SOS3003
Applied data analysis for social science
Lecture note 02-2009

Erling Berge
Department of sociology and political science
NTNU

Today

Lecture

- Multiple regression
  - Hamilton Ch 3 p65-101

Seminar

- Writing term paper
- Choosing your dependent variable

Recall:
Bivariate regression: Modelling a sample

- $Y_i = b_0 + b_1 x_{1i} + e_i$
  - i=1,...,n   n = # cases in the sample

- e_i is usually called the residual (mot the error term as in the population model)
- Y and X must be defined unambiguously, and Y must be interval scale (or ratio scale) in ordinary regression (OLS regression)

Recall:
Bivariate regression: Modelling a population

- $Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$
  - i=1,...,n   n = # cases in the population
  - ε_i is the error term for case no i
- Y and X must be defined unambiguously, and Y must be interval scale (or ratio scale) in ordinary regression (OLS regression)

Summary on bivariate regression

- In bi-variate regression the OLS method finds the ”best” LINE or CURVE in a two dimensional scatter plot
- Scatter-plot and analysis of residuals are tools for diagnosing problems in the regression
- Transformations are a general tool helping to mitigate several types of problems, such as
  - Curvilinearity
  - Heteroscedasticity
  - Non-normal distributions of residuals
  - Case with too high influence
- Regression with transformed variables are always curvilinear. Results can most easily be interpreted by means of graphs

Multiple regression: model (1)

- The goal of multiple regression is to find the net impact of one variable controlled for the impact of all other variables
- Let K= number of parameters in the model (this means that K-1 is the number of variables)
- Then the population model can be written
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + ... + \beta_{K-1} x_{i,K-1} + \varepsilon_i$

Multiple regression: model (2)

- This can also be written
  $y_i = E[y_i] + \varepsilon_i$,

  this means that
- $E[y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + ... + \beta_{K-1} x_{i,K-1}$
  $E[y_i]$ is read as “the expected value of y_i”

Multiple regression: model (3)

- We will find the OLS estimates of the model parameters as the b-values in
  $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + ... + b_{K-1} x_{i,K-1}$
  ($\hat{y}_i$ is read as ”estimated” or ”predicted” value of y_i)
  That minimizes the squared sum of the residuals

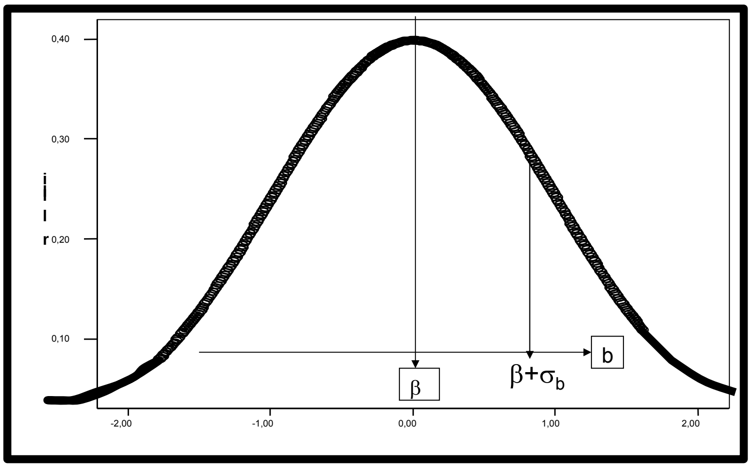
$$RSS = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} e_i^2$$

Estimation methods

- OLS: parameters are found by minimizing RSS
- But this is not the only method for finding suitable b-values. Two alternatives are:
  - WLS: Weighted least squates
  - ML: maximum likelihood

More on testing hypotheses

- We can draw many samples from a population
- In every new sample we can estimate new values (a new b-value) of the same regression parameter (β)
- If we make a histogram of the many estimates of e.g. b_1 we will see that b_1 has a distribution. This distribution is called the sampling distribution of β_1
- Different types of parameters have different types of sampling distributions
- Regression parameters (β-as) have a t-distribution



Sampling distribution of the regression parameter b:

E[b]= β

On partial effects (1)

- Example with 2 variables
- If we estimate a model

  $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$

  it will in principle involve 3 different correlations:
  - Between y and x_1
  - Between y and x_2
  - Between x_1 and x_2

On partial effects (2)

- This might have been represented by 3 different bivariate regressions where the third variable was kept constant

(1) $y = a_{yIx1} + b_{yIx1}x_1 + e_{yIx1}$ x_2 constant

(2) $y = a_{yIx2} + b_{yIx2}x_2 + e_{yIx2}$ x_1 constant

(3) $x_1 = a_{x1Ix2} + b_{x1Ix2}x_2 + e_{x1Ix2}$ y constant

the index ”yIx1” is read ”from the regression of y on x1”

- Equations (2) and (3) can be rewritten as:

On partial effects (3)

(2) $e_{yIx2} = y - (a_{yIx2} + b_{yIx2}x_2)$

(3) $e_{x1Ix2} = x_1 - (a_{x1Ix2} + b_{x1Ix2}x_2)$

We may interprete this as a removal of the effect of x_2 from y and from x_1

We also see that e_{yIx2} and e_{x1Ix2} become new variables where the effect of x_2 has been removed

On partial effects (4)

- If we based on this make a new regression
  $\hat{e}_{y|x2} = a + b\ e_{x1|x2}$
  we find that
  a = 0
  b = b_1 from the regression
  $y_i = b_0 + b_1\ x_{i1} + b_2\ x_{i2} + e_i$
- b_1 is in other words the effect of x_1 on y after we have removed the effect of x_2

Experiments and partial effects

- Experiments investigate the causal connection between two variables controlled for all other causal impacts
- Multiple regression is a kind of half-way replication of experiments – the next best solution – and is a close relative of quasi-experimental research designs

Partial effects

A leverage plot for y and x_k is a plot where

- y-axis is the residual from the regression of y on all x-variables except x_k , and
- x-axis is the residual from regression of x_k on all the other x-variables

The regression line in such a plot will always go through y=0 and will ahve a slope coefficient equal to b_k

An example with 2 independent variables

Table 2.2 Dependent: **Summer 1981 Water Use**	**B**	Std. Error	t	Sig.
(Constant)	**1201.124**	123.325	9.740	.000
Income in Thousands	**47.549**	4.652	10.221	.000

Table 3.1 Dependent: **Summer 1981 Water Use**	**B**	Std. Error	t	Sig.
(Constant)	**203.822**	94.361	2.160	.031
Income in Thousands	**20.545**	3.383	6.072	.000
Summer 1980 Water Use	**.593**	.025	23.679	.000

From the table 2.2 (p46) and 3.1 (p68) in Hamilton. In the tables in the book the constant is on the last line. SPSS put it on the first line.
Question: What does it mean that the coefficient of income declines when we add a new variable?

On the addition of new variables

- It is not common that existing theory will give precise prescriptions for what variables to include in a model. Usually there is an element of trial and error in developing a model
- When new variables are added to a model several things happen
  - The explanatory force increase: R^2 increase, but will the increase be significant?
  - The coefficient of the regression shows the effect on y. Is this effect significantly different from 0?
  - If the coefficient is significantly different from 0, is it also so big that it is of substantial interest?
  - Spurious coefficients can decline. Do the new variable change the interpretation of the effect of the other variables?

Parsimony

- Parsimony is what might be called an aesthetic criterion of a good model. We want to explain as much as possible of the variation in y by means of as few variables as possible
- The adjusted coefficient of determination, Adjusted R^2, is based on parsimony in the sense that it takes into consideration the complexity of the data relative to the complexity of the model by the difference between n and K

  (n-K is the degrees of freedom in the residual,

  n = number of observations, K = number of estimated parameters)

Irrelevant variable

- Including irrelevant variables
  - A variable is irrelevant if the real effect (β) is 0; or more pragmatically, if it si so small that it has no substantive interest
  - **Inclusion of an irrelevant variable** makes the model unnecessarily complex and will have the consequence that coefficient estimates on all variables have larger variance (coefficients varies more form sample to sample)
- Including an irrelevant variable is probably the least damaging error we can do

Relevant variable

- A variable is relevant if
  1. Its real effect (β) is significantly different from 0, and
  2. Large enough to have substantive interest, and
  3. Is **correlated with other included x-variables**
- If we exclude a relevant variable all results from our regression will be unreliable. The model is unrealistically simple
- Not including a relevant variable is the most damaging error we can do. But consider requirement 2 and 3. This makes it a lot easier to avoid this problem.

Sample specific results?

- Choice of variables is a trade-off among risks. Which risk is worse depends on the purpose of the study and the strength of relations
- With a test level of 0.05 one may easily find sample specific results. In about 5% of all samples a coefficient that show up as not significantly different from 0 will in "reality" be different from 0 ($\beta \neq 0$) and vice versa for those we find to be significantly different from 0
- The best defence against this is the theoretical argument for finding an effect different from 0

Hamilton (s74) example

y_i	Postshortage water use (1981)
x_{i1}	Household income, in thousands of dollars
x_{i2}	Preshortage water use, in cubic feet (1980)
x_{i3}	Education of household head, in years
x_{i4}	retirement (coded 1 if household head is retired and 0 otherwise)
x_{i5}	Number of people living in household at time of water shortage (summer 1981)
x_{i6}	Change in number of people, summer 1981 minus summer 1980

Table 3.2 (Hamilton p74)

Dependent Variable: Summer 1981 Water Use	B	Std. Error	t	Sig.	Beta
(Constant)	242.220	206.864	1.171	.242	
Income in Thousands	20.967	3.464	6.053	.000	.184
Summer 1980 Water Use	.492	.026	18.671	.000	.584
Education in Years	-41.866	13.220	-3.167	.002	-.087
Head of house retired?	189.184	95.021	1.991	.047	.058
# of People Resident, 1981	248.197	28.725	8.641	.000	.277
Increase in # of People	96.454	80.519	1.198	.232	.031

How do we interprete the coefficient of ”Increase in # of People” ?

What leads to less water use after the crisis?

Standardized coefficients

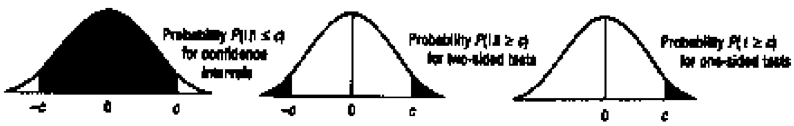
- Standardized variables (z-scores)
  $z_{ix} = (x_i – \bar{x})/s_x$
  (means unit of measurement is standad deviation)
- Standardized regression coefficients (beta-weights, or path coefficients)
  $b_k^s = b_k(s_k/s_y)$ (varies between -1 and +1)
- Predicted standard score of y_i (z_{iy} with hat) =
  $0.18z_{i1} + 0.58z_{i2} – 0.09z_{i3} + 0.06z_{i4} + 0.28z_{i5} + 0.03z_{i6}$

t-test

- The difference between the observed coefficient (b_k) and the unobserved coefficient (β_k) standardized by the standard deviation of the observed coefficient (SE_{b_k}) will usually be very close to zero if the observed b_k is close to the population value. This means that we in the formula
- $t = (b_k - \beta_k)/ SE_{bk}$ substitutes H_0: $\beta_k = 0$ and find that "t" is small we will believe that the population value β_k in reality equals 0
- How big "t" has to be before we stop believing that $\beta_k = 0$ we can find from knowing the sampling distribution of b_k and SE_{b_k}

360 Appendix 4 Statistical Tables

Table A4.1 Critical values for student's t-distribution



Probability

	.50	.80	.90	.95	.98	.99	.995	.998	.999	Confidence Intervals
	.50	.20	.10	.05	.02	.01	.005	.002	.001	Two-Sided Tests
df	.25	.10	.05	.025	.01	.005	.0025	.001	.0005	One-Sided Tests
1	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62	
2	.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598	
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924	
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	
5	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	
7	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408	
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	

"t" has a sampling distribution called the t-distribution The t-distribution varies with the number of degrees of freedom (n-K) and is listed according to level of significance α

Confidence interval for β

- Chose a t_α-value from the table of the t-distribution with n-K degrees of freedom
- Then if $H_0 : \beta_k = b_k$ is correct, a two tailed test will have a probability of α to reject H_0 when H_0 in reality is correct (type I error)
- This means that there is a probability of α that β_k in reality is outside the interval

  $< b_k - t_\alpha(SE_{b_k}) , b_k + t_\alpha(SE_{b_k}) >$
- This is equivalent to saying that

  $b_k - t_\alpha(SE_{b_k}) \le \beta_k \le b_k + t_\alpha(SE_{b_k})$

  is correct with probability 1 – α

F-test: big model against small

RSS{*} = residual sum of squares with index {*} where
* stands for number of parameters in the model

  - Big model: RSS{K}
  - Small model: RSS{K-H}
  - H equals the difference in the number of parameters in the two models
- Define:

$$F^H_{n-K} = \frac{(RSS\{K-H\} - RSS\{K\})/H}{(RSS\{K\})/(n-K)} = F[H, n-K]$$

F[H, n-K] will have the sampling distribution called the F-distribution with H and n-K degrees of freedom

Example (Hamilton table 3.1 and 3.2)

Liten modell Table 3.1	**Sum of Squares**	df	Mean Square	F	Sig.
Regression (Model) (Explained)	671025350.237	2	335512675.119	391.763	.000(a)
Residual	422213359.440	493	856416.551		
Total	1093238709.677	495			

Stor model Tabell 3.2	Sum of Squares	df	Mean Square	F	Sig.
Regression	740477522.059	**K - 1 = 6**	123412920.343	171.076	.000(a)
Residual	352761187.618	**n - K = 489**	721393.022		
Total	1093238709.677	**n - 1 = 495**			

Test if the big model (7 parameters) is better than the small (3 parameters)

Notes to the example

- K = number of parameters of the big model (6 variablar pluss konstant) = 7
- H = K – [number of parameters in the small model (2 variables plus constant)] = 7 – 3 = 4
- RSS{K-H} = 422213359.440
- RSS{K} = 352761187.618
- n = 496
- n – K = 496 – 7 = 489
- (RSS{K-H} – RSS{K})/H = (422213359.440 - 352761187.618)/4 = 17363042.9555
- RSS{K)/(n-K) = 352761187.618/489 = 721393.0217

Testing all parameters in one test

- If the big model has K parameters and we let the small model be as small as possible with only 1 parameter (the constant = the mean) our test will have H=K-1. Inserting this into our formula we have

$$F_{\{K-1,\ n-K\}} = \frac{ESS/(K-1)}{RSS/(n-K)}$$

This is the F-value we find in the ANOVA tables from SPSS

Multicollinearity (1)

- Multicollinearity only involves the x-variables, not y, and is about linear relationships between two or more x-variables
- If there is a perfect correlation between 2 explanatory variables, e.g. x and w (r_{xw} = 1) the multiple regression model breaks down
- The same will happen if there is perfect correlation between two groups of x-variables

Multicollinearity (2)

- Perfect correlation is rarely a practical problem
- But high correlations between different x-variables or between groups of x-variables will make estimates of their effect unreliable. The regression coefficients will have a very large standard deviation and t-tests will practically speaking have no interest whatsoever
- F-tests of groups of variables will not be affected by this

Search strategies

- There are methods for authomatic searches for explanatory variables in a large set of data
- The best advice to give on this is to avoid using it
- One problem is that the p-values of the tests from such searches are wrong and too ”kind” (the problem of multiple comparisons)
- Another problem is that such searches do not work well if the variables are highly correlated

Dummy variables: group differences

- Dichotomous variables taking the values of 0 or 1 are called dummy variables
- In the example in table 3.2 (p74) x_{i4} is (Head of house retired?) a dummy variable
- First put into the equation $x_{i4} = 1$ then $x_{i4} = 0$
  $y_i = 242 + 21x_{i1} + 0.49x_{i2} - 42x_{i3} + 189x_{i4} + 248x_{i5} + 96x_{i6}$ og
- Explain what the two equations tell us

Interaction

- There is interaction between two variables if the effect of one variable changes or varies depending on the value of the other variable

Interaction effects in regression (1)

- If we do a non-linear transformation of y all estimated effects will implicitly be interaction effects
- Simple additive interaction effects can be included in a linear model by means of product terms where two x-variables are multiplied
- $\hat{y}_i = b_0 + b_1x_i + b_2w_i + b_3x_iw_i$
- Conditional effect plots will be able to illustrate what interaction means

Interaction effects in regression (2)

- An interaction effect involving x and w can be included in a regression model by means of an auxiliary variable equal to the product of the two variables, i.e.
- Auxiliary variable H=x*w
- $y_i = b_0 + b_1{*}x_i + b_2{*}w_i + b_3{*}H_i + e_i$
- $y_i = b_0 + b_1{*}x_i + b_2{*}w_i + b_3{*}x_i{*}w_i + e_i$

Example from Hamilton(p85-91)

Let

- y = natural logarithm of chloride concentration
- x = depth of well (1=deep, 0=shallow)
- w = natural logarithm of distance from road
- xw = interaction term between distance and depth (product x*w). Then
- $\hat{y}_i = b_0 + b_1 x_i + b_2 w_i + b_3 x_i w_i$

First take a look at the simple models without interaction

Figures 3.3 and 3.4 (Hamilton p85-86)

Figure 3.3 is based on

Dependent Variable: lnChlorideConcentra	B	Std. Error	Beta	t	Sig.
(Constant)	3.775	.429		8.801	.000
x= BEDROCK OR SHALLOW WELL?	-.706	.477	-.205	-1.479	.145

Figure 3.4 is based on

Dependent Variable: lnChlorideConcentra	B	Std. Error	Beta	t	Sig.
(Constant)	4.210	.961		4.381	.000
w= lnDistanceFromRoad	-.091	.180	-.071	-.506	.615
x= BEDROCK OR SHALLOW WELL?	-.697	.481	-.202	-1.449	.154

Figure 3.3

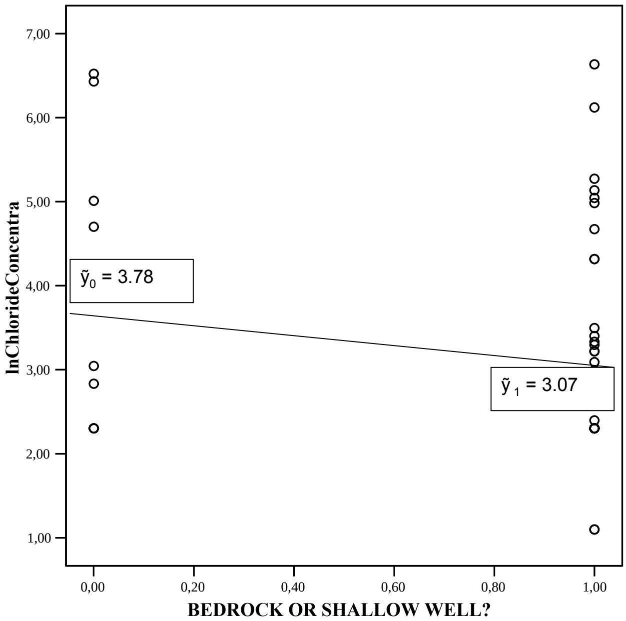
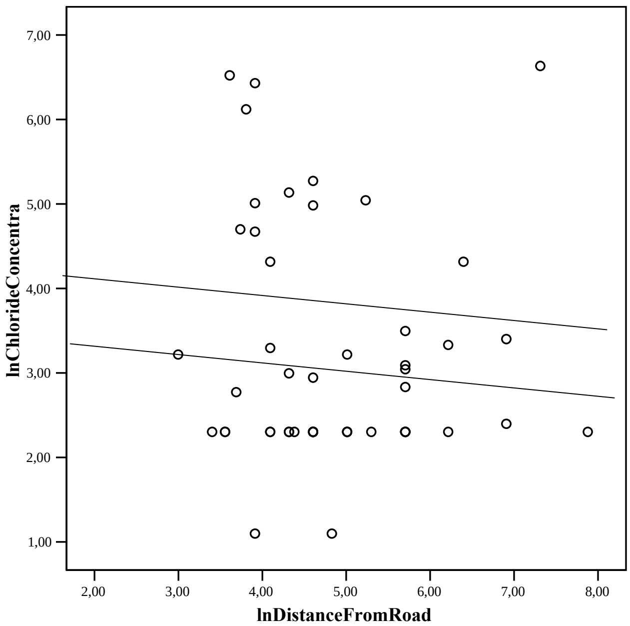


Figure 3.4



Figures 3.5 and 3.6 (Hamilton p89-91)

Figure 3.5 is based on

Dependent Variable: lnChlorideConcentra	B	Std. Error	Beta	t	Sig.
(Constant)	3.666	.905		4.050	.000
w= lnDistanceFromRoad	-.029	.202	-.022	-.144	.886
x*w= lnDroadDeep	-.081	.099	-.128	-.819	.417

Figure 3.6 is based on

Also see Table 3.4 in Hamilton p90 Dependent Variable: lnChlorideConcentra	B	Std. Error	Beta	t	Sig.
(Constant)	9.073	1.879		4.828	.000
w= lnDistanceFromRoad	-1.109	.384	-.862	-2.886	.006
x= BEDROCK OR SHALLOW WELL?	-6.717	2.095	-1.948	-3.207	.002
x*w= lnDroadDeep	1.256	.427	1.979	2.942	.005

Figure 3.5

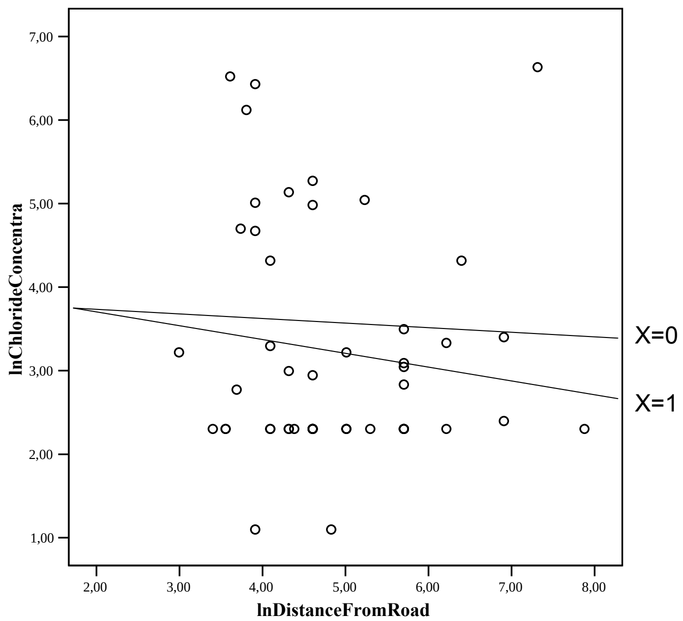
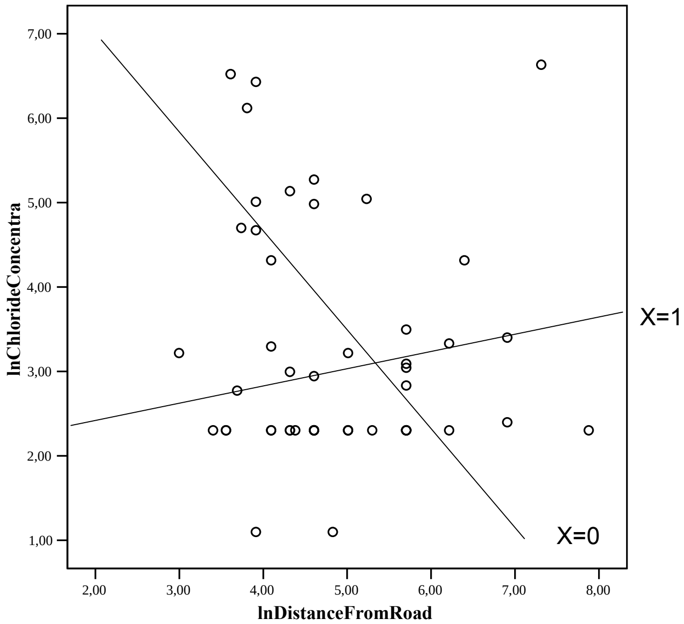


Figure 3.6



Multicollinearity

- Taking all three variables, x, w, and x*w will introduce an element of multicollinearity. This means that we cannot trust tests of single coefficients
- But as shown in the previous example one can not drop any one of the variables without dropping a relevant variable
- F-test of e.g. w and z*w simultaneously circumvents the test problem, and with some experimentation with different models one may see if excluding w or x*w changes the relations substantially

Nominal scale variable

- Can be included in regression models by the use of new auxiliary variables: one for each category of the nominal scale variable. J categories implies H(j), j=1,…,J new auxiliary variables
- If the dependent variable is interval scale and the the only independent variable is nominal scale analysis of variance (ANOVA) is the most common approach to analysis
- By introducing auxiliary variables the same type of analysis can be done in a regression model

Analysis of variance - ANOVA

- Analysing an interval scale dependent variable with one or more nominal scale independent variables, often called factors
  - One way ANOVA uses one nominal scale variable
  - Two way ANOVA uses two nominal scale variable
  - And so on …
- Tests of differences between groups are based on an evaluation of whether the variation within a group (defined by the ”factors”) is large compared to the variation between groups

Nominal scale variables in regression (1)

- If the nominal scale has J categories a maximum of J-1 auxiliary variables can enter the regression

  If H(j), j=1, ... , J-1 are included H(J) have to be excluded
- The excluded auxiliary variable is called the **reference category** and is the most important category in the interpretation of the results from the regression

Nominal scale variables in regression (2)

Dummy coding of a nominal scale variable

- The auxiliary variable H(j) is coded 1 for a person if the person belongs to category j on the nominal scale variable, it is coded 0 if theperson do not belong to category j
- NB: The mean of a dummy coded variable is the proportion in the sample with value 1 (i.e. that belongs in the category)

Nominal scale variables in regression (3)

The reference category

(the excluded auxiliary variable)

- The chosen reference category ought to be large and clearly defined
- The estimated effect of an included auxiliary variable measures the effect of being in the included category relative to being in the reference category

Nominal scale variables in regression (4)

- This means that the regression parameter for an included dummy coded auxiliary variable tells us about additions or subtractions from the expected Y-value a person gets by being in this category rather than in the reference category

Nominal scale variables in regression (5)

Testing I

- Testing if a regression coefficient for an included auxiliary variable equals 0 answers the question whether the persons in this group have a mean Y value different from the mean value of the persons in the reference category

Nominal scale variables in regression (6)

Testing II

- Testing whether a Nominal scale variable contributes significantly to a regression model have to be done by testing if all auxiliary variables in sum contributes significantly to the regression
- For this we use the F-test, applying formula 3.28 in Hamilton (p80)

Nominal scale variables in regression (7)

Interaction

- When dummy coded nominal scale variables are entered into an interaction all included auxiliary variables have to be multiplied with the variable suspected of interacting with it

On terminology (1)

- **Dummy coding** of nominal scale variables are called different names in different textbooks. For example it is
  1. Dummy coding in Hamilton, Hardy, and Weisberg
  2. Indicator coding in Menard (and also Weisberg)
  3. Reference coding or partial method in Hosmer&Lemeshow

On terminology (2)

- To reproduce results from the analysis of variance (ANOVA) by means of regression techniques Hamilton introduces a coding of the auxiliary variables he calls effect coding. Other authors call it differently:
  - It is called effect coding by Hardy
  - It is called deviance coding by Menard
  - It is called the marginal method or deviance method by Hosmer&Lemeshow
- To highlight particular group comparisons Hardy (Ch5) introduces a coding scheme called contrast coding

Ordinal scale variables

- Can be included as an interval scale if the unobserved theoretical dimension is continuous and distance mesures seems resonable
- Also it may be used directly as dependent variable if the program allows ordinal dependent variables
  - In that case parameters are estimated for every level above the lowest as cumulative effects relative to the lowest level

Nominal scale variables

TYPE OF GROUP	Frequency	Percent	Valid Percent	Cumulative Percent
POL	48	12.6	12.6	12.6
FARMER	132	34.7	34.7	47.4
O. PEOPLE	200	52.6	52.6	100.0
Total	**380**	**100.0**	**100.0**	

Example of dummy coding

Nominal scale			Auxiliary	variables	H (*)	
Type of group	Code	N	H(1)= Pol	H(2)= Farmer	H(3)= People	
Politicians	1	48	1	0	0	
Farmers	2	132	0	1	0	
Other People	3	200	0	0	1	**Reference category**

A variable with 3 categories leads to 2 dummy coded variables in a regression with the third used as reference

Example of effect coding

Nominal scala			Auxiliary variable			
Type of group	Code	N	H(1)= Pol	H(2)= Farmer		
Politicians	1	48	1	0		
Farmers	2	132	0	1		
Other People	3	200	-1	-1		Reference category

In effect coding the reference category is coded -1. Effect coding make if possible to duplicate all F-tests of ordinary ANOVA analyses.

Contrast coding

- Is used to present just those comparisons that are of the highest theoretical interest
- Contrast coding requires
  - That with J categories there have to be J-1 contrasts
  - The values of the codes on each auxiliary variable have to sum to 0
  - The values of the codes on any two auxiliary variables have to be orthogonal (their vector product has to be 0)

Use of dummy coded variables(1)

Dependent Variable: I. of political contr. of sales of agric. est.	B	Std. Error	Beta	t	Sig.
(Constant)	4.106	.152		26.991	.000
Pol	.914	.337	.147	2.711	.007
Farmer	.421	.240	.096	1.758	.080

- The constant shows the mean of the dependent variable for those who belong to the reference category
- The mean of the dependent variable for politicians are 0.91 opinion score points above the mean of the reference category
- The mean on the dependent variable for farmers are 0.42 opinion score points above the mean of the reference category

Use of dummy coded variables (2)

Dependent Variable: I. of political control of sales of agricultural estates	B	Std. Error	t	Sig.
(Constant)	4.264	.186	22.954	.000
Number of dekar land Owned	.000	.000	2.176	.030
Pol	.566	.382	1.482	.139
Farmer	-.309	.338	-.913	.362

Compare this table with the previous. What has changed?

How do we interpret the coefficient on ”Pol” and ”Farmer”?

Recall:

Multiple regression: modell

Let K = number of parameters in the model

(then K-1 = number of variables)

Population model

- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} +...+ \beta_{K-1} x_{i,K-1} + \varepsilon_i$
  i = 1, ... ,N; where N = number of case in the population

Sample model

- $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} +...+ b_{K-1} x_{i,K-1} + e_i$
  i = 1, ... ,n; where n = number of case in the sample

Conclusions (1)

- Linear regression can easily be extended to use 2 or more explanatory variables
- If the assumptions of the regression is satisfied (that the error terms are normally distributed with independent and identically distributed errors – normal I.i.d. errors) the regression will be a versatile and strong tool for analytical studies of the connection between a dependent and one or more independent variables

Conclusions (2)

- The most common method of estimating coefficients for a regression model is called OLS (ordinary least squares)
- Coefficients computed based on a sample are seen as estimates of the population coefficient
- Using the t-test we can judge how good such coefficient estimates are
- Using the F-test we may evaluate several coefficient estimates in one test

Conclusions (3)

- Dummy variables are useful in several ways
  - A single dummy coded x-variable will give a test of the difference in means for two groups (0 and 1 groups)
  - Nominal scale variables with more than 2 categories can be recoded by means of dummy coding and included in regression anlysis
  - By using effect coding we can perform analysis of variance of the ANOVA type