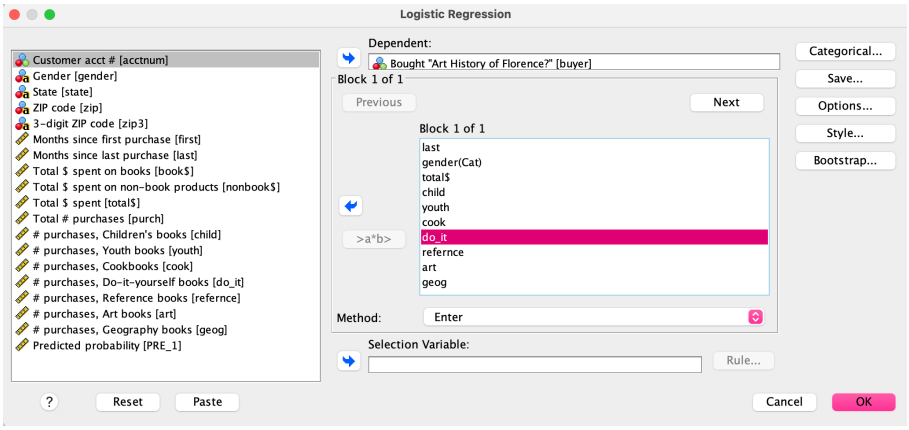
Part I: Logistic Regression

1. Estimate a logistic regression model using BUYER as the dependent variable and the following as predictor variables: (Use 'Analyze/Regression/Binary Logistic" in SPSS. Save the predicted probabilities by clicking on the 'Save' button and then on 'Probabilities' under 'Predicted Values').



2. Summarize and interpret the results (so that a marketing manager can understand them). Which variables are significant? Which seem to be 'important'? Interpret the coefficients for each of the predictors.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Months since last purchase	-.095	.003	1150.401	1	<.001	.910
	Gender(1)	.761	.036	452.515	1	<.001	2.140
	Total $ spent	.001	.000	31.701	1	<.001	1.001
	# purchases, Children's books	-.186	.017	116.097	1	<.001	.830
	# purchases, Youth books	-.113	.026	18.724	1	<.001	.893
	# purchases, Cookbooks	-.270	.017	249.075	1	<.001	.763
	# purchases, Do-it-yourself books	-.539	.027	399.777	1	<.001	.583
	# purchases, Reference books	.235	.027	78.087	1	<.001	1.265
	# purchases, Art books	1.156	.022	2723.273	1	.000	3.176
	# purchases, Geography books	.574	.019	950.087	1	<.001	1.776
	Constant	-2.361	.049	2293.523	1	.000	.094

a. Variable(s) entered on step 1: Months since last purchase, Gender, Total $ spent, # purchases, Children's books, # purchases, Youth books, # purchases, Cookbooks, # purchases, Do-it-yourself books, # purchases, Reference books, # purchases, Art books, # purchases, Geography books.

As the results showed, P-values of all variables are less than 0.05, indicating all variables are significant for the dependent variable.
(Based on all other predictors are the same)
Along with length of **time since last purchase** increasing 1 unite, the odds of purchasing the book decreases a factor of 0.91.
For every increase in **total dollars spent**, the odds of making a purchase increase by 10.01%.
For **Gender (1)**, it increases the odds of purchasing by a factor of 2.14.

For every increase in the total number of **children's books** purchased, it decrease the odds of purchasing by a factor of 0.83.
For every increase in the total number of **youth books** purchased, it decreases the odds of purchasing by a factor of 0.893.
For every increase in the total number of **cookbooks** purchased, it decreases the odds of purchasing by a factor of 0.763.
For every increase in the total number of **do-it-yourself books** purchased, it decreases the odds of purchasing by a factor of 0.583.
For every increase in the total number of **reference books** purchased, it increases the odds of purchasing by a factor of 1.265.
The odds of buying a copy of The Art History of Florence change by a factor of 3.176 with **art books** purchased.
For every increase in the total number of **geography books** purchased, it increases the odds of purchasing by a factor of 1.776.

Gender and the variable related whether the customer purchased the art books are seems more important. They can lead to a 2 to 3 times probability for customers purchasing the 'The Art History of Florence'.

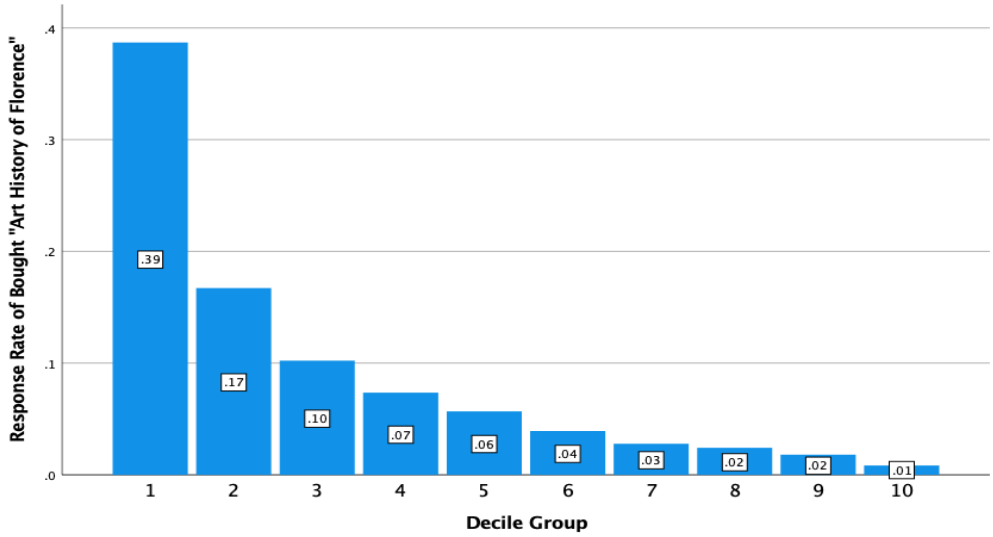
Part II: Decile Analysis of Logistic Regression Results

1. Assign each customer to a decile based on his or her predicted probability of purchase.

PRE_1	NPRE_1
.02003	8
.01661	9
.01583	9
.07688	4
.02012	8
.04695	6
.01109	10
.08707	3
.39124	1
.03298	7
.05267	5
.03636	7
.02025	8
.11323	3
.00681	10
.06426	4
.05766	5
.13945	2

PRE_1	NPRE_1
.02700	8
.25937	1
.01990	9
.00334	10
.01198	9
.08997	3
.12084	3
.01865	9
.04524	6
.03929	6
.00987	10
.01513	9
.02385	8
.44132	1
.12550	3
.05370	5
.10948	3
.08150	4
.04192	6

2. Create a bar chart plotting response rate by decile.



3. Generate a report showing number of customers, the number of buyers of "The Art History of Florence' and the response rate to the offer by decile.

Case Summaries

Bought "Art History of Florence?"

Percentile Group of PRE_1	N	Sum	Mean	% of Total Sum	% of Total N
1	5000	1935	.39	42.8%	10.0%
2	5000	836	.17	18.5%	10.0%
3	5000	511	.10	11.3%	10.0%
4	5000	368	.07	8.1%	10.0%
5	5000	284	.06	6.3%	10.0%
6	5000	196	.04	4.3%	10.0%
7	5001	139	.03	3.1%	10.0%
8	4999	121	.02	2.7%	10.0%
9	5000	90	.02	2.0%	10.0%
10	5000	42	.01	0.9%	10.0%
Total	50000	4522	.09	100.0%	100.0%

Values in N column are the number of customers.
Values in Sum is the number of buyers of 'The Art History of Florence'.
The customers response rate is the value in Mean.

4. Generate a report showing the mean values of the following variables by probability of purchase decile:
Total $ spent
Months since last purchase, and
Number of books purchased for each of the seven categories (i.e., children, youth, cookbooks, do-it-yourself, reference, art and geography).

Case Summaries

Mean

Percentile Group of PRE_1	Total $ spent	Months since last purchase	# purchases, Children's books	# purchases, Youth books	# purchases, Cookbooks	# purchases, Do-it-yourself books	# purchases, Reference books	# purchases, Art books	# purchases, Geography books
1	257.3526	7.19	1.06	.51	1.07	.47	.56	1.50	1.33
2	224.8692	7.96	.84	.39	.85	.39	.40	.75	.89
3	214.2284	8.62	.79	.37	.80	.37	.38	.48	.70
4	207.6430	8.78	.75	.36	.80	.34	.31	.30	.54
5	199.1118	9.57	.76	.33	.82	.37	.27	.22	.46
6	199.1302	10.94	.75	.36	.86	.39	.26	.16	.39
7	191.3457	12.37	.76	.35	.84	.42	.23	.13	.29
8	191.5499	14.42	.81	.36	.91	.45	.21	.11	.25
9	193.6108	17.86	.96	.41	1.12	.65	.25	.13	.32
10	204.3416	25.87	1.07	.46	1.31	.77	.25	.07	.29
Total	208.3183	12.36	.85	.39	.94	.46	.31	.39	.55

5. Summarize and interpret the decile analysis results. Are the patterns in the decile analysis consistent with your conclusions from the logistic regression?

The deciles from 1 to 10 are descending by the probability of purchase. In the first group, with the highest probability, customers spent most on books, about total $257, but have the longest interval since last purchase. As to the associations between different book categories and the purchase probability, Reference, Art and Geography books have a positive correlation with it. The rest of book categories, such as children's book, youth books and cookbooks purchased with a higher volume in the first group. With the decreased probability in next groups, the numbers of each category decrease and then increase, even to the largest sales.

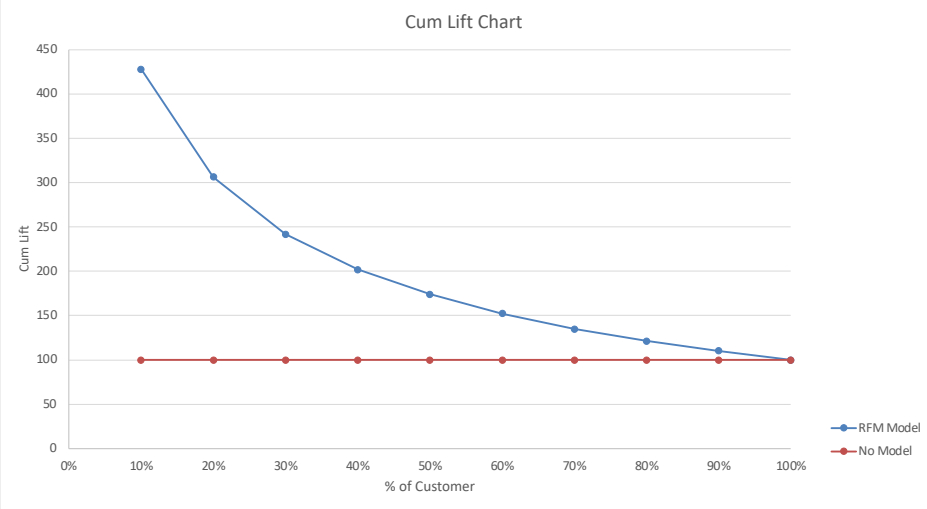
The patterns in the decile analysis are consistent with the conclusions from the logistic regression.

Part III: Lifts and Gains

1. Use the information from the report in 2c) above to create a chart showing the lift and cumulative lift for each decile. You may want to use Excel for these calculations.

Percentile Groups	# of Customer	Cum of Customer	# of Response	Cum of Response	Response Rate	Cum of Total Response Rate	Lift	Cum Lift
1	5000	5000	1935	1935	38.7%	38.7%	428	428
2	5000	10000	836	2771	16.7%	27.7%	185	306
3	5000	15000	511	3282	10.2%	21.9%	113	242
4	5000	20000	368	3650	7.4%	18.3%	81	202
5	5000	25000	284	3934	5.7%	15.7%	63	174
6	5000	30000	196	4130	3.9%	13.8%	43	152
7	5001	35001	139	4269	2.8%	12.2%	31	135
8	4999	40000	121	4390	2.4%	11.0%	27	121
9	5000	45000	90	4480	1.8%	10.0%	20	110
10	5000	50000	42	4522	0.8%	9.0%	9	100
Total	50000	50000	4522	4522	9.0%		100	0

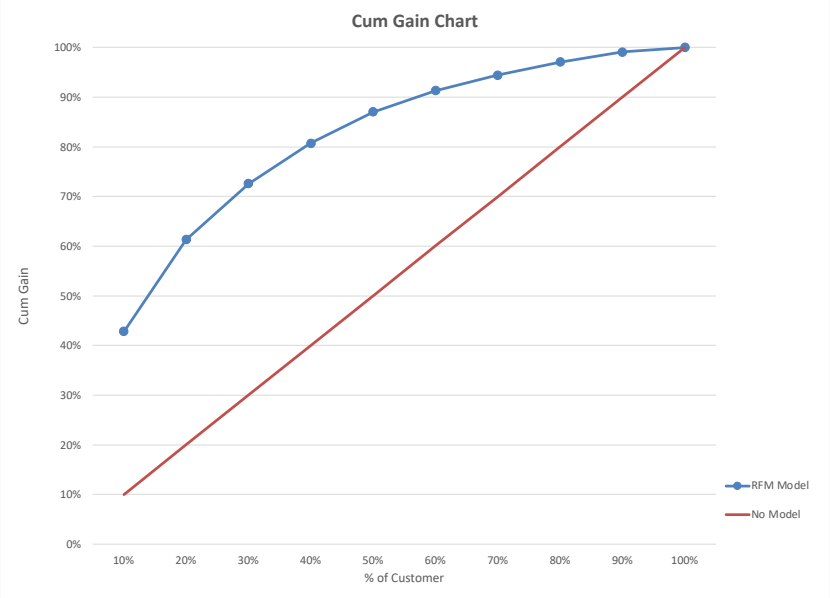
2. Create a chart showing the cumulative lift by decile.



3. Use the information from the report in 2c) above to create a chart showing the gains and cumulative gains for each decile. You may want to use Excel for these calculations.

Percentile Groups	# of Customer	Cum of Customer	# of Response	Cum of Response	Response Rate	Cum of Total Response Rate	Lift	Cum Lift	Gain	Cum Gain
1	5000	5000	1935	1935	38.7%	38.7%	428	428	42.8%	42.8%
2	5000	10000	836	2771	16.7%	27.7%	185	306	18.5%	61.3%
3	5000	15000	511	3282	10.2%	21.9%	113	242	11.3%	72.6%
4	5000	20000	368	3650	7.4%	18.3%	81	202	8.1%	80.7%
5	5000	25000	284	3934	5.7%	15.7%	63	174	6.3%	87.0%
6	5000	30000	196	4130	3.9%	13.8%	43	152	4.3%	91.3%
7	5001	35001	139	4269	2.8%	12.2%	31	135	3.1%	94.4%
8	4999	40000	121	4390	2.4%	11.0%	27	121	2.7%	97.1%
9	5000	45000	90	4480	1.8%	10.0%	20	110	2.0%	99.1%
10	5000	50000	42	4522	0.8%	9.0%	9	100	0.9%	100.0%
Total	50000	50000	4522	4522	9.0%		100	0	100.0%	100.0%

4. Create a chart showing the cumulative gains by decile along with a reference line corresponding to 'no model'.



Part IV: Profitability Analysis

Use the following cost information to assess the profitability of using logistic regression to

determine which customers should receive a specific offer:
Cost to mail offer to customer: $.50
Selling price (shipping included): $18.00
Wholesale price paid by BookBinders: $9.00
Shipping costs: $3.00
1. What is the breakeven response rate?
Breakeven Response Rate = $.5 / ($18-$9-$3) * 100% = 8.3%

2. Create a new variable (call it TARGET) with a value of 1 if the customer's predicted probability is greater than or equal to the breakeven response rate and 0 otherwise.

PRE_1	NPRE_1	Target
.02003	8	0
.01661	9	0
.01583	9	0
.07688	4	0
.02012	8	0
.04695	6	0
.01109	10	0
.08707	3	1
.39124	1	1
.03298	7	0
.05267	5	0
.03636	7	0
.02025	8	0
.11323	3	1
.00681	10	0
.06426	4	0
.05766	5	0
.13945	2	1

3. Generate a report summarizing the number of customers, the number of buyers of 'The Art History of Florence' and the response rate to the offer by the TARGET variable.

Case Summaries

Bought "Art History of Florence?"

Compare with Breakeven response rate	N	Sum	Mean	% of Total Sum	% of Total N
smaller than BER	34390	1195	.03	26.4%	68.8%
larger than BER	15610	3327	.21	73.6%	31.2%
Total	50000	4522	.09	100.0%	100.0%

4. What would the gross profit (in dollars, and also as a percentage of gross sales) and the return on marketing expenditures have been if BookBinders had mailed the offer to buy "The Art History of Florence" only to customers with a predicted probability of buying that was greater than or equal to the breakeven rate?

Cost to mail offer to customer: $.50
Selling price (shipping included): $18.00
Wholesale price paid by BookBinders: $9.00
Shipping costs: $3.00

Gross Profit: ($18-$9-$3) * 3327 - $.5 * 15610 = $19,962 – $7,805 = $12,157
ROI = $12,157 / $.5*15610 * 100% = 156%