Using SPSS for Multiple Regression

UDP 520 Lab 7 Lin Lin December 4th , 2007

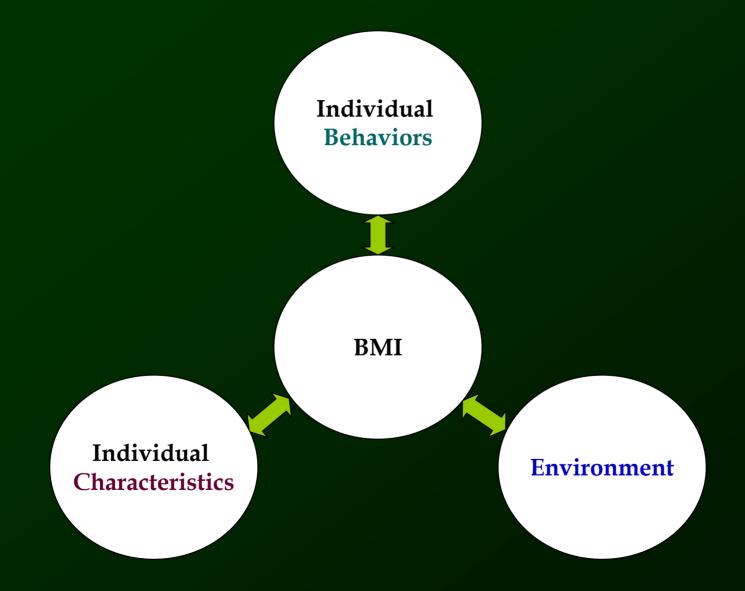


Step 1 — **Define Research Question**

• What factors are associated with BMI?

• Predict BMI.

Step 2 — Conceptualizing Problem (Theory)



Step 2 — Conceptualizing Problem (Theory)

- Individual behaviors are associated with BMI.
- Individual characteristics are associated with BMI.

• Environment is associated with BMI.

Step 3 & 4 — Operationalizing and Hypothesizing

- Individual behaviors are associated with BMI.
 - Eating behavior: daily calorie intake is positively associated with BMI
 - Exercising behavior: level of exercise is negatively associated with BMI.
- Individual characteristics are associated with BMI.
 - Sex
 - Income
 - Education level
 - Occupation
- Environment is associated with BMI.
 - Physical environment
 - Social environment

Step 5 – Collecting Data

- 1000 adults aged 18+ (males and females) were recruited to study factors associated with BMI (BMI)
- Variables
 - BMI (before WLTP)
 - Sex (female=1) individual characteristics
 - Calorie (calorie intake daily) individual behaviors
 - Exercise (minutes of exercise per week) individual behaviors
 - Income (monthly salary in dollars \$) individual characteristics
 - Expenditure on food (monthly food expense in dollars \$) individual behaviors
 - Education (education level in years) individual characteristics
 - Residential density (high, median, low) physical environment

Step 6 – Developing OLS Equation

Multiple regression

$$Y_{\text{BMI}} = \beta_0 + \beta_1 x_{\text{calorie}} + \beta_2 x_{\text{exercise}} + \beta_3 x_{\text{sex}} + \beta_4 x_{\text{income}} + \beta_5 x_{\text{education}} + \beta_6 x_{\text{built environment}} + \varepsilon$$

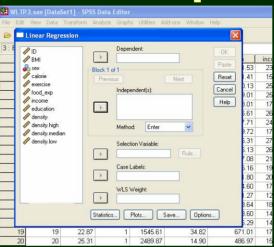
OLS Equation for SPSS

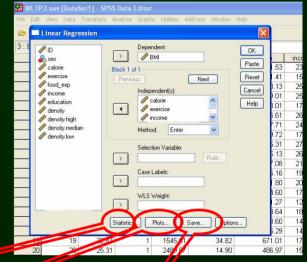
• Multiple regression Model 1

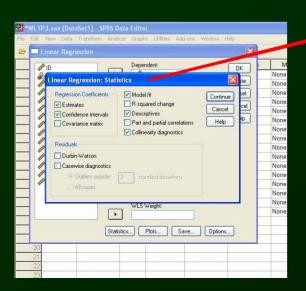
$$Y_{\text{BMI}} = \beta_0 + \beta_1 x_{\text{calorie}} + \beta_2 x_{\text{exercise}} + \beta_4 x_{\text{income}} + \beta_5 x_{\text{education}} + \varepsilon$$

Using SPSS for Multiple Regression

🛃 WLTP	3.sav [Data	Set1] - SPS	S Data Editor						
File Edit	View Data	Transform	Analyze Graph	ns Utilities	Add-ons	Window	Help		
🗁 🔲	🖲 📴 🚸	• • •	Reports Descriptive S	► tatistics	V 🗐	¥ 🔕 🛛			
3 : BMI		2	Tables	•					
	ID	BMI	Compare Mea		exe	ercise	food	ехр	income
1	1	22.5	General Linea		4	18.24		731.53	2334.58
2	2	23.6	Mixed Models Correlate	:)	7	24.80		601.41	1590.08
3	3	24.4	Regression		Linear.	10.00	1	670.13	2541.44
4	4	24.5	Loglinear	•		stimation.		709.01	2597.41
5	5	25.8	Classify	•				- 478.01	1773.81
6	6	23.3	Data Reducti	on 🔸		.ogistic		786.61	2619.12
7	7	24.7	Scale	•	Ordinal	mial Logisti	c	567.71	2416.39
8	8	25.1	Nonparametr	ic Tests 🔸	Prohit			529.72	1765.04
9	9	23.5	Time Series	•				- 335.31	2717.12
10	10	25.2	Survival		Nonline			675.13	2672.94
11	11	25.4	Multiple Resp	onse 🔸	-	Estimation		577.08	2155.30
12	12	24.57	0	2200	2-Stage	e Least Squ	ares	506.16	1999.59
13	13	24.38	1	2099.5	51	20.18		631.80	2054.10







🕅 WI 119	a service of the service of the	t1] - SPSS Da	CARL DECK OF					
File Edit	View Data T	ransform Analy	ze Graph	s Utilities Ad	d.e. window	Help		
🗁 🗖	Linear Regres	sion				×		
3 : E	D ID		Deper	ndent:	_	ОК		
	inear Regress	U. Diata			X	Paste		inc
	inear Regree	Joh: Plots				Paste	1.53	23
	DEPENDNT	Scatter 1 of	1		Continue	Reset	1.41	16
	*ZPRED	Previous		Next		Cancel	0.13	25
	"ZRESID "DRESID				Cancel		9.01	25
	*ADJPRED				Help	Help	8.01	17
	*SRESID *SDRESID		3				6.61	28
	SURESID	► ×					7.71	24
	Standardized R	soldual Plata	_				9.72	17
		esidual i lots	Produc	e all partial plots	5		5.31	27
	🗹 Histogram						5.13	28
	Normal proba	ability plot					7.08	21
							6.16	19
						31	1.80	20
			WLS \	√eight:			1.60	17
		- F					1.27	12
							8.64	18
		Statis	tics]	Plots] [S	ave] [Option:	s	8.60	14
							6.29	14
19	19	22.87	1	1545.61	34.82		1.01	17
20	20	25.31	1	2489.87	14.90	48	6.97	15

🖬 w	/LTP	3.sav [Data	Set1] - SPS	ss p	Editor						
File	Edit	View	Data	Transform	lyze	Graphs	Utilities	Add-ons	Window	Help		
B		Linear	Regr	ession							X	
3 : 6	H	Linear	Dear	ession. Sa								-
0.0			-		ve	20 8 30					К	
	1	Predic				Residu			Contin	ue	ste	.53
				ardized			andardize	d	Cano		set	.41
		Sta					dardized					0.13
		Ad					lentized		Help		ncel	9.01
		S.E	. of m	ean prediction	ns	Dele					elp	8.01
		Distar	nces			Stuc	lentized de	sleted				6.61
		Ma	ihalan	obis		Influen	e Statistic	s				7.71
		Co	ok's			DfB	sta(s)					9.72
		Le	verage	e values		Star	dardized [OfBeta(s)				5.31
		Predic	tion l	ntervals		DiFi						5.13
						Star	dardized [DifFit				7.08
				Interval:	95 %	Cov	ariance rat	io				6.16
					00 %							1.80
				statistics								1.60 1.27
		Cre	eate c	pefficient stati	stics							3.64
			Dreate	a new datas	et							8.60
				Dataset nam	re:							6.29
	19	0	Write .	a new data file	R.						67	1.01
-	20		-	le							48	6.97
	21		- FI	Ie							61	8.84
	22	Expor	t mod	el information	to XML file	e					71	7.48
	23						В	rowse			58	8.48
	24			he covariance							72	5.54
1	25	1 mc	audeit	rie cuvaliario	e maux							3.68
	26	-	20	24.3	24	1	2094.0	1	20.34	2		7.13
	27		27	23.66		1	1833.8	2022	24.71			5.78
	28		28	21.43	2	0	1166.7	7	8.70	1	63	8.04

SPSS Output Tables

Descriptive Statistics

	Mean	Std. Deviation	Ν
BMI	24.0674	1.28663	1000
calorie	2017.7167	513.71981	1000
exercise	21.7947	7.66196	1000
income	2005.1981	509.49088	1000
education	19.95	3.820	1000

Correlations

		BMI	calorie	exercise	income	education
Pearson Correlation	BMI	1.000	.784	310	.033	.011
	calorie	.784	1.000	193	009	.004
	exercise	310	193	1.000	030	046
	income	.033	009	030	1.000	.069
	education	.011	.004	046	.069	1.000
Sig. (1-tailed)	BMI		.000	.000	.148	.361
	calorie	.000		.000	.391	.451
	exercise	.000	.000		.175	.072
	income	.148	.391	.175		.014
	education	.361	.451	.072	.014	
Ν	BMI	1000	1000	1000	1000	1000
	calorie	1000	1000	1000	1000	1000
	exercise	1000	1000	1000	1000	1000
	income	1000	1000	1000	1000	1000
	education	1000	1000	1000	1000	1000

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	education, calorie, income, exercise(a)		Enter

a All requested variables entered.b Dependent Variable: BMI

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.801(a)	.642	.641	.77095

a Predictors: (Constant), education, calorie, income, exercise b Dependent Variable: BMI

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1062.377	4	265.594	446.853	.000(a)
	Residual	591.394	995	.594		
	Total	1653.771	999			

a Predictors: (Constant), education, calorie, income, exercise b Dependent Variable: BMI

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		В	B Std. Error				Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	20.693	.208		99.404	.000	20.285	21.102		
	calorie	.002	.000	.753	38.969	.000	.002	.002	.962	1.039
	exercise	027	.003	163	-8.434	.000	034	021	.960	1.042
	income	8.82E-005	.000	.035	1.837	.067	.000	.000	.994	1.006
	education	001	.006	002	086	.932	013	.012	.993	1.007

a Dependent Variable: BMI

Collinearity Diagnostics(a)

					Variance Proportions						
Model	Dimension	Eigenvalue	Condition Index	(Constant)	calorie	exercise	income	education			
1	1	4.778	1.000	.00	.00	.00	.00	.00			
	2	.110	6.584	.00	.10	.72	.02	.01			
	3	.060	8.924	.00	.41	.03	.56	.00			
	4	.041	10.842	.01	.21	.05	.26	.55			
	5	.011	21.197	.99	.28	.19	.16	.44			

a Dependent Variable: BMI

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	Ν
Predicted Value	21.8115	26.9475	24.0674	1.03123	1000
Residual	-3.36145	4.91952	.00000	.76941	1000
Std. Predicted Value	-2.188	2.793	.000	1.000	1000
Std. Residual	-4.360	6.381	.000	.998	1000

a Dependent Variable: BMI

Step 7 – Checking for Multicollinearity

		Corre	lations			
		BMI	calorie	exercise	income	education
Pearson Correlation	BMI	1.000	.784	.310	.033	.011
	calorie	.784	1.000	193	009	.004
	exercise	310	193	1.000	030	046
	income	.033	009	030	1.000	.069
	education	.011	.004	046	.069	1.000
Sig. (1-tailed)	BMI		.000	.000	.1 <mark>4</mark> 8	.361
	calorie	.000		.000	.3 <mark>9</mark> 1	.451
	exercise	.000	.000		.175	.072
	income	.148	.391	.175		.014
	education	.361	.451	.072	.014	
Ν	BMI	1000	1000	1000	1000	1000
	calorie	1000	1000	1000	1000	1000
	exercise	1000	1000	1000	1000	1000
	income	1000	1000	1000	1000	1000
	education	1000	1000	1000	1000	1000
-						

Check multicollinearity of independent variables. If the absolute value of Pearson correlation is greater than 0.8, collinearity is very likely to exist. If the absolute value of Pearson correlation is close to 0.8 (such as 0.7±0.1), collinearity is likely to exist.

Step 7 – Checking for Multicollinearity (cont.)

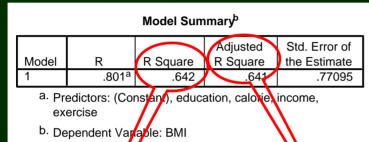
Collinearity Diagnostics								
	Condition			Variance Proportions				
Model	Dimension	Eigenvalue	Index	(Constant)	calorie	exercise	income	education
1	1	4.778	1.000	.00	.00	.00	.00	.00
	2	.110	6.584	.00	.10	.72	.02	.01
	3	.060	8.924	.00	.41	.03	.56	.00
	4	.041	10.842	.01	.21	.05	.26	.55
	5	.011	21.197	.99	.28	.19	.16	.44
a. Dependent Variable: BMI								

A condition index greater than 15 indicates a possible problem

An index greater than 30 suggests a serious problem with collinearity.

Step 8 – Statistics

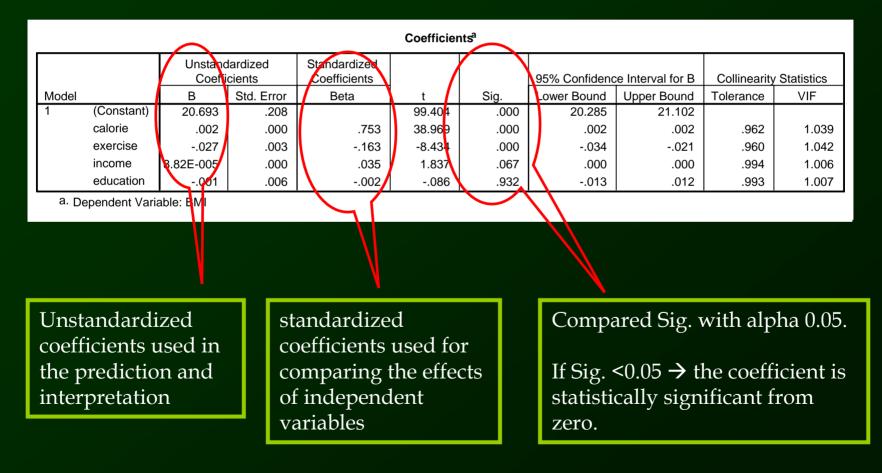
Goodness of fit of model



 $R^2 = 0.642$ It means that 64.2% of variation is explained by the model. The adjusted *R*² adjusts for the number of explanatory terms (independent variables) in a model and increases only if the new independent variable(s) improve(s) the model more than would be expected by chance.

Step 8 – Statistics (cont.)

• Coefficient of each independent variable



Step 9 Interpreting Estimated Coefficient

 $Y_{\rm BMI} = 20.693 + 0.002x_{\rm calorie} + (-0.027)x_{\rm exercise} + 0.0000882x_{\rm income} + (-0.001)x_{\rm education}$

- Controlling for other variables constant, if a person increase 1 calorie intake per day, the BMI of the person will increase by 0.002.
- *Please explain the estimated coefficient of exercise.*

Steps on Model Development and Model Selection

- First, include the theoretically important variables
- Second, include variables that are strongly associated with the dependent variable (to identify independent variables that are strongly associated with the dependent variable, Pearson r test could be used for interval-ratio variables with the dependent variable).
- Third, adjusted *R*² need to be compared to determine if the new independent variables improve the model. At the mean time, multicollinearity needs to be checked.

Notes on Regression Model

- It is VERY important to have theory before starting developing any regression model.
- If the theory tells you certain variables are too important to exclude from the model, you should include in the model even though their estimated coefficients are not significant. (Of course, it is more conservative way to develop regression model.)

BMI data

http://courses.washington.edu/urbdp520/UDP520/BMI.sav

For exercise, you can develop your own conceptual frameworks (theories), create different OLS models, and examine different independent variables.