Individual-Level Logistic Regression: Employment Probability
With Interaction Terms (Individual-Level Advantage)

The LOGISTIC Procedure

Model Information	
Data Set	WORK.INDIVIDUAL_DATA
Response Variable	employed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	494755
Number of Observations Used	494755

Response Profile		
Ordered Value	employed	Total Frequency
1	0	188586
2	1	306169

Probability modeled is employed=0.

Class Level Information								
Class	Value	Design Variables						
education_level	H10	-1	-1	-1	-1	-1	-1	
	H20_35	1	0	0	0	0	0	
	H40	0	1	0	0	0	0	
	H50_60	0	0	1	0	0	0	
	H70	0	0	0	1	0	0	
	H80	0	0	0	0	1	0	
	H90	0	0	0	0	0	1	
age_group	16-29	-1	-1	-1				
	30-49	1	0	0				
	50-66	0	1	0				
	67+	0	0	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	657664.21	461565.91
SC	657675.32	461710.36
-2 Log L	657662.21	461539.91

Individual-Level Logistic Regression: Employment Probability
With Interaction Terms (Individual-Level Advantage)

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	196122.307	12	<.0001
Score	179269.115	12	<.0001
Wald	113952.760	12	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	1	3440.3113	<.0001
education_level	6	22175.6848	<.0001
age_group	3	97481.4590	<.0001
male_high_edu	1	105.7883	<.0001
young_male	1	720.8242	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2402	0.0101	566.1320	<.0001
sex		1	-0.6203	0.0106	3440.3113	<.0001
education_level	H20_35	1	-0.0285	0.0115	6.1496	0.0131
education_level	H40	1	-0.4494	0.0176	651.1936	<.0001
education_level	H50_60	1	-0.3295	0.0116	811.6479	<.0001
education_level	H70	1	-0.7388	0.0142	2691.3406	<.0001
education_level	H80	1	-0.8593	0.0414	431.1558	<.0001
education_level	H90	1	1.4037	0.0246	3244.1875	<.0001
age_group	30-49	1	-1.0693	0.00673	25232.6167	<.0001
age_group	50-66	1	-0.6687	0.00636	11048.8117	<.0001
age_group	67+	1	2.5555	0.00846	91344.9635	<.0001
male_high_edu		1	0.1690	0.0164	105.7883	<.0001
young_male		1	0.4455	0.0166	720.8242	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sex	0.538	0.527	0.549
education_level H20_35 vs H10	0.357	0.350	0.363
education_level H40 vs H10	0.234	0.225	0.244
education_level H50_60 vs H10	0.264	0.257	0.271
education_level H70 vs H10	0.175	0.170	0.181
education_level H80 vs H10	0.155	0.141	0.171

Individual-Level Logistic Regression: Employment Probability
With Interaction Terms (Individual-Level Advantage)

The LOGISTIC Procedure

Odds Ratio Estimates		
Effect	**Point Estimate**	**95% Wald Confidence Limits**
education_level H90 vs H10	1.494	1.411	1.583
age_group 30-49 vs 16-29	0.777	0.758	0.798
age_group 50-66 vs 16-29	1.160	1.132	1.190
age_group 67+ vs 16-29	29.166	28.314	30.044
male_high_edu	1.184	1.147	1.223
young_male	1.561	1.511	1.613

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	**Somers' D**	0.660
Percent Discordant	15.9	**Gamma**	0.675
Percent Tied	2.2	**Tau-a**	0.311
Pairs	57739187034	**c**	0.830

Multinomial Logistic Regression: All Employment Statuses
Individual-Level Advantage: Modeling Multiple Outcomes

The LOGISTIC Procedure

Model Information	
Data Set	WORK.INDIVIDUAL_DATA
Response Variable	employment_status
Number of Response Levels	5
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	494755
Number of Observations Used	494755

Response Profile		
Ordered Value	employment_status	Total Frequency
1	Employed	306169
2	In Education or Training	22086
3	Not in Labor Force	17349
4	On Pension	126359
5	Unemployed	22792

Logits modeled use employment_status='Not in Labor Force' as the reference category.

Class Level Information								
Class	Value	Design Variables						
education_level	H10	-1	-1	-1	-1	-1	-1	
	H20_35	1	0	0	0	0	0	
	H40	0	1	0	0	0	0	
	H50_60	0	0	1	0	0	0	
	H70	0	0	0	1	0	0	
	H80	0	0	0	0	1	0	
	H90	0	0	0	0	0	1	
age_group	16-29	-1	-1	-1				
	30-49	1	0	0				
	50-66	0	1	0				
	67+	0	0	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Multinomial Logistic Regression: All Employment Statuses
Individual-Level Advantage: Modeling Multiple Outcomes

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1032712.7	655362.03
SC	1032757.2	655850.95
-2 Log L	1032704.7	655274.03

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	377430.712	40	<.0001
Score	381013.710	40	<.0001
Wald	166682.377	40	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	4	4082.9750	<.0001
education_level	24	28902.3651	<.0001
age_group	12	147632.686	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		employment_status	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		Employed	1	2.4934	0.0176	20176.0963	<.0001
Intercept		In Education or Training	1	-1.6512	0.0443	1389.4019	<.0001
Intercept		On Pension	1	0.8025	0.0235	1164.4118	<.0001
Intercept		Unemployed	1	-0.7661	0.0431	316.4846	<.0001
sex		Employed	1	0.1253	0.0159	62.0245	<.0001
sex		In Education or Training	1	-0.1695	0.0211	64.4014	<.0001
sex		On Pension	1	-0.5066	0.0187	730.3595	<.0001
sex		Unemployed	1	-0.2910	0.0205	200.4710	<.0001
education_level	H20_35	Employed	1	0.5495	0.0166	1089.9575	<.0001
education_level	H20_35	In Education or Training	1	0.6912	0.0340	414.2272	<.0001
education_level	H20_35	On Pension	1	0.6770	0.0225	907.3793	<.0001
education_level	H20_35	Unemployed	1	0.5836	0.0276	448.3367	<.0001
education_level	H40	Employed	1	0.4418	0.0315	196.9929	<.0001
education_level	H40	In Education or Training	1	-0.0155	0.0598	0.0668	0.7961
education_level	H40	On Pension	1	0.0812	0.0399	4.1521	0.0416
education_level	H40	Unemployed	1	0.2550	0.0460	30.7932	<.0001
education_level	H50_60	Employed	1	0.4672	0.0198	559.0340	<.0001
education_level	H50_60	In Education or Training	1	0.5543	0.0377	215.6371	<.0001

Multinomial Logistic Regression: All Employment Statuses
Individual-Level Advantage: Modeling Multiple Outcomes

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates							
Parameter		employment_status	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
education_level	H50_60	On Pension	1	0.2465	0.0261	89.1449	<.0001
education_level	H50_60	Unemployed	1	0.1849	0.0316	34.1988	<.0001
education_level	H70	Employed	1	0.3565	0.0218	268.4494	<.0001
education_level	H70	In Education or Training	1	-1.1209	0.0583	369.9250	<.0001
education_level	H70	On Pension	1	-0.5139	0.0304	285.3290	<.0001
education_level	H70	Unemployed	1	-0.00951	0.0347	0.0749	0.7844
education_level	H80	Employed	1	-0.1842	0.0489	14.1967	0.0002
education_level	H80	In Education or Training	1	-0.6398	0.1506	18.0389	<.0001
education_level	H80	On Pension	1	-1.8766	0.0884	450.7654	<.0001
education_level	H80	Unemployed	1	-1.6595	0.1182	197.2431	<.0001
education_level	H90	Employed	1	-1.7016	0.0331	2641.6136	<.0001
education_level	H90	In Education or Training	1	-0.6078	0.0691	77.4492	<.0001
education_level	H90	On Pension	1	-0.1290	0.0420	9.4147	0.0022
education_level	H90	Unemployed	1	-0.5117	0.0510	100.8155	<.0001
age_group	30-49	Employed	1	0.0543	0.0155	12.2314	0.0005
age_group	30-49	In Education or Training	1	0.2565	0.0375	46.7563	<.0001
age_group	30-49	On Pension	1	-1.2383	0.0197	3946.4938	<.0001
age_group	30-49	Unemployed	1	0.9677	0.0380	648.4192	<.0001
age_group	50-66	Employed	1	0.2192	0.0166	173.8364	<.0001
age_group	50-66	In Education or Training	1	-1.9270	0.0615	982.2744	<.0001
age_group	50-66	On Pension	1	0.4148	0.0189	482.6795	<.0001
age_group	50-66	Unemployed	1	0.7297	0.0389	352.6147	<.0001
age_group	67+	Employed	1	0.0938	0.0329	8.1066	0.0044
age_group	67+	In Education or Training	1	-0.4722	0.0825	32.7836	<.0001
age_group	67+	On Pension	1	3.9978	0.0335	14244.8630	<.0001
age_group	67+	Unemployed	1	-1.8877	0.1063	315.5480	<.0001

Odds Ratio Estimates				
Effect	employment_status	Point Estimate	95% Wald Confidence Limits	
sex	Employed	1.134	1.099	1.169
sex	In Education or Training	0.844	0.810	0.880
sex	On Pension	0.603	0.581	0.625
sex	Unemployed	0.748	0.718	0.778
education_level H20_35 vs H10	Employed	1.614	1.547	1.684
education_level H20_35 vs H10	In Education or Training	0.639	0.607	0.673
education_level H20_35 vs H10	On Pension	0.433	0.412	0.454
education_level H20_35 vs H10	Unemployed	0.563	0.535	0.594
education_level H40 vs H10	Employed	1.449	1.344	1.564

Multinomial Logistic Regression: All Employment Statuses
Individual-Level Advantage: Modeling Multiple Outcomes

The LOGISTIC Procedure

Odds Ratio Estimates				
Effect	**employment_status**	**Point Estimate**	**95% Wald Confidence Limits**	
education_level H40 vs H10	In Education or Training	0.315	0.278	0.358
education_level H40 vs H10	On Pension	0.238	0.218	0.261
education_level H40 vs H10	Unemployed	0.406	0.367	0.449
education_level H50_60 vs H10	Employed	1.487	1.414	1.563
education_level H50_60 vs H10	In Education or Training	0.558	0.522	0.596
education_level H50_60 vs H10	On Pension	0.281	0.265	0.298
education_level H50_60 vs H10	Unemployed	0.378	0.355	0.403
education_level H70 vs H10	Employed	1.331	1.260	1.406
education_level H70 vs H10	In Education or Training	0.104	0.092	0.118
education_level H70 vs H10	On Pension	0.132	0.123	0.141
education_level H70 vs H10	Unemployed	0.311	0.289	0.335
education_level H80 vs H10	Employed	0.775	0.690	0.870
education_level H80 vs H10	In Education or Training	0.169	0.120	0.239
education_level H80 vs H10	On Pension	0.034	0.027	0.041
education_level H80 vs H10	Unemployed	0.060	0.046	0.079
education_level H90 vs H10	Employed	0.170	0.157	0.184
education_level H90 vs H10	In Education or Training	0.174	0.150	0.202
education_level H90 vs H10	On Pension	0.193	0.176	0.213
education_level H90 vs H10	Unemployed	0.188	0.168	0.211
age_group 30-49 vs 16-29	Employed	1.524	1.467	1.584
age_group 30-49 vs 16-29	In Education or Training	0.152	0.143	0.161
age_group 30-49 vs 16-29	On Pension	6.931	6.440	7.460
age_group 30-49 vs 16-29	Unemployed	2.176	2.071	2.287
age_group 50-66 vs 16-29	Employed	1.798	1.725	1.873
age_group 50-66 vs 16-29	In Education or Training	0.017	0.015	0.020
age_group 50-66 vs 16-29	On Pension	36.202	33.701	38.889
age_group 50-66 vs 16-29	Unemployed	1.715	1.625	1.810
age_group 67+ vs 16-29	Employed	1.586	1.451	1.734
age_group 67+ vs 16-29	In Education or Training	0.073	0.059	0.090
age_group 67+ vs 16-29	On Pension	>999.999	>999.999	>999.999
age_group 67+ vs 16-29	Unemployed	0.125	0.095	0.166

Multinomial Logistic Regression: All Employment Statuses
Individual-Level Advantage: Modeling Multiple Outcomes

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	INDIVIDUAL_DATA
Random Number Seed	12345
Sampling Rate	0.7
Sample Size	346329
Selection Probability	0.700001
Sampling Weight	1.428569
Output Data Set	TRAINING_DATA

Training Model on Individual-Level Data

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TRAINING_DATA
Response Variable	employed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	346329
Number of Observations Used	346329

Response Profile		
Ordered Value	employed	Total Frequency
1	0	132398
2	1	213931

Probability modeled is employed=0.

Class Level Information								
Class	Value	Design Variables						
education_level	H10	-1	-1	-1	-1	-1	-1	
	H20_35	1	0	0	0	0	0	
	H40	0	1	0	0	0	0	
	H50_60	0	0	1	0	0	0	
	H70	0	0	0	1	0	0	
	H80	0	0	0	0	1	0	
	H90	0	0	0	0	0	1	
age_group	16-29	-1	-1	-1				
	30-49	1	0	0				
	50-66	0	1	0				
	67+	0	0	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	460740.03	323434.64
SC	460750.79	323552.95
-2 Log L	460738.03	323412.64

Training Model on Individual-Level Data

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	137325.391	10	<.0001
Score	125539.904	10	<.0001
Wald	80175.8704	10	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	1	2519.9928	<.0001
education_level	6	17263.2123	<.0001
age_group	3	68937.0233	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2141	0.0115	348.3125	<.0001
sex		1	-0.4482	0.00893	2519.9928	<.0001
education_level	H20_35	1	-0.0829	0.0120	47.4134	<.0001
education_level	H40	1	-0.4364	0.0209	434.7173	<.0001
education_level	H50_60	1	-0.3060	0.0138	488.2422	<.0001
education_level	H70	1	-0.7255	0.0169	1833.9884	<.0001
education_level	H80	1	-0.8769	0.0500	307.4337	<.0001
education_level	H90	1	1.4730	0.0292	2538.1636	<.0001
age_group	30-49	1	-1.1274	0.00769	21490.1424	<.0001
age_group	50-66	1	-0.7246	0.00725	9981.9727	<.0001
age_group	67+	1	2.4956	0.00964	67011.4950	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sex	0.639	0.628	0.650
education_level H20_35 vs H10	0.354	0.347	0.362
education_level H40 vs H10	0.249	0.238	0.260
education_level H50_60 vs H10	0.283	0.276	0.291
education_level H70 vs H10	0.186	0.180	0.193
education_level H80 vs H10	0.160	0.143	0.180
education_level H90 vs H10	1.679	1.573	1.793
age_group 30-49 vs 16-29	0.616	0.602	0.631
age_group 50-66 vs 16-29	0.922	0.902	0.943
age_group 67+ vs 16-29	23.086	22.428	23.764

Training Model on Individual-Level Data

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	Somers' D	0.661
Percent Discordant	15.9	Gamma	0.676
Percent Tied	2.2	Tau-a	0.312
Pairs	28324036538	c	0.830

Validation on Individual-Level Holdout Sample

The LOGISTIC Procedure

Model Information	
Data Set	WORK.VALIDATION_DATA
Response Variable	employed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	494755
Number of Observations Used	494755

Response Profile		
Ordered Value	**employed**	**Total Frequency**
1	0	188586
2	1	306169

Probability modeled is employed=0.

Class Level Information								
Class	**Value**	**Design Variables**						
education_level	H10	-1	-1	-1	-1	-1	-1
	H20_35	1	0	0	0	0	0
	H40	0	1	0	0	0	0
	H50_60	0	0	1	0	0	0
	H70	0	0	0	1	0	0
	H80	0	0	0	0	1	0
	H90	0	0	0	0	0	1
age_group	16-29	-1	-1	-1			
	30-49	1	0	0			
	50-66	0	1	0			
	67+	0	0	1			

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	657664.21	462317.25
SC	657675.32	462439.48
-2 Log L	657662.21	462295.25

Validation on Individual-Level Holdout Sample

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	195366.963	10	<.0001
Score	178937.844	10	<.0001
Wald	114288.390	10	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	1	3485.8109	<.0001
education_level	6	24271.5074	<.0001
age_group	3	98485.6863	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2172	0.00952	519.8849	<.0001
sex		1	-0.4409	0.00747	3485.8109	<.0001
education_level	H20_35	1	-0.0884	0.01000	78.2298	<.0001
education_level	H40	1	-0.4243	0.0174	595.5797	<.0001
education_level	H50_60	1	-0.3127	0.0115	735.3236	<.0001
education_level	H70	1	-0.7176	0.0141	2593.9230	<.0001
education_level	H80	1	-0.8346	0.0413	408.6886	<.0001
education_level	H90	1	1.4301	0.0244	3434.9625	<.0001
age_group	30-49	1	-1.1220	0.00643	30440.5934	<.0001
age_group	50-66	1	-0.7185	0.00607	14016.0726	<.0001
age_group	67+	1	2.4940	0.00806	95858.3475	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sex	0.643	0.634	0.653
education_level H20_35 vs H10	0.355	0.349	0.361
education_level H40 vs H10	0.254	0.244	0.263
education_level H50_60 vs H10	0.284	0.277	0.290
education_level H70 vs H10	0.189	0.184	0.195
education_level H80 vs H10	0.168	0.153	0.185
education_level H90 vs H10	1.620	1.534	1.711
age_group 30-49 vs 16-29	0.626	0.614	0.639
age_group 50-66 vs 16-29	0.937	0.919	0.955
age_group 67+ vs 16-29	23.279	22.723	23.849

Validation on Individual-Level Holdout Sample

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	Somers' D	0.659
Percent Discordant	15.9	Gamma	0.674
Percent Tied	2.2	Tau-a	0.311
Pairs	57739187034	c	0.830

Validation on Individual-Level Holdout Sample

The LOGISTIC Procedure

Model Information	
Data Set	WORK.INDIVIDUAL_DATA
Response Variable	employed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	494755
Number of Observations Used	494755

Response Profile		
Ordered Value	employed	Total Frequency
1	0	188586
2	1	306169

Probability modeled is employed=0.

Class Level Information								
Class	Value	Design Variables						
education_level	H10	-1	-1	-1	-1	-1	-1	
	H20_35	1	0	0	0	0	0	
	H40	0	1	0	0	0	0	
	H50_60	0	0	1	0	0	0	
	H70	0	0	0	1	0	0	
	H80	0	0	0	0	1	0	
	H90	0	0	0	0	0	1	
age_group	16-29	-1	-1	-1				
	30-49	1	0	0				
	50-66	0	1	0				
	67+	0	0	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	657664.21	462317.25
SC	657675.32	462439.48
-2 Log L	657662.21	462295.25

Validation on Individual-Level Holdout Sample

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	195366.963	10	<.0001
Score	178937.844	10	<.0001
Wald	114288.390	10	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
sex	1	3485.8109	<.0001
education_level	6	24271.5074	<.0001
age_group	3	98485.6863	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.2172	0.00952	519.8849	<.0001
sex		1	-0.4409	0.00747	3485.8109	<.0001
education_level	**H20_35**	1	-0.0884	0.01000	78.2298	<.0001
education_level	**H40**	1	-0.4243	0.0174	595.5797	<.0001
education_level	**H50_60**	1	-0.3127	0.0115	735.3236	<.0001
education_level	**H70**	1	-0.7176	0.0141	2593.9230	<.0001
education_level	**H80**	1	-0.8346	0.0413	408.6886	<.0001
education_level	**H90**	1	1.4301	0.0244	3434.9625	<.0001
age_group	**30-49**	1	-1.1220	0.00643	30440.5934	<.0001
age_group	**50-66**	1	-0.7185	0.00607	14016.0726	<.0001
age_group	**67+**	1	2.4940	0.00806	95858.3475	<.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
sex	0.643	0.634	0.653
education_level H20_35 vs H10	0.355	0.349	0.361
education_level H40 vs H10	0.254	0.244	0.263
education_level H50_60 vs H10	0.284	0.277	0.290
education_level H70 vs H10	0.189	0.184	0.195
education_level H80 vs H10	0.168	0.153	0.185
education_level H90 vs H10	1.620	1.534	1.711
age_group 30-49 vs 16-29	0.626	0.614	0.639
age_group 50-66 vs 16-29	0.937	0.919	0.955
age_group 67+ vs 16-29	23.279	22.723	23.849

Validation on Individual-Level Holdout Sample

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	Somers' D	0.659
Percent Discordant	15.9	Gamma	0.674
Percent Tied	2.2	Tau-a	0.311
Pairs	57739187034	c	0.830

Individual-Level Outlier Analysis
Identifying Unusual Individual Patterns

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of outlier_flag by sex		
	sex		
outlier_flag	**0**	**1**	**Total**
0	236005 47.70 52.32 94.20	215114 43.48 47.68 88.08	451119 91.18
1	14527 2.94 33.29 5.80	29109 5.88 66.71 11.92	43636 8.82
Total	250532 50.64	244223 49.36	494755 100.00

Statistics for Table of outlier_flag by sex

Statistic	DF	Value	Prob
Chi-Square	1	5760.8529	<.0001
Likelihood Ratio Chi-Square	1	5855.2667	<.0001
Continuity Adj. Chi-Square	1	5760.0919	<.0001
Mantel-Haenszel Chi-Square	1	5760.8413	<.0001
Phi Coefficient		0.1079	
Contingency Coefficient		0.1073	
Cramer's V		0.1079	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	236005
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 494755

Individual-Level Risk Segmentation
Employment Rates by Individual Risk Segments

The MEANS Procedure

Analysis Variable : actual_employed						
segment	N Obs	N	Mean	Std Dev	Minimum	Maximum
Low Risk	293322	293322	0.6170284	0.4861123	0	1.0000000
Medium R	201433	201433	0.6214523	0.4850263	0	1.0000000

Individual-Level Risk Segmentation
Employment Rates by Individual Risk Segments

The LOGISTIC Procedure

Model Information	
Data Set	WORK.INDIVIDUAL_DATA
Response Variable	employed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	494755
Number of Observations Used	494755

Response Profile		
Ordered Value	employed	Total Frequency
1	0	188586
2	1	306169

Probability modeled is employed=0.

Class Level Information								
Class	Value	Design Variables						
education_level	H10	-1	-1	-1	-1	-1	-1	
	H20_35	1	0	0	0	0	0	
	H40	0	1	0	0	0	0	
	H50_60	0	0	1	0	0	0	
	H70	0	0	0	1	0	0	
	H80	0	0	0	0	1	0	
	H90	0	0	0	0	0	1	
age_group	16-29	-1	-1	-1				
	30-49	1	0	0				
	50-66	0	1	0				
	67+	0	0	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	657664.21	462317.25
SC	657675.32	462439.48
-2 Log L	657662.21	462295.25

Individual-Level Risk Segmentation
Employment Rates by Individual Risk Segments

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	195366.963	10	<.0001
Score	178937.844	10	<.0001
Wald	114288.390	10	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	1	3485.8109	<.0001
education_level	6	24271.5074	<.0001
age_group	3	98485.6863	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2172	0.00952	519.8849	<.0001
sex		1	-0.4409	0.00747	3485.8109	<.0001
education_level	H20_35	1	-0.0884	0.01000	78.2298	<.0001
education_level	H40	1	-0.4243	0.0174	595.5797	<.0001
education_level	H50_60	1	-0.3127	0.0115	735.3236	<.0001
education_level	H70	1	-0.7176	0.0141	2593.9230	<.0001
education_level	H80	1	-0.8346	0.0413	408.6886	<.0001
education_level	H90	1	1.4301	0.0244	3434.9625	<.0001
age_group	30-49	1	-1.1220	0.00643	30440.5934	<.0001
age_group	50-66	1	-0.7185	0.00607	14016.0726	<.0001
age_group	67+	1	2.4940	0.00806	95858.3475	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sex	0.643	0.634	0.653
education_level H20_35 vs H10	0.355	0.349	0.361
education_level H40 vs H10	0.254	0.244	0.263
education_level H50_60 vs H10	0.284	0.277	0.290
education_level H70 vs H10	0.189	0.184	0.195
education_level H80 vs H10	0.168	0.153	0.185
education_level H90 vs H10	1.620	1.534	1.711
age_group 30-49 vs 16-29	0.626	0.614	0.639
age_group 50-66 vs 16-29	0.937	0.919	0.955
age_group 67+ vs 16-29	23.279	22.723	23.849

Individual-Level Risk Segmentation
Employment Rates by Individual Risk Segments

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	Somers' D	0.659
Percent Discordant	15.9	Gamma	0.674
Percent Tied	2.2	Tau-a	0.311
Pairs	57739187034	c	0.830

Individual-Level Marginal Effects
How Variables Affect Each Individual Differently

The MEANS Procedure

education_level	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
H10	117235	male_effect edu_premium	117235 117235	0.0912935 0	0.0448567 0	0.0183976 0	0.1249956 0
H20_35	195769	male_effect edu_premium	195769 195769	0.0746017 0	0.0172219 0	0.0453178 0	0.1001473 0
H40	24667	male_effect edu_premium	24667 24667	0.0636509 0.0765963	0.0130701 0.0751994	0.0447411 0.0298060	0.0843892 0.2593273
H50_60	91343	male_effect edu_premium	91343 91343	0.0689422 0.0927460	0.0135769 0.0825376	0.0488712 0.0329394	0.0897400 0.2630947
H70	53742	male_effect edu_premium	53742 53742	0.0541919 0.0584683	0.0161990 0.0660870	0.0351210 0.0228063	0.0928211 0.2478754
H80	4782	male_effect edu_premium	4782 4782	0.0459206 0.0410291	0.0155425 0.0485810	0.0317746 0.0204601	0.0982426 0.2426432
H90	7217	male_effect edu_premium	7217 7217	0.0954424 0.1900928	0.0464854 0.0609380	0.0116952 0.1240114	0.1249106 0.2928106

Individual-Level Policy Simulation
Simulating Education Policy Impact on Individual Employment

The MEANS Procedure

education_level	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
H10	117235	original_employed simulated_employed	117235 117235	0.4201049 0.5349170	0.4935776 0.4987814	0 0	1.0000000 1.0000000
H20_35	195769	original_employed simulated_employed	195769 195769	0.6405764 0.7129832	0.4798327 0.4523707	0 0	1.0000000 1.0000000
H40	24667	original_employed simulated_employed	24667 24667	0.7446791 0.7446791	0.4360503 0.4360503	0 0	1.0000000 1.0000000
H50_60	91343	original_employed simulated_employed	91343 91343	0.6908466 0.6908466	0.4621471 0.4621471	0 0	1.0000000 1.0000000
H70	53742	original_employed simulated_employed	53742 53742	0.8051059 0.8051059	0.3961229 0.3961229	0 0	1.0000000 1.0000000
H80	4782	original_employed simulated_employed	4782 4782	0.8632371 0.8632371	0.3436328 0.3436328	0 0	1.0000000 1.0000000
H90	7217	original_employed simulated_employed	7217 7217	0.3663572 0.3663572	0.4818421 0.4818421	0 0	1.0000000 1.0000000

Individual-Level Prediction Intervals
Uncertainty for Each Individual Prediction

The LOGISTIC Procedure

Model Information	
Data Set	WORK.INDIVIDUAL_DATA
Response Variable	employed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	494755
Number of Observations Used	494755

Response Profile		
Ordered Value	**employed**	**Total Frequency**
1	0	188586
2	1	306169

Probability modeled is employed=0.

Class Level Information								
Class	**Value**	**Design Variables**						
education_level	**H10**	-1	-1	-1	-1	-1	-1
	H20_35	1	0	0	0	0	0
	H40	0	1	0	0	0	0
	H50_60	0	0	1	0	0	0
	H70	0	0	0	1	0	0
	H80	0	0	0	0	1	0
	H90	0	0	0	0	0	1
age_group	**16-29**	-1	-1	-1			
	30-49	1	0	0			
	50-66	0	1	0			
	67+	0	0	1			

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	657664.21	462317.25
SC	657675.32	462439.48
-2 Log L	657662.21	462295.25

Individual-Level Prediction Intervals
Uncertainty for Each Individual Prediction

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	195366.963	10	<.0001
Score	178937.844	10	<.0001
Wald	114288.390	10	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	1	3485.8109	<.0001
education_level	6	24271.5074	<.0001
age_group	3	98485.6863	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2172	0.00952	519.8849	<.0001
sex		1	-0.4409	0.00747	3485.8109	<.0001
education_level	H20_35	1	-0.0884	0.01000	78.2298	<.0001
education_level	H40	1	-0.4243	0.0174	595.5797	<.0001
education_level	H50_60	1	-0.3127	0.0115	735.3236	<.0001
education_level	H70	1	-0.7176	0.0141	2593.9230	<.0001
education_level	H80	1	-0.8346	0.0413	408.6886	<.0001
education_level	H90	1	1.4301	0.0244	3434.9625	<.0001
age_group	30-49	1	-1.1220	0.00643	30440.5934	<.0001
age_group	50-66	1	-0.7185	0.00607	14016.0726	<.0001
age_group	67+	1	2.4940	0.00806	95858.3475	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sex	0.643	0.634	0.653
education_level H20_35 vs H10	0.355	0.349	0.361
education_level H40 vs H10	0.254	0.244	0.263
education_level H50_60 vs H10	0.284	0.277	0.290
education_level H70 vs H10	0.189	0.184	0.195
education_level H80 vs H10	0.168	0.153	0.185
education_level H90 vs H10	1.620	1.534	1.711
age_group 30-49 vs 16-29	0.626	0.614	0.639
age_group 50-66 vs 16-29	0.937	0.919	0.955
age_group 67+ vs 16-29	23.279	22.723	23.849

Individual-Level Prediction Intervals
Uncertainty for Each Individual Prediction

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	Somers' D	0.659
Percent Discordant	15.9	Gamma	0.674
Percent Tied	2.2	Tau-a	0.311
Pairs	57739187034	c	0.830

Individual-Level Decision Tree
Machine Learning Approach to Individual Prediction

The HPSPLIT Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	**Engine**	**Role**	**Path**
WORK.INDIVIDUAL_DATA	V9	Input	On Client

Model Information	
Split Criterion Used	Variance
Pruning Method	Entropy
Subtree Evaluation Criterion	Full Tree
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	8
Tree Depth	8
Number of Leaves Before Pruning	56
Number of Leaves After Pruning	56

Number of Observations Read	494755
Number of Observations Used	494755

Individual-Level Decision Tree
Machine Learning Approach to Individual Prediction

The HPSPLIT Procedure

Regression Tree for employed

Individual-Level Decision Tree
Machine Learning Approach to Individual Prediction

The HPSPLIT Procedure

Subtree Starting at Node=0

Node 0
N 494755
Avg 0.6...

age_group
67+

Node 1
N 109649
Avg 0.1...

age_group
16–29,30–49,50–66

Node 2
N 385106
Avg 0.7...

sex
< 0.010

Node 3
N 59433
Avg 0.0...

sex
>= 0.010

Node 4
N 50216
Avg 0.1...

education_level
H10,H90

Node 5
N 88591
Avg 0.5...

education_level
H20_35,H40,H50

Node 6
N 296515
Avg 0.8...

education
H10,H2...

Node 7
N 57012
Avg 0.0...

education
H70,H80

Node 8
N 2421
Avg 0.1...

education
H10,H20

Node 9
N 45768
Avg 0.1...

education
H70,H80

Node A
N 4448
Avg 0.2...

sex
< 0.010

Node B
N 38638
Avg 0.4...

sex
>= 0.010

Node C
N 49953
Avg 0.6...

age_group
16–29,50–66

Node D
N 171019
Avg 0.7...

age_gr...
30–49

| Node E |
|---|
| N 125496 |
| Avg 0.8... |
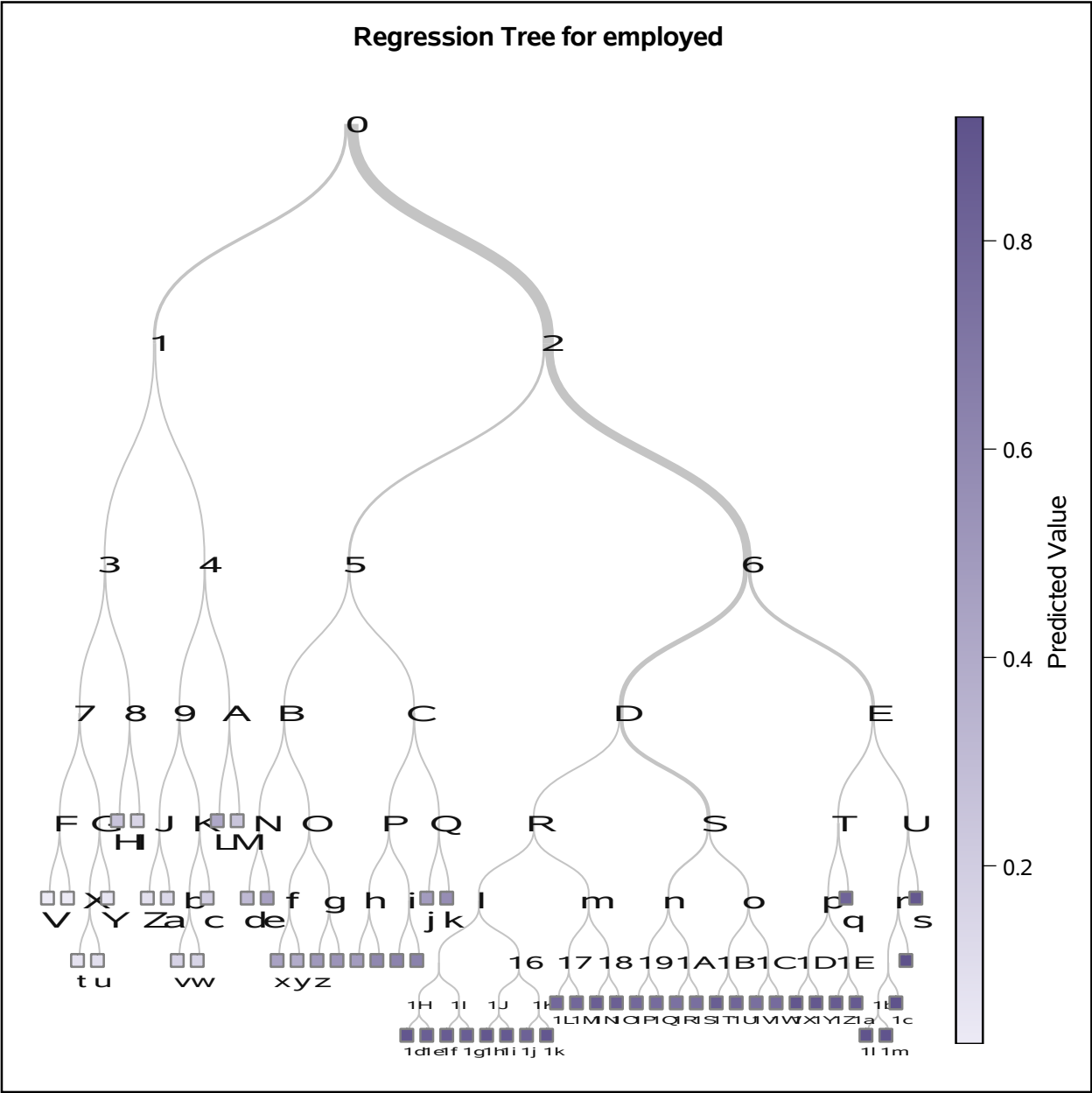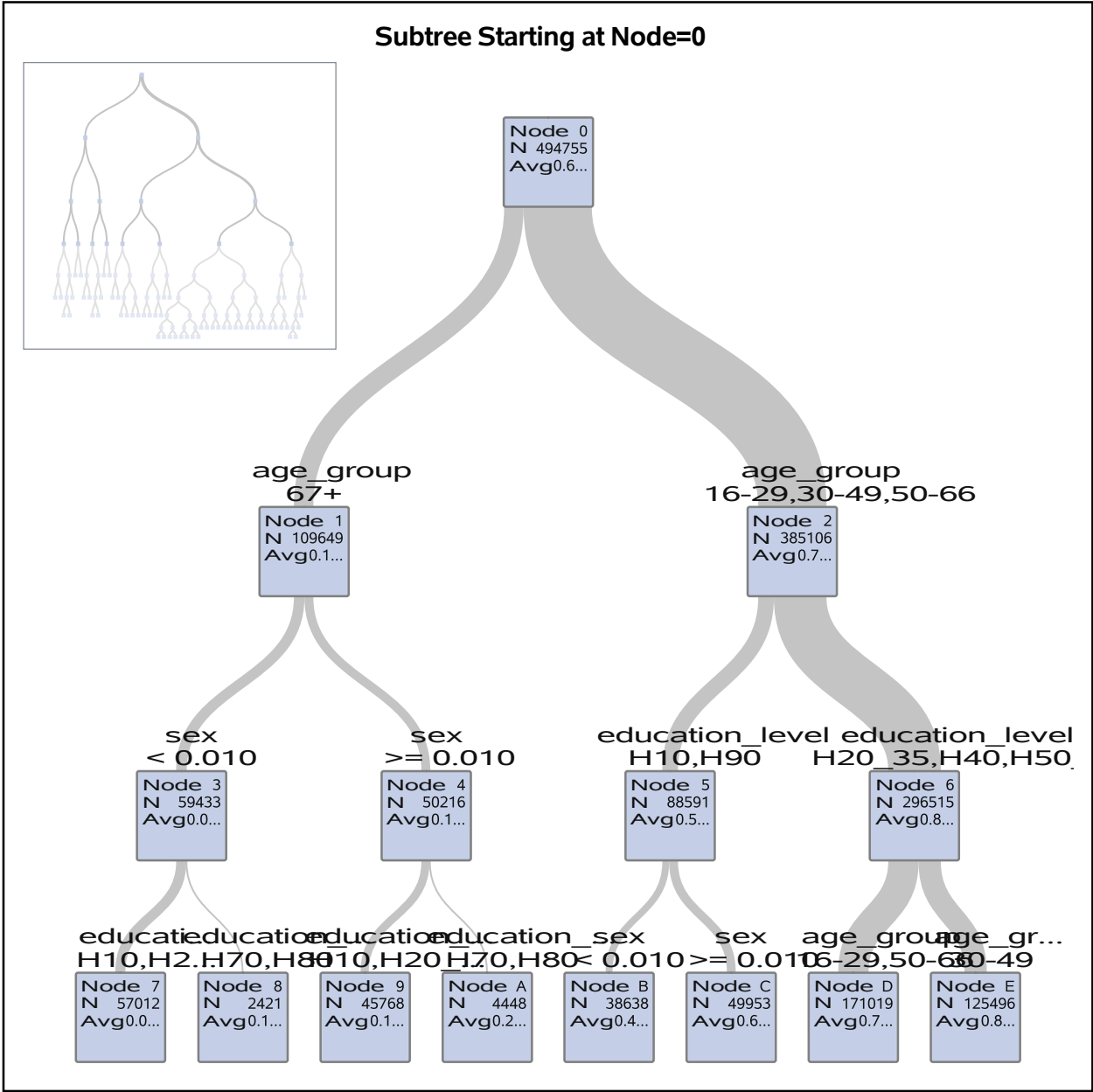
Individual-Level Decision Tree
Machine Learning Approach to Individual Prediction

The HPSPLIT Procedure

Model-Based Fit Statistics for Selected Tree		
N Leaves	ASE	RSS
56	0.1484	73412.7

Variable Importance			
	Training		
Variable	Relative	Importance	Count
age_group	1.0000	192.8	16
education_level	0.3820	73.6688	33
sex	0.1348	26.0035	6

Full Tree Pruning for employed
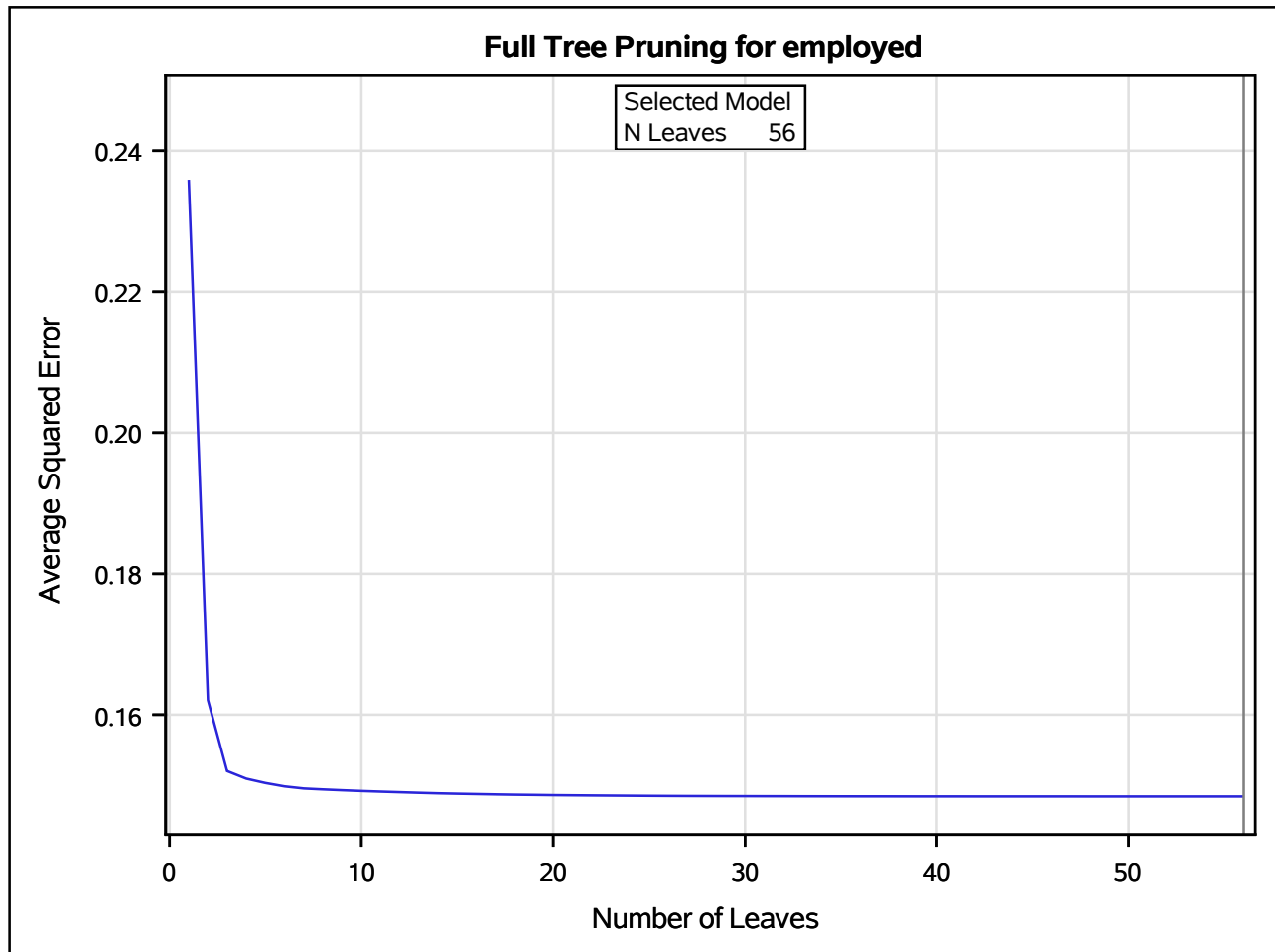
Individual-Level Transition Analysis
Simulating Individual Employment Transitions

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of employed by next_period_employed		
	next_period_employed		
employed	0	1	Total
0	160601 32.46 85.16 73.62	27985 5.66 14.84 10.12	188586 38.12
1	57539 11.63 18.79 26.38	248630 50.25 81.21 89.88	306169 61.88
Total	218140 44.09	276615 55.91	494755 100.00

Statistics for Table of employed by next_period_employed

Statistic	DF	Value	Prob
Chi-Square	1	208526	<.0001
Likelihood Ratio Chi-Square	1	224681	<.0001
Continuity Adj. Chi-Square	1	208523	<.0001
Mantel-Haenszel Chi-Square	1	208525	<.0001
Phi Coefficient		0.64921	
Contingency Coefficient		0.54452	
Cramer's V		0.64921	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	160601
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 494755