```
* Encoding: UTF-8.


* SPSS CHAPTER 8

* Table 8.4

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=smoker
  /ORDER=ANALYSIS.
```

Frequencies

Statistics

1 if smokes, 0 otherwise

N	Valid	1122
	Missing	0

1 if smokes, 0 otherwise

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	695	61.9	61.9	61.9
	yes	427	38.1	38.1	100.0
	Total	1122	100.0	100.0	

```
* Table 8.5

DESCRIPTIVES VARIABLES=educ age income pcigs79 smoker
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
education in years	1122	.00	18.00	12.1912	3.29595
age in years	1122	17	88	41.88	17.081
household income in dollars	1122	500	30000	19359.18	9052.516
price of a pack in cents	1122	46.30	69.80	60.9719	4.85213
1 if smokes, 0 otherwise	1122	0	1	.38	.486
Valid N (listwise)	1122				

```
* Table 8.6

SUMMARIZE
  /TABLES=educ age income pcigs79 BY smoker
  /FORMAT=NOLIST TOTAL
  /TITLE='Case Summaries'
  /MISSING=VARIABLE
  /CELLS=COUNT MEAN STDDEV.
```

Summarize

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
education in years * 1 if smokes, 0 otherwise	1122	100.0%	0	0.0%	1122	100.0%
age in years * 1 if smokes, 0 otherwise	1122	100.0%	0	0.0%	1122	100.0%
household income in dollars * 1 if smokes, 0 otherwise	1122	100.0%	0	0.0%	1122	100.0%
price of a pack in cents * 1 if smokes, 0 otherwise	1122	100.0%	0	0.0%	1122	100.0%

Case Summaries

1 if smokes, 0 otherwise		education in years	age in years	household income in dollars	price of a pack in cents
no	N	695	695	695	695
	Mean	12.4273	43.72	19492.81	61.2145
	Std. Deviation	3.51747	18.146	9192.072	4.80040
yes	N	427	427	427	427
	Mean	11.8068	38.89	19141.69	60.5770
	Std. Deviation	2.86197	14.727	8826.999	4.91525
Total	N	1122	1122	1122	1122
	Mean	12.1912	41.88	19359.18	60.9719
	Std. Deviation	3.29595	17.081	9052.516	4.85213

```
* Table 8.7 and Figure 8.27


REGRESSION
```

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT smoker
/METHOD=ENTER educ age income pcigs79.
```

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	price of a pack in cents, household income in dollars, age in years, education in years[b]	.	Enter

a. Dependent Variable: 1 if smokes, 0 otherwise

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.195[a]	.038	.035	.477

a. Predictors: (Constant), price of a pack in cents, household income in dollars, age in years, education in years

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.062	4	2.516	11.044	.000[b]
	Residual	254.434	1117	.228		
	Total	264.496	1121			

a. Dependent Variable: 1 if smokes, 0 otherwise

b. Predictors: (Constant), price of a pack in cents, household income in dollars, age in years, education in years

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.181	.194		6.078	.000
	education in years	-.020	.005	-.133	-4.141	.000
	age in years	-.005	.001	-.161	-5.357	.000
	household income in dollars	7.948E-7	.000	.015	.468	.640
	price of a pack in cents	-.006	.003	-.063	-2.143	.032

a. Dependent Variable: 1 if smokes, 0 otherwise

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT smoker
  /METHOD=ENTER educ age pcigs79 inc_10k.
```

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	hhld income/ 10,000, price of a pack in cents, age in years, education in years^b	.	Enter

a. Dependent Variable: 1 if smokes, 0 otherwise

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.195^a	.038	.035	.477

a. Predictors: (Constant), hhld income/ 10,000, price of a pack in cents, age in years, education in years

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.062	4	2.516	11.044	.000[b]
	Residual	254.434	1117	.228		
	Total	264.496	1121			

a. Dependent Variable: 1 if smokes, 0 otherwise

b. Predictors: (Constant), hhld income/ 10,000, price of a pack in cents, age in years, education in years

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.181	.194		6.078	.000
	education in years	-.020	.005	-.133	-4.141	.000
	age in years	-.005	.001	-.161	-5.357	.000
	price of a pack in cents	-.006	.003	-.063	-2.143	.032
	hhld income/ 10,000	.008	.017	.015	.468	.640

a. Dependent Variable: 1 if smokes, 0 otherwise

```
* Table 8.9 and 8.10 and Figure 8.29

LOGISTIC REGRESSION VARIABLES smoker
  /METHOD=ENTER educ age income pcigs79
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1122	100.0
	Missing Cases	0	.0
	Total	1122	100.0
Unselected Cases		0	.0
Total		1122	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
no	0
yes	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			1 if smokes, 0 otherwise		Percentage Correct
Observed			no	yes	
Step 0	1 if smokes, 0 otherwise	no	695	0	100.0
		yes	427	0	.0
	Overall Percentage				61.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.487	.061	62.763	1	.000	.614

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	education in years	9.384	1	.002
		age in years	21.167	1	.000
		household income in dollars	.398	1	.528
		price of a pack in cents	4.570	1	.033
	Overall Statistics		42.685	4	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	43.475	4	.000
	Block	43.475	4	.000
	Model	43.475	4	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1447.310[a]	.038	.052

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			1 if smokes, 0 otherwise		Percentage Correct
	Observed		no	yes	
Step 1	1 if smokes, 0 otherwise	no	629	66	90.5
		yes	370	57	13.3
	Overall Percentage				61.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.
Step 1[a]	education in years	-.087	.021	16.741	1	.000
	age in years	-.020	.004	27.557	1	.000
	household income in dollars	.000	.000	.249	1	.618
	price of a pack in cents	-.027	.013	4.536	1	.033
	Constant	2.993	.855	12.241	1	.000

Variables in the Equation

		Exp(B)
Step 1[a]	education in years	.917
	age in years	.980
	household income in dollars	1.000
	price of a pack in cents	.973
	Constant	19.937

a. Variable(s) entered on step 1: education in years, age in years, household income in dollars, price of a pack in cents.

```
LOGISTIC REGRESSION VARIABLES smoker
  /METHOD=ENTER educ age pcigs79 inc_10k
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1122	100.0
	Missing Cases	0	.0
	Total	1122	100.0
Unselected Cases		0	.0
Total		1122	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
no	0
yes	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			1 if smokes, 0 otherwise		Percentage Correct
	Observed		no	yes	
Step 0	1 if smokes, 0 otherwise	no	695	0	100.0
		yes	427	0	.0
	Overall Percentage				61.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.487	.061	62.763	1	.000	.614

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	education in years	9.384	1	.002
		age in years	21.167	1	.000
		price of a pack in cents	4.570	1	.033
		hhld income/ 10,000	.398	1	.528
	Overall Statistics		42.685	4	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	43.475	4	.000
	Block	43.475	4	.000
	Model	43.475	4	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1447.310[a]	.038	.052

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			1 if smokes, 0 otherwise		Percentage
	Observed		no	yes	Correct
Step 1	1 if smokes, 0 otherwise	no	629	66	90.5
		yes	370	57	13.3
	Overall Percentage				61.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	education in years	-.087	.021	16.741	1	.000	.917
	age in years	-.020	.004	27.557	1	.000	.980
	price of a pack in cents	-.027	.013	4.536	1	.033	.973
	hhld income/ 10,000	.037	.075	.249	1	.618	1.038
	Constant	2.993	.855	12.241	1	.000	19.937

a. Variable(s) entered on step 1: education in years, age in years, price of a pack in cents, hhld income/ 10,000.

```
* Figure 8.9 and  8.31

* Generalized Linear Models.
GENLIN smoker (REFERENCE=0) WITH educ age inc_10k pcigs79
  /MODEL educ age inc_10k pcigs79 INTERCEPT=YES
 DISTRIBUTION=BINOMIAL LINK=PROBIT
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100 MAXSTEPHALVI
NG=5
    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD) CILEVEL=95
 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.
```

Generalized Linear Models

Model Information

Dependent Variable	1 if smokes, 0 otherwise[a]
Probability Distribution	Binomial
Link Function	Probit

a. The procedure models yes as the response, treating no as the reference category.

Case Processing Summary

	N	Percent
Included	1122	100.0%
Excluded	0	0.0%
Total	1122	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable	1 if smokes, 0 otherwise	no	695	61.9%
		yes	427	38.1%
		Total	1122	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	education in years	1122	.00	18.00	12.1912	3.29595
	age in years	1122	17	88	41.88	17.081
	hhld income/ 10,000	1122	.05	3.00	1.9359	.90525
	price of a pack in cents	1122	46.30	69.80	60.9719	4.85213

Goodness of Fit[a]

	Value	df	Value/df
Deviance	1384.808	1078	1.285
Scaled Deviance	1384.808	1078	
Pearson Chi-Square	1070.590	1078	.993
Scaled Pearson Chi-Square	1070.590	1078	
Log Likelihood[b]	-707.771		
Akaike's Information Criterion (AIC)	1425.542		
Finite Sample Corrected AIC (AICC)	1425.595		
Bayesian Information Criterion (BIC)	1450.656		
Consistent AIC (CAIC)	1455.656		

Dependent Variable: 1 if smokes, 0 otherwise
Model: (Intercept), education in years, age in years, hhld income/ 10,000, price of a pack in cents

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
43.934	4	.000

Dependent Variable: 1 if smokes, 0 otherwise

Model: (Intercept), education in years, age in years, hhld income/ 10,000, price of a pack in cents

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	12.398	1	.000
education in years	17.103	1	.000
age in years	28.290	1	.000
hhld income/ 10,000	.221	1	.639
price of a pack in cents	4.529	1	.033

Dependent Variable: 1 if smokes, 0 otherwise

Model: (Intercept), education in years, age in years, hhld income/ 10,000, price of a pack in cents

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test
			Lower	Upper	Wald Chi-Square
(Intercept)	1.852	.5259	.821	2.882	12.398
education in years	-.054	.0130	-.079	-.028	17.103
age in years	-.013	.0024	-.017	-.008	28.290
hhld income/ 10,000	.021	.0458	-.068	.111	.221
price of a pack in cents	-.017	.0079	-.032	-.001	4.529
(Scale)	1[a]				

Parameter Estimates

Parameter	Hypothesis Test	
	df	Sig.
(Intercept)	1	.000
education in years	1	.000
age in years	1	.000
hhld income/ 10,000	1	.639
price of a pack in cents	1	.033
(Scale)		

Dependent Variable: 1 if smokes, 0 otherwise

Model: (Intercept), education in years, age in years, hhld income/ 10,000, price of a pack in cents

a. Fixed at the displayed value.

```
* Table 8.16/ Figure 8.20

DATASET ACTIVATE DataSet2.
PROBIT n_transit OF n_trips WITH hh_autos hh_size d_cbd
  /LOG NONE
  /MODEL LOGIT
  /PRINT NONE
  /CRITERIA ITERATE(20) STEPLIMIT(.1).
```

Probit Analysis

[DataSet2] C:\Users\Murtaza\Google Drive\AEBE\Data\ch.08\SPSS_08\Stata_grouped _logit.sav

Data Information

		N of Cases
Valid		2707
Rejected	Missing	0
	Number of Responses > Number of Subjects	0
Control Group		9

Convergence Information

	Number of Iterations	Optimal Solution Found
LOGIT	13	Yes

Parameter Estimates

	Parameter	Estimate	Std. Error	Z	Sig.	95% ... Lower Bound
LOGIT[a]	Average autos per household	-1.693	.003	-551.376	.000	-1.699
	Average persons per household	.090	.002	44.277	.000	.086
	Distance from CBD	-.006	.000	-27.387	.000	-.007
	Intercept	.974	.005	189.505	.000	.969

Parameter Estimates

	Parameter	95% Confidence Upper Bound
LOGIT[a]	Average autos per household	-1.687
	Average persons per household	.094
	Distance from CBD	-.006
	Intercept	.979

a. LOGIT model: LOG(p/(1-p)) = Intercept + BX

Covariances and Correlations of Parameter Estimates

		Average autos per household	Average persons per household	Distance from CBD
LOGIT	Average autos per household	.000	-.056	-.551
	Average persons per household	.000	.000	-.275
	Distance from CBD	.000	.000	.000

Covariances (below) and Correlations (above).

Chi-Square Tests

		Chi-Square	df[a]	Sig.
LOGIT	Pearson Goodness-of-Fit Test	241844.755	2703	.000

a. Statistics based on individual cases differ from statistics based on aggregated cases.

```
* Figure 8.25
```

```
DATASET ACTIVATE DataSet1.
GRAPH
  /HISTOGRAM=age
  /PANEL ROWVAR=smoker ROWOP=CROSS.
```

Graph

[DataSet1] C:\Users\Murtaza\Google Drive\AEBE\Data\ch.08\SPSS_08\Stata_smoking
_v12.sav