IEOR E4650 Business Analytics

Session 2: Lending Analytics
Nomis (B): Analyzing the Value Proposition

Fall 2018

Today

Nomis (B)

- e-Car data
- Market analysis
- Pricing opportunity

Logistic regression

- Predicting customers' decisions
- Analytics-driven APR

E-Car Data

	A	B	C	D	E	F	G	H	I	J	K	L
1	Tier	FICO	Approve Date	Term	Amount	Previous APR	Car Type	Competition APR	Accept?	APR	Cost of Funds	Partner Bin
2	2	702	7/2/2002	60	22000		U	5.85	0	6.19	1.84	3
3	2	710	7/3/2002	60	21000		U	5.85	0	6.19	1.84	1
4	3	693	7/6/2002	60	19598		U	5.85	1	7.29	1.84	1
5	3	696	7/6/2002	60	23071		U	5.85	0	7.29	1.84	3
6	3	697	7/8/2002	60	21578		U	5.80	1	7.29	1.84	2
7	2	702	7/8/2002	60	20211		U	5.80	1	6.19	1.84	2
8	2	709	7/10/2002	60	21051		U	5.80	1	6.19	1.84	1
9	3	690	7/12/2002	60	21498		U	5.80	0	7.29	1.84	1
10	3	684	7/13/2002	60	18000		U	5.80	1	7.29	1.84	3
11	2	706	7/15/2002	60	24025		U	5.80	0	6.19	1.84	1
12	2	704	7/15/2002	60	23000		U	5.80	0	6.19	1.84	3
13	3	693	7/16/2002	60	19620		U	5.80	1	6.59	1.84	2
14	2	702	7/17/2002	60	24227		U	5.80	1	5.35	1.84	1
15	2	705	7/18/2002	60	21262		U	5.80	0	5.69	1.84	1
16	3	692	7/19/2002	60	25000		U	5.80	0	6.59	1.84	1

- Rows: observations (208,085)
- Columns: variables (12)
  - Categorical variables: tier, car type, accept?, partner bin
  - Continuous variables: FICO, term, amount, competition APR, APR, prime

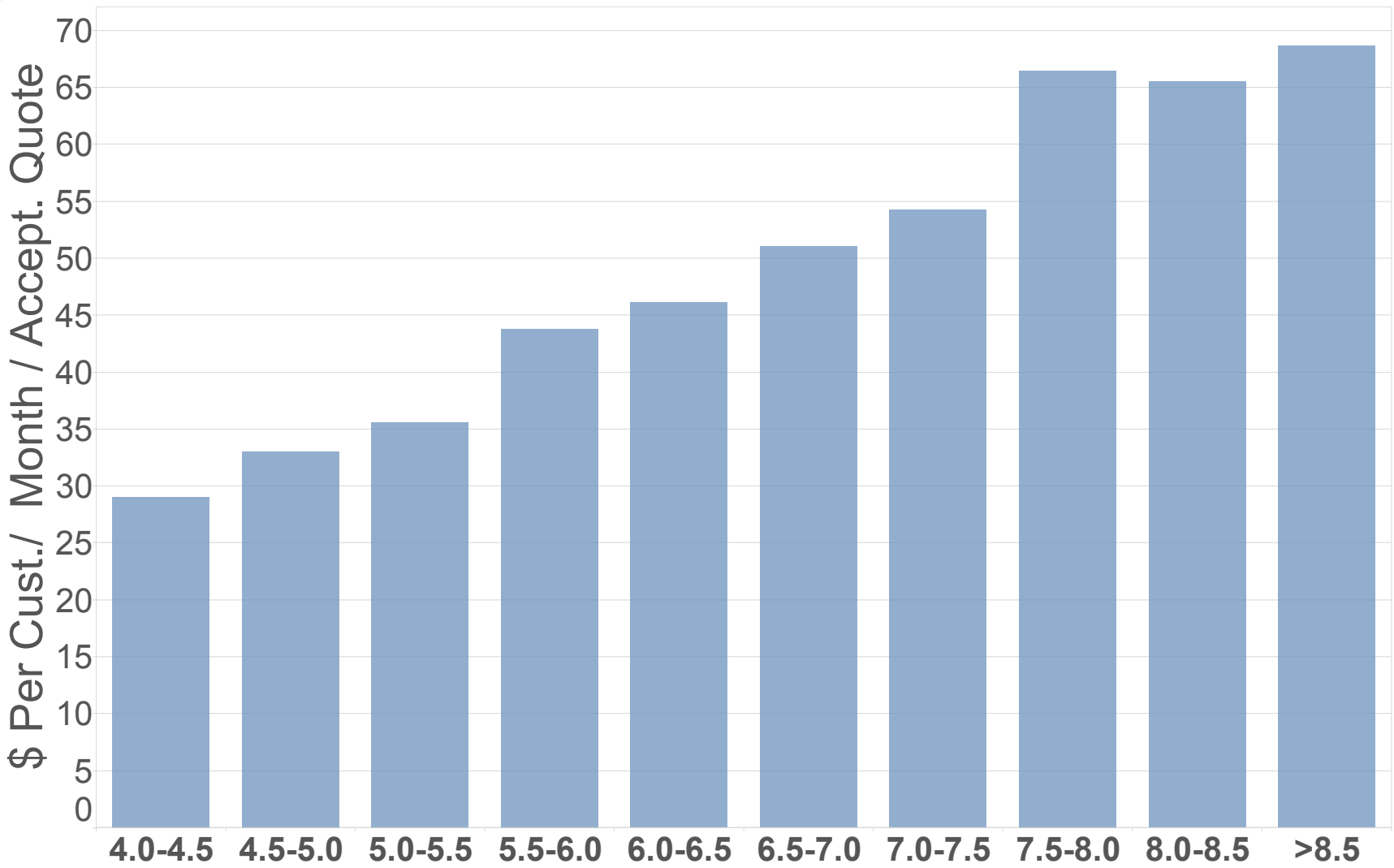
The e-Car Environment: Observations

Net Revenue Analysis: Segment Level

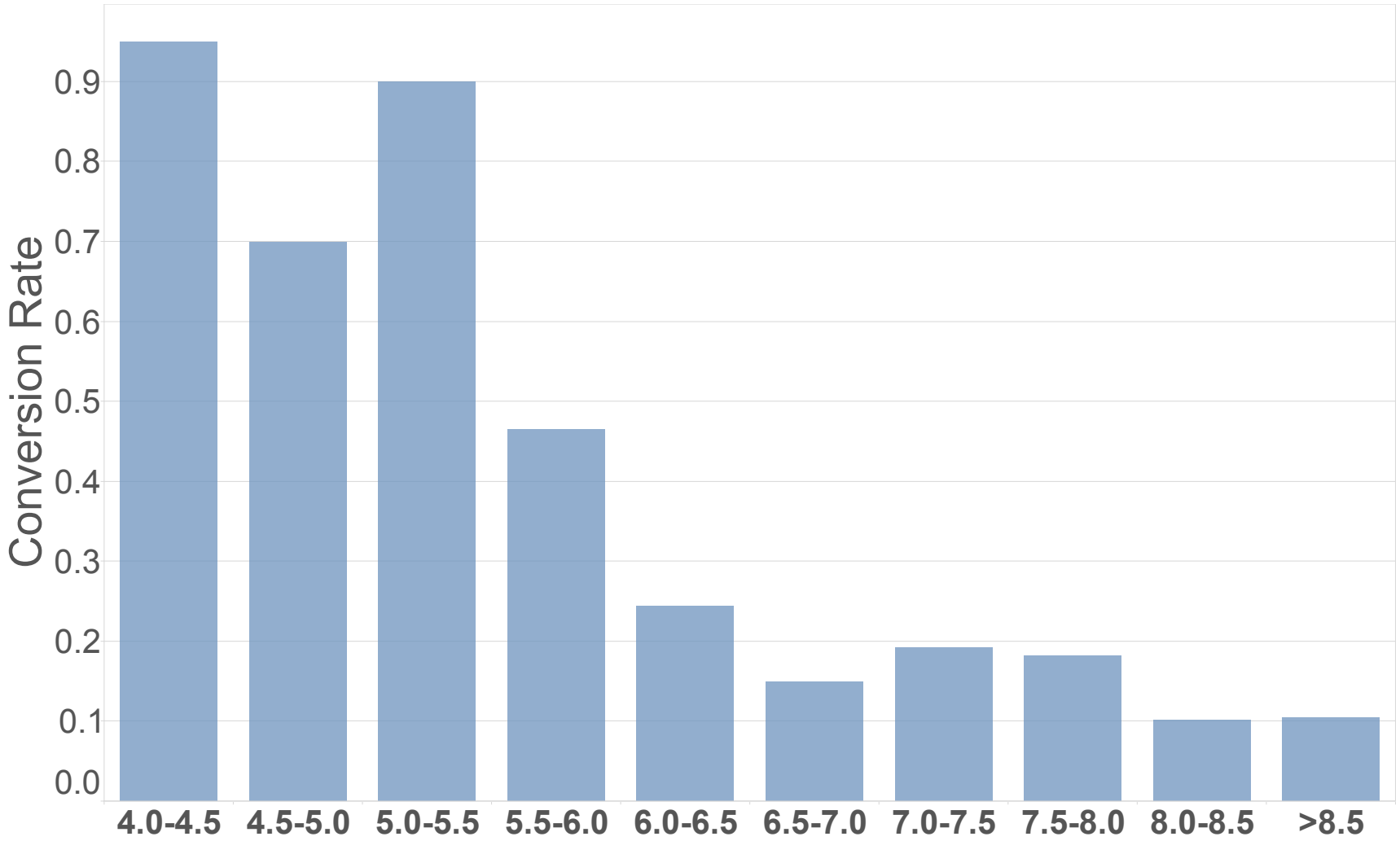
Focus on one segment

- USED cars
- FICO scores 684-712
- Term = 60 months
- Amount between 17.8K and 25K

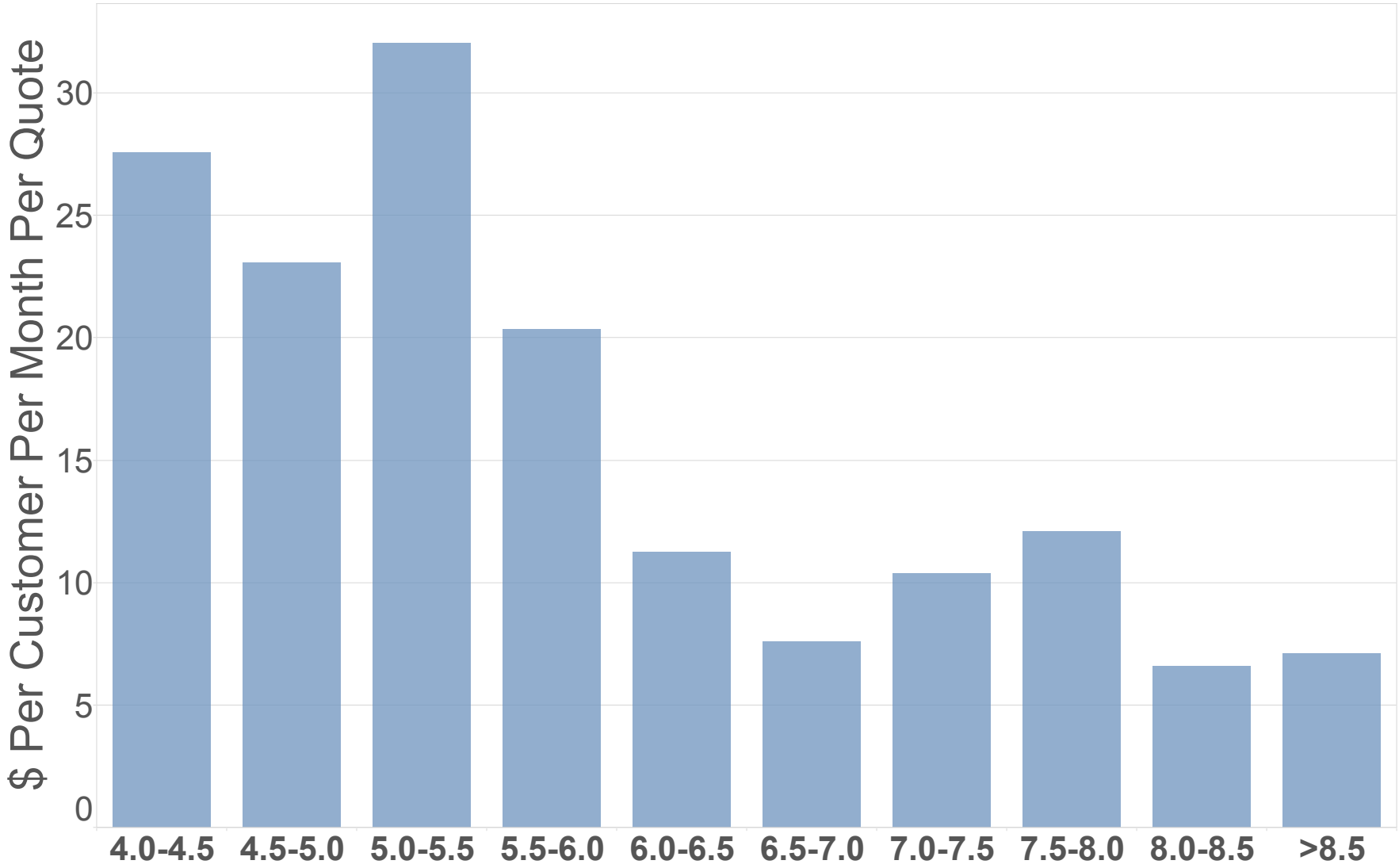
Segment Analysis: Avg Net Rev per Cust per Month per Accepted Quote vs. APR



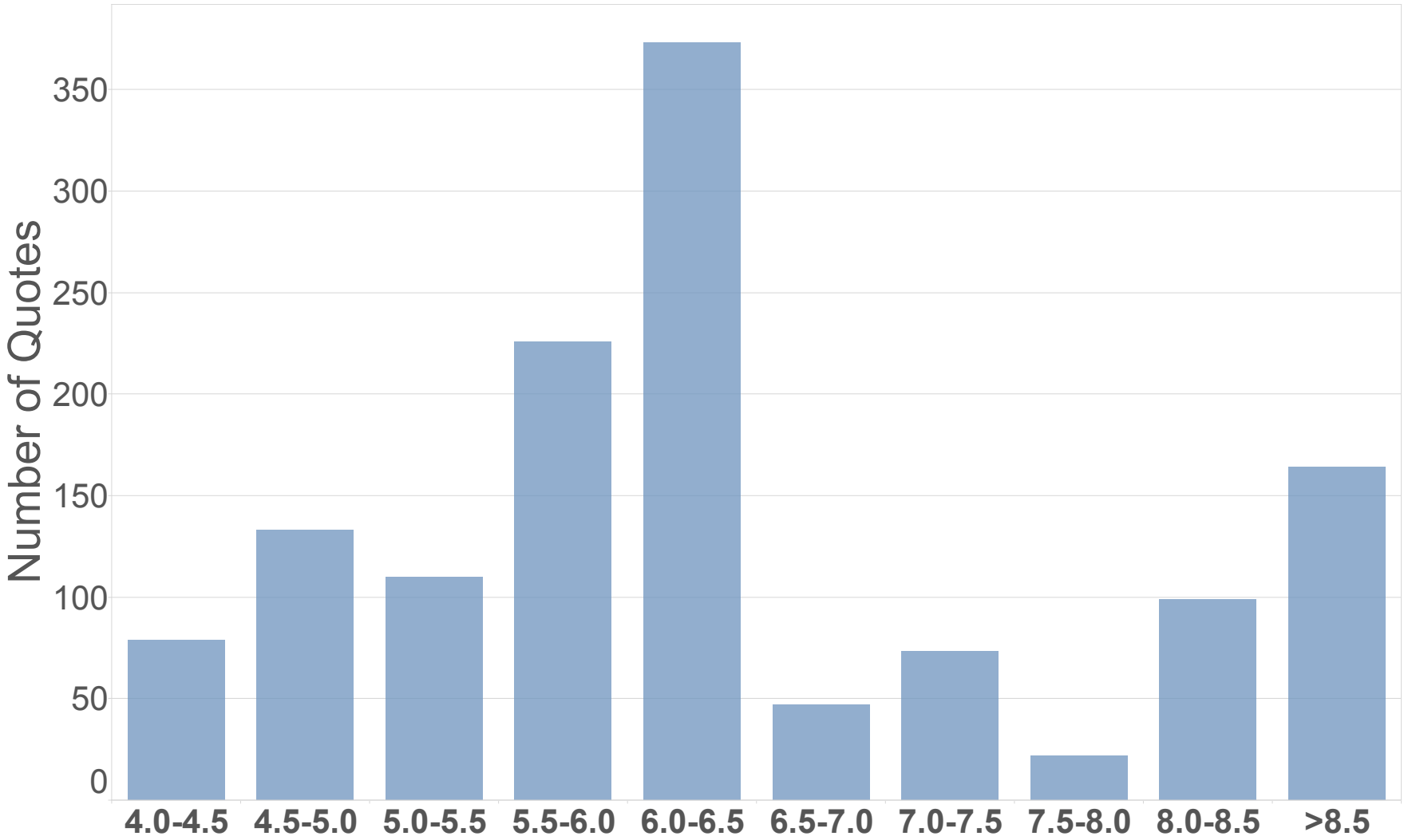
Segment Analysis: Conversion Rate versus APR



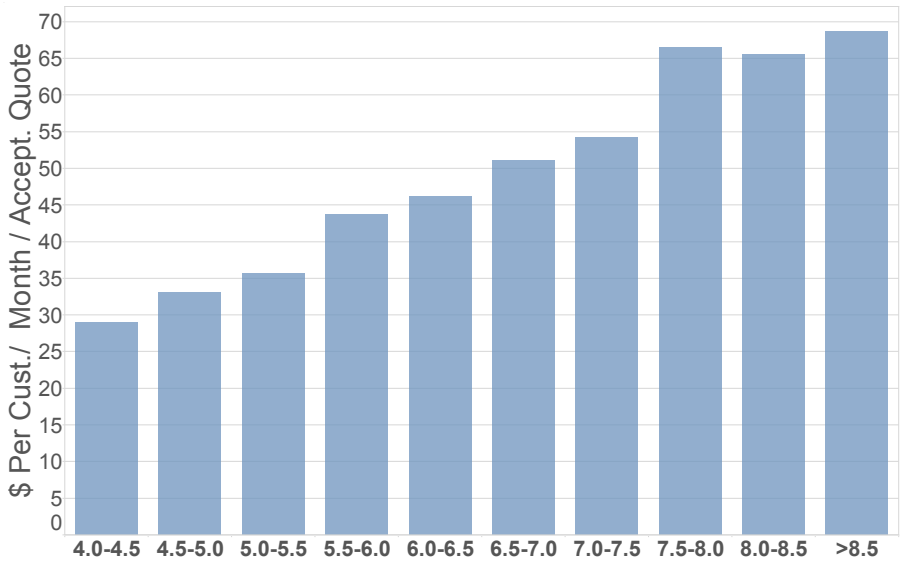
Segment Analysis: Average Net Revenue vs. APR

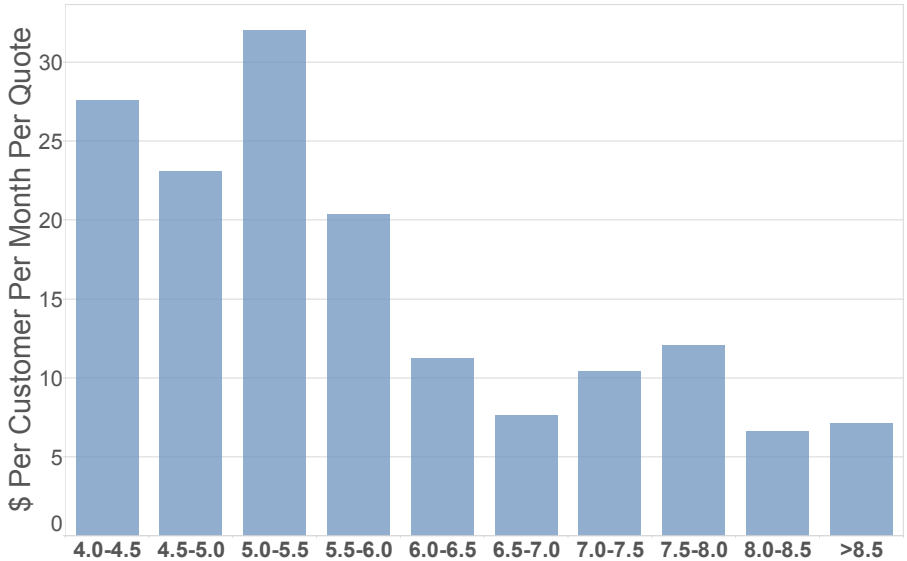


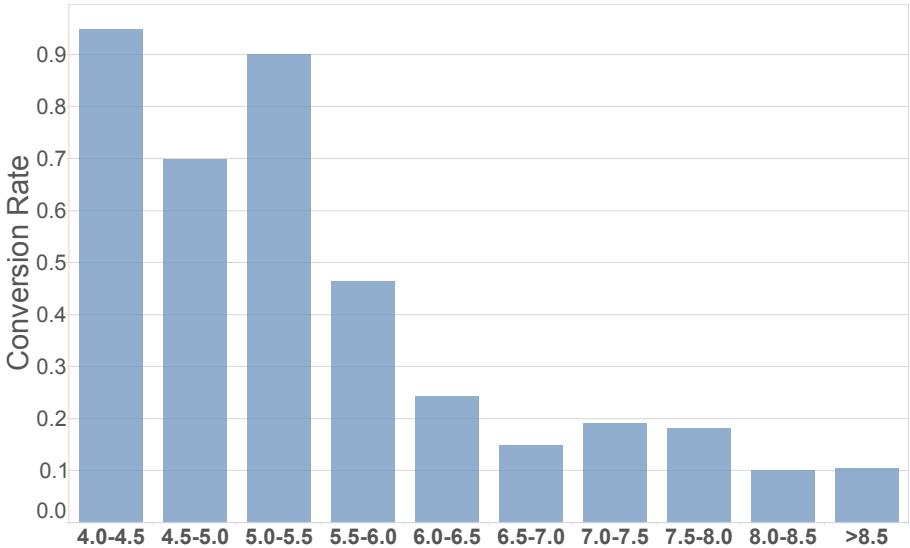
Segment Analysis: APR Quoted (Frequency)

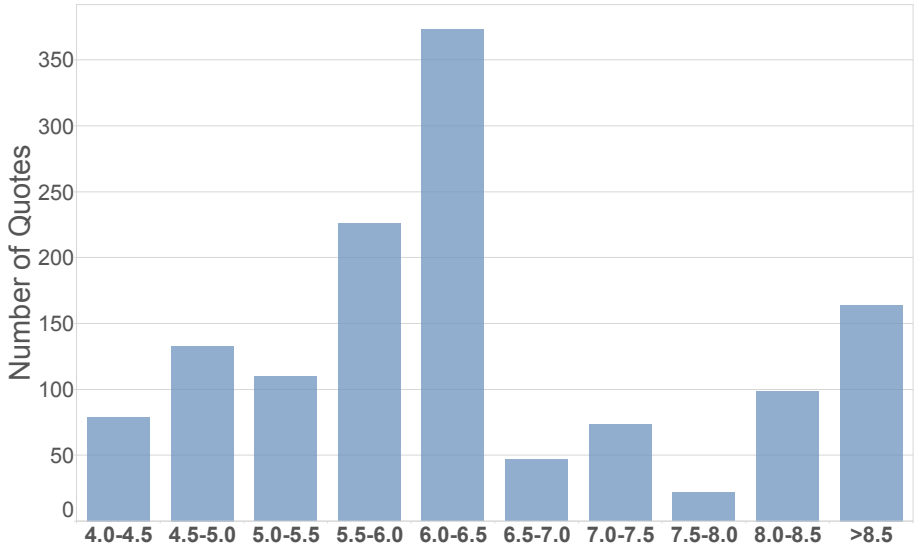


Segment Analysis









From Data to Decisions

The Goal

Data available for 208,085 past customer interactions:

- customer characteristics (FICO)
- loan characteristics (amount, term)
- APR quotes
- responses (accept / reject)

Objective: Find the "best" APR to set for future customers

Framing the Problem

Framing the Problem

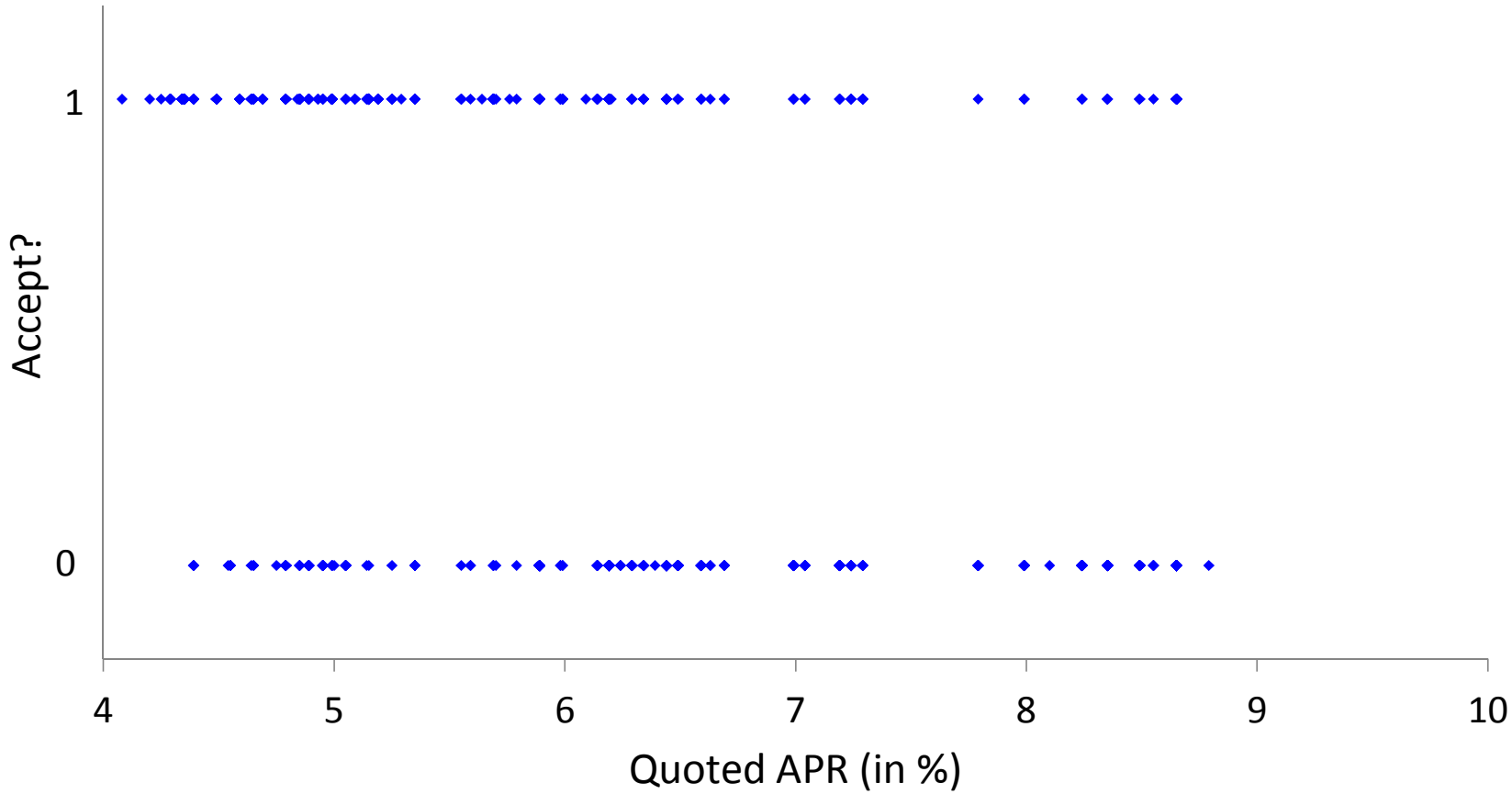
Maximize

(Net Revenue for loan) $\times$ (Probability of accept given APR quoted)

Two key inputs:

- Net Revenue
  $\Rightarrow$ calculate from cost information Typically, profit per loan is Net revenue minus adjustment for customer risk
- Prob. of accept given APR quoted
  $\Rightarrow$ need to "predict" customer decisions
  $\rightarrow$ In economic terms, we need to estimate the *demand curve* from the data.

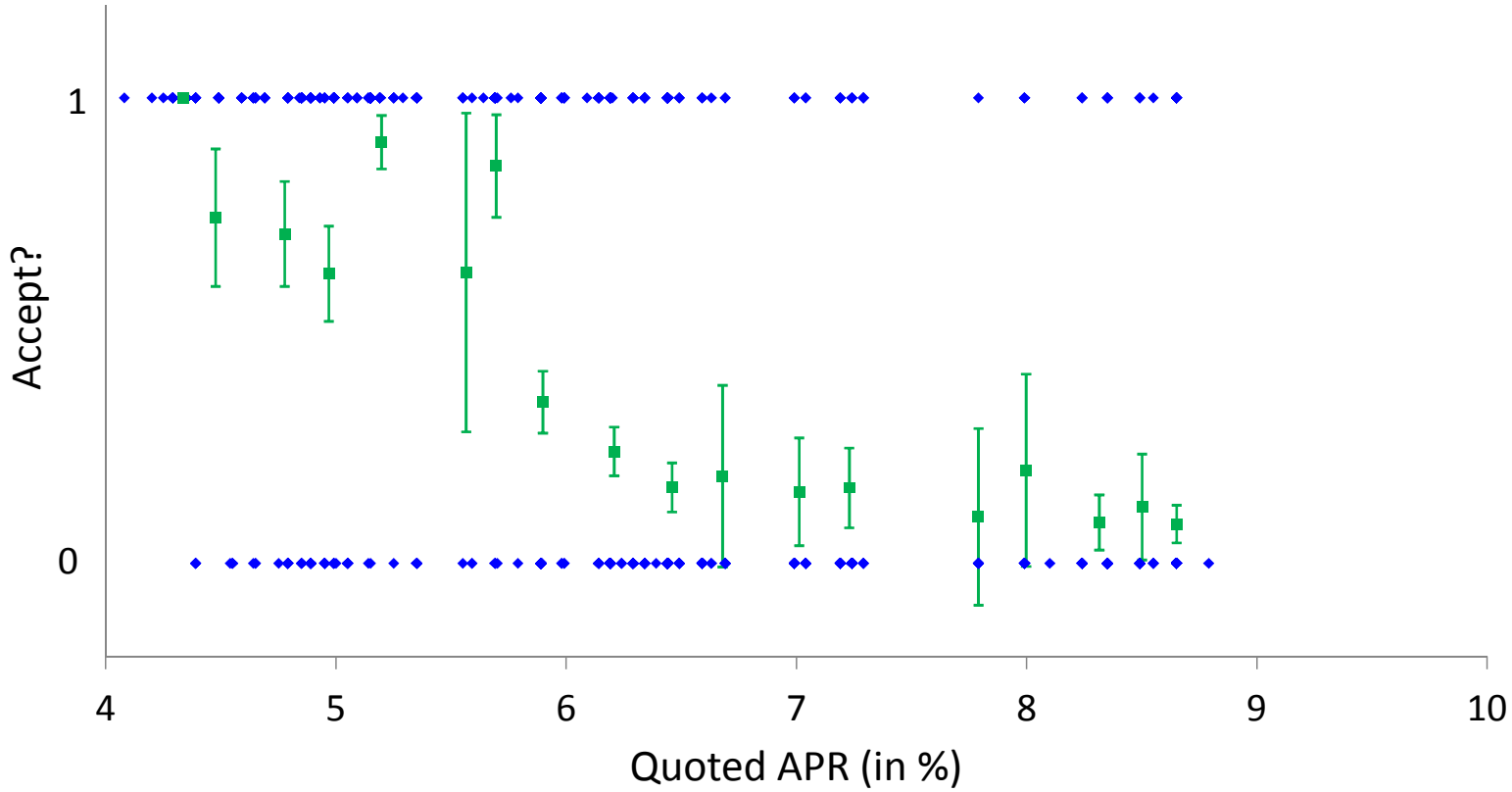
Accept? versus Quoted APR



Illustrative segment for all future analysis

- USED cars, FICO: 684-712
- Term: 60 months, Amount: between 17.8K and 25K

Accept? versus Quoted APR: Bucketed Data



Bucketed data for easier visualization

- APR buckets: 4.125-4.375, 4.375-4.625, etc.
- Average in each bucket plotted with two standard errors

How to Predict "Accept?"

Linear regression

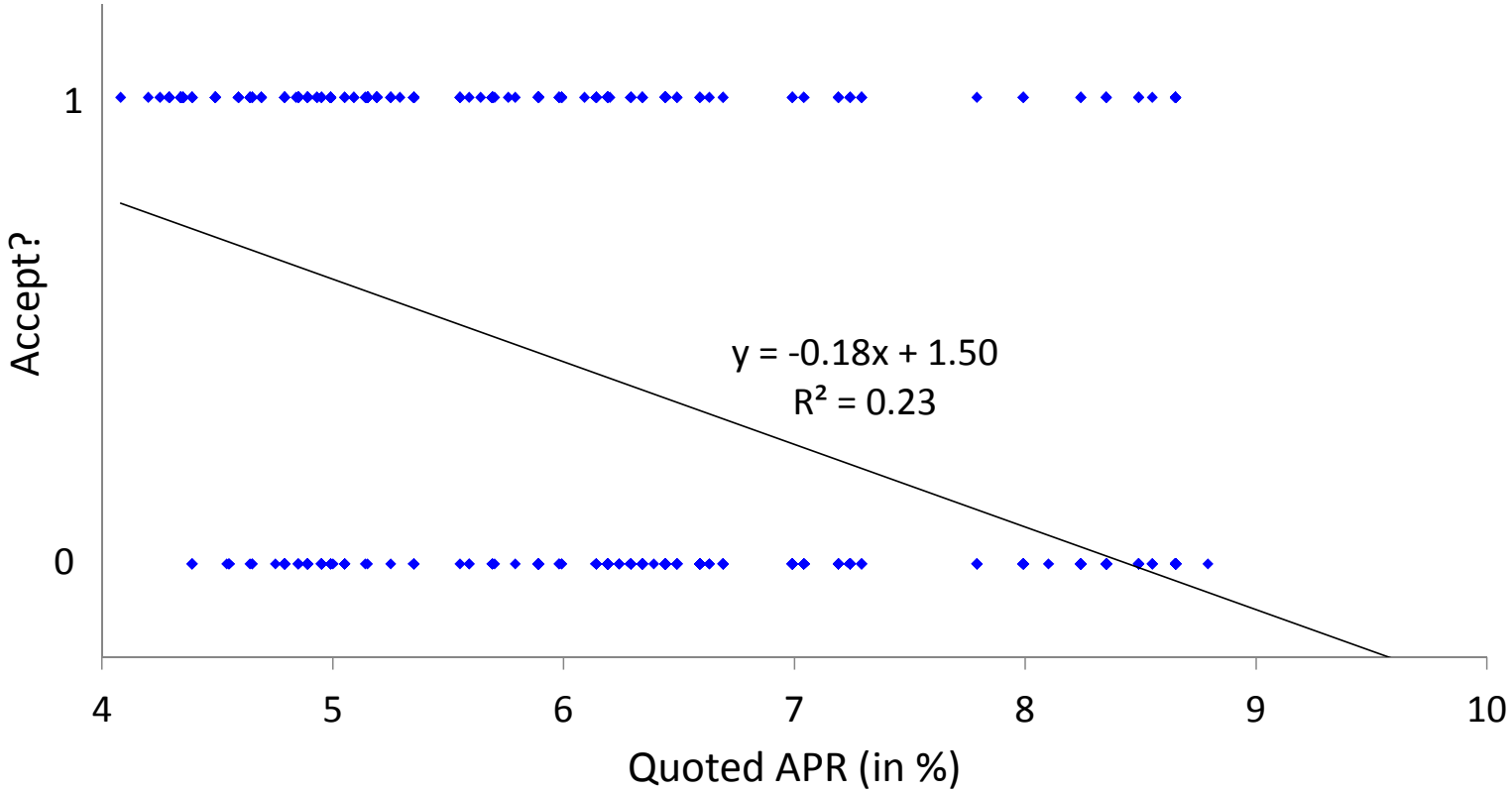
$$\text{Accept}_j = a + b\,\text{APR}_j + \varepsilon_j$$

where APR_j is the APR quoted to customer j and

$$\text{Accept}_j = \begin{cases} 1 & \text{if customer } j \text{ accepted the quote} \\ 0 & \text{otherwise} \end{cases}$$

- Focus on a given profile of customer / loan characteristics (FICO bucket, loan term, amount bucket)
- Linear regression allows loan acceptance to vary with the quoted rate

Accept? versus Quoted APR: Linear Regression



Does this look reasonable?

Accept? versus Quoted APR: Linear Regression

Issues with linear regression

- Response function reflects probabilities: it should be between 0 and 1
- For a binary response variable, there are only two cases
  - Violates the normality assumption
  - Different variance at different levels of quoted APR

Logistic Regression

Logistic regression is a technique for fitting a curve to data in which the dependent variable is binary

Applications

- Response to medical treatment:
  worked (coded as 1) or did not work (coded as 0)
- Customized pricing: bought (1) or not (0)
- Sponsored search: user clicked (1) or did not click (0)

Logistic Regression for Predicting Customer Response

$$\text{Probability of accept given quoted APR} = \frac{\exp(w)}{1+\exp(w)}$$

$$\text{with } w = a + b\,\text{APR}$$

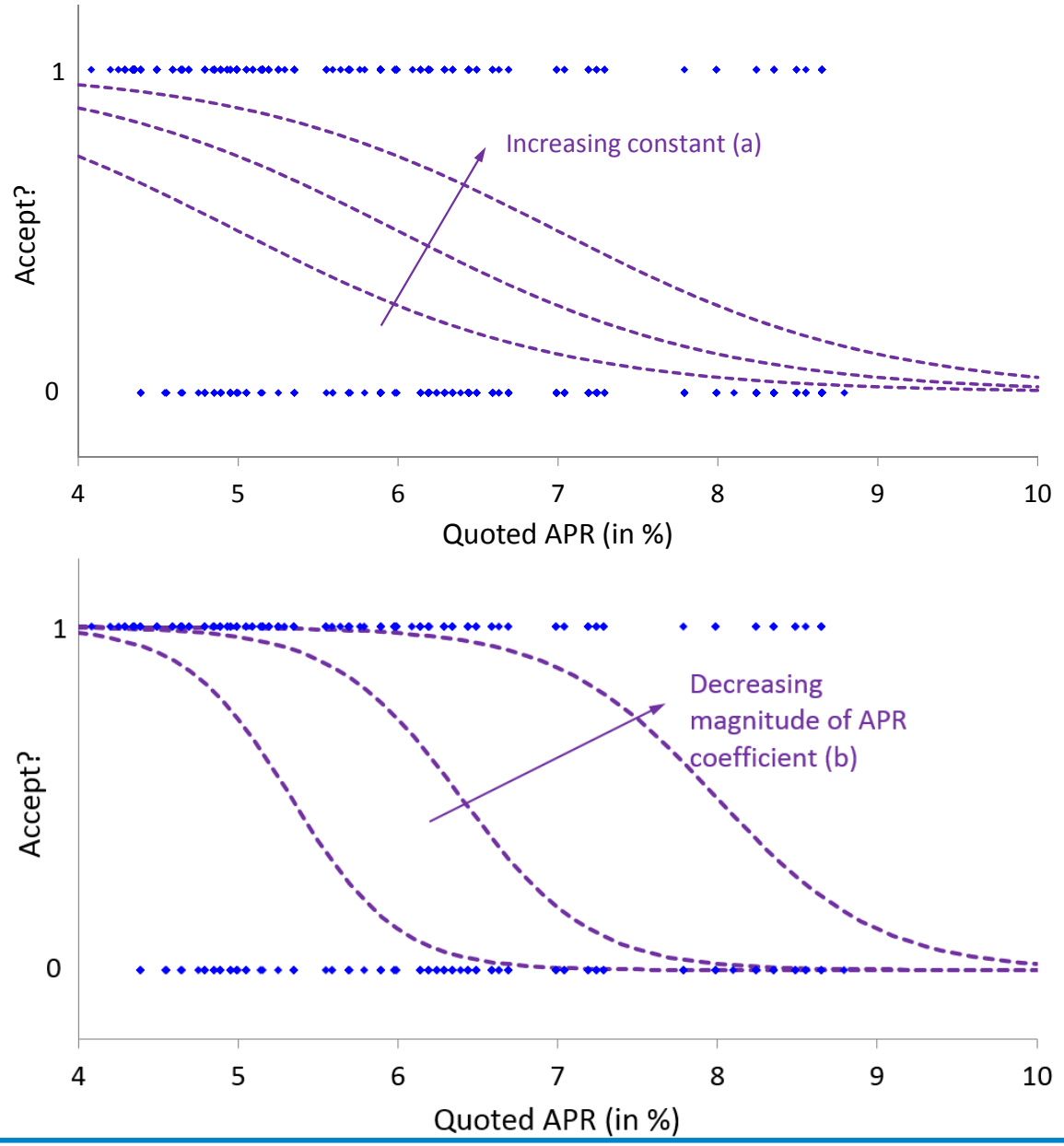
Logistic Regression for Predicting Customer Response

$$\text{Probability of accept given quoted APR} = \frac{\exp(w)}{1+\exp(w)}$$

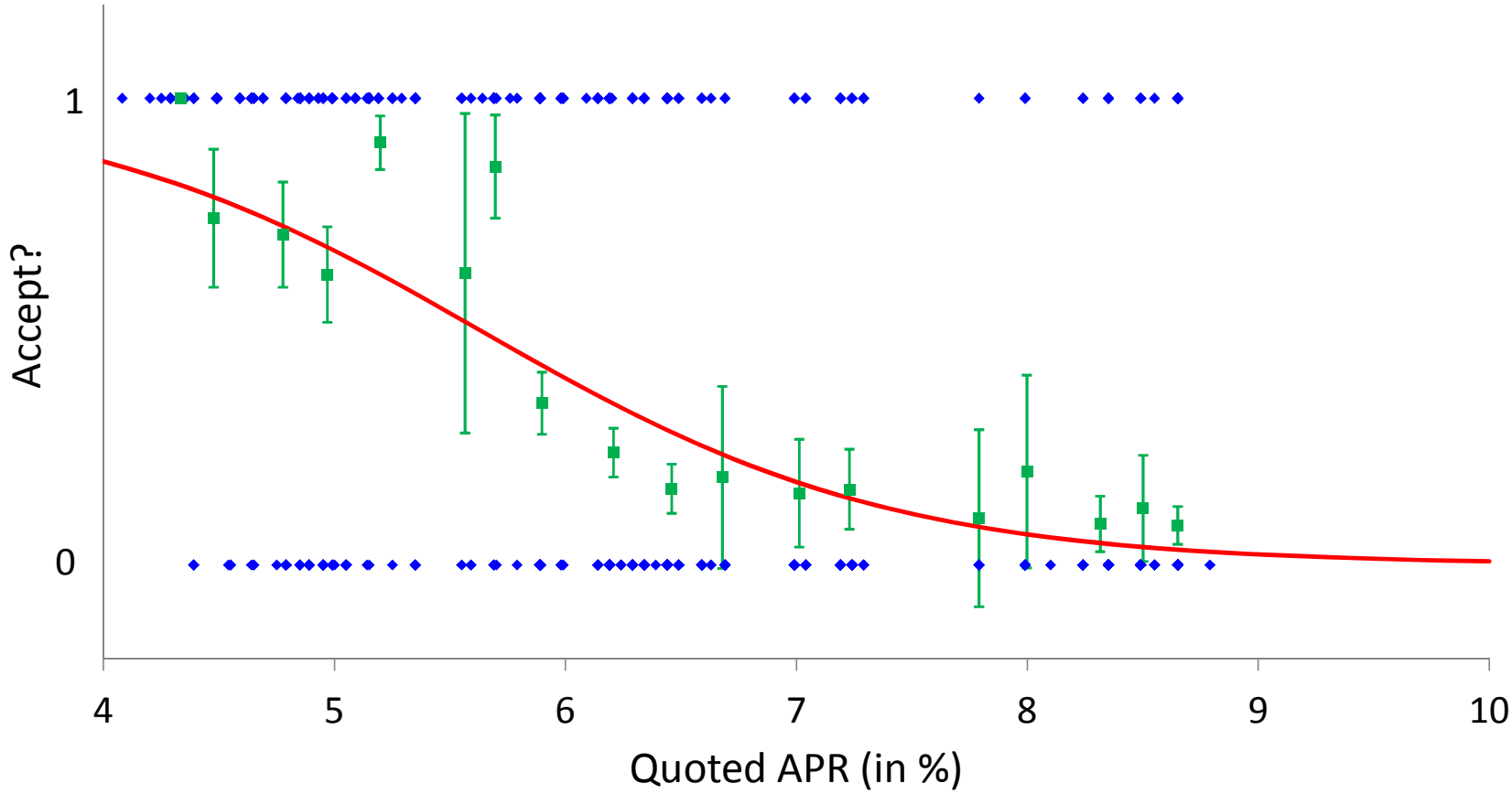
$$\text{with } w = a + b\,\text{APR}$$

- Always between 0 and 1
- Allows for flexible non-linear shapes
- Parameters a and b need to be estimated based on the data
  pick the parameters that "best" explain the data we have

Logistic Regression: Model Flexibility



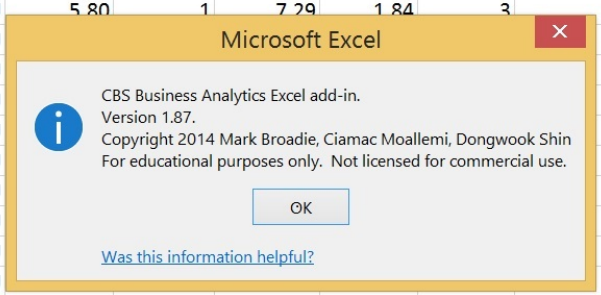
Accept? versus Quoted APR: Logistic Regression



Logistic Regression: Estimating the Parameters

Logistic Regression in Excel: Step 1

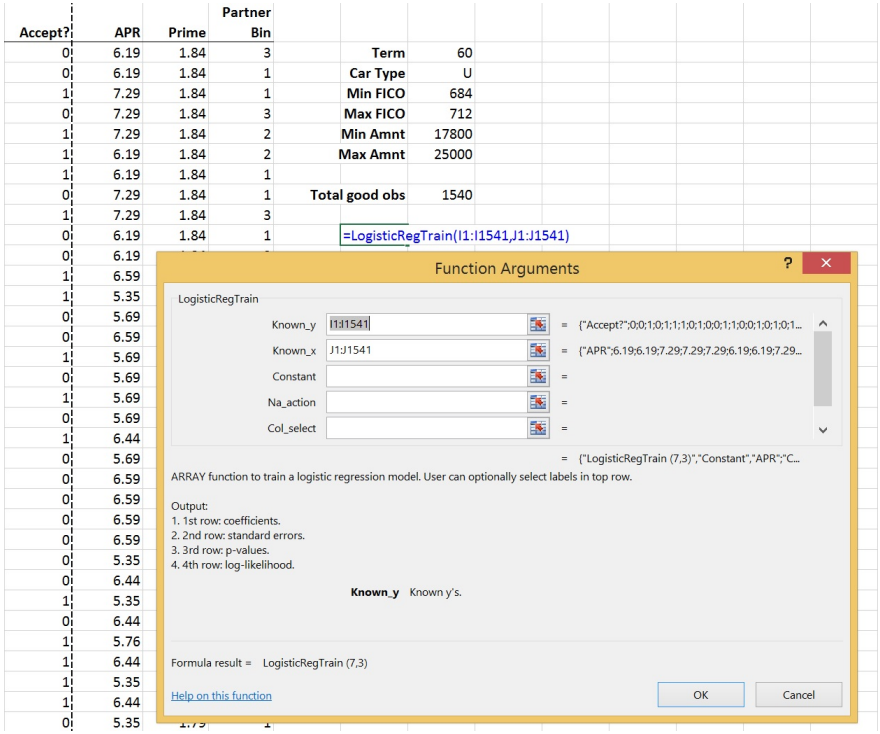
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Tier	FICO	Approve Date	Term	Amount	Previous APR	Car Type	Competition APR	Accept?	APR	Cost of funds	Partner Bin	
2	2	702	7/2/2002	60	22000		U	5.85	0	6.19	1.84	3	
3	2	710	7/3/2002	60	21000		U	5.85	0	6.19	1.84	1	
4	3	693	7/6/2002	60	19598		U	5.85	1	7.29	1.84	1	
5	3	696	7/6/2002	60	23071		U	5.85	0	7.29	1.84	3	
6	3	697	7/8/2002	60	21578		U	5.80	1	7.29	1.84	2	
7	2	702	7/8/2002	60	20211		U	5.80	1	6.19	1.84	2	
8	2	709	7/10/2002	60	21051		U	5.80	1	6.19	1.84	1	
9	3	690	7/12/2002	60	21498		U	5.80	0	7.29	1.84	1	Tota
10	3	684	7/13/2002	60	18000		U	5.80	1	7.29	1.84	3	
11	2	706	7/15/2002	60	24025		U						
12	2	704	7/15/2002	60	23000		U						
13	3	693	7/16/2002	60	19620		U						
14	2	702	7/17/2002	60	24227		U						
15	2	705	7/18/2002	60	21262		U						
16	3	692	7/19/2002	60	25000		U						
17	2	702	7/19/2002	60	19071		U						
18	2	705	7/22/2002	60	20000		U						
19	2	709	7/22/2002	60	21678		U						



Step 1: Load the CBS Business Analytics add-in

- Double-click cbs_ba.xll (or use Excel options | add-ins to automatically load every time)

Logistic Regression in Excel: Step 2



Step 2

- Type `=LogisticRegTrain` and press the f_x (Insert function) icon
- Select input variable column (APR) and output variable column (Accept?)
- Select 7 rows and 3 columns; press F2 (edit) and CNTL-SHIFT-ENTER
- See videos on Canvas for more detail

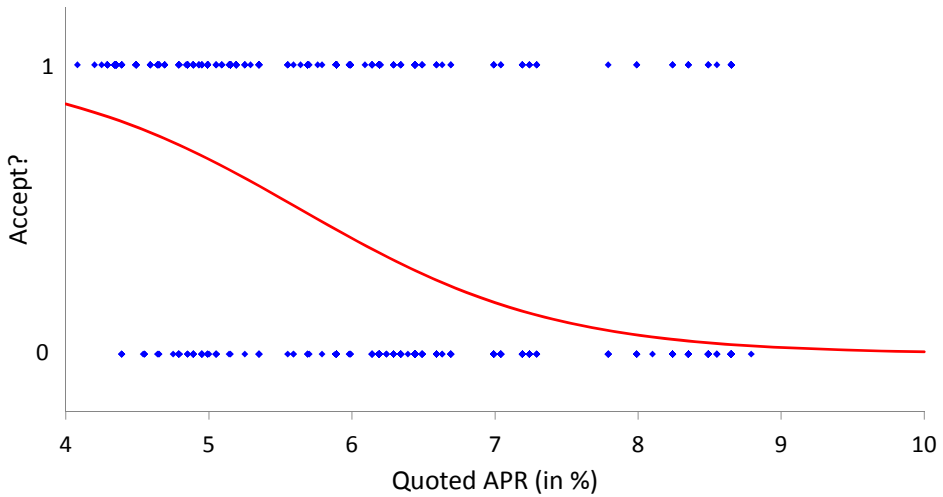
Logistic Regression in Excel: Step 3 (Output)

LogisticRegTrain (7,3)	Constant	APR
Coefficients	6.36	-1.13
Std error	0.42	0.07
p-value	1.3E-51	7.8E-59
Log-likelihood	-786.3	
Number valid obs	1540	
Total obs	1540	

Interpreting the output:

Prob. of accept given quoted APR

$$= \frac{\exp(6.36 - 1.13 \times \text{APR})}{1 + \exp(6.36 - 1.13 \times \text{APR})}$$



Logistic Regression in R

```
>nomis_log_reg = glm( formula = Accept. ~ APR, family = binomial)
>summary(nomis_log_reg)
Call:
glm(formula = Accept. ~ APR, family = binomial, data = nomis_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8044  -0.8909  -0.3032   0.8859   3.8847

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.36032    0.42079   15.12   <2e-16 ***
APR         -1.12777    0.06973  -16.17   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2000.3  on 1539  degrees of freedom
Residual deviance: 1572.6  on 1538  degrees of freedom
AIC: 1576.6

Number of Fisher Scoring iterations: 5
```

Moving to Decisions: Setting the APR for Used Car Loans

e-Car's objective: assign "best" rate to each incoming customer

Maximize

$$(\text{Net Revenue}(\text{APR})) \times (\text{Probability accept quoted APR})$$

Exercise

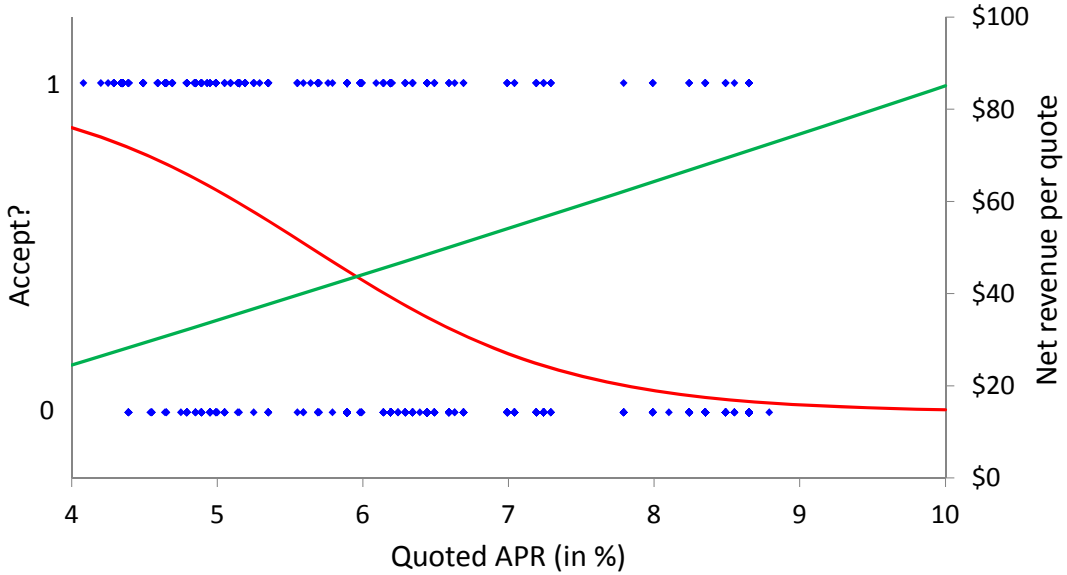
Assume prime = 1.40%, loan amount = $21,000

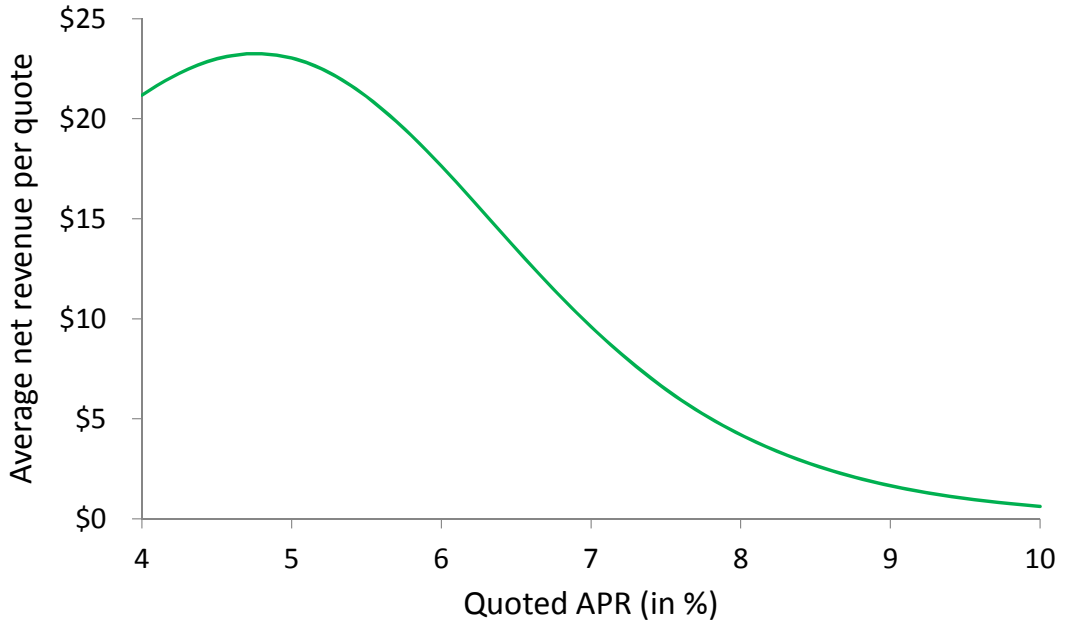
Compute expected net revenue per customer per month for:

- APR = 6.00 (mode of rates offered in segment)
  Payments per month from customer to e-Car if accept: $406
  Cost per month to e-Car: $363

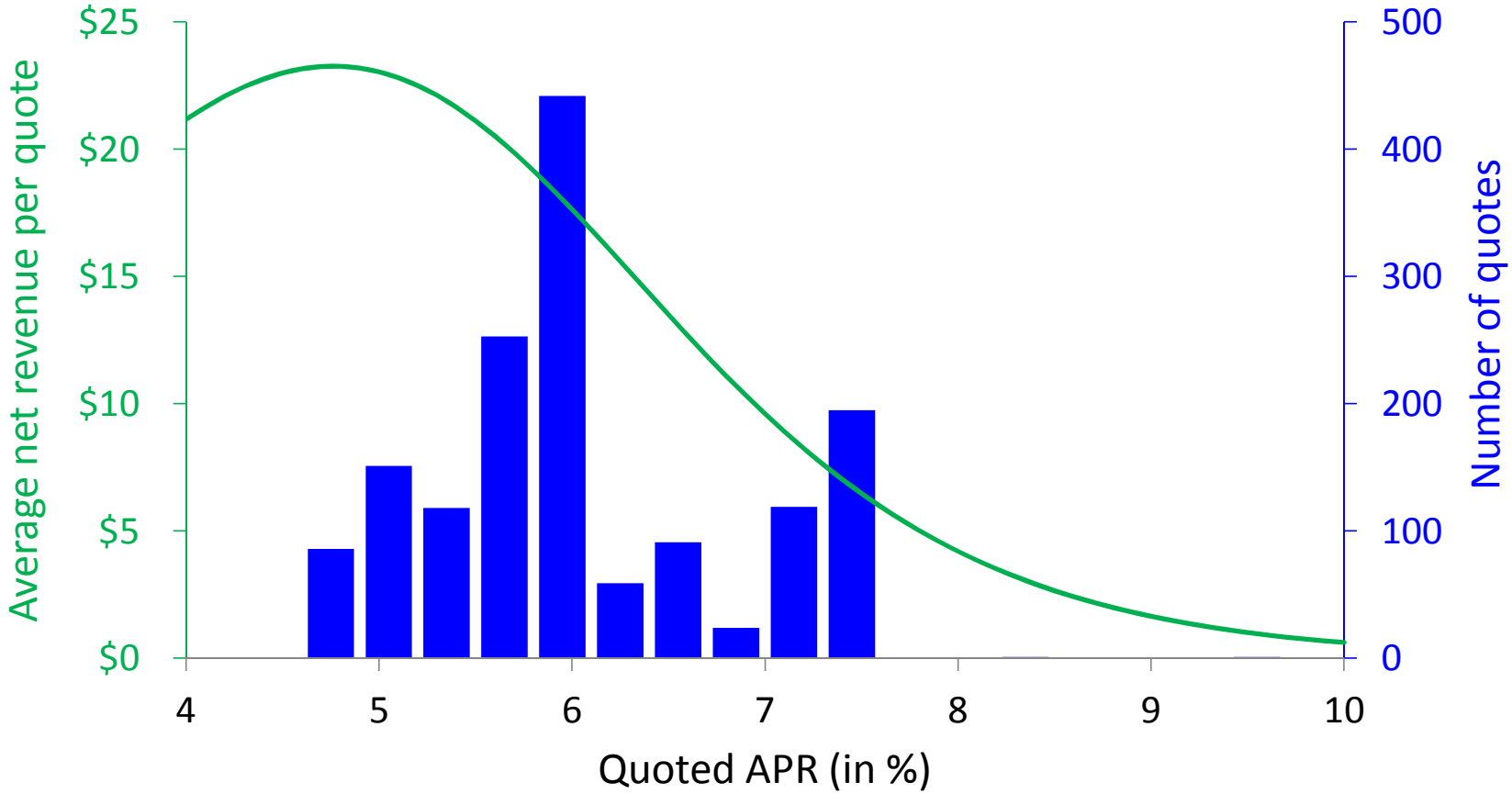
- APR = 4.75 (alternative rate)
  Payments per month from customer to e-Car if accept: $393
  Cost per month to e-Car: $363

Average Net Revenue per Quote vs. Quoted APR





The Opportunity



Logistic Regression: General Formula

In general, there could be multiple independent variables $(x_1, ..., x_k)$ (as in linear regression)

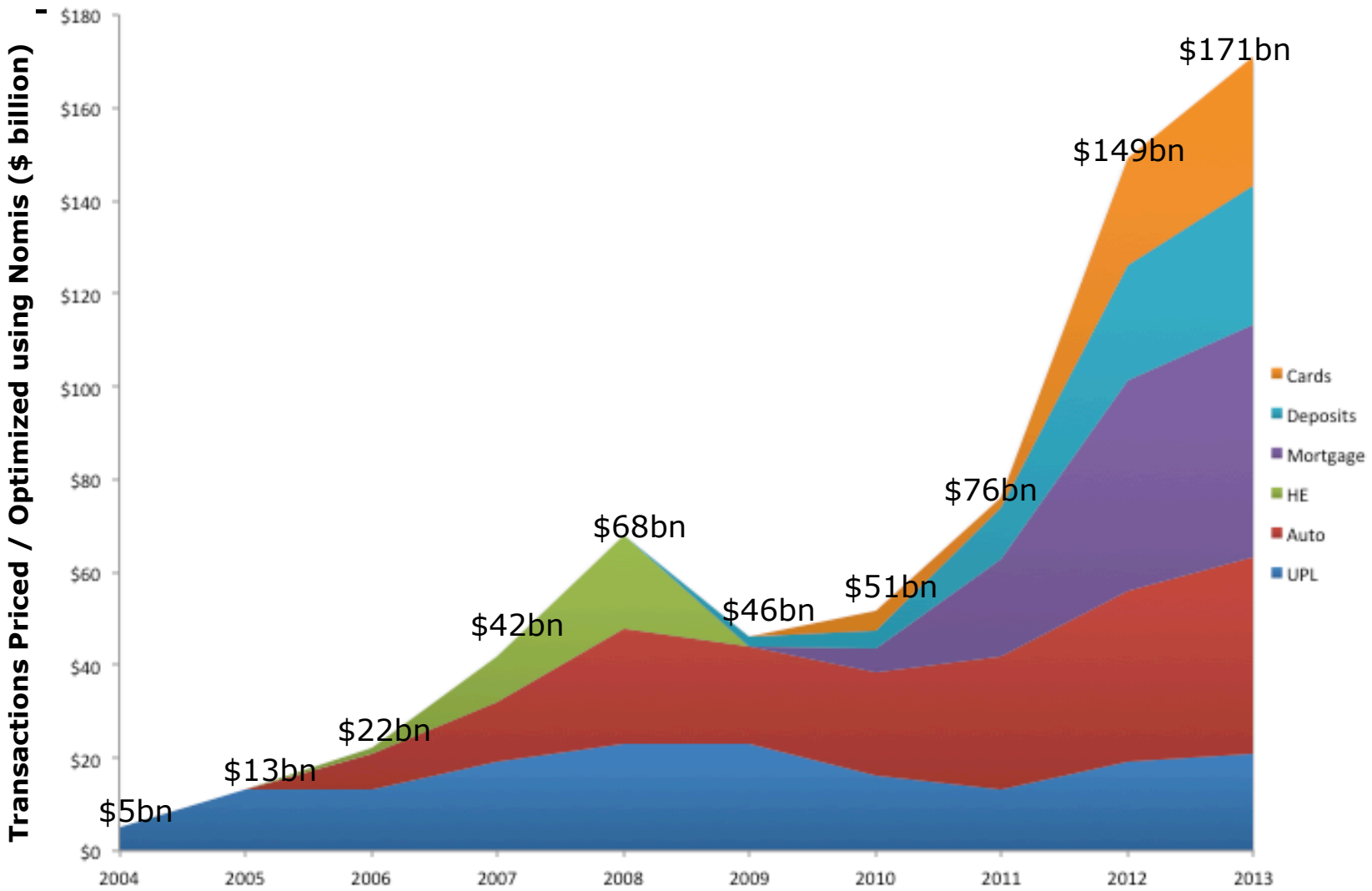
Probability of success given $(x_1, ..., x_k)$ is modeled as

$$p(x_1, ..., x_k) = \frac{\exp(w)}{1 + \exp(w)}$$

$$\text{with } w = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k$$

Parameters $a, b_1, ..., b_k$ need to be estimated

Nomis Solutions Today



In 2015, more than $1 trillion transactions priced

Summary

Predictive Analytics

- Attempt to capture and exploit relationships in the data to:
  - Predict customer decisions
  - Improve decision making

Logistic regression

- Model probabilistic behavior of customers with 0/1 outcomes
- Select the parameters that maximize the likelihood of observing the data

Many business models are supported by or rely on predictive analytics

Other Applications of Logistic Regression

Online advertising (e.g., Google / `www.google.com`)

Business model:

- Sponsored search: cost per click (CPC)
- Key question: what drives clicks (quality of ad,...)

Online Content Recommendations (e.g., Outbrain, Inc. / `www.outbrain.com`)

Business model:

- Users "sold" from site to another
- Revenue sharing between operator and source publisher
- Key revenue driver of recommendation operator: clicks
- Key question: what drives clicks (context, exploration, clicks of other users, etc.)