11. Multiple Regression

- *y* – response variable

  $x_1, x_2, \ldots, x_k$ -- a set of explanatory variables

In this chapter, all variables assumed to be *quantitative*. Chapters 12-14 show how to incorporate categorical variables also in a regression model.

Multiple regression equation (population):

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

Parameter Interpretation

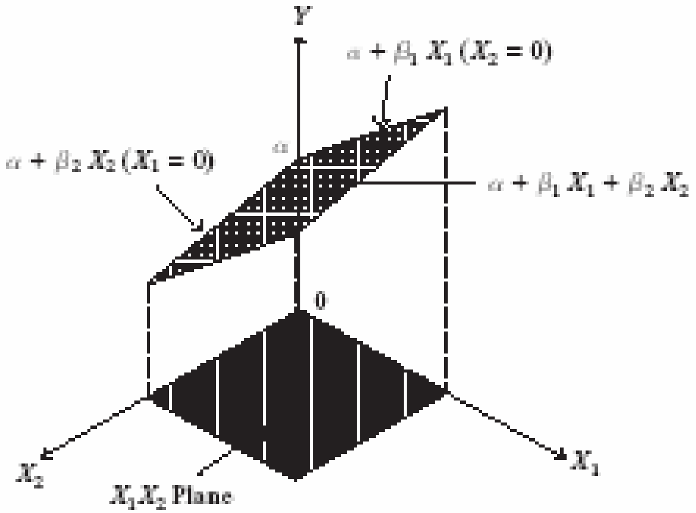
- $\alpha = E(y)$ when $x_1 = x_2 = \ldots = x_k = 0$.

- $\beta_1, \beta_2, \ldots, \beta_k$ are called *partial regression coefficients.*

Controlling for other predictors in model, there is a linear relationship between $E(y)$ and x_1 with slope β_1.

i.e., if x_1 goes up 1 unit with other *x's* held constant, the change in $E(y)$ is

$$[\alpha + \beta_1(x_1 + 1) + \beta_2 x_2 + \ldots. + \beta_k x_k] - [\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots. + \beta_k x_k]$$
$$= \beta_1.$$



Y
$\alpha + \beta_1 X_1 (X_2 = 0)$
α
$\alpha + \beta_2 X_2 (X_1 = 0)$
$\alpha + \beta_1 X_1 + \beta_2 X_2$
0
X_2
X_1
$X_1 X_2$ Plane

Prediction equation

- With sample data, we get “least squares” estimates of parameters by minimizing

SSE = sum of squared prediction errors (residuals)

= Σ(observed *y* – predicted *y*)^2

to get a sample prediction equation

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k$$

Example: Mental impairment study

- *y* = mental impairment (summarizes extent of psychiatric symptoms, including aspects of anxiety and depression, based on questions in "Health opinion survey" with possible responses hardly ever, sometimes, often)

  Ranged from 17 to 41 in sample, mean = 27, *s* = 5.

- x_1 = life events score (composite measure of number and severity of life events in previous 3 years)

  Ranges from 0 to 100, sample mean = 44, *s* = 23

- x_2 = socioeconomic status (composite index based on occupation, income, and education)

  Ranges from 0 to 100, sample mean = 57, *s* = 25

Data (*n* = 40) at www.stat.ufl.edu/~aa/social/data.html and p. 327 of text

Other explanatory variables in study (not used here) include age, marital status, gender, race

- Bivariate regression analyses give prediction equations:

$$\hat{y} = 23.3 + 0.090x_1$$

$$\hat{y} = 32.2 - 0.086x_2$$

- Correlation matrix

Correlations

		mental	life_events	ses
mental	Pearson Correlation	1.000	.372	-.399
	Sig. (2-tailed)		.018	.011
	N	40.000	40	40
life_events	Pearson Correlation	.372	1.000	.123
	Sig. (2-tailed)	.018		.448
	N	40	40.000	40
ses	Pearson Correlation	-.399	.123	1.000
	Sig. (2-tailed)	.011	.448	
	N	40	40	40.000

Prediction equation for multiple regression analysis is:

$$\hat{y} = 28.23 + 0.103x_1 - 0.097x_2$$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	28.230	2.174		12.984	.000	23.824	32.635
	life_events	.103	.032	.428	3.177	.003	.037	.169
	ses	-.097	.029	-.451	-3.351	.002	-.156	-.039

a. Dependent Variable: mental

Predicted mental impairment:

- increases by 0.103 for each 1-unit increase in life events, controlling for SES.
- decreases by 0.097 for each 1-unit increase in SES, controlling for life events.

(e.g., decreases by 9.7 when SES goes from minimum of 0 to maximum of 100, which is relatively large since sample standard deviation of *y* is 5.5)

- Can we compare the estimated partial regression coefficients to determine which explanatory variable is "most important" in the predictions?
- These estimates are *unstandardized* and so depend on units.
- "*Standardized coefficients*" presented in multiple regression output refer to partial effect of a standard deviation increase in a predictor, keeping other predictors constant. (Sec. 11.8).
- In bivariate regression, standardized coefficient = correlation. In multiple regression, std. coeff. relates algebraically to "partial correlations" (Sec. 11.7).
- We skip or only briefly cover Sec. 11.7, 11.8 (lack of time), but I've included notes at end of this chapter on these topics.

Predicted values and residuals

- One subject in the data file (p. 327) has
  y = 33, x_1 = 45 (near mean), x_2 = 55 (near mean)

This subject has predicted mental impairment

$$\hat{y} = 28.23 + 0.103(45) - 0.097(55) = 27.5$$

(near mean)

The prediction error (residual) is 33 – 27.5 = 5.5

i.e., this person has mental impairment 5.5 higher than predicted given his/her values of life events, SES.

SSE = 768.2 smaller than SSE for either bivariate model or for any other linear equation with predictors x_1 , x_2.

Comments

- Partial effects in multiple regression refer to *controlling* other variables in model, so differ from effects in bivariate models, which ignore *all* other variables.
- *Partial effect* of x_1 (controlling for x_2) is same as *bivariate effect* of x_1 when correlation = 0 between x_1 and x_2

  (as is true in most designed experiments).
- Partial effect of a predictor in this multiple regression model is identical at all fixed values of other predictors in model

Example: $\hat{y} = 28.23 + 0.103x_1 - 0.097x_2$

At $x_2 = 0$, $\hat{y} = 28.23 + 0.103x_1 - 0.097(0) = 28.23 + 0.103x_1$

At $x_2 = 100$, $\hat{y} = 28.23 + 0.103x_1 - 0.097(100) = 18.5 + 0.103x_1$

- This parallelism means that this model assumes *no interaction* between predictors in their effects on *y.* (i.e., effect of x_1 does not depend on value of x_2)
- Model is inadequate if, in reality

(insert graph)

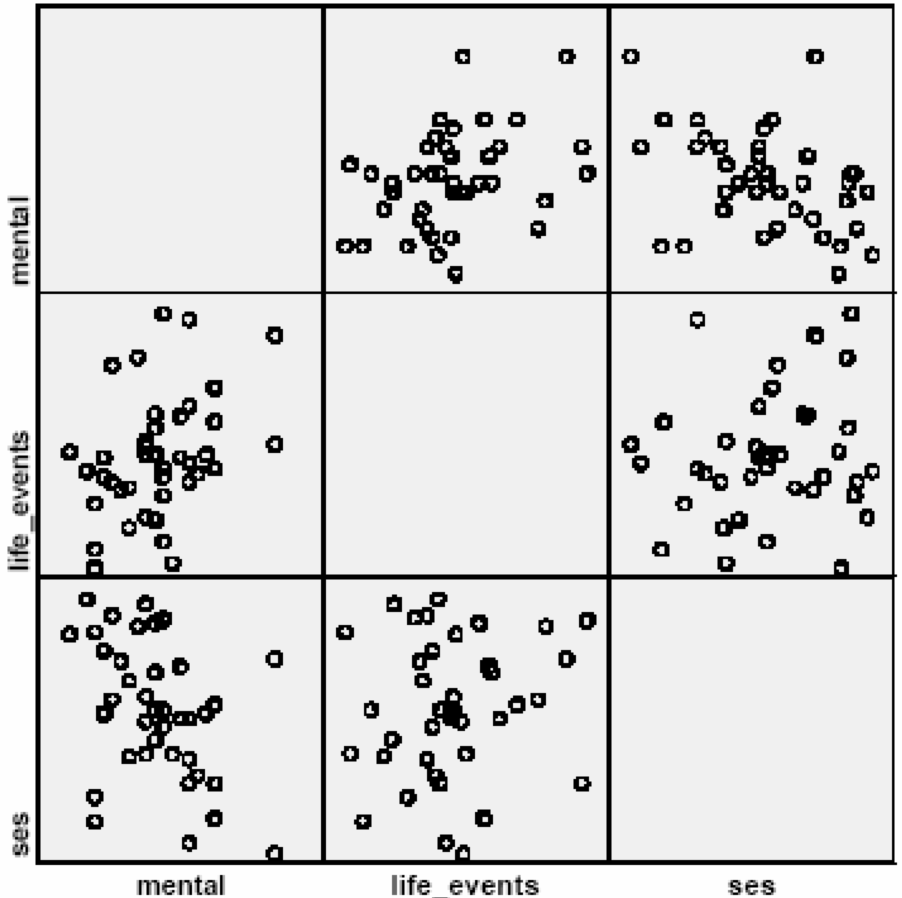
- The model $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ is equivalently expressed as

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

where $\varepsilon = y - E(y)$ = "error" having $E(\varepsilon) = 0$ is population analog of residual $e = y$ – predicted *y.*

Graphics for multiple regression

- *Scatterplot matrix:* Scatterplot for each pair of variables



- *Partial regression plots:* One plot for each predictor, shows its partial effect controlling for other predictors

Example: With two predictors, show partial effect of x_1 on y (i.e., controlling for x_2) by using residuals after

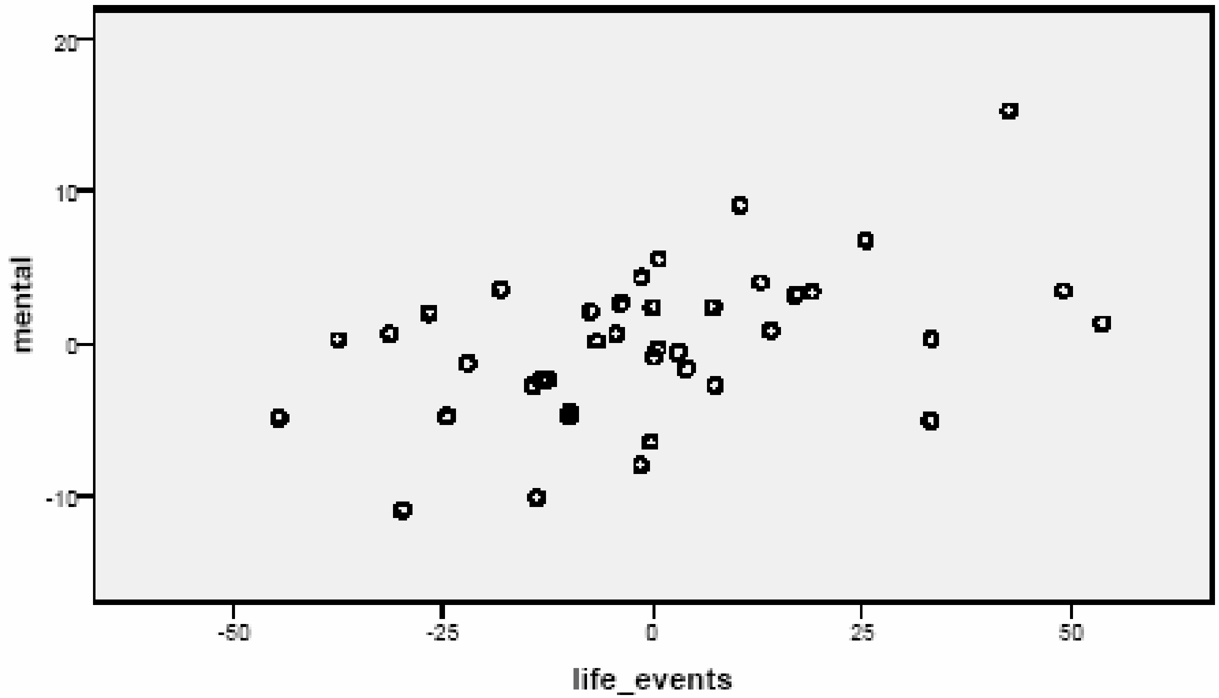
- Regressing y on x_2
- Regressing x_1 on x_2

Partial regression plot is a scatterplot with residuals from regressing y on x_2 on vertical axis and residuals from regressing x_1 on x_2 on horizontal axis.

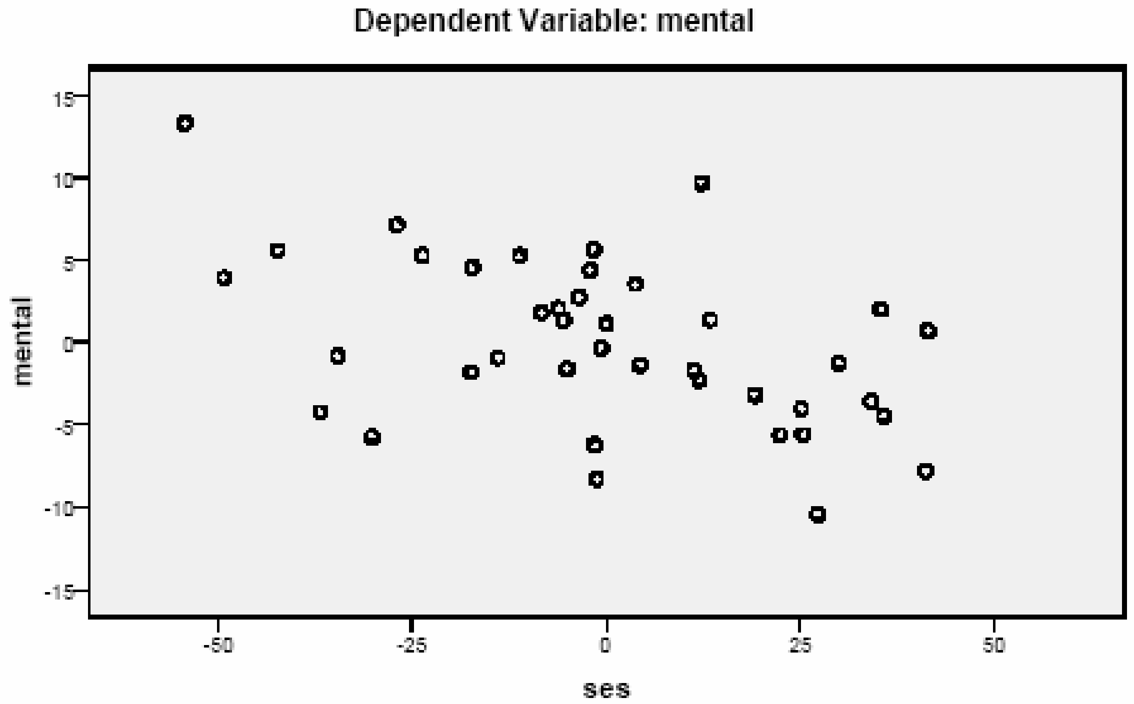
The prediction equation for these points has the same slope as the effect of x_1 in the prediction equation for the multiple regression model.

Partial Regression Plot

Dependent Variable: mental



Partial Regression Plot



Multiple correlation and R^2

- How well do the explanatory variables in the model predict *y*, using the prediction equation?

- The ***multiple correlation,*** denoted by *R,* is the correlation between the observed *y*-values and predicted values

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k$$

from the prediction equation.

i.e., it is the ordinary correlation between *y* and an artificial variable whose values for the *n* subjects in the sample are the predicted values from the prediction equation.

Example: Mental impairment predicted by life events and SES

The multiple correlation is the correlation between the n = 40 pairs of values of observed y and predicted y values:

Subject	y	Predicted y $\hat{y} = 28.23 + 0.103x_1 - 0.097x_2$
1	17	24.8 = 28.23 + 0.103(46) – 0.097(84)
2	19	22.8 = 28.23 + 0.103(39) – 0.097(97)
3	20	28.7 = 28.23 + 0.103(27) – 0.097(24)

……

Software reports R = 0.58

(bivariate correlations with y were 0.37 for x_1, -0.40 for x_2)

- The ***coefficient of multiple determination R^2*** is the proportional reduction in error obtained by using the prediction equation to predict *y* instead of using $\bar{y}$ to predict *y*

$$R^2 = \frac{TSS - SSE}{TSS} = \frac{\Sigma(y-\bar{y})^2 - \Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2}$$

Example:

Predictor	TSS	SSE	R^2
x_1	1162.4	1001.4	0.14
x_2	1162.4	977.7	0.16
x_1, x_2	1162.4	768.2	0.34

For the multiple regression model,

$$R^2 = \frac{TSS - SSE}{TSS} = \frac{\Sigma(y-\bar{y})^2 - \Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2} = \frac{1162.4 - 768.2}{1162.4} = 0.339$$

Software provides an ANOVA table with the sums of squares used in *R*-squared and a Model Summary table with values of *R* and *R*-squared.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	394.238	2	197.119	9.495	.000[a]
	Residual	768.162	37	20.761		
	Total	1162.400	39			

a. Predictors: (Constant), ses, life_events

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.582[a]	.339	.303	4.556

a. Predictors: (Constant), ses, life_events

- There is a 34% reduction in error when we use life events and SES together to predict mental impairment (via the prediction equation), compared to using $\bar{y}$ to predict mental impairment.

- This is sometimes expressed as “34% of the variation in mental impairment is explained by life events and SES.”

- The multiple correlation is *R* = $\sqrt{0.339} = 0.582$

  = the correlation between the 40 values of *y* and the 40 corresponding predicted *y*-values from the prediction equation for the multiple regression model.

Properties of R and R^2

- $0 \le R^2 \le 1$
- $R = +\sqrt{R^2}$ so $0 \le R \le 1$ (i.e., it can't be negative)
- The larger their values, the better the set of explanatory variables predict y
- $R^2 = 1$ when observed y = predicted y, so SSE = 0
- $R^2 = 0$ when all predicted $y = \bar{y}$ so TSS = SSE.

When this happens, $b_1 = b_2 = \ldots = b_k = 0$ and the correlation $r = 0$ between y and each predictor.
- R^2 cannot decrease when predictors added to model
- With single predictor, $R^2 = r^2$, $R = |r|$

- The numerator of R^2, which is TSS – SSE, is called the *regression sum of squares.* This represents the variability in *y* "explained" by the model.

- R^2 is *additive* (i.e., it equals the sum of r^2 values from bivariate regressions of *y* with each *x*) when each pair of explanatory variables is uncorrelated; this is true in many designed experiments, but we don't expect it in observational studies.

- Sample R^2 tends to be a biased estimate (upwards) of population value of R^2 , more so for small *n* (e.g., extreme case -- consider $n = 2$ in bivariate regression!) Software also reports *adjusted* R^2, a less biased estimate (p. 366, Exer. 11.61)

- In observational studies with *many* predictors, explanatory var's are often sufficiently correlated that we can predict one of them well using the others – called *multicollinearity* – consider later.

  (picture)

Inference for multiple regression

Based on assumptions

- Model $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots. + \beta_k x_k$ is (nearly) correct
- Population conditional distribution of *y* is normal, at each combination of predictor values
- Standard deviation σ of conditional dist. of responses on *y* is same at each combination of predictor values

(The estimate *s* of σ is the square root of MSE. It is what SPSS calls "Std. error of the estimate" in the Model Summary table!)

- Sample is randomly selected

Two-sided inference about β parameters is robust to normality and common σ assumptions

Collective influence of explanatory var's

- To test whether explanatory variables collectively have effect on *y*, we test

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

(i.e., *y* independent of all the explanatory variables)

$$H_a: \text{At least one } \beta_i \neq 0$$

(at least one explanatory variable has an effect on *y*, controlling for the others in the model)

Equivalent to testing

H_0: population multiple correlation = 0 (or popul. $R^2 = 0$)

vs. H_a: population multiple correlation > 0

- Test statistic (with *k* explanatory variables)

$$F = \frac{R^2 / k}{(1 - R^2)/[n - (k+1)]}$$

- When H_0 true, *F* values follow the *F* distribution (R. A. Fisher)

- For given *n*, larger *R* gives larger *F* test statistic, more evidence against null hypothesis.

- Since larger *F* gives stronger evidence against null, *P*-value = right-tail probability above observed value

Properties of *F* distribution

- *F* can take only nonnegative values
- Distribution is skewed right
- Exact shape depends on two *df* values:

  $df_1 = k$ (number of explanatory variables in model)
  $df_2 = n - (k + 1)$ (sample size – no. model parameters)

- Mean is approximately 1 (closer to 1 for large *n*)
- *F* tables report *F*-scores for right-tail probabilities such as 0.05, 0.01, 0.001 (one table for each tail prob.)

Example: Is mental impairment independent of life events and SES?

$H_0: \beta_1 = \beta_2 = 0$

(i.e., y independent of x_1 and x_2)

$H_a: \beta_1 \neq 0$ or $\beta_2 \neq 0$ or both

Test statistic

$$F = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} = \frac{0.339/2}{(1-0.339)/[40-(2+1)]} = 9.5$$

$df_1 = 2$, $df_2 = 37$, P-value = 0.000 (i.e., $P < 0.001$)

(From F table, $F = 8.4$ has P-value = 0.001)

- Software provides ANOVA table with result of *F* test about all regression parameters

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	394.238	2	197.119	9.495	.000[a]
	Residual	768.162	37	20.761		
	Total	1162.400	39			

a. Predictors: (Constant), ses, life_events

b. Dependent Variable: mental

- There is very strong evidence that at least one of the explanatory variables is associated with mental impairment.
- Alternatively, can calculate F as ratio of *mean squares* from the ANOVA table.

Example: $F = 197.12/20.76 = 9.5$

- When only k=1 predictor,

$$F = \frac{r^2/1}{(1-r^2)/(n-2)} = \left(\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \right)^2 = t^2$$

the square of the t stat. = *b/(se)* for testing H_0: $\beta = 0$.

(If t is a statistic having a t distribution, then t^2 has the F dist with $df_1 = 1$, $df_2 = df$ for the t statistic.)

Inferences for individual regression coefficients (Need all predictors in model?)

- To test partial effect of x_i controlling for the other explan. var's in model, test H_0: $\beta_i = 0$ using test stat.

  $$t = (b_i - 0)/se, \; df = n - (k + 1)$$

  which is df_2 from the F test (and in *df* column of ANOVA table in *Residual* row)

- CI for β_i has form $b_i \pm t(se)$, with *t*-score from *t*-table also having

  $df = n - (k + 1)$, for the desired confidence level

- Software provides estimates, standard errors, *t* test statistics, *P*-values for tests (2-sided by default)

- In SPSS, check “confidence intervals” under “Statistics” in Linear regression dialog box to get CI’s for regression parameters (95% by default)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	28.230	2.174		12.984	.000	23.824	32.635
	life_events	.103	.032	.428	3.177	.003	.037	.169
	ses	-.097	.029	-.451	-3.351	.002	-.156	-.039

a. Dependent Variable: mental

Example: Effect of SES on mental impairment, controlling for life events

- H_0: $\beta_2 = 0$, H_a: $\beta_2 \neq 0$

  Test statistic $t = b_2/se = -0.097/0.029 = -3.35$,

  $df = n - (k + 1) = 40 – 3 = 37$.

  Software reports *P*-value = 0.002

Conclude there is very strong evidence that SES has a *negative* effect on mental impairment, controlling for life events. (We would reject H_0 at standard significance levels, such as 0.05.)

Likewise for test of H_0: $\beta_1 = 0$ (*P*-value = 0.003), but life events has positive effect on mental impairment, controlling for SES.

A 95% CI for β_2 is $b_2 \pm t(se)$, which is

$$-0.097 \pm 2.03(0.029), \text{ or } (-0.16, -0.04)$$

- This does not contain 0, in agreement with rejecting H_0 for two-sided H_a at 0.05 significance level

- Perhaps simpler to interpret corresponding CI of (-16, -4) for the change in mean mental impairment for an increase of 100 units in SES (from minimum of 0 to maximum of 100).

(relatively wide CI because of relatively small $n = 40$)

Why bother with *F* test? Why not go right to the *t* tests?

A caution: “Overlapping variables” (multicollinearity)

- It is possible to get a small *P*-value in *F* test of
$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ yet not get a small *P*-value for *any* of the *t* tests of individual H_0: $\beta_i = 0$
- Likewise, it is possible to get a small *P*-value in a bivariate test for a predictor but not for its partial test controlling for other variables.
- This happens when the partial variability explained uniquely by a predictor is small. (i.e., each x_i can be predicted well using the other predictors) (picture)

Example (purposely absurd)**:** y = height

x_1 = length of right leg, x_2 = length of left leg

- When multicollinearity occurs,

- *se* values for individual b_i may be large (and individual *t* statistics not significant)

- R^2 may be nearly as large when drop some predictors from model

- It is advisable to simplify the model by dropping some “nearly redundant” explanatory variables.

Modeling interaction between predictors

- Recall that the multiple regression model

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots. + \beta_k x_k$$

assumes the partial slope relating *y* to each x_i is the same at all values of other predictors (i.e., assumes "no interaction" between pairs of predictors)

(recall picture showing parallelism)

For a model allowing interaction between x_1 and x_2 the effect of x_1 may *change* as x_2 changes.

Simplest interaction model: Introduce cross product terms for predictors

Ex: k = 2 explan var's: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2)$
is special case of the multiple regression model

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

with $x_3 = x_1 x_2$ (create x_3 in "transform" menu with "compute variable" option in SPSS)

Example: For mental impairment data, we get

$$\hat{y} = 26.0 + 0.156 x_1 - 0.060 x_2 - 0.00087 x_1 x_2$$

SPSS output for interaction model

(need more decimal places! Highlight table and repeatedly click on the value.)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	26.037	3.949		6.594	.000	18.028	34.045
	life_events	.156	.085	.646	1.826	.076	-.017	.329
	ses	-.060	.063	-.280	-.965	.341	-.188	.067
	cross	.000	.001	-.307	-.668	.509	-.003	.002

a. Dependent Variable: mental

Fixed x_2	Prediction equation for y and x_1
0	$26.0 + 0.156x_1 - 0.060(0) - 0.00087\, x_1(0)$
	$= 26.0 + 0.16x_1$
50	$26.0 + 0.156x_1 - 0.060(50) - 0.00087\, x_1(50)$
	$= 23.0 + 0.11x_1$
100	$26.0 + 0.156x_1 - 0.060(100) - 0.00087\, x_1(100)$
	$= 20.0 + 0.07x_1$

The higher the value of SES, the weaker the relationship between y = mental impairment and x_1 = life events (plausible for these variables)

(picture)

Comments about interaction model

- Note that $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

$$= (\alpha + \beta_2 x_2) + (\beta_1 + \beta_3 x_2) x_1$$

i.e, E(y) is a linear function of x_1

E(y) = (constant with respect to x_1) + (coeff. of x_1)x_1

where coefficient of x_1 is $(\beta_1 + \beta_3 x_2)$.

For fixed x_2 the slope of the relationship between E(y) and x_1 *depends* on the value of x_2.

- To model interaction with $k > 2$ explanatory variables, take cross product for each pair; e.g., $k = 3$:

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3$$

- To test H_0: no interaction in model $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, test H_0: $\beta_3 = 0$ using test statistic

$$t = b_3 / se.$$

Example: $t = -0.00087/0.0013 = -0.67$, $df = n - 4 = 36$.

P-value = 0.51 for $H_a : \beta_3 \neq 0$

Insufficient evidence to conclude that interaction exists.

(It is significant for the entire data set, with $n > 1000$)

- With several predictors, often some interaction terms are needed but not others. E.g., could end up using model such as $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2$
- Be careful not to misintepret “main effect” terms when there is interaction between them in the model.

Comparing two regression models

- How to test whether a model gives a better fit than a simpler model containing only a subset of the predictors?

Example: Compare

$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3$

to

$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

to test H_0: no interaction by testing H_0: $\beta_4 = \beta_5 = \beta_6 = 0$.

- An F test compares the models by comparing their SSE values, or equivalently, their R^2 values.
- The more complex ("complete") model is better if its SSE is sufficiently smaller (or equivalently if its R^2 value is sufficiently larger) than the SSE (or R^2) value for the simpler ("reduced") model.
- Denote the SSE values for the complete and reduced models by SSE_c and SSE_r. Denote the R^2 values by R^2_c and R^2_r.
- The test statistic for comparing the models is

$$F = \frac{(SSE_r - SSE_c)/df_1}{SSE_c/df_2} = \frac{(R_c^2 - R_r^2)/df_1}{(1-R_c^2)/df_2}$$

df_1 = number of extra parameters in complete model,

$df_2 = n-(k+1) = df_2$ for F test that all β terms in complete model = 0

(e.g., $df_2 = n - 7$ for model above)

Example: Mental impairment study (n = 40)

Reduced model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2$

for x_1 = life events score, x_2 = SES

Complete model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

with x_3 = religious attendance (number of times subject attended religious service in past year)

$R^2_r = 0.339$, $R^2_c = 0.358$

Test comparing models has H_0: $\beta_3 = 0$.

Test statistic

$$F = \frac{(R_c^2 - R_r^2)/df_1}{(1 - R_c^2)/df_2} = \frac{(0.358 - 0.339)/1}{(1 - 0.358)/(40 - 4)} = 1.07$$

with $df_1 = 1$, $df_2 = 36$.

P-value = 0.31.

We cannot reject H_0 at the usual significance levels (such as 0.05). The simpler model is adequate.

Note: Since only one parameter in null hypo., the F test statistic is the square of $t = b_3/se$ for testing H_0: $\beta_3 = 0$. The t test also gives P-value = 0.31, for H_a: $\beta_3 \neq 0$

Partial Correlation

- Used to describe association between *y* and an explanatory variable, while controlling for the other explanatory variables in the model

- Notation:

$$r_{yx_1 . x_2}$$

denotes the partial correlation between *y* and x_1 while controlling for x_2.

Formula in text (p. 347), which we'll skip, focusing on interpretation and letting software do the calculation. ("correlate" on "Analyze" menu has a "partial correlation" option)

Properties of partial correlation

- Falls between -1 and +1.
- The larger the absolute value, the stronger the association, controlling for the other variables
- Does not depend on units of measurement
- Has same sign as corresponding partial slope in the prediction equation
- Can regard $r_{yx_1 . x_2}$ as approximating the ordinary correlation between y and x_1 at a fixed value of x_2.
- Equals ordinary correlation found for data points in the corresponding partial regression plot
- Squared partial correlation has a proportional reduction in error (PRE) interpretation for predicting *y* using that predictor, controlling for other explan. var's in model.

Example: Mental impairment as function of life events and SES

- The ordinary correlations are:

  0.372 between *y* and life events

  -0.399 between *y* and SES

- The partial correlations are:

  0.463 between *y* and life events, controlling for SES

  -0.483 between *y* and SES, controlling for life events

Notes:

- Since partial correlation = 0.463 between *y* and life events, controlling for SES,
  and since $(0.463)^2 = 0.21$,

"Controlling for SES, 21% of the variation in mental impairment is explained by life events."

or

"Of the variability in mental impairment unexplained by SES, 21% is explained by life events."

- Test of H_0 : population partial correlation = 0

  is equivalent to *t* test of
  H_0 : population partial slope = 0

For the corresponding regression parameter.

e.g., model $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

H_0 : population partial correlation = 0 between *y* and x_2 controlling for x_1 and x_3

is equivalent to test of

$H_0: \ \beta_2 = 0$

Standardized Regression Coefficients

- Recall for a bivariate model, the correlation is a "standardized slope," reflecting what the slope would be if *x* and *y* had equal standard dev's.
- In multiple regression, there are also *standardized regression coefficients* that describe what the partial regression coefficients would equal if all variables had the same standard deviation.

Def: The *standardized regression coefficient* for an explanatory variable represents the change in the mean of *y* (in *y* std. dev's) for a 1-std.-dev. increase in that variable, controlling for the other explanatory variables in the model

- $r = b(s_x/s_y)$ (bivariate) generalizes to

$$b_1^* = b_1 (s_{x1} / s_y)$$
$$b_2^* = b_2 (s_{x2} / s_y),\ \text{etc.}$$

Properties of standardized regression coeff's:

- Same sign as unstandardized coefficients, but do not depend on units of measurement
- Represent what the partial slopes would be if the standard deviations were equal
- Equals 0 when unstandardized coeff. = 0, so a new test is not needed about significance
- SPSS reports in same table as unstandardized coefficients

Example: Mental impairment by life events and SES

- We found unstandardized equation

$$\hat{y} = 28.23 + 0.103x_1 - 0.097x_2$$

- Standard deviations $s_y = 5.5$, $s_{x1} = 22.6$, $s_{x2} = 25.3$

The standardized coefficient for the effect of life events is

$b_1^* = 0.103(22.6/5.5) = 0.43$. Likewise $b_2^* = -0.45$

- We estimate that E(y) increases by 0.43 standard deviations for a 1 std. dev. increase in life events, controlling for SES.

Note:

An alternative form for prediction equations uses standardized regression coefficients as coefficients of standardized variables

(text, p. 352)

Some multiple regression review questions

Predicted cost = -19.22 + 0.29(food) + 1.58(décor) + 0.90(service)

a. The correct interpretation of the estimated regression coefficient for décor is: for every 1-point increase in the décor score, the cost of dining increases by $1.58
b. Décor is the most important predictor
c. Décor has a positive correlation with cost
d. Ignoring other predictors, it is impossible that décor could have a negative correlation with cost
e. The *t* statistic for testing the partial effect of décor against the alternative of a positive effect could have a *P*-value above 0.50
f. None of the above

- (T/F) For the multiple regression model with two predictors, if the correlation between x_1 and y is 0.50 and the correlation between x_2 and y is 0.50, it is possible that the multiple correlation R = 0.35.

- For the previous exercise in which case would you expect *R* to be larger: when the correlation between x_1 and x_2 is 0, or when it is 1.0?

- (T/F) For every *F* test, there is always an equivalent *t* test.

- Explain with an example what it means for there to be (a) no interaction, (b) interaction between two quantitative explanatory variables in their effects on a quantitative response variable.

- Give an example of interaction when the three variables are categorical (see back in Chap. 10).

- Approximately what value do you expect to see for a *F* test statistic when a null hypothesis is true?

- (T/F) If you get a small *P*-value in the *F* test that all regression coefficients = 0, then the *P*-value will be small in at least one of the *t* tests for the individual regression coefficients.

- Why do we need the *F* (chi-squared) distribution? Why can't we use the *t* (normal) distribution for all hypotheses involving quantitative (categorical) var's?

- What is the connection between (a) *t* and *F*,

  (b) *z* and chi-squared?