

Appendix

Table 1: Variables of adult.data

Variable	Description	Example
"age"	Participant's age	20,21
"workclass"	Participant's type of work	Private, self-emp-inc
"fnlwgt"	Final weight: It represents how many participant with same chrematistic in the real world	10,11
"education"	Participant's education level	Bachelors, Some-college
"education_num"	Represent Participant's education level by number	13, 10
"marital_status"	Participant's marital status	Divorced, Never-married
"occupation "	Participant's occupation	Tech-support, Craft-repair
"relationship"	Participant's social relationship	Wife, Own-child
"race"	Participant's race	White, Asian-Pac-Islander
"sex"	Participant's sex	Female, Male
"capital_gain"	Participant's increase in capital assets	20, 21
"capital_loss"	Participant's loss in capital assets	20, 21
"hours_per_week"	Participant's work how many hours per week	20, 21
"native_country"	Participant's native country	United-States, Cambodia
"Earning"	Participant's income per year	<=50K, >50K
"numeric_earning"	Represent Participant's earning level by binary number	0, 1
"age_group"	Group different age range	1, 2

Table 2: One-way ANOVA by Levene's Test for equal variance

The SAS System

The GLM Procedure

Levene's Test for Homogeneity of numeric_earning Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
education_num	15	108.5	7.2345	161.31	<.0001
Error	32545	1459.6	0.0448		

Table 3: Overall test by two-way ANOVA

Dependent Variable: numeric_earning

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	99	1196.674327	12.087619	82.50	<.0001
Error	32461	4756.137933	0.146519		
Corrected Total	32560	5952.812260			

Table 4: Two-way ANOVA table with interaction

Source	DF	Type I SS	Mean Square	F Value	Pr > F
occupation	14	737.1268059	52.6519147	359.35	<.0001
age_group	6	358.3797940	59.7299657	407.66	<.0001
occupation*age_group	79	101.1677271	1.2806041	8.74	<.0001

Table5: Multiple linear regression output

Parameter	Estimate		Standard Error	t Value	Pr > t	95% Confidence Limits	
Intercept	0.7409200969	B	0.01956097	37.88	<.0001	0.7025798799	0.7792603139
education_num 1	-.7409200969	B	0.05900166	-12.56	<.0001	-.8565655251	-.6252746686
education_num 2	-.7052058111	B	0.03637674	-19.39	<.0001	-.7765055586	-.6339060636
education_num 3	-.6928720488	B	0.02927774	-23.67	<.0001	-.7502575083	-.6354865893
education_num 4	-.6790005922	B	0.02504506	-27.11	<.0001	-.7280898299	-.6299113546
education_num 5	-.6883909140	B	0.02626931	-26.21	<.0001	-.7398797263	-.6369021017
education_num 6	-.6744677925	B	0.02349482	-28.71	<.0001	-.7205184965	-.6284170884
education_num 7	-.6898562671	B	0.02274032	-30.34	<.0001	-.7344281399	-.6452843943
education_num 8	-.6647076257	B	0.02734208	-24.31	<.0001	-.7182991118	-.6111161397
education_num 9	-.5814114786	B	0.01994192	-29.16	<.0001	-.6204983772	-.5423245800
education_num 10	-.5506855611	B	0.02010735	-27.39	<.0001	-.5900967154	-.5112744069
education_num 11	-.4797044673	B	0.02229300	-21.52	<.0001	-.5233995707	-.4360093639
education_num 12	-.4925602093	B	0.02303769	-21.38	<.0001	-.5377149357	-.4474054830
education_num 13	-.3261675292	B	0.02030127	-16.07	<.0001	-.3659587675	-.2863762908
education_num 14	-.1843327492	B	0.02177952	-8.46	<.0001	-.2270214121	-.1416440863
education_num 15	-.0065450969	B	0.02563169	-0.26	0.7985	-.0567841529	0.0436939592
education_num 16	0.0000000000	B

Figure 1: Box-plot of different education level's income

The GLM Procedure

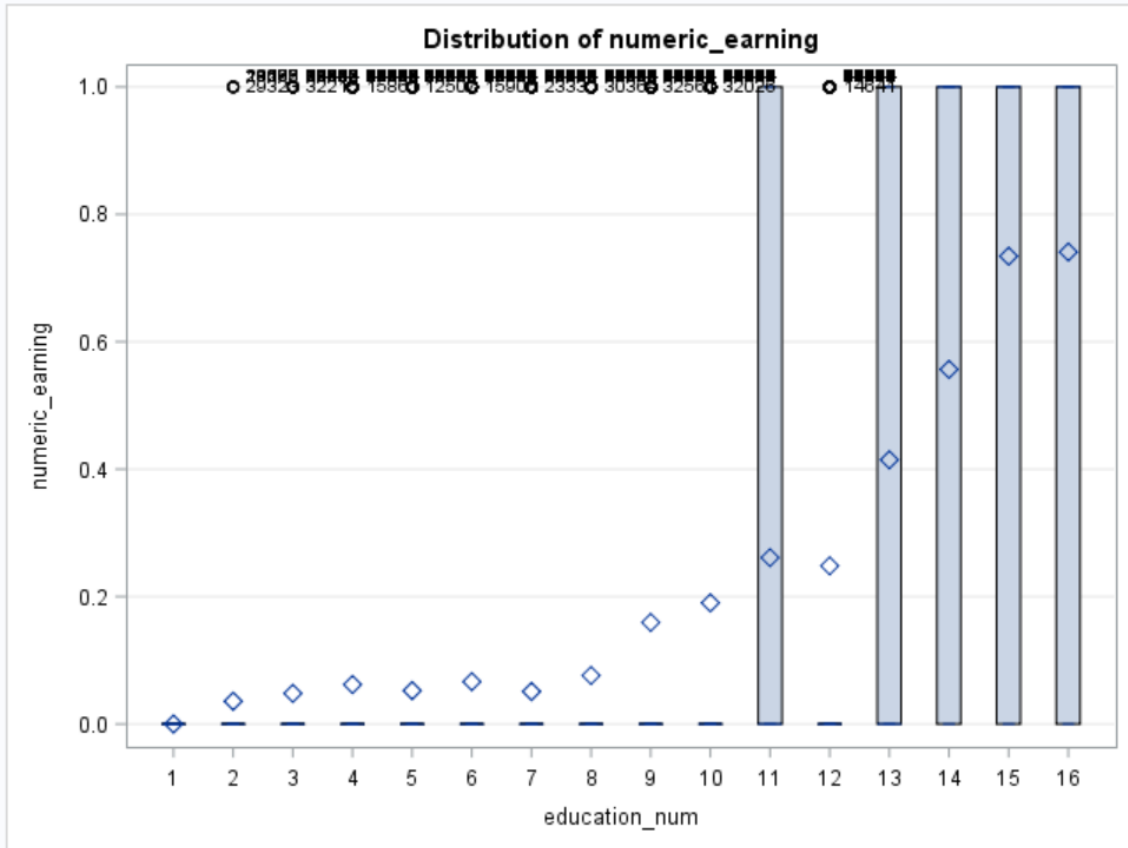
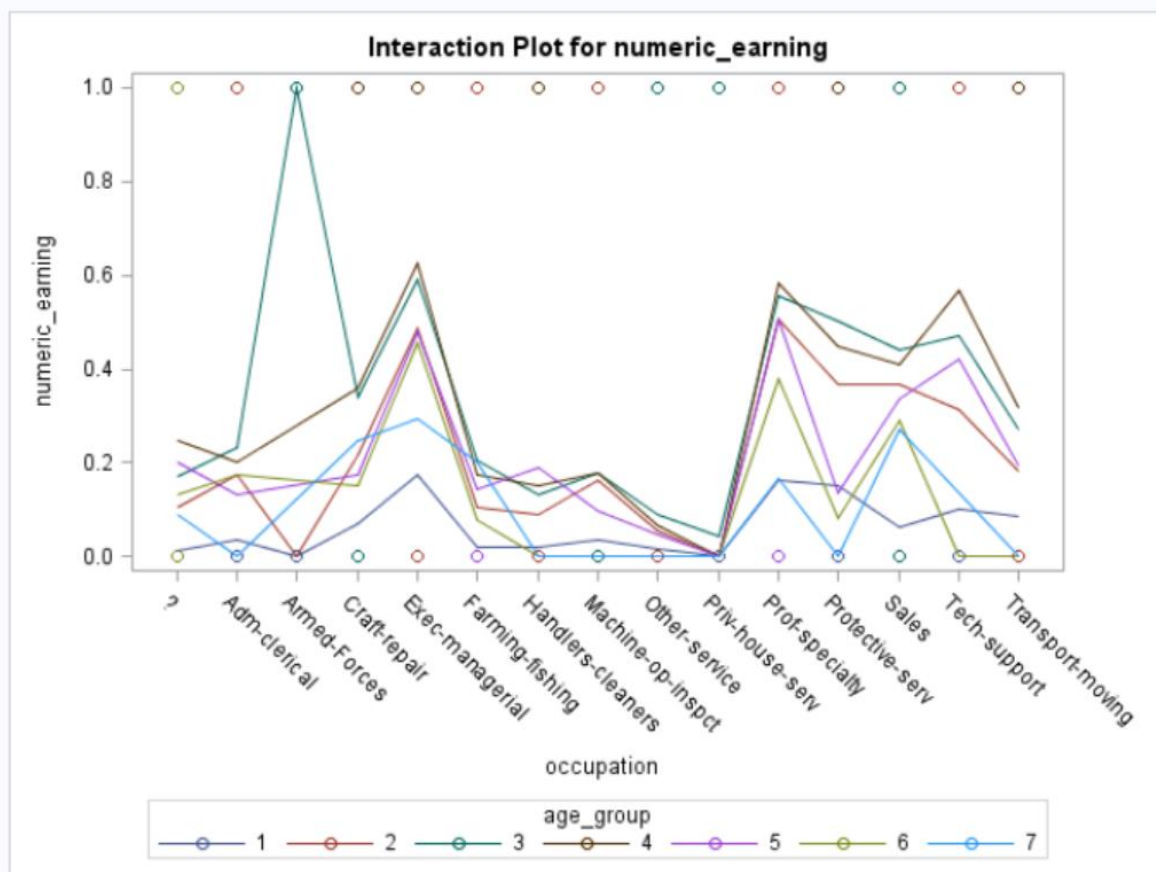


Figure 2: interaction plot for numeric_earning by different occupation and different age groups



Reference:

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.