STATS 3A03
Applied Regression Analysis with SAS

Lecture 11: Practice Problems on Simple Linear Regression

Senay Asma

3A03, McMaster University

Oct 7, 2009

PRACTICE PROBLEM 1

Researchers interested in determining if there is a relationship between death anxiety and religiosity conducted the following study.

Subjects completed a death anxiety scale (high score = high anxiety) and also completed a checklist designed to measure an individual's degree of religiosity

(belief in a particular religion, regular attendance at religious services, number of times per week they regularly pray, etc.)

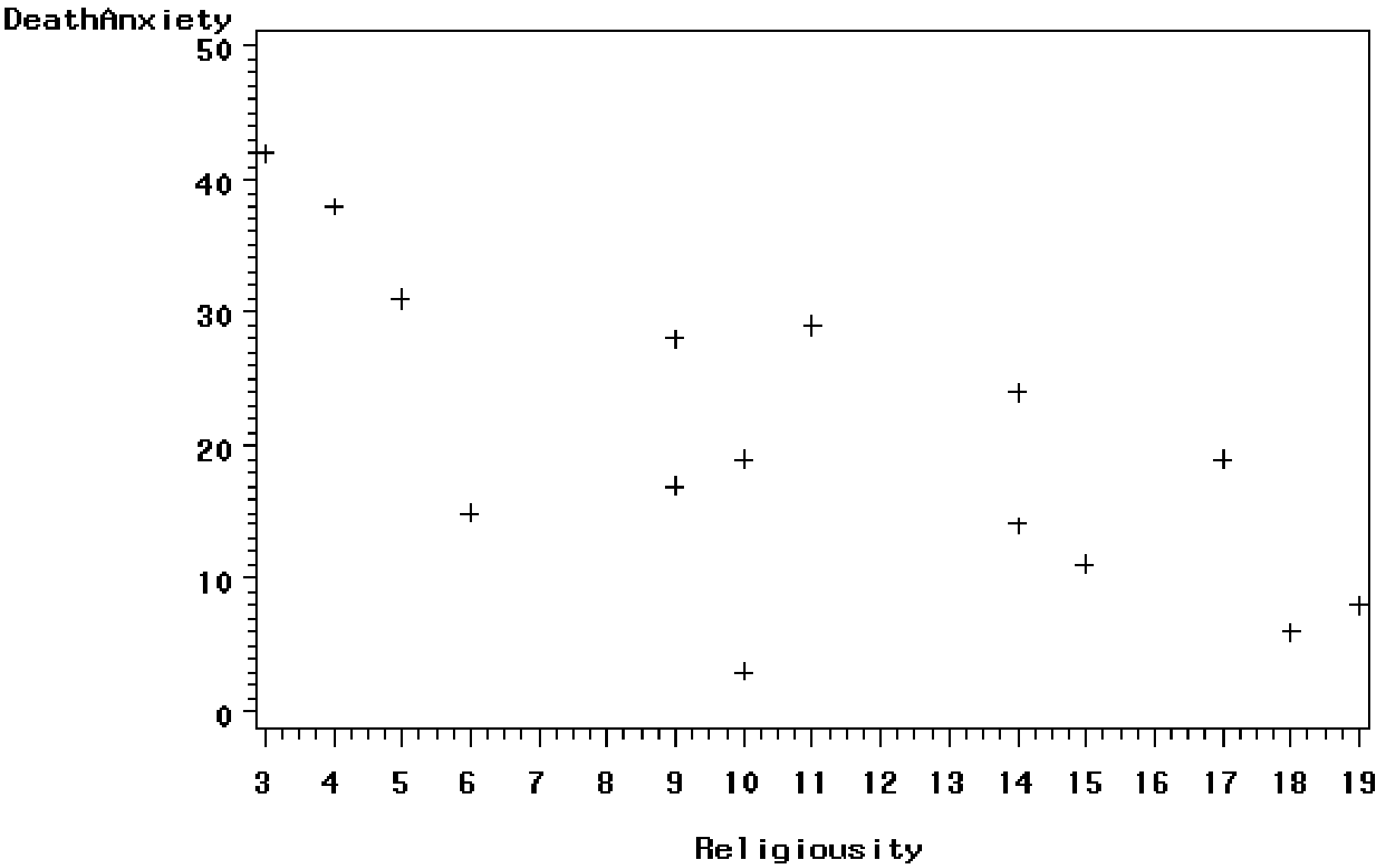
(high score = greater religiosity .

Death Anxiety	Religiosity
38	4
42	3
29	11
31	5
28	9
15	6
24	14
17	9
19	10
11	15
8	19
19	17
3	10
14	14
6	18

QUESTIONS

- What is the appropriate regression model?
- What does this statistic mean concerning the relationship between death anxiety and religiosity?
- What percent of the variability is accounted for by the relation of these two variables?

Regression Analysis
Fof
Practice Problem 1


DeathAnxiety
50
40
30
20
10
0
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Religiousity

The GLM Procedure

Dependent Variable: DeathAnxiety

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	906.489838	906.489838	12.22	0.0039
Error	13	964.443496	74.187961		
Corrected Total	14	1870.933333			

R-Square	Coeff Var	Root MSE	DeathAnxiety Mean
0.484512	42.49956	8.613243	20.26667

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	37.45130814	5.39578183	6.94	<.0001
Religiousity	-1.57176599	0.44964849	-3.50	0.0039

The GLM Procedure

Dependent Variable: DeathAnxiety

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	906.489838	906.489838	12.22	0.0039
Error	13	964.443496	74.187961		
Corrected Total	14	1870.933333			

Ho: E[Y|X]=beta0
H1: E[Y|X]=beta0+beta1*X

$\hat{\sigma}=8.6132$

R-Square	Coeff Var	Root MSE	DeathAnxiety Mean
0.484512	42.49956	8.613243	20.26667

48% of variability of Death Anxiety is explained by Religiousity

Corr(Death Anxiety,Religiousity)= -0.69

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	37.45130814	5.39578183	6.94	<.0001
Religiousity	-1.57176599	0.44964849	-3.50	0.0039

Ho: beta0=0
H1: beta0≠0

Ho: beta1=0
H1: beta1≠0

E(DeathAnxiety|Religiousity)=37.4513-1.5717*Religiousity

Q1: What is the appropriate regression model?

Answer:

Death Anxiety=37.45130814-1.57176599* Religiousity

Q2: What does this statistic mean concerning the relationship between death anxiety and religiosity?

Answer:

- This means 1 unit increasement in religiousity make the death anxiety to be decrease by -1.57 unit.
- P-value for F statistics is 0.0039 and that is smaller than significance level 0.05 so there is an evidence to accept simple linear regression model given as

  E(Death Anxiety| Religiousity)=37.45130814-1.57176599 Religiousity*

  is statistically significant with 95% confidence level. Also we can see from the SAS output that p-value for t statistics is 0.0039 and that is also smaller than significance level 0.05 (t.calc= - 3.50 < t.critical(0.025,13)= -2.160369) that means the hypothesis that claims beta 1 is not equal to 0 is accepted with 95% confidence level and there is an evidence to say Death Anxiety depend on Religiousity.
- R-square=0.484512 so Correlation(X,Y)= - 0.696069. This means there is negative relation between Death Anxiety and Religiousity (look to scatterplot)

Q 3: What percent of the variability is accounted for by the relation of these two variables?

Answer:

48% of the variability is accounted.

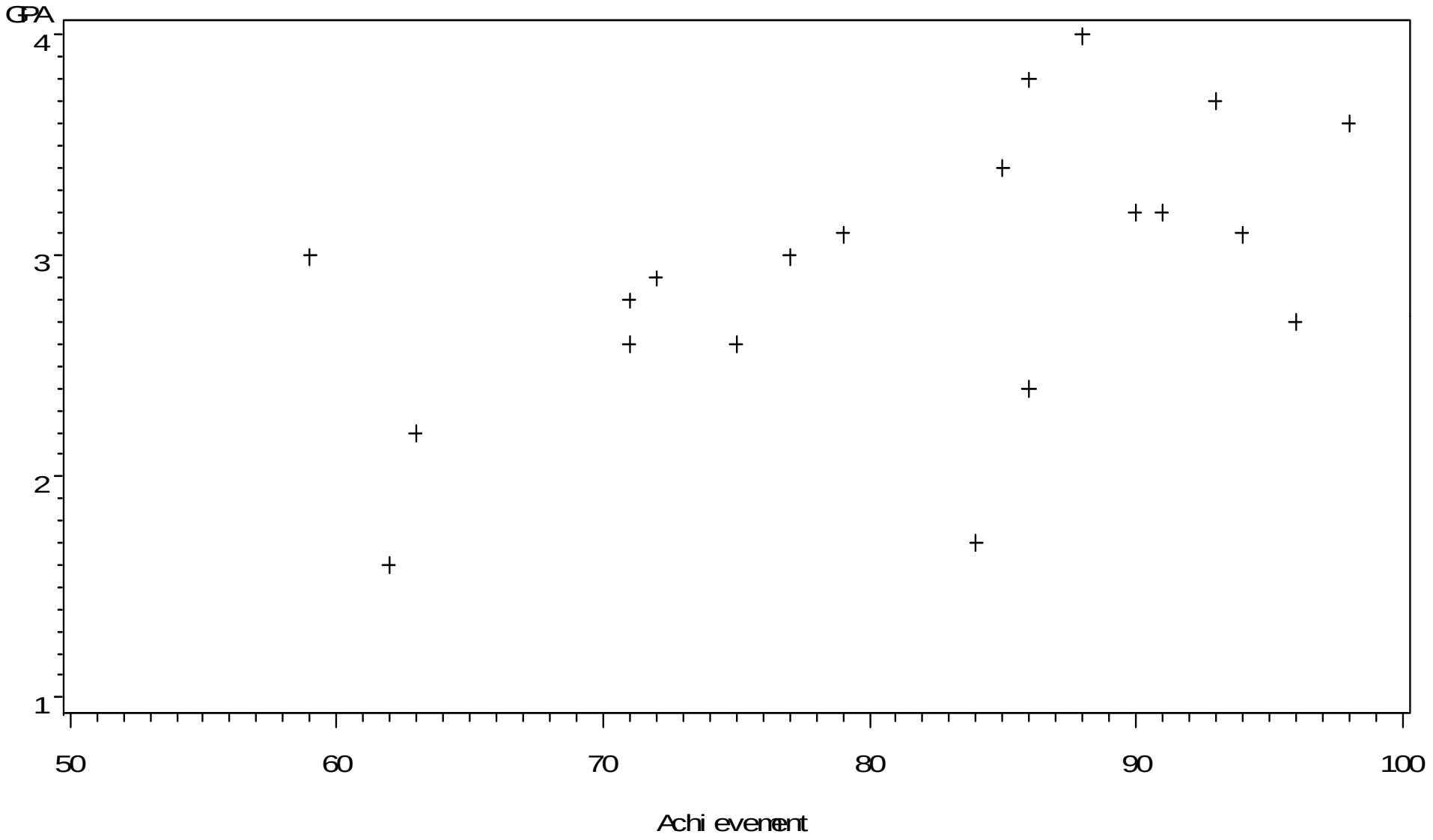
PRACTICE PROBLEM 2

It is assumed that achievement test scores should be correlated with student's classroom performance. One would expect that students who consistently perform well in the classroom (tests, quizes, etc.) would also perform well on a standardized achievement test (0 - 100 with 100 indicating high achievement).

A teacher decides to examine this hypothesis. At the end of the academic year, she computes a correlation between the student’s achievement test scores (she purposefully did not look at this data until after she submitted students grades) and the overall g.p.a. for each student computed over the entire year. The data for her class are provided below.

Achievement	G.P.A.
98	3.6
96	2.7
94	3.1
88	4.0
91	3.2
77	3.0
86	3.8
71	2.6
59	3.0
63	2.2
84	1.7
79	3.1
75	2.6
72	2.9
86	2.4
85	3.4
71	2.8
93	3.7
90	3.2
62	1.6

Regression Analysis for Practice Problem 2


GPA
4
3
2
1
50
60
70
80
90
100
Achi evenent

The GLM Procedure

Dependent Variable: GPA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.13229719	2.13229719	6.82	0.0177
Error	18	5.62970281	0.31276127		
Corrected Total	19	7.76200000			

R-Square	Coeff Var	Root MSE	GPA Mean
0.274710	19.08705	0.559251	2.930000

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.6271190296	0.89079189	0.70	0.4904
Achievement	0.0284306293	0.01088853	2.61	0.0177

The GLM Procedure

Dependent Variable: GPA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.13229719	2.13229719	6.82	0.0177
Error	18	5.62970281	0.31276127		
Corrected Total	19	7.76200000			

R-Square	Coeff Var	Root MSE	GPA Mean
0.274710	19.08705	0.559251	2.930000

$R^2 = 0.27471;$

$r_{XY} = \sqrt{0.27471} = 0.524128$

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.6271190296	0.89079189	0.70	0.4904
Achievement	0.0284306293	0.01088853	2.61	0.0177

Achievement(X)	G.P.A.(Y)	YY	XY	XX
98	3.6	0.4489	11.39	289
96	2.7	0.0529	-3.45	225
94	3.1	0.0289	2.21	169
88	4	1.1449	7.49	49
91	3.2	0.0729	2.7	100
77	3	0.0049	-0.28	16
86	3.8	0.7569	4.35	25
71	2.6	0.1089	3.3	100
59	3	0.0049	-1.54	484
63	2.2	0.5329	13.14	324
84	1.7	1.5129	-3.69	9
79	3.1	0.0289	-0.34	4
75	2.6	0.1089	1.98	36
72	2.9	0.0009	0.27	81
86	2.4	0.2809	-2.65	25
85	3.4	0.2209	1.88	16
71	2.8	0.0169	1.3	100
93	3.7	0.5929	9.24	144
90	3.2	0.0729	2.43	81
62	1.6	1.7689	25.27	361

$$\bar{x} = 2.93 \qquad \bar{y} = 81$$

$$SXX = 2638; SXY = 75; SYY = 7.762;$$

$$R^2 = \frac{(SXY)\text{^}2}{(SXX)(SYY)} = \frac{75^2}{2638 * 7.762} = 0.27471$$

$$r_{XY} = \sqrt{R^2} = \sqrt{0.27471} = 0.524128$$

Q 1: Compute the correlation coefficient.

Answer:

$$r_{XY} = \sqrt{R^2} = \sqrt{0.27471} = 0.524128$$

Q 2: What does this statistic mean concerning the relationship between a chievement test performance and g.p.a.?

Answer:

There is positive relation between achievement test performance and g.p.a.
So as the g.p.a increases the achievement test performance will increase or vice versa.

Q 3: What percent of the variability is accounted for by the relationship between the two variables and what does this statistic mean?

Answer:

%27 of the variability is explained by GPA. That means there can be different predictors that explain the variability better than GPA.

Q 4: What would be the slope and y-intercept for a regression line based on this data?

Answer:

The GLM Procedure

Dependent Variable: GPA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.13229719	2.13229719	6.82	0.0177
Error	18	5.62970281	0.31276127		
Corrected Total	19	7.76200000			

R-Square	Coeff Var	Root MSE	GPA Mean
0.274710	19.08705	0.559251	2.930000

beta_0.hat

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.6271190296	0.89079189	0.70	0.4904
Achievement	0.0284306293	0.01088853	2.61	0.0177

beta_1.hat

Q 5: If a student scored a 93 on the achievement test, what would be their predicted G.P.A.? If they scored a 74? A 88?

- E[GPA|Achievement test=93]=0.6271+0.02843*93
- E[[GPA|Achievement test=**93**]=**3.27109** (fitted value)
- E[GPA|Achievement test=74]=0.6271+0.02843*74
- E[GPA|Achievement test=**74**]= **2.73092** (predicted value)
- E[GPA|Achievement test=88]=0.6271+0.02843*88
- E[GPA|Achievement test=**88**]= **3.12894** (fitted value)

PRACTICE PROBLEM 3

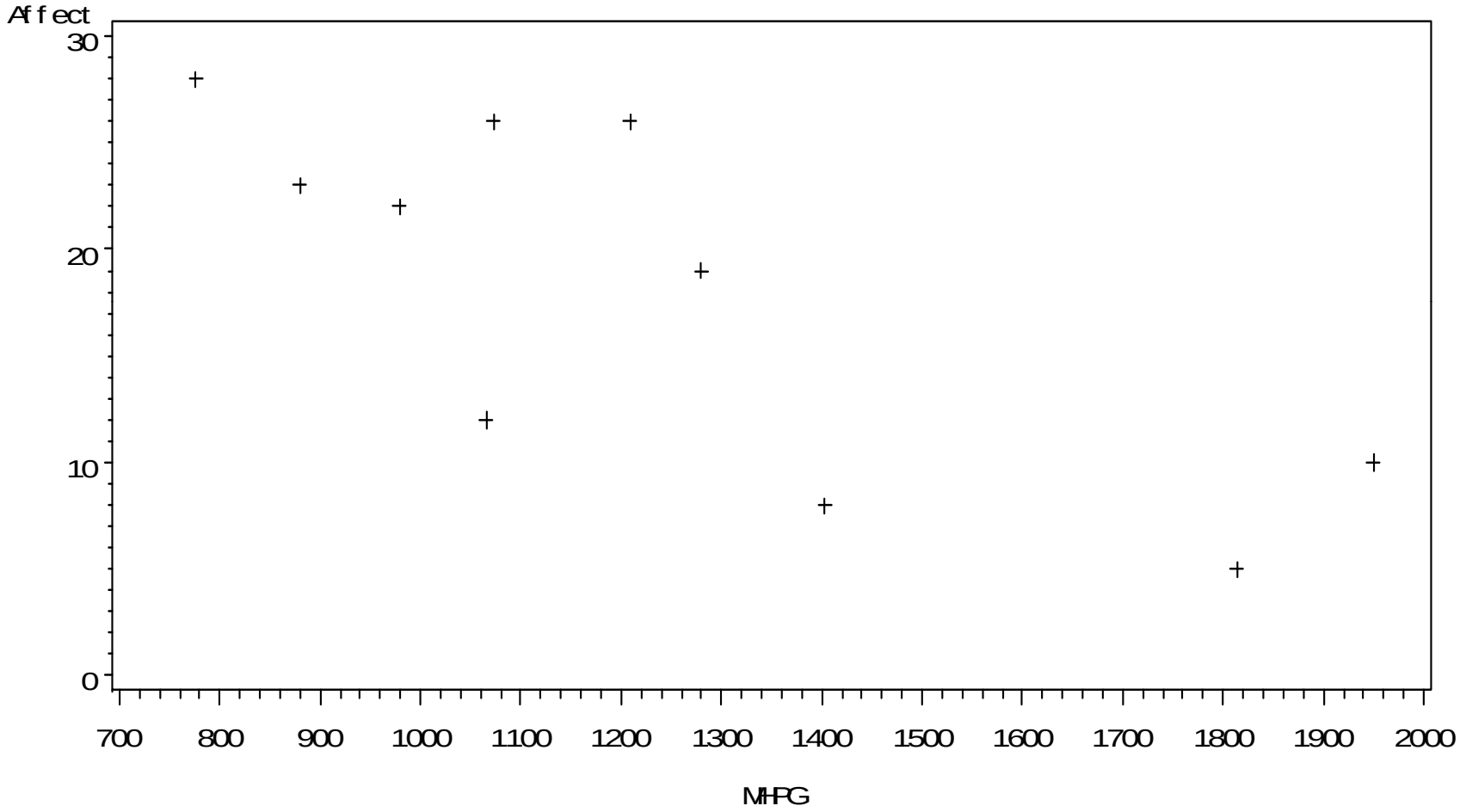
It is hypothesized that there are fluctuations in norepinephrine (NE) levels which accompany fluctuations in affect with bipolar affective disorder (manic-depressive illness). Thus, during depressive states, NE levels drop; during manic states, NE levels increase. To test this relationship, researchers measured the level of NE by measuring the metabolite 3-methoxy-4-hydroxyphenylglycol (MHPG in micro gram per 24 hour) in the patient's urine experiencing varying levels of mania/depression. Increased levels of MHPG are correlated with increased metabolism (thus higher levels) of central nervous system NE. Levels of mania/depression were also recorded on a scale with a low score indicating increased mania and a high score increased depression. The data is provided below.

MHPG	Affect
980	22
1209	26
1403	8
1950	10
1814	5
1280	19
1073	26
1066	12
880	23
776	28

QUESTIONS

- Compute the correlation coefficient.
- What does this statistic mean concerning the relationship between MHPG levels and affect?
- What percent of the variability is accounted for by the relationship between the two variables?
- What would be the slope and y-intercept for a regression line based on this data?
- What would be the predicted affect score if the individual had an MHPG level of 1100? of 950? of 700?

Regression Analysis
for
Practice Problem 3


Affect
30
20
10
0
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
MHPG

The GLM Procedure

Dependent Variable: Affect

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	388.1525136	388.1525136	12.38	0.0079
Error	8	250.7474864	31.3434358		
Corrected Total	9	638.9000000			

R-Square	Coeff Var	Root MSE	Affect Mean
0.607532	31.27665	5.598521	17.90000

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	39.11036137	6.28189836	6.23	0.0003
MHPG	-0.01706247	0.00484857	-3.52	0.0079

Dependent Variable: Affect

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	388.1525136	388.1525136	12.38	0.0079
Error	8	250.7474864	31.3434358		
Corrected Total	9	638.9000000			

R-Square	Coeff Var	Root MSE	Affect Mean
0.607532	31.27665	5.598521	17.90000

beta_0.hat

beta_1.hat

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	39.11036137	6.28189836	6.23	0.0003
MHPG	-0.01706247	0.00484857	-3.52	0.0079

Q 1: Compute the correlation coefficient.

Answer:

$$r_{XY} = \sqrt{R^2} = \sqrt{0.607532} = -0.7794434$$

Q 2: What does this statistic mean concerning the relationship between MHPG levels and affect?

Answer:

The statistics in Q1 for this problem

$$r_{XY} = Corr(Affect, MHPG\ level) = -0.7794434$$

means there is negative relationship between MHPG levels and affect.
But this relationship is not so strong.

Q 3: What percent of the variability is accounted for by the relationship between the two variables?

Answer:

61% of the variability
is explained by MHPG level.

Q 4: What would be the slope and y-intercept for a regression line based on this data?

Answer:

E[Affect| MHPG]= 39.11036137 - 0.01706247* MHPG

An estimated slope of - 0.01706247 means that for every increase of 1 unit of MHPG level, affect decrease by - 0.017.

An estimated y-intercept of 39.11 means that the estimated affect for MHPG level =0 which is meaningless measurement.

Q 5: What would be the predicted affect score if the individual had an MHPG level of 1100? of 950? of 700?

Answer:

- E[Affect| MHPG=1100]= 39.11036137 - 0.01706247* (1100)
- E[Affect| MHPG=**1100**]= **20.34164**
- E[Affect| MHPG=950]= 39.11036137 - 0.01706247* (950)
- E[Affect| MHPG=**950**]= **22.90101**
- E[Affect| MHPG=700]= 39.11036137 - 0.01706247* (700)
- E[Affect| MHPG=**700**]= **27.16663**

GOOD LUCK

FOR

YOUR MIDTERM EXAM