyses can be re-opened using the Load Saved Analysis option on the Workspace tab.
Automine	Also known as "Automated Mining" is a tab within the Model Subtab that allows users to quickly evaluate all variables against the Y-variable to find any statistical relationships that exist.
Binary Data	A data type consisting of only two possible categories, "0" or "1". Predictive models can make use of Binary data as either the **Y-variable** or a **coefficient** variable if related to the outcome.

©2022 by EAB. All Rights Reserved. eab.com

1. Back to Basics Series
2. Terminology– Slides & Live Demo
3. **Preview of Next Session**
4. Audience Q & A

©2022 by EAB. All Rights Reserved. **eab.com**

Back to Basics in Construct and Predict Series

Register for the Final Session and Encourage Colleagues to Attend!

Six-Part Series

Sessions occur on Tuesdays from 2:00 to 3:00 pm Eastern Time

Construct Part 1: Nodes in the Workspace

On Demand

Learn tips and tricks to work with nodes.

Construct Part 2: Inside the Nodes

Today's Session ☺

Explore node configuration windows and learn where to look for features.

Construct Part 3: Definitions in Construct

August 16

Learn definitions of data and Construct-specific terminology.

Predict Part 1: Basic How-To's

October 11

Explore data visuals, as well as building, saving, and adding predictive models.

Predict Part 2: Definitions in Predict

November 15

Learn modeling process and other relevant terminology.

Navigating Help Resources

December 6

Learn how to access self-service resources for quick answers to your questions.

©2022 by EAB. All Rights Reserved. **eab.com**

Poll

Would you like to be registered for the December 6th *Navigating Help Resources* session?

©2022 by EAB. All Rights Reserved. **eab.com**

Rapid Insight Office Hours

Sessions Occur Once a Month

- Next session: Thursday November 17th, 1:00 – 1:45pm Eastern Time
- Register for sessions online (linked in the Chat)

©2022 by EAB. All Rights Reserved. **eab.com**

1. Back to Basics Series
2. Terminology– Slides & Live Demo
3. Preview of Next Session
4. **Audience Q & A**

©2022 by EAB. All Rights Reserved. **eab.com**

Q & A: Submit a Question Using the Q&A Button

Lily Brennan
Strategic Leader, Data and Analytics

Earl Sires
Senior Analyst, Product Marketing

Additional Questions?

Email **RI-Support@eab.com** with technical questions and **RapidInsight@eab.com** for all other inquiries.

©2022 by EAB. All Rights Reserved. **eab.com**

Quick Poll

How was today's session?

Please take a few minutes to complete the follow-up survey (linked in the Chat) to provide additional feedback!

©2022 by EAB. All Rights Reserved. **eab.com**

Washington DC | Richmond | Birmingham | Minneapolis

202-747-1000 | **eab.com**

@eab **@eab_** **@WeAreEAB** **@eab.life**