

Regression of Log10 Dry Weight (Z) on Age (X)

Root MSE

0.02807

R-Square

0.9983

PARAMETER ESTIMATES

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-2.68925	0.03064	-87.78	<.0001
X	1	0.19589	0.00268	73.18	<.0001

OUTPUT STATISTICS

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict	
1	-1.5380	-1.5139	0.0158	-1.5497	-1.4781	-1.5868	-1.4410
2	-1.2840	-1.3180	0.0136	-1.3489	-1.2871	-1.3886	-1.2474
3	-1.1020	-1.1221	0.0117	-1.1485	-1.0957	-1.1909	-1.0534
4	-0.9030	-0.9262	0.0100	-0.9489	-0.9036	-0.9937	-0.8588
5	-0.7420	-0.7303	0.008878	-0.7504	-0.7103	-0.7970	-0.6637
6	-0.5830	-0.5345	0.008465	-0.5536	-0.5153	-0.6008	-0.4681
7	-0.3720	-0.3386	0.008878	-0.3586	-0.3185	-0.4052	-0.2720
8	-0.1320	-0.1427	0.0100	-0.1653	-0.1200	-0.2101	-0.0752
9	0.0530	0.0532	0.0117	0.0268	0.0796	-0.0156	0.1220
10	0.2750	0.2491	0.0136	0.2182	0.2800	0.1785	0.3197
11	0.4490	0.4450	0.0158	0.4092	0.4808	0.3721	0.5179

2. The following table gives the systolic blood pressure (SBP), body size (QUET),⁷ age (AGE), and smoking history (SMK = 0 if a nonsmoker, SMK = 1 if a current or previous smoker) for a hypothetical sample of 32 white males over 40 years old from the town of Angina.

Person	SBP	QUET	AGE	SMK	Person	SBP	QUET	AGE	SMK	Person	SBP	QUET	AGE	SMK
1	135	2.876	45	0	12	138	4.032	51	1	23	137	3.296	53	0
2	122	3.251	41	0	13	152	4.116	64	0	24	132	3.210	50	0
3	130	3.100	49	0	14	138	3.673	56	0	25	149	3.301	54	1
4	148	3.768	52	0	15	140	3.562	54	1	26	132	3.017	48	1
5	146	2.979	54	1	16	134	2.998	50	1	27	120	2.789	43	0
6	129	2.790	47	1	17	145	3.360	49	1	28	126	2.956	43	1
7	162	3.668	60	1	18	142	3.024	46	1	29	161	3.800	63	0
8	160	3.612	48	1	19	135	3.171	57	0	30	170	4.132	63	1
9	144	2.368	44	1	20	142	3.401	56	0	31	152	3.962	62	0
10	180	4.637	64	1	21	150	3.628	56	1	32	164	4.010	65	0
11	166	3.877	59	1	22	144	3.751	58	0					

⁷ QUET stands for "quetelet index," a measure of size defined by $QUET = 100 (\text{weight}/\text{height}^2)$.

- b. (1) Determine the least-squares estimates of the slope (β_1) and intercept (β_0) for the straight-line regression of SBP (Y) on QUET (X).
- (2) Sketch the estimated regression line on the scatter diagram involving SBP and QUET. Compare this new line with the line you drew in part (a).
- (3) Test the null hypothesis of zero slope; be sure to interpret the result.
- (4) Based on your test in part (b)(3), would you conclude that blood pressure increases as body size increases?
- (5) Find and sketch 95% prediction bands on the appropriate scatter diagram.
- (6) Using your answer to part (b)(5), find an approximate 95% prediction interval for an individual with QUET = 3.4 (the sample mean value of QUET). Interpret your answer.
- (7) Are any assumptions for straight-line regression clearly not satisfied in this example?
- c. Repeat questions (1) through (4) in part (b) for the regression of QUET on AGE.
- d. Repeat questions (1) through (4) in part (b) for the regression of SBP on AGE.
- e. (1) Determine the least-squares estimates of the slope and intercept for the straight-line regression of SBP (Y) on SMK (X).
- (2) Compare the value of $\hat{\beta}_0$ with the mean SBP for nonsmokers. Compare the values of $\hat{\beta}_0 + \hat{\beta}_1$ with the mean SBP for smokers. Explain the results of these comparisons.
- (3) Test the null hypothesis that the true slope is 0; be sure to interpret the result.
- (4) Is the test in part (e)(3) equivalent to the usual two-sample t test for the equality of two population means, assuming equal but unknown variances? Demonstrate your answer numerically.

Edited SAS Output (PROC REG) for Problem 2

Regression of SBP (Y) on QUET (Z)

[Portion of output omitted]

Root MSE

9.81160

R-Square

0.5506

PARAMETER ESTIMATES

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	70.57640	12.32187	5.73	<.0001
QUET	1	21.49167	3.54515	6.06	<.0001

OUTPUT STATISTICS

	Dependent Variable	Predicted Values	Std Error	95% CI	95% PI
1	135.0000	132.3864	2.6499	111.6306	153.1423
2	122.0000	140.4458	1.8608	120.0507	160.8409
3	130.0000	137.2006	2.1144	116.7026	157.6985

(continued)

OUTPUT STATISTICS

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Predict		Residual
4	148.0000	151.5570	2.0860	131.0712	172.0428	-3.5570
5	146.0000	134.6001	2.3858	113.9782	155.2219	11.3999
6	129.0000	130.5382	2.8873	109.6506	151.4257	-1.5382
7	162.0000	149.4078	1.9119	128.9930	169.8227	12.5922
8	160.0000	148.2043	1.8372	127.8181	168.5905	11.7957
9	144.0000	121.4687	4.1810	99.6873	143.2501	22.5313
10	180.0000	170.2333	4.5807	148.1191	192.3475	9.7667
11	166.0000	153.8996	2.3230	133.3077	174.4915	12.1004
12	138.0000	157.2308	2.7197	136.4373	178.0243	-19.2308
13	152.0000	159.0361	2.9552	138.1090	179.9632	-7.0361
14	138.0000	149.5153	1.9194	129.0975	169.9331	-11.5153
15	140.0000	147.1297	1.7866	126.7623	167.4972	-7.1297
16	134.0000	135.0084	2.3401	114.4085	155.6084	-1.0084
17	145.0000	142.7884	1.7581	122.4313	163.1455	2.2116
18	142.0000	135.5672	2.2792	114.9957	156.1387	6.4328
19	135.0000	138.7265	1.9812	118.2841	159.1689	-3.7265
20	142.0000	143.6696	1.7403	123.3189	164.0203	-1.6696
21	150.0000	148.5482	1.8567	128.1546	168.9418	1.4518
22	144.0000	151.1917	2.0531	130.7197	171.6636	-7.1917
23	137.0000	141.4129	1.8091	121.0372	161.7887	-4.4129
24	132.0000	139.5647	1.9182	119.1473	159.9820	-7.5647
25	149.0000	141.5204	1.8042	121.1465	161.8943	7.4796
26	132.0000	135.4168	2.2954	114.8378	155.9958	-3.4168
27	120.0000	130.5167	2.8901	109.6275	151.4058	-10.5167
28	126.0000	134.1058	2.4425	113.4563	154.7553	-8.1058
29	161.0000	152.2447	2.1511	131.7309	172.7586	8.7553
30	170.0000	159.3800	3.0013	138.4255	180.3344	10.6200
31	152.0000	155.7264	2.5335	135.0312	176.4216	-3.7264
32	164.0000	156.7580	2.6601	135.9967	177.5193	7.2420

Regression of QUET (Y) on AGE (X)

[Portion of output omitted]

Root MSE

0.30131

R-Square

0.6444

PARAMETER ESTIMATES

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.38645	0.41769	0.93	0.3622
AGE	1	0.05736	0.00778	7.37	<.0001

(continued)

Regression of SBP (Y) on AGE (X)

[Portion of output omitted]

Root MSE 9.24543 R-Square 0.6009

PARAMETER ESTIMATES					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	59.09163	12.81626	4.61	<.0001
AGE	1	1.60450	0.23872	6.72	<.0001

Regression of SBP (Y) on SMK (X)

[Portion of output omitted]

Root MSE 14.18082 R-Square 0.0612

PARAMETER ESTIMATES					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	140.80000	3.66147	38.45	<.0001
SMK	1	7.02353	5.02350	1.40	0.1723

3. For married couples with one or more offspring, a demographic study was conducted to determine the effect of the husband's annual income (at marriage) on the time (in months) between marriage and the birth of the first child. The following table gives the husband's annual income (INC) and the time between marriage and the birth of the first child (TIME) for a hypothetical sample of 20 couples.

INC	TIME	INC	TIME
5,775	16.20	4,608	9.70
9,800	35.00	24,210	20.00
13,795	37.20	19,625	38.20
4,120	9.00	18,000	41.25
25,015	24.40	13,000	44.00
12,200	36.75	5,400	9.20
7,400	31.75	6,440	20.00
9,340	30.00	9,000	40.20
20,170	36.00	18,180	32.00
22,400	30.80	15,385	39.20