

CMC Satisfaction

Applied Regression Analysis – Columbia Business School

Daniel Busby

Christian Dunne

Tullio Purtill

Gabriel Souza Adde

Sean Tallon

Travis Vollmer

Jennifer Zhou

February 28th, 2017

Executive Summary

- Proposition
- Methodology
- Representativeness of collected data
- Basic Statistics Highlights
- A Full Regression Model with Dummy Variables
- Confidence Intervals
- Ordinal Logistic Regression
- Extra: Discriminant Analysis
- Conclusions
- Appendix


“Word-of-mouth” implies students dislike CMC

- Upon talking to current MBA students, we feel there is a general level of dissatisfaction with the services provided by the CMC;
- We intend to understand where that dissatisfaction comes from. In other words:

Is there a statistically sound way to determine who is more prone to liking/disliking CMC services?


A Survey in which students rate CMC and provide details about their lives at CBS

Applied Regression Analysis - CBS and your Career

1. Grades prior to CBS 

GMAT

Undergraduate GPA (if you're international, please make your best effort to convert to a scale of 0.0 to 4.0)

2. Hometown Geography 

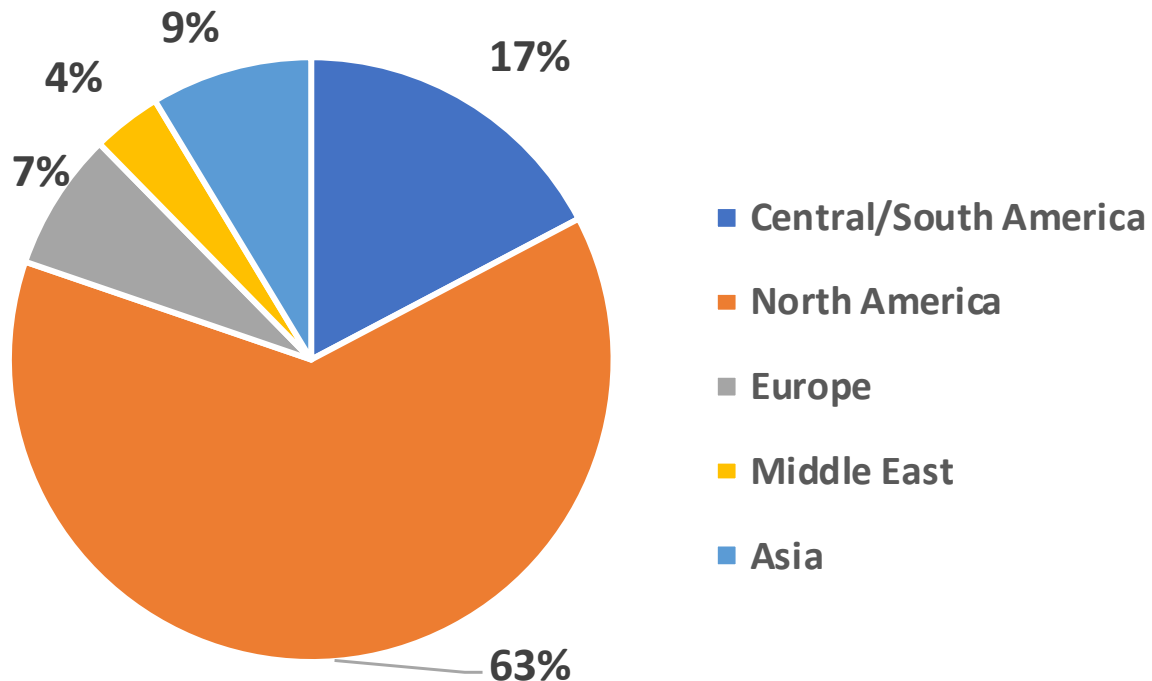
☐ North America

☐ Central/South America

☐ Africa

**81 respondents in
around a week**

Aggregate data representative of broad CBS community



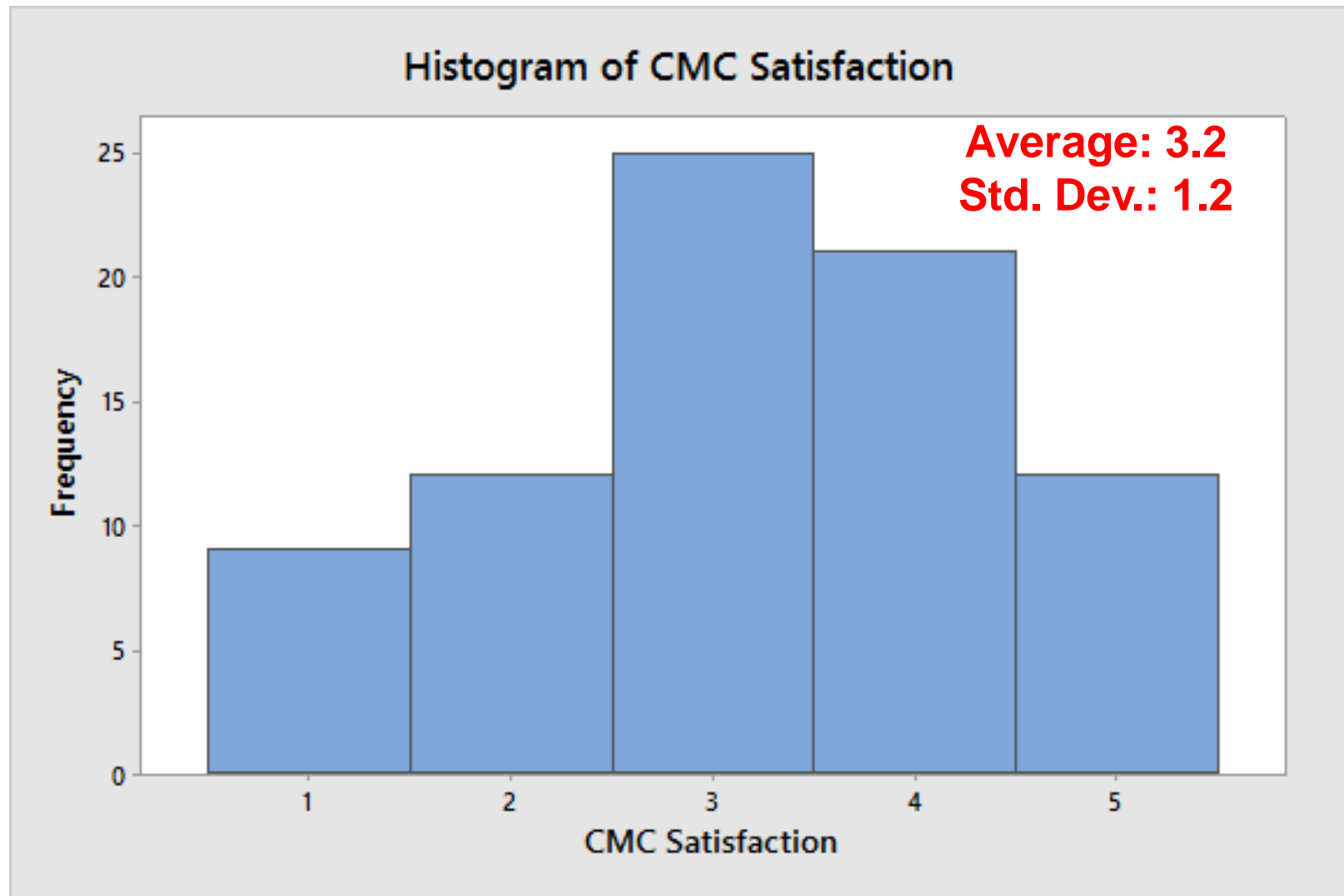
80% GMAT Range:
[690-760]

Average GPA:
3.6

- **73 Fall-Term Students**
- **61 Students from Class of 2017**

Collected Data Highlights

CMC Satisfaction: Please rate, on a scale of 1 to 5, your overall satisfaction with CMC



Collected Data Highlights

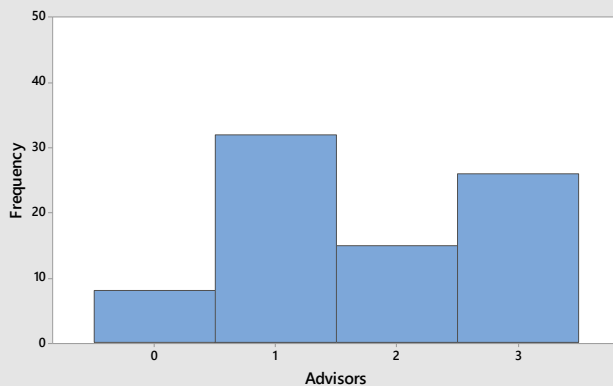
CF: # of interactions with Career Fellows

Workshops: # of CMC workshops one participated

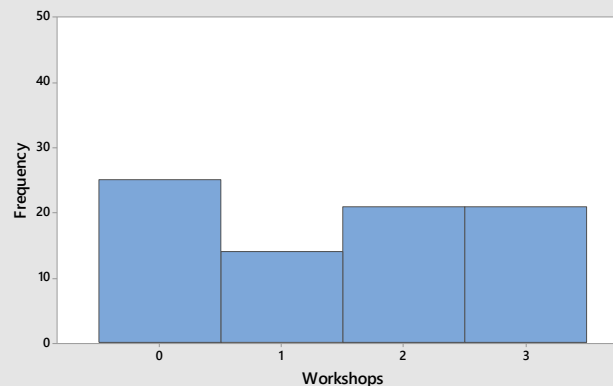
Coaches/EIR: # of interactions with Coached or Executives-in-Residence

Advisors: # of interactions with CMC Advisors

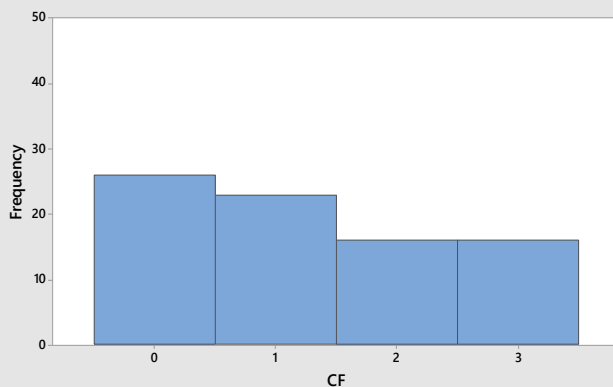
Histogram of Advisors



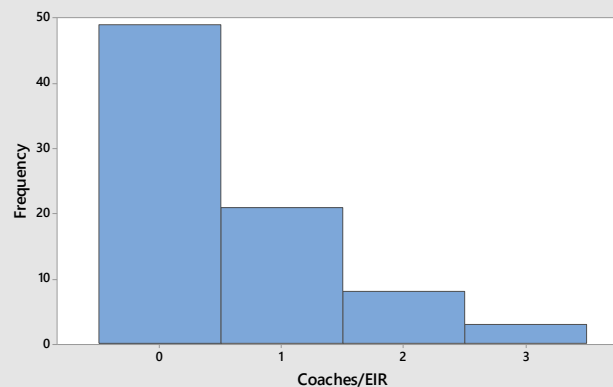
Histogram of Workshops



Histogram of CF



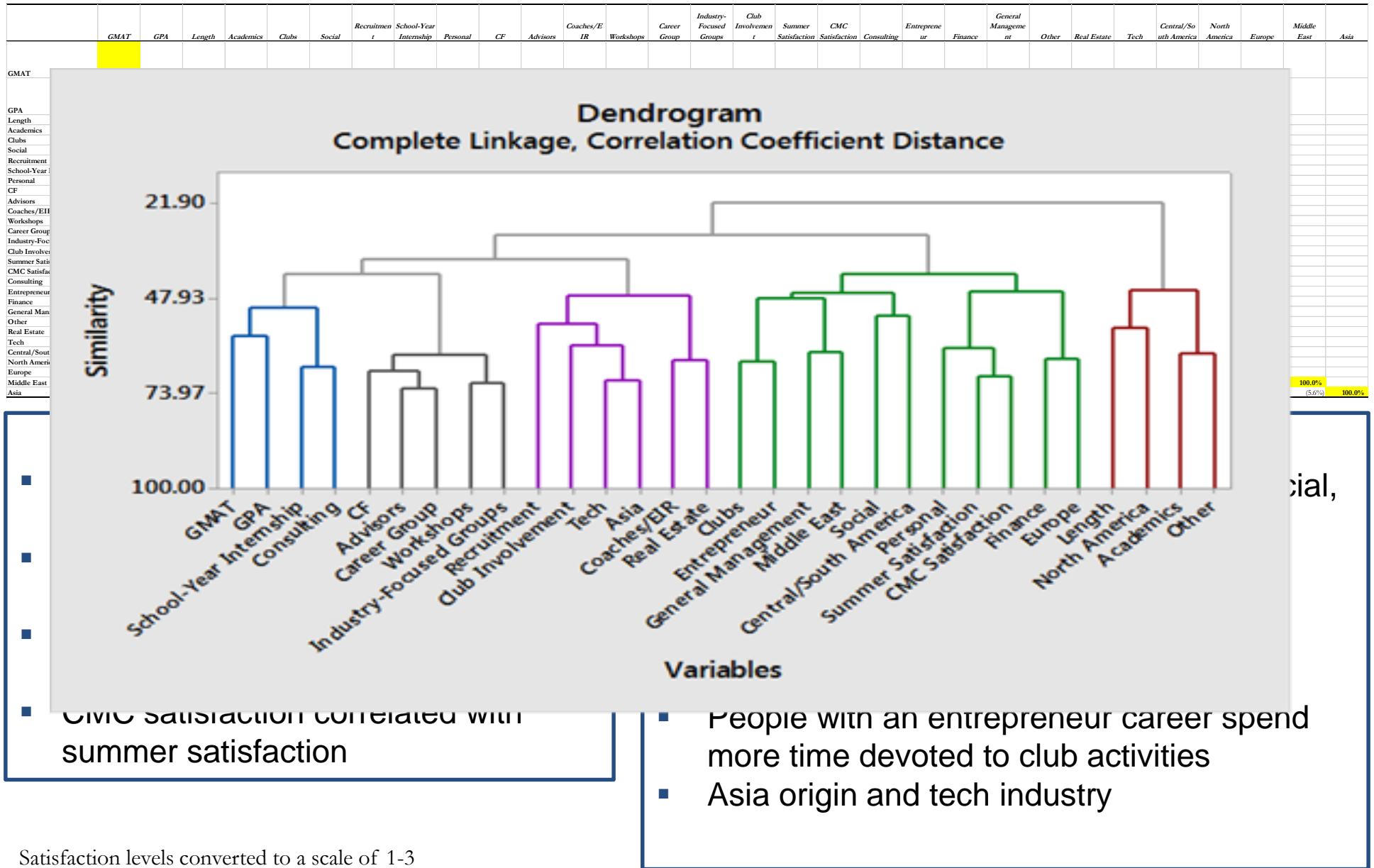
Histogram of Coaches/EIR



Findings: Out of all the resources the CMC provides, the most heavily used were the Advisors with 73 of 81 total students using this service.

Insights: This could imply that the students value the resource most and potentially are scoring the CMC based off of this interaction.

Collected Data Highlights – Correlation Matrix



Satisfaction levels converted to a scale of 1-3
Disregards incomplete surveys

CMC seems to start with Positive Goodwill...

Regression Analysis: CMC Satisfaction versus FT Satisfaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	9.2083	9.2083	8.55	0.006
FT Satisfaction	1	9.2083	9.2083	8.55	0.006
Error	33	35.5345	1.0768		
Lack-of-Fit	4	0.8774	0.2193	0.18	0.945
Pure Error	29	34.6571	1.1951		
Total	34	44.7429			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.03769	20.58%	18.17%	13.40%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.746	0.620	2.82	0.008	
FT Satisfaction	0.417	0.143	2.92	0.006	1.00

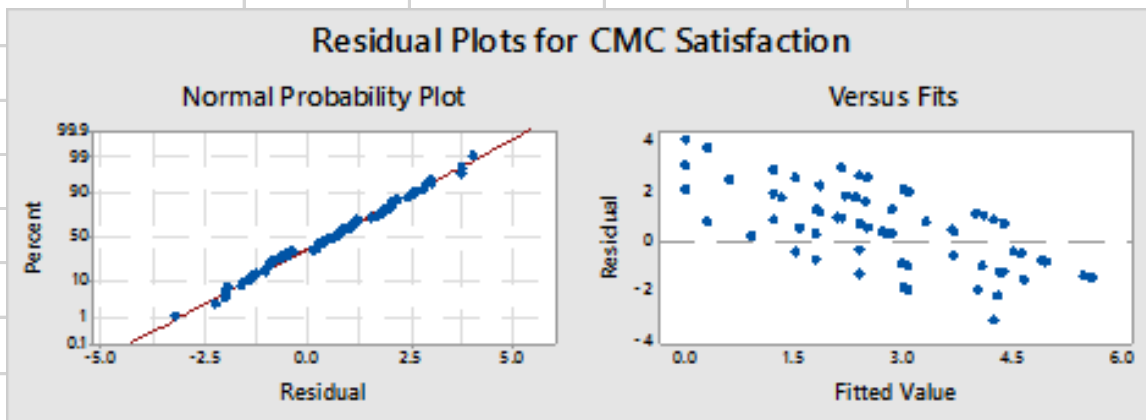
Regression Equation

CMC Satisfaction = 1.746 + 0.417 FT Satisfaction

FT Satisfaction:
Please rate on a
scale of 1 to 5 your
satisfaction with
your Full Time offer

A full-model to explain CMC Satisfaction...

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R	0.8729					
R Square	0.7619					
Adjusted R Square	0.7344					
Standard Error	1.7216					
Observations	76					
ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	5	673.5563	134.7113	45.4492	0.0000	
Residual	71	210.4437	2.9640			
Total	76	884				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
Recruitment	0.0597	0.0100	5.9807	0.0000	0.0398	0.0796
Active	1.5041	0.6370	2.3613	0.0210	0.2340	2.7741
Very Active	2.4771	0.5267	4.7030	0.0000	1.4269	3.5273
On the Board	1.2471	0.4801	2.5977	0.0114	0.2899	2.2044
2018	1.3050	0.4689	2.7834	0.0069	0.3701	2.2399



Disregards incomplete surveys

A full-model to explain CMC Satisfaction...

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R						
R Square						
Adjusted R Square						
Standard Error						
Observations						
ANOVA						
	<i>df</i>				<i>Significance F</i>	
Regression					0.0000	
Residual						
Total						
		<i>Standard Error</i>	<i>t Stat</i>		<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept		#N/A	#N/A		#N/A	#N/A
Recruitment		0.0100	5.9807	0.0000	0.0398	0.0796
Active	1.5041	0.6370	2.3613	0.0210	0.2340	2.7741
Very Active	2.4771	0.5267	4.7030	0.0000	1.4269	3.5273
On the Board	1.2471	0.4801	2.5977	0.0114	0.2899	2.2044
2018	1.3050	0.4689	2.7834	0.0069	0.3701	2.2399

**INTERCEPT HAD
BEEN SET TO ZERO
WHEN IN FACT IT
WAS
STATISTICALLY
SIGNIFICANT**

Disregards incomplete surveys

A full-model to explain CMC Satisfaction... 2nd attempt

■ Categorical Data converted to dummy variables

#	CMC Satisfaction	Recruitment	Club Involvement	Very Active Member	On the Board	Active Member	Class	2018	Geography	Central/South America	North America	Europe	Middle East
1	5	25	Very Active Member	1	0	0	2017	0	Central/South America	1	0	0	0
3	2	30	On the Board	0	1	0	2017	0	North America	0	1	0	0
4	3	25	On the Board	0	1	0	2018	1	North America	0	1	0	0
5	4	20	Very Active Member	1	0	0	2017	0	North America	0	1	0	0
6	2	5	On the Board	0	1	0	2017	0	North America	0	1	0	0
7	3	40	On the Board	0	1	0	2017	0	North America	0	1	0	0
8	4	35	Active Member	0	0	1	2018	1	North America	0	1	0	0
9	2	40	Passive Member	0	0	0	2017	0	North America	0	1	0	0
10	1	50	On the Board	0	1	0	2017	0	North America	0	1	0	0
11	4	0	Passive Member	0	0	0	2017	0	Central/South America	1	0	0	0
12	4	0	Active Member	0	0	1	2018	1	Central/South America	1	0	0	0
13	2	0	Passive Member	0	0	0	2017	0	Central/South America	1	0	0	0
14	3	15	On the Board	0	1	0	2017	0	Europe	0	0	1	0
15	4	30	Active Member	0	0	1	2017	0	North America	0	1	0	0
16	2	50	Passive Member	0	0	0	2017	0	Middle East	0	0	0	1
17	1	25	Passive Member	0	0	0	2017	0	North America	0	1	0	0
18	3	40	Passive Member	0	0	0	2017	0	Asia	0	0	0	0
19	5	25	Very Active Member	1	0	0	2017	0	North America	0	1	0	0
20	5	50	On the Board	0	1	0	2017	0	Asia	0	0	0	0
21	1	15	Passive Member	0	0	0	2017	0	Europe	0	0	1	0
23	5	15	On the Board	0	1	0	2017	0	North America	0	1	0	0
24	5	30	On the Board	0	1	0	2017	0	Asia	0	0	0	0
25	3	20	Passive Member	0	0	0	2017	0	North America	0	1	0	0
26	4	5	Passive Member	0	0	0	2017	0	North America	0	1	0	0
27	4	40	Very Active Member	1	0	0	2017	0	North America	0	1	0	0
28	4	25	Passive Member	0	0	0	2017	0	Europe	0	0	1	0
29	1	25	Passive Member	0	0	0	2017	0	North America	0	1	0	0
30	3	10	On the Board	0	1	0	2017	0	North America	0	1	0	0
31	4	20	Passive Member	0	0	0	2017	0	Europe	0	0	1	0
32	5	10	Very Active Member	1	0	0	2017	0	North America	0	1	0	0
33	1	5	Passive Member	0	0	0	2017	0	North America	0	1	0	0
34	3	30	Passive Member	0	0	0	2017	0	North America	0	1	0	0

A full-model to explain CMC Satisfaction... 2nd attempt

- Used Best Subsets function with # of variables equal to 30... Took a long time!

Regression Analysis: CMC Satisfac versus Recruitment, Very Active , Active Membe, 2018, ...

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	15	61.925	4.1283	4.98	0.000
Recruitment	1	6.006	6.0063	7.25	0.009
Very Active Member	1	23.386	23.3858	28.21	0.000
Active Member	1	3.179	3.1792	3.84	0.055
2018	1	1.709	1.7093	2.06	0.156
Tech	1	6.299	6.2990	7.60	0.008
Entrepreneur	1	3.693	3.6935	4.46	0.039
Length	1	5.759	5.7589	6.95	0.011
Academics	1	11.215	11.2153	13.53	0.000
Clubs	1	3.665	3.6655	4.42	0.039
Social	1	8.394	8.3938	10.13	0.002
School-Year Internship	1	11.217	11.2166	13.53	0.000
Industry-Focused Groups	1	2.737	2.7369	3.30	0.074
Moderately Satisfied_FT	1	1.582	1.5818	1.91	0.172
Dissatisfied_FT	1	4.856	4.8561	5.86	0.018
On the Board	1	1.636	1.6360	1.97	0.165
Error	63	52.227	0.8290		
Total	78	114.152			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.910494	54.25%	43.35%	*

A full-model to explain CMC Satisfaction... 2nd attempt

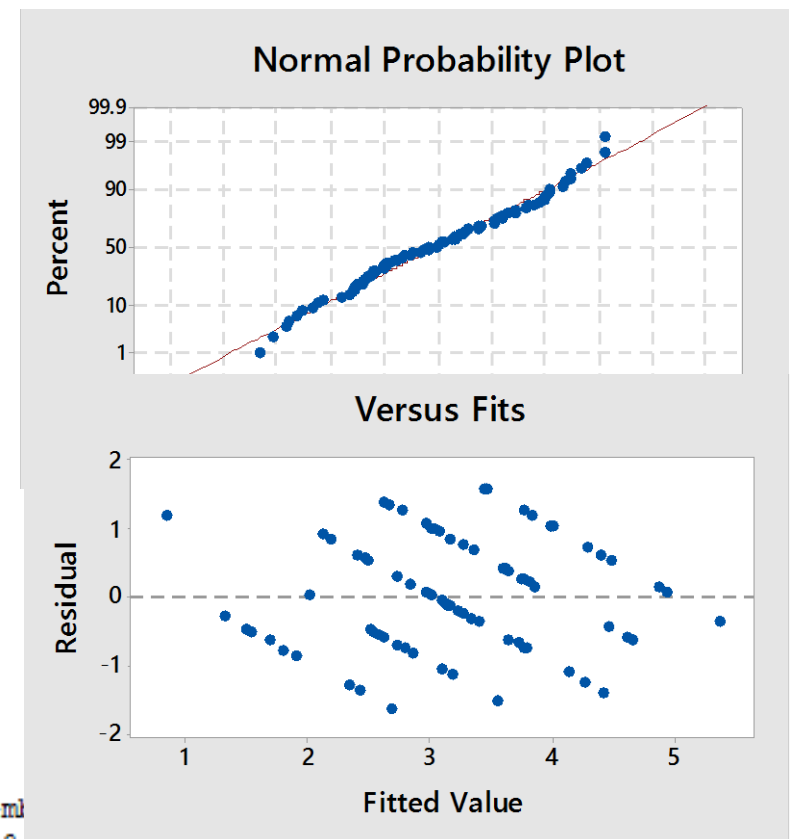
- Used Best Subsets function with # of variables equal to 30... Took a long time!

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	7.74	1.19	6.48	0.000	
Recruitment	-0.0343	0.0127	-2.69	0.009	3.20
Very Active Member	1.767	0.333	5.31	0.000	1.45
Active Member	0.727	0.371	1.96	0.055	1.45
2018	0.393	0.273	1.44	0.156	1.30
Tech	-1.416	0.514	-2.76	0.008	1.21
Entrepreneur	-2.16	1.02	-2.11	0.039	1.25
Length	-0.2089	0.0792	-2.64	0.011	1.25
Academics	-0.0521	0.0142	-3.68	0.000	3.86
Clubs	-0.0404	0.0192	-2.10	0.039	2.07
Social	-0.0516	0.0162	-3.18	0.002	2.27
School-Year Internship	-0.0631	0.0171	-3.68	0.000	1.72
Industry-Focused Groups	0.199	0.110	1.82	0.074	1.13
Moderately Satisfied_FT	-0.582	0.421	-1.38	0.172	1.37
Dissatisfied_FT	-1.669	0.690	-2.42	0.018	1.12
On the Board	0.383	0.272	1.40	0.165	1.38

Regression Equation

CMC Satisfaction = 7.74 - 0.0343 Recruitment + 1.767 Very Active Mem
+ 0.393 2018 - 1.416 Tech - 2.16 Entrepreneur - 0.2089 Length
- 0.0521 Academics - 0.0404 Clubs - 0.0516 Social
- 0.0631 School-Year Internship + 0.199 Industry-Focused Groups
- 0.582 Moderately Satisfied_FT - 1.669 Dissatisfied_FT
+ 0.383 On the Board



A full-model to explain CMC Satisfaction... 3rd attempt

- Used Backward Regression and required $p < 10\%$

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	10	52.136	5.2136	5.72	0.000
Recruitment	1	4.135	4.1353	4.53	0.037
Very Active Member	1	20.401	20.4010	22.37	0.000
2018	1	4.904	4.9040	5.38	0.023
Tech	1	7.457	7.4570	8.18	0.006
Entrepreneur	1	6.262	6.2620	6.87	0.011
Length	1	2.565	2.5652	2.81	0.098
Academics	1	9.831	9.8312	10.78	0.002
Social	1	8.711	8.7114	9.55	0.003
School-Year Internship	1	8.669	8.6695	9.51	0.003
Dissatisfied_FT	1	5.721	5.7207	6.27	0.015
Error	68	62.016	0.9120		
Lack-of-Fit	67	61.516	0.9181	1.84	0.537
Pure Error	1	0.500	0.5000		
Total	78	114.152			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.954984	45.67%	37.68%	*

A full-model to explain CMC Satisfaction... 3rd attempt

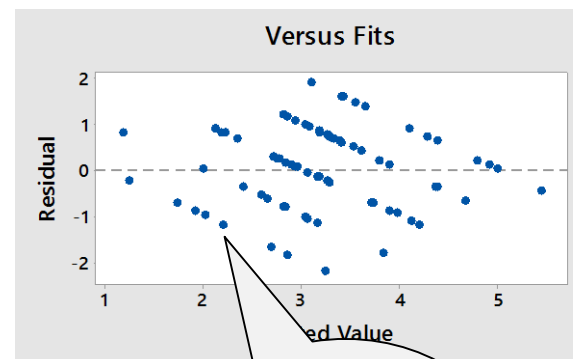
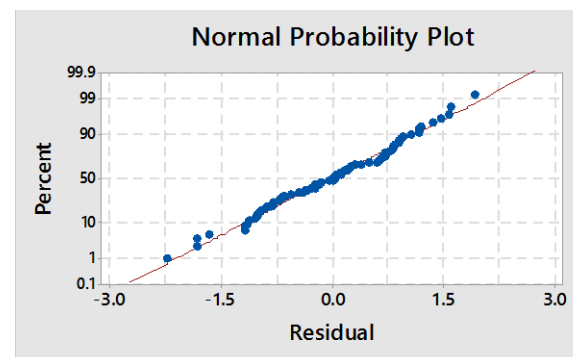
- Used Backward Regression and required $p < 10\%$

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	6.449	0.919	7.02	0.000	
Recruitment	-0.0238	0.0112	-2.13	0.037	2.24
Very Active Member	1.515	0.320	4.73	0.000	1.22
2018	0.620	0.267	2.32	0.023	1.13
Tech	-1.492	0.522	-2.86	0.006	1.13
Entrepreneur	-2.73	1.04	-2.62	0.011	1.18
Length	-0.1328	0.0792	-1.68	0.098	1.13
Academics	-0.0381	0.0116	-3.28	0.002	2.36
Social	-0.0484	0.0157	-3.09	0.003	1.92
School-Year Internship	-0.0498	0.0162	-3.08	0.003	1.39
Dissatisfied_FT	-1.787	0.713	-2.50	0.015	1.09

Regression Equation

CMC Satisfaction = 6.449 - 0.0238 Recruitment + 1.515 Very Active Member
 - 1.492 Tech - 2.73 Entrepreneur - 0.1328 Length - 0.0381 Academics
 - 0.0484 Social - 0.0498 School-Year Internship - 1.787 Dissatisfied_FT



More
on this
later...

Confidence Intervals for grading differences

- Using pooled differences, we computed 95% CI for the grading differences among different groups of people:

	#	MEAN	STDEV
Candidates who got a summer job	68	3.13	1.22
Candidates who didn't	11	3.55	1.13
Difference		(0.41)	0.37
95% CI for Difference		(1.14)	0.31

$$Std. Dev (\bar{X} - \bar{Y}) = \sqrt{\frac{Var(X)}{n_x} + \frac{Var(Y)}{n_y}}$$

Confidence Intervals for grading differences

- Using pooled differences, we computed 95% CI for the grading differences among different groups of people:

Summer Outcome	#	MEAN	STDEV
Dissatisfied	6	1.83	0.75
Others	73	3.30	1.17
Difference		(1.468)	0.337
95% CI for Difference		(2.13)	(0.81)

$$Std. Dev (\bar{X} - \bar{Y}) = \sqrt{\frac{Var(X)}{n_x} + \frac{Var(Y)}{n_y}}$$

Confidence Intervals for grading differences

- Using pooled differences, we computed 95% CI for the grading differences among different groups of people:

Summer Outcome	#	MEAN	STDEV
Totally Satisfied	37	3.49	1.15
Others	42	2.93	1.22
Difference		0.558	0.266
95% CI for Difference		0.04	1.08

$$Std. Dev (\bar{X} - \bar{Y}) = \sqrt{\frac{Var(X)}{n_x} + \frac{Var(Y)}{n_y}}$$

Ordinal Logistic Regression

- Due to the nature of our dataset, it seemed proper to conduct an Ordinal Logistic Regression
- Think of a Binary Logistic Regression (as we learned in class), but one in which the output can take discrete values other than 0,1

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
CMC Satisfaction	1	9
	2	12
	3	23
	4	20
	5	12
	Total	76

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	4.40045	5.88209	0.75	0.454			
Const(2)	5.98107	5.89625	1.01	0.310			
Const(3)	8.07131	5.93406	1.36	0.174			
Const(4)	10.3470	5.98297	1.73	0.084			
GMAT	-0.0201635	0.0074874	-2.69	0.007	0.98	0.97	0.99
Academics	0.0909988	0.0320438	2.84	0.005	1.10	1.03	1.17
Clubs	0.0816244	0.0394737	2.07	0.039	1.09	1.00	1.17
Social	0.130411	0.0374892	3.48	0.001	1.14	1.06	1.23
Recruitment	0.0927185	0.0278801	3.33	0.001	1.10	1.04	1.16
School-Year Internship	0.162392	0.0432947	3.75	0.000	1.18	1.08	1.28
Workshops	-0.912855	0.225515	-4.05	0.000	0.40	0.26	0.62
Entrepreneur	6.56630	2.18769	3.00	0.003	710.73	9.76	51745.48
Tech	2.31319	1.17580	1.97	0.049	10.11	1.01	101.26
Very Active Member	-3.84689	0.823698	-4.67	0.000	0.02	0.00	0.11

Ordinal Logistic Regression

Log-Likelihood = -87.698

Test that all slopes are zero: G = 59.987, DF = 10, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	301.871	290	0.304
Deviance	175.397	290	1.000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	1842	82.3	Somers' D	0.65
Discordant	388	17.3	Goodman-Kruskal Gamma	0.65
Ties	9	0.4	Kendall's Tau-a	0.51
Total	2239	100.0		

Extra: Discriminant Analysis for A New Variable: Got FT?

Predictors: GMAT, GPA, Central/South America, North America, Europe, Middle East, 2018, J-Term, Finance, Tech, Consulting, Real Estate, General Management, Entrepreneur, Length, Academics, Clubs, Social, Recruitment, School-Year Internship, CF, Very Active Member, Active Member, On the Board, Totally Satisfied, Dissatisfied, Moderately Satisfied, Sponsored

Group	0	1
Count	37	39

Summary of classification

	True Group	
Put into Group	0	1
0	32	2
1	5	37
Total N	37	39
N correct	32	37
Proportion	0.865	0.949

N = 76

N Correct = 69

Proportion Correct = 0.908

Discriminant Analysis

Linear Discriminant Function for Groups

	0	1
Constant	-0.8862	-0.7976
GMAT	-0.2130	0.2021
GPA	-0.2293	0.2175
Central/South America	-0.3497	0.3318
North America	-0.0719	0.0682
Europe	0.1035	-0.0982
Middle East	0.3579	-0.3395
2018	1.5236	-1.4455
J-Term	0.0003	-0.0003
Finance	0.2131	-0.2022
Tech	0.2375	-0.2253
Consulting	-0.0423	0.0402
Real Estate	-0.1132	0.1074
General Management	0.1371	-0.1301
Entrepreneur	0.5316	-0.5043
Length	-0.6775	0.6427
Academics	-0.1120	0.1062
Clubs	-0.4837	0.4589
Social	0.2380	-0.2258
Recruitment	-1.0951	1.0389
School-Year Internship	0.7104	-0.6740
CF	-0.1443	0.1369
Very Active Member	0.2293	-0.2176
Active Member	0.4615	-0.4378
On the Board	0.3168	-0.3006
Totally Satisfied	-0.2100	0.1993
Dissatisfied	0.4567	-0.4333
Moderately Satisfied	0.4188	-0.3973
Sponsored	-0.6995	0.6636

Conclusions

- CMC satisfaction can be depicted by a left-skewed normal distribution
- CMC satisfaction is most highly correlated with effort expended on recruiting-related activities, such as proportion of time spent, career fellow meetings, advisor meetings, and most importantly, career workshops
 - In terms of professional clubs, Very Active members were the most satisfied with the CMC
 - Further club involvement showed diminishing marginal returns
- Time spent on Academics is negatively correlated with CMC Satisfaction, but group of people who got FT jobs had a higher amount of time dedicated to that category
- CMC satisfaction / job satisfaction is also highly correlated to GMAT scores
 - **Food for thought: Is our satisfaction more a reflection of ourselves or based on the contributions of the CMC?**
- Lastly, CMC satisfaction does not show a bifurcated difference between candidates who have found jobs and those who have not, but level of satisfaction with summer job matters

Appendix

■ List of Variables

- GMAT
- GPA
- Geography
- Class
- Term
- Industry/Length
- Breakdown Time in 1st Year*
- Interactions with CMC*
- Club Involvement
- Summer Satisfaction
- Full Time Satisfaction
- Sponsorship
- CMC Satisfaction

*Group of Variables