SOC 3811/5811: BASIC SOCIAL STATISTICS

Multiple Regression

Multiple Regression with *k* Independent Variables

We've seen how to estimate regression models that include two continuous predictor variables (X_1 and X_2) and a continuous response variable (Y)

Extension #1: Models with *k* continuous predictor variables (X_1 through X_k) and a continuous response variable (Today)

Extension #2: Models with *k* predictor variables—some of which are continuous and some of which are discrete—and a continuous response variable (Bonus Material)

Extension #3: Models with *k* predictor variables and a discrete response variable (SOC 5811 and 8811)

Interpreting Multiple Regression Coefficients

How are a and b_k interpreted in the equation:

$$\hat{Y}_i = a + \sum_{j=1}^{k} b_{kj} X_{ji}$$

Intercept *a*:

The intercept, *a*, equals the predicted value of Y when each of the k predictor variables (X_1 through X_k) equal 0

Multiple regression coefficient (or slope) b_k:

Multiple regression coefficient b_k represents the expected change in Y associated with a one unit increase in X_k, controlling for all other predictors in the model

Coefficient of Determination

As before, we can use R^2 to express the proportion of variation in Y that is accounted for by the predictor variables

Testing Hypotheses about $\rho^2_{Y\bullet X1\ldots Xk}$

Do the k predictor variables collectively explain any of the variation in Y?

We use $R^2_{Y\bullet X1\ldots Xk}$ to estimate $\rho^2_{Y\bullet X1\ldots Xk}$

As before, another way to express $R^2_{Y\bullet X1\ldots Xk}$ is:

$$R^2_{Y\bullet X_1\ldots X_k} = \frac{SS_{REGRESSION}}{SS_{TOTAL}}$$

where $SS_{REGRESSION} = SS_{TOTAL} - SS_{ERROR}$

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

The hypothesis test for $\rho^2_{Y \bullet X1 \ldots Xk}$ is an F test with $df_{NUM}=k$ and $df_{DENOM}=N—k—1$

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

The F statistic when there are k predictors is

$$F_{k,N-k-1} = \frac{SS_{REGRESSION}/k}{SS_{ERROR}/N-k-1} = \frac{MS_{REGRESSION}}{MS_{ERROR}}$$

Computationally:

$$SS_{TOTAL} = (s_Y^2)(N-1)$$

$$SS_{REGRESSION} = (R^2_{Y \bullet X_1 \ldots X_k})(SS_{TOTAL})$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION}$$

Testing Hypotheses about β_k

Can we conclude that β_k is different from 0?

We use b_k to estimate β_k

In the model with k predictor variables the variance of the sampling distribution of b_k is

$$s_{b_k}^2 = \frac{MS_{ERROR}}{\left(s_{X_k}^2\right)(N-1)\left(1-R_{X_k \bullet X_1 \ldots X_{k-1}}^2\right)}$$

Testing Hypotheses about β_k

The hypothesis test for β_k is a t tests with N-k-1 degrees of freedom (because MS_{ERROR} has N-k-1 degrees of freedom)

Testing Hypotheses about β_k

The t statistic for β_k is

$$t_{N-k-1} = \frac{b_k - 0}{s_{b_1}} = \frac{b_k - 0}{\sqrt{\dfrac{MS_{ERROR}}{\left(s^2_{X_k}\right)(N-1)\left(1-R^2_{X_k \bullet X_1 \ldots X_{k-1}}\right)}}}$$

WISCONSIN LONGITUDINAL STUDY

Variable	Obs	Mean	Std. Dev.	Min	Max
NetWorth	4,393	673.1551	1216.293	-40	11325
Beauty	4,393	.0518019	1.245267	-3.408672	4.000115
Weight	4,393	6.230875	1.189261	4	9.7
Height	4,393	65.8892	3.63164	60	74
IQ	4,393	102.7539	14.58032	61	145
MomEd	4,393	10.59026	2.814809	0	21

(obs=4,393)

	NetWorth	Beauty	Weight	Height	IQ	MomEd
NetWorth	1.0000					
Beauty	0.0883	1.0000				
Weight	-0.0433	-0.1597	1.0000			
Height	0.1445	0.0506	-0.0557	1.0000		
IQ	0.1881	0.0734	-0.1007	0.0747	1.0000	
MomEd	0.1135	0.1123	-0.0765	0.0875	0.2267	1.0000

Number of obs =	4,393
R-squared =	0.0605
Adj R-squared =	0.0594
Root MSE =	1179.6

NetWorth	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
Beauty	61.27388	14.57228			
Weight	-5.852128	15.24806			
Height	41.49432	4.935395			
IQ	13.37102	1.260568			
MomEd	25.42056	6.544364			
_cons	-3670.718	361.0301			

Source	SS	df	MS
Model	393152293	5	78630458.7
Residual	6.1042e+09	4,387	1391436.97
Total	6.4974e+09	4,392	1479368.46

Number of obs	=	4,393
F(5, 4387)	=	56.51
Prob > F	=	0.0000
R-squared	=	0.0605
Adj R-squared	=	0.0594
Root MSE	=	1179.6

NetWorth	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Beauty	61.27388	14.57228	4.20	0.000	32.70485	89.84291
Weight	-5.852128	15.24806	-0.38	0.701	-35.74602	24.04176
Height	41.49432	4.935395	8.41	0.000	31.81845	51.17019
IQ	13.37102	1.260568	10.61	0.000	10.89967	15.84237
MomEd	25.42056	6.544364	3.88	0.000	12.5903	38.25082
_cons	-3670.718	361.0301	-10.17	0.000	-4378.519	-2962.916

What did we learn about the world???

MEN

NetWorth	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Beauty	29.39734	26.28224	1.12	0.263	-22.14585	80.94054
Weight	14.8509	26.28152	0.57	0.572	-36.69087	66.39267
Height	33.49823	12.40042	2.70	0.007	9.179243	57.81721
IQ	17.60229	2.158089	8.16	0.000	13.36998	21.83461
MomEd	24.60574	11.66347	2.11	0.035	1.732035	47.47945
_cons	-3613.391	882.2239	-4.10	0.000	-5343.557	-1883.226

NetWorth	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Beauty	84.95515	15.71286	5.41	0.000	54.14274	115.7676
Weight	-21.09401	17.07663	-1.24	0.217	-54.58073	12.39271
Height	-3.925945	8.301222	-0.47	0.636	-20.20437	12.35248
IQ	9.927091	1.383271	7.18	0.000	7.214542	12.63964
MomEd	26.57815	6.893811	3.86	0.000	13.05961	40.09669
_cons	-404.2691	548.3631	-0.74	0.461	-1479.591	671.0532

NOW what did we learn about the world???