

Regression Diagnostics Example

Salary, years since Ph.D., and number of publications example from Cohen, Cohen. West, & Aiken, Table 3.2.1.

Syntax

list vars=all.

```

regression vars=time pubs salary
  /descriptives =mean stddev
  /statistics=r coeff ses anova
  /dependent=salary
  /method=enter time pubs
  /casewise=all sresid sdresid mahal cook lever sdbeta sdfit
  /scatterplot (*zresid *zpred).
    
```

Results with Original Data

R-square = .530, F(2,12) = 6.78, p = .011.

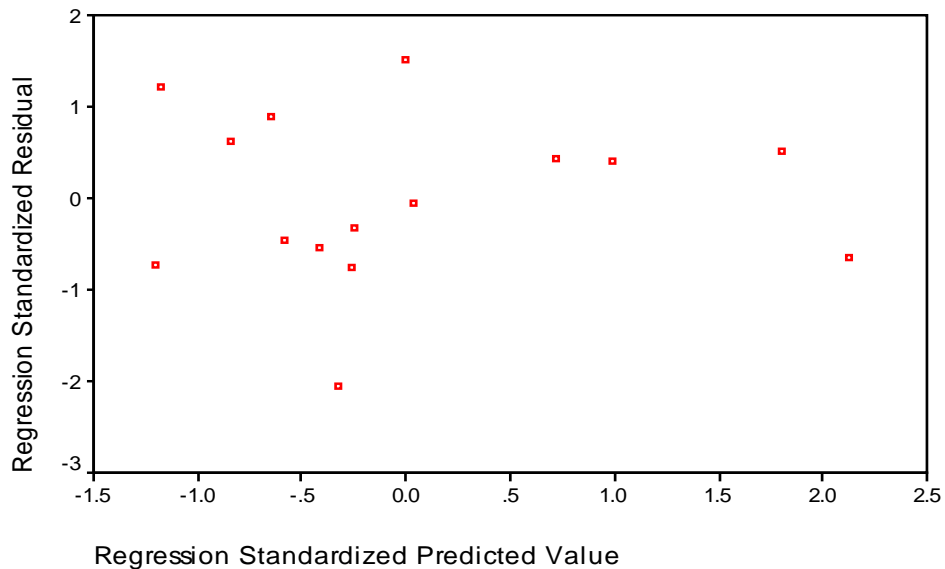
List				Model							
ID	TIME	PUBS	SALARY	Unstandardized Coefficients		Standardized Coefficients		t	Sig.		
				B	Std. Error	Beta	Std. Error				
1	3	18	51876	1	(Constant)	43082.394	3099.493			13.900	.000
2	6	3	54511		TIME years since PhD	982.867	452.057	.570	.262	2.174	.050
3	3	2	53425		PUBS number of publications	121.801	149.699	.213	.262	.814	.432
4	8	17	61863								
5	9	11	52926								
6	6	6	47034								
7	16	38	66432								
8	10	48	61100								
9	2	9	41934								
10	5	22	47454								
11	5	30	49832								
12	6	21	47047								
13	7	10	39115								
14	11	27	59677								
15	18	37	61458								

†
 a Dependent Variable: SALARY annual salary in dollars

Number of cases read: 15
 Number of cases listed: 15

Scatterplot

Dependent Variable: annual salary in dollars



Casewise Diagnostics(a)

Case Number	Std. Residual	SALARY annual salary in dollars	Stud. Residual	Stud. Deleted Residual	Centered Leverage Value	Mahal. Distance	Cook's Distance	Std. DFFIT	Std. DFBETA		
									(Constant)	PUBS number of publications	TIME years since PhD
1	.626	51876	.689	.673	.109	1.532	.034	.311	.232	.139	-.244
2	.885	54511	.988	.987	.132	1.842	.080	.491	.317	-.385	.172
3	1.225	53425	1.361	1.417	.124	1.733	.145	.687	.640	-.350	-.094
4	1.515	61863	1.576	1.694	.009	.124	.068	.484	.229	-.162	.132
5	-.059	52926	-.064	-.061	.094	1.318	.000	-.027	-.009	.020	-.017
6	-.458	47034	-.497	-.481	.084	1.172	.015	-.202	-.141	.142	-.055
7	.513	66432	.615	.599	.238	3.336	.055	.397	-.211	.028	.245
8	.401	61100	.539	.523	.380	5.315	.078	.469	-.089	.422	-.205
9	-.721	41934	-.794	-.781	.110	1.533	.045	-.361	-.347	-.007	.219
10	-.552	47454	-.590	-.574	.060	.837	.017	-.218	-.131	-.116	.148
11	-.311	49832	-.359	-.345	.179	2.509	.014	-.197	-.066	-.156	.149
12	-.769	47047	-.806	-.793	.022	.308	.021	-.247	-.151	-.093	.122
13	-2.066	39115	-2.199	-2.724	.050	.704	.213	-.991	-.606	.640	-.335
14	.427	59677	.451	.436	.038	.532	.008	.149	-.019	.005	.064
15	-.655	61458	-.874	-.864	.372	5.204	.199	-.764	.419	.101	-.591

With an outlier on X2 - NUMPUB.

Overall R-square = .363, $F(2,12) = 3.42$, $p = .067$.

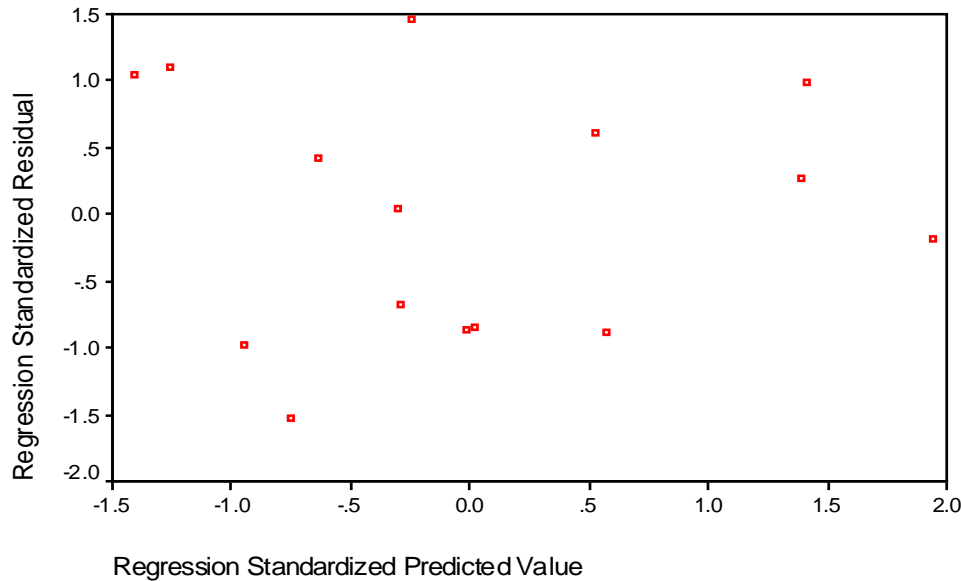
ID	TIME	PUBS	SALARY
1	3	18	51876
2	6	3	54511
3	3	2	53425
4	8	17	61863
5	9	11	52926
6	36	6	47034
7	16	38	66432
8	10	48	61100
9	2	9	41934
10	5	22	47454
11	5	30	49832
12	6	21	47047
13	7	10	39115
14	11	27	59677
15	18	37	61458

Number of cases read: 15
 Number of cases listed: 15

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
		B	Std. Error	Beta	Std. Error		
1	(Constant)	45334.772	3631.295			12.484	.000
	TIME years since PhD	121.069	212.659	.132	.232	.569	.580
	PUBS number of publications	328.118	132.174	.575	.232	2.482	.029

Scatterplot

Dependent Variable: annual salary in dollars



Casewise Diagnostics(a)

Case Number	Std. Residual	SALARY annual salary in dollars	Stud. Residual	Stud. Deleted Residual	Centered Leverage Value	Mahal. Distance	Cook's Distance	Std. DFFIT	Std. DFBETA		
									(Constant)	PUBS number of publications	TIME years since PhD
1	.040	51876	.042	.041	.043	.606	.000	.014	.011	-.001	-.009
2	1.097	54511	1.212	1.239	.114	1.595	.108	.582	.550	-.435	-.113
3	1.039	53425	1.175	1.196	.151	2.108	.128	.630	.623	-.443	-.235
4	1.467	61863	1.523	1.624	.005	.075	.060	.452	.318	-.087	-.078
5	.425	52926	.447	.432	.030	.418	.007	.141	.114	-.078	-.002
6	-.680	47034	-1.828	-2.060	.795	11.127	6.925	-5.137	.559	1.947	-4.704
7	.984	66432	1.110	1.122	.149	2.082	.113	.588	-.264	.420	.207
8	-.176	61100	-.220	-.211	.296	4.149	.009	-.159	.065	-.144	.011
9	-.970	41934	-1.057	-1.063	.093	1.296	.071	-.463	-.447	.219	.254
10	-.839	47454	-.879	-.870	.024	.337	.026	-.275	-.155	-.050	.137
11	-.875	49832	-.939	-.934	.065	.912	.045	-.364	-.066	-.210	.166
12	-.868	47047	-.905	-.898	.014	.196	.024	-.266	-.157	-.030	.109
13	-1.521	39115	-1.610	-1.741	.041	.574	.104	-.605	-.545	.341	.119
14	.610	59677	.638	.622	.019	.272	.013	.191	.008	.087	.018
15	.265	61458	.302	.290	.161	2.251	.009	.157	-.075	.101	.075

With an outlier on Y - Salary

R-square = .011, $F(2,12) < 1$, ns.

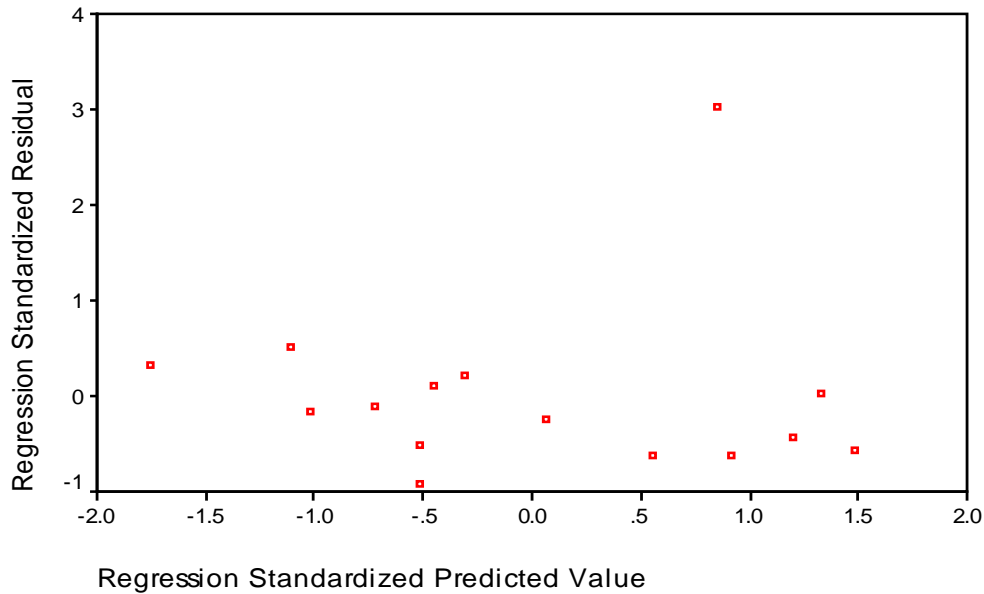
ID TIME PUBS SALARY				Coefficients(a)							
1	3	18	51876.0	Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
2	6	3	54511.0			B	Std. Error	Beta	Std. Error		
3	3	2	53425.0	1	(Constant)	59775.198	10585.421			5.647	.000
4	8	17	61863.0		TIME years since PhD	-556.096	1543.870	-.137	.381	-.360	.725
5	9	11	52926.0		PUBS number of publications	137.368	511.252	.102	.381	.269	.793
6	6	6	47034.0								
7	16	38	66432.0								
8	10	48	61100.0								
9	2	9	120000.0								
10	5	22	47454.0								
11	5	30	49832.0								
12	6	21	47047.0								
13	7	10	39115.0								
14	11	27	59677.0								
15	18	37	61458.0								

Number of cases read:
15 Number of cases listed: 15

Charts

Scatterplot

Dependent Variable: annual salary in dollars



Casewise Diagnostics(a)

Case Number	Std. Residual	SALARY annual salary in dollars	Stud. Residual	Stud. Deleted Residual	Centered Leverage Value	Mahal. Distance	Cook's Distance	Std. DFFIT	Std. DFBETA		
									(Constant)	PUBS number of publications	TIME years since PhD
1	-.436	51876.0	-.481	-.465	.109	1.532	.016	-.215	-.160	-.096	.168
2	-.117	54511.0	-.131	-.126	.132	1.842	.001	-.062	-.040	.049	-.022
3	-.249	53425.0	-.276	-.265	.124	1.733	.006	-.129	-.120	.066	.018
4	.211	61863.0	.219	.210	.009	.124	.001	.060	.028	-.020	.016
5	-.168	52926.0	-.184	-.176	.094	1.318	.002	-.077	-.025	.057	-.049
6	-.513	47034.0	-.556	-.540	.084	1.172	.018	-.227	-.158	.160	-.062
7	.518	66432.0	.622	.605	.238	3.336	.057	.401	-.213	.029	.247
8	.015	61100.0	.020	.019	.380	5.315	.000	.017	-.003	.015	-.007
9	3.014	120000	3.320	11.154	.110	1.533	.786	5.159	4.951	.096	-3.129
10	-.630	47454.0	-.674	-.658	.060	.837	.022	-.250	-.151	-.133	.170
11	-.566	49832.0	-.652	-.635	.179	2.509	.046	-.363	-.122	-.288	.275
12	-.616	47047.0	-.645	-.628	.022	.308	.013	-.196	-.120	-.074	.097
13	-.910	39115.0	-.968	-.965	.050	.704	.041	-.351	-.215	.227	-.119
14	.116	59677.0	.122	.117	.038	.532	.001	.040	-.005	.001	.017
15	.331	61458.0	.442	.427	.372	5.204	.051	.377	-.207	-.050	.292

ϕ
a Dependent Variable: SALARY annual salary in dollars

Outlier on X and Y - an influential data point.

R-square = .840, F(2,12) = 31.41, p < .001

ID	TIME	PUBS	SALARY
1	3	18	51876.0
2	6	3	54511.0
3	3	2	53425.0
4	8	17	61863.0
5	9	11	52926.0
6	6	6	47034.0
7	16	38	66432.0
8	10	48	61100.0
9	32	9	120000
10	5	22	47454.0
11	5	30	49832.0
12	6	21	47047.0
13	7	10	39115.0
14	11	27	59677.0
15	18	37	61458.0

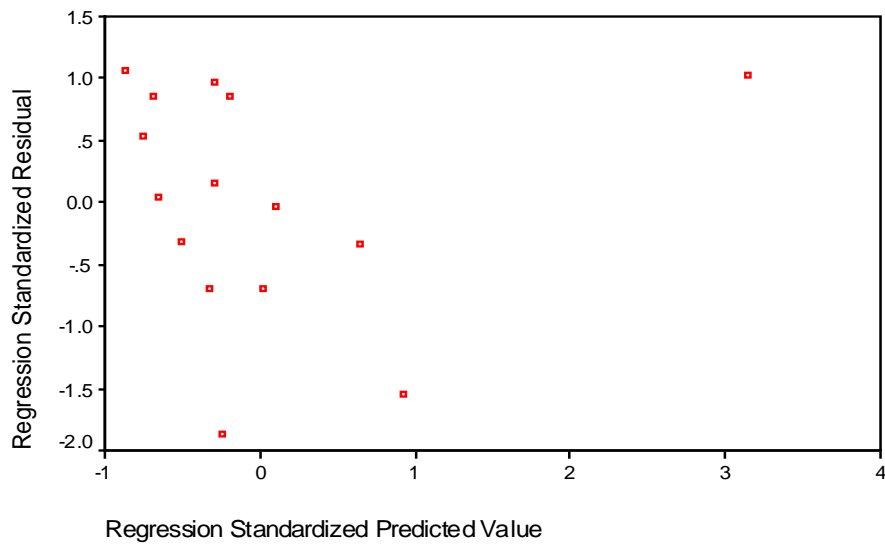
Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
		B	Std. Error	Beta	Std. Error		
1	(Constant)	40079.314	4352.857			9.208	.000
	TIME years since PhD	2294.502	289.531	.930	.117	7.925	.000
	PUBS number of publications	-201.145	157.676	-.150	.117	-1.276	.226

Number of cases read: 15
 Number of cases listed: 15

☞ a Dependent Variable: SALARY annual salary in dollars

Scatterplot

Dependent Variable: annual salary in dollars



Casewise Diagnostics(a)

Case Number	Std. Residual	SALARY annual salary in dollars	Stud. Residual	Stud. Deleted Residual	Centered Leverage Value	Mahal. Distance	Cook's Distance	Std. DFFIT	Std. DFBETA		
									(Constant)	PUBS number of publications	TIME years since PhD
1	1.063	51876.0	1.135	1.150	.056	.785	.060	.430	.332	.005	-.287
2	.158	54511.0	.174	.167	.113	1.578	.002	.078	.073	-.057	-.014
3	.855	53425.0	.967	.964	.152	2.132	.087	.511	.504	-.339	-.195
4	.853	61863.0	.885	.877	.006	.080	.020	.245	.172	-.043	-.046
5	-.696	52926.0	-.733	-.718	.030	.418	.019	-.235	-.183	.129	-.005
6	-.698	47034.0	-.756	-.741	.080	1.116	.033	-.307	-.287	.201	.068
7	-.338	66432.0	-.382	-.368	.150	2.097	.013	-.193	.088	-.131	-.069
8	.963	61100.0	1.211	1.238	.302	4.221	.285	.945	-.342	.855	-.131
9	1.034	120000	2.447	3.310	.755	10.565	9.176	7.098	-1.273	-2.779	6.599
10	.041	47454.0	.043	.041	.032	.453	.000	.014	.008	.003	-.008
11	.537	49832.0	.581	.565	.079	1.105	.019	.233	.055	.139	-.124
12	-.321	47047.0	-.335	-.323	.019	.264	.004	-.099	-.060	-.015	.046
13	-1.870	39115.0	-1.979	-2.309	.041	.571	.157	-.801	-.713	.436	.152
14	-.026	59677.0	-.027	-.026	.019	.269	.000	-.008	.000	-.004	-.001
15	-1.554	61458.0	-1.776	-1.980	.168	2.345	.321	-1.095	.536	-.640	-.548

†
a Dependent Variable: SALARY annual salary in dollars