PubH 7405: BIOSTATISTICS REGRESSION, 2011

PRACTICE PROBLEMS FOR SIMPLE LINEAR REGRESSION

(Some are new & Some from Old exams; last 4 are from 2010 Midterm)

Problem 1:

The Pearson Correlation Coefficient (r) between two variables X and Y can be expressed in several equivalent forms; one of which is

$$r(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Where x-bar (y-bar) is the sample mean and s_x (s_y) the sample standard deviation of X (Y).

(1) If a and c are two positive constants and b and d are any two constants, prove that:
$r(aX+b, cY+d) = r(X,Y)$

(2) Is the result in (1) still true if we do not assume that a and c are positive?

(3) For a group of men, if the Correlation Coefficient between Weight in pounds and Height in inches is r=.29; what is the value of that Correlation Coefficient if Weight is measured in kilograms and Height in centimeters? Explain your answer.

(4) Body Temperature (BT) can be measured at many locations in your body. Suppose, for certain group of children with fever, the Correlation Coefficient between oral BT and rectal BT is r=.91 when BT is measured in Fahrenheit scale (^{0}F); what is the value of that Correlation Coefficient if BT is measured in Celsius scale (^{0}C)? Explain your answer.

Problem 2:

Let X and Y be two variables in a study; the regression line that can be used to predict Y from X values is:

$$\text{Predicted } y = b_0 + b_1 x$$

The estimated intercept and slope can be expressed in several equivalent forms; one of which is

$$b_1 = r\frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

Where x-bar is sample mean and s_x is the sample standard deviation of X.

(1) If a and c are two positive constants and b and d are any two constants, consider the data transformation:

$U = aX + b$

$V = cY + d$

And let denote the estimated intercept and slope of the regression line predicting V from U as B_0 and B_1. Express B_0 and B_1 as function of a, b, c, d, and b_0 and b_1

(2) What would be the results of (1) in the special case that a=c and b=d=0? What would be the results of (1) in the special case that a=1 and b=d=0?

(3) During some operations, it would be more convenient to measure Blood Pressure (BP) from the patient's leg than from a cuff on the arm. Let X = leg BP and Y = arm BP, the results for a group under going orthopedic surgeries are b_0=9.052 and b_1=0.761 when BP is measure in millimeters of mercury (Hg); what would be these results if BP is measured in centimeters of Hg? Explain your answer.

(4) Apgar score was devised in 1952 by Dr. Virginia Apgar as a simple method to quickly assess the health of the newborn. Let X = Apgar score and Y= Birth Weight, the results for a group of newborns are b_0=1.306 and b_1=0.205 when Birth Weight is measured in kilograms; what would be these results if Birth Weight is measured in pounds? Explain your answer.

Problem 3:

Let X and Y be two variables in a study.

(1) Investigator #1 is interested in predicting Y from X, and fits and computes a regression line for this purpose. Investigator #2 is interested in predicting X from Y, and computes his regression line for that purpose (note that in the real problem of "parallel-line bioassays, with X=log(dose) and Y=response, we have both of these steps – the first for the Standard Preparation and the second for the Test preparation). Are these two regression lines the same? If so, shy? If not, compute the ratio and the product of the two slopes as function of standard statistics.

(2) Let X = Height and Y = Weight, we have for a group of 409 men:

$$\sum x = 28{,}359 \text{ inches}$$

$$\sum y = 64{,}938 \text{ pounds}$$

$$\sum x^2 = 1{,}969{,}716 \text{ inches}^2$$

$$\sum y^2 = 10{,}517{,}079 \text{ pounds}^2$$

$$\sum xy = 4{,}513{,}810 \text{ (inch)(pound)s}$$

(a) Calculate the Coefficient of Correlation
(b) Calculate the Slopes, the product, and the ratio of slopes in question (1)
(c) Calculate the Intercept for Investigator #2
(d) Calculate 95 percent Confidence Interval for the Slope for Investigator #1

Problem 4:

Let X and Y be two variables in a study; the regression line that can be used to predict Y from X values is:

$$\text{Predicted } y = b_0 + b_1 x$$

So that the "error" of the prediction is:

$$\text{Error} = \text{y - Predicted } y$$

$$e = y - (b_0 + b_1 x)$$

(1) From the Sum of Squared Errors:

$$S = \sum e^2$$

$$S = [y - (b_0 + b_1 x)]^2$$

Derive the two "normal equations

(2) Use the two normal equations in (1) to prove that (2.1) the average error is zero, and (2.2) the errors of prediction and the values of the Predictor are uncorrelated (the coefficient of correlation is zero, r(e,X)=0).

(3) Recall that if a, b, c, and are constants then r(aX+b,cY+d) = r(X,Y); use this and the result in (2.2) to show that the errors of prediction and the predicted values of the Response are uncorrelated (the coefficient of correlation is zero, r(Predicted y,e)=0).

(4) Prove that Var(y) = Var(Predicted y) + Var(e)

(5) (BONUS) From the result of (4), prove that
Var(e) = $(1-r^2)$Var(e); hence, $-1 \leq r \leq 1$

Problem 5:

From a sample of n=15 readings on X = Traffic Volume (cars per hour) and Y = Carbon Monoxide Concentration (PPM) taken at certain metropolitan air quality sampling site, we have these statistics:

$\sum x = 3{,}550$

$\sum y = 167.8$

$\sum x^2 = 974{,}450$

$\sum y^2 = 1{,}915.36$

$\sum xy = 41{,}945$

(1) Compute the sample Correlation Coefficient r.

(2) Test for H_0: $\rho = 0$ at the .05 level of significance and state your conclusion in context of this problem (ρ is the Population Coefficient of Correlation).

(3) Determine either the exact p-value for the test or its upper bound

(4) Construction the 95 percent Confidence Interval for ρ via Fisher's transformation.

Problem 6:

Consider the regression line/model without intercept,

Predicted y = bx

(1) Minimize $S = \sum(y-bx)^2$ to verify that the estimated slope of the regression line for predicting Y from X is given by $b_1 = \sum xy/\sum x^2$.

(2) Consider another alternative estimate of the slope, the ratio of the sample means, $b_2 = \sum y/\sum x$. Show that if Var(Y) is constant then $Var(b_1) \leq Var(b_2)$. (However, if the variance Var(Y) is proportional to x, $Var(b_2) \leq Var(b_1)$; an example of this situation would occur in a radioactivity counting experiment where the same material is observed for replicate periods of different lengths; counts are distributed as Poisson).

Problem 7:

The data below show the consumption of alcohol (X, liters per year per person, 14 years or older) and the death rate from cirrhosis, a liver disease (Y, death per 100,000 population) in 15 countries (each country is an observation unit).

Country	Alc. Consumption	Death Rate from Cirrhosis	x^2	y^2	xy
France	24.7	46.1	610.09	2125.21	1138.67
Italy	15.2	23.6	231.04	556.96	358.72
Germany	12.3	23.7	151.29	561.69	291.51
Australia	10.9	7	118.81	49	76.3
Belgium	10.8	12.3	116.64	151.29	132.84
USA	9.9	14.2	98.01	201.64	140.58
Canada	8.3	7.4	68.89	54.76	61.42
England	7.2	3.0	51.84	9	21.6
Sweden	6.6	7.2	43.56	51.84	47.52
Japan	5.8	10.6	33.64	112.36	61.48
Netherland	5.7	3.7	32.49	13.69	21.09
Ireland	5.6	3.4	31.36	11.56	19.04
Norway	4.2	4.3	17.64	18.49	18.06
Finland	3.9	3.6	15.21	12.96	14.04
Ireal	3.1	5.4	9.61	29.16	16.74
Total	134.2	175.5	1630.12	3959.61	2419.61

(1) Draw a Scatter Diagram to show the association, if any, between these two variables; can you draw any conclusion/observation without doing any calculation?

(2) Calculate the Coefficient of Correlation and its 95% Confidence Interval using the Fisher's transformation; then state your interpretation.
(3) Form the regression line by calculating the estimate Intercept and Slope; if the model holds, what would be the death rate from Cirrhosis for a country with alcohol consumption rate of 11.0 liters per year per person?
(4) What fraction of the total variability of Y is explained by its relationship to X? Form the ANOVA Table.
(5) Test for H_0: Slope = 0 at the .05 level of significance and state your conclusion in term of this problem description

Problem 8:

When a patient is diagnosed as having cancer of the prostate, an important question in deciding on treatment strategy for the patient is whether or not the cancer has spread to the neighboring lymph nodes. The question is so critical in prognosis and treatment that it is customary to operate on the patient (i.e., perform a laparotomy) for the sole purpose of examining the nodes and removing tissue samples to examine under the microscope for evidence of cancer. However, certain variables that can be measured without surgery are predictive of the nodal involvement; and the purpose of the study presented here was to examine the data for 53 prostate cancer patients receiving surgery, to determine which of five preoperative variables are predictive of nodal involvement. For each of the 53 patients, there are information on patients' age and four other potential independent variables, the level of serum acid phosphatase (the factor of primary interest), and three binary variables, X-ray reading, pathology reading (grade) of a biopsy of the tumor obtained by needle before surgery, and a rough measure of the size and location of the tumor (stage) obtained by palpation with the fingers via the rectum. The primary outcome of interest, or dependent variable, represents the finding at surgery which is binary indicating nodal involvement or no nodal involvement found at surgery.

The analysis, with some results included here, is not about the main objective of predicting nodal involvement; it's a side analysis focusing on a possible confounder , age. The objective here is to see if the level of serum acid phosphatase and the patient's age are related.

Computer Program (SAS):

```
options ls=79;
data Pcancer;
input Xray Stage Grade Age Acid Nodes;
cards;
0 0 0 66  48 0
0 0 0 68  56 0
…..
1 1 0 64  89 1
1 1 1 68 126 1
;
Proc UNIVARIATE data=Pcancer;
    Var Age Acid;
run;
Proc CORR data=Pcancer;
run;
Proc REG data=Pcancer;
    model Acid = Age/COVB CLM;
    plot r.*Age="+" r.*p.="*";
run;
```

Computer Output/results

PART A: **Univariate Procedure**

Variable=AGE

Moments

N	53	Sum Wgts	53
Mean	59.37736	Sum	3147
Std Dev	6.168239	Variance	38.04717
Skewness	-0.49481	Kurtosis	-0.69677

Quantiles

100% Max	68	99%	68
75% Q3	65	95%	67
50% Med	60	90%	67
25% Q1	56	10%	51
0% Min	45	5%	49

Variable=ACID

Moments

N	53	Sum Wgts	53
Mean	69.41509	Sum	3679
Std Dev	26.20146	Variance	686.5167
Skewness	2.251881	Kurtosis	7.29481

Quantiles

100% Max	187	99%	187
75% Q3	78	95%	126
50% Med	65	90%	98
25% Q1	50	10%	48
0% Min	40	5%	46

PART B: **Correlation Analysis**

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 53

	XRAY	STAGE	GRADE	AGE	ACID	NODES
XRAY	1.00000	0.19761	0.20217	-0.00453	0.14973	0.46140
	0.0	0.1561	0.1466	0.9743	0.2846	0.0005
STAGE	0.19761	1.00000	0.37463	-0.01970	-0.02939	0.37463
	0.1561	0.0	0.0057	0.8887	0.8345	0.0057
GRADE	0.20217	0.37463	1.00000	-0.04808	-0.08294	0.27727
	0.1466	0.0057	0.0	0.7324	0.5549	0.0444
AGE	-0.00453	-0.01970	-0.04808	1.00000	0.05399	-0.14365
	0.9743	0.8887	0.7324	0.0	0.7010	0.3048
ACID	0.14973	-0.02939	-0.08294	0.05399	1.00000	0.24252
	0.2846	0.8345	0.5549	0.7010	0.0	0.0802
NODES	0.46140	0.37463	0.27727	-0.14365	0.24252	1.00000
	0.0005	0.0057	0.0444	0.3048	0.0802	0.0

PART C: **Regression Analysis, Dependent Variable: ACID**

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model		104.04189			0.7010
Error		35594.82603			
C Total					

Root MSE — R-square 0.0029
Dep Mean 69.41509

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	55.798699	35.45303425	1.574	0.1217
AGE	1	0.229320	0.59394400	0.386	0.7010

Covariance of Estimates

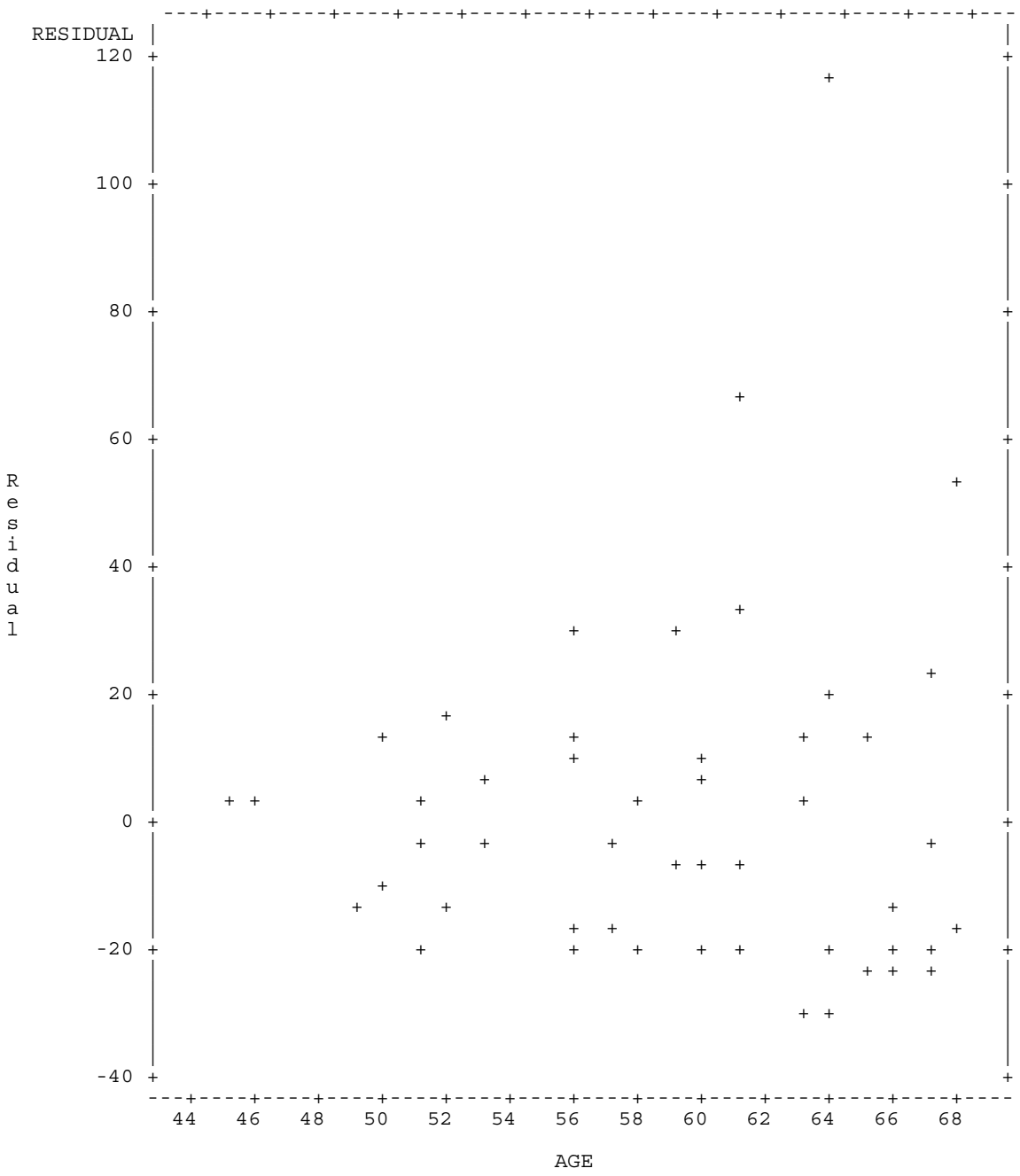
COVB	INTERCEP	AGE
INTERCEP	1256.9176378	-20.94651955
AGE	-20.94651955	0.3527694745

Obs	Dep Var ACID	Predict Value	Std Err Predict	Lower95% Mean	Upper95% Mean	Residual
1	48.0000	70.9338		60.1898	81.6778	-22.9338
2	56.0000	71.3924	6.277	58.7914	83.9934	-15.3924
3	50.0000	70.9338	5.352	60.1898	81.6778	-20.9338
4	52.0000	68.6406	4.146	60.3164	76.9648	-16.6406
5	50.0000	69.0992	3.720	61.6312	76.5673	-19.0992
6	49.0000	69.5579	3.648	62.2349	76.8809	-20.5579
7	46.0000	70.7045	4.932	60.8038	80.6052	-24.7045
8	62.0000		3.64	62.2349	76.8809	
9	56.0000	67.2647	6.648	53.9193	80.6101	-11.2647
10	55.0000	67.0354	7.152	52.6761	81.3946	-12.0354
.....		etc.....				
44	76.0000	69.5579	3.648	62.2349	76.8809	6.4421
45	70.0000	66.1181	9.278	47.4909	84.7453	3.8819
46	78.0000	68.6406	4.146	60.3164	76.9648	9.3594
47	70.0000	66.3474	8.735	48.8114	83.8834	3.6526
48	67.0000	71.1631	5.802	59.5146	82.8116	-4.1631
49	82.0000	70.2458	4.219	61.7763	78.7154	11.7542
50	67.0000	68.8699	3.894	61.0526	76.6872	-1.8699
51	72.0000	67.4940	6.158	55.1305	79.8575	4.5060
52	89.0000	70.4752	4.550	61.3397	79.6106	18.5248
53	126.0000	71.3924	6.277	58.7914	83.9934	54.6076

Sum of Residuals 0
Sum of Squared Residuals 35594.8260

PART D: Graphs

Graph #1



(1) From the problem's description in section 1.1, if the main objective is to predict Nodal Involvement from Age, Serum acid phosphatase, X-ray, Grade, and Stage, then why we may be interested to see if Serum acid phosphatase and Age are related?

(2) What does the SAS computer program in section 1.2 suppose to give you?

(3) Using only the computer results/output in PART A (Univariate Procedure) can we calculate the coefficient of correlation between Age and Serum acid phosphatase? Why or why not? If not, what else do you need? Can you get what you want (using your calculator if needed) to obtain the needed piece of information from all data given here?

(4) From the computer results/output in PART B, do you think it is reasonable to conclude that Serum acid phosphatase and the patient's Age are not related?

(5) What is the model we usually assume in performing regression analysis in PART C? is it true that the assumptions of the model are only about the distribution of the level of Serum acid phosphatase? Do we make any assumptions about the distribution of age?

(6) Fill in the blanks to complete the ANOVA Table given in PART C (Degrees of freedom, Means squares, F statistic, Root MSE); can you get F statistic without MSE?

(7) Fixing a value of Age, the values of Serum acid phosphatase from the sub-population of patients with that age form a distribution with variance σ^2, use the results you just filled in (in question 1.6) to provide a point estimate for this variance.

(8) Re-calculate that point estimate of using only the results in the last 2 rows preceding the graph#1; does this agree with the previous estimate?

(9) If you are given only the results in PART C – but not the results in PART B – can you find the value of the coefficient of correlation? Does this agree with the result given in PART B?

(10) Does the F-test result in PART C agree with the result of the corresponding t-test in PART B, why or why not?

(11) Fill in the two (2) blanks (Predicted value and Residual) for observation/patient #8 in the long Table preceding the graph.

(12) Fill in the blank (Standard error of the Predicted value) for observation/patient #1.

(13) What is the average/mean Serum acid phosphatase for patients of 40 years of age?

(14) If we treat the "Predicted value" in that Table as an estimate of a new observation, can we calculate its standard error? Why or why not? If not, what else do you need?

(15) What does the graph tell you about the model's assumption(s) in question (5)

Problem 9:

Let FEV (a measure of Lung Health) be the dependent variable and Age is a potential predictor; and we have the following results (computer output) using data n subjects. Suppose we fit the (Model FEV = Age) and obtain these two tables:

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	**[A]**	**[B]**	**[E]**	0.09101
Residual	**[C]**	16.685	**[D]**		
Total	24	18.942			

	Coefficients	*Std Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.253	0.921	5.703	<0.00001	3.34776	7.15847
Age	-0.040	0.023	-1.764	0.09101	-0.08708	0.00692

(1) *Calculate* the quantities A, B, C, D, and E (in order to complete the above ANOVA table); what was the *sample size*?

(2) Is it reasonable to conclude that FEV and the subject's Age are not related?

(3) What is the model we usually assume in performing the above regression analysis? Do we make any assumptions about the distribution of age?

(4) Fixing a value of Age, the values of FEV from the sub-population of subjects with that age form a distribution with variance σ^2, provide a point estimate for this variance.

(5) Suppose you are also given, as part of the computer output for regression analysis, $R^2 = .119$. Give your interpretation of this number and use it (and any results from the above computer output) to calculate the Coefficient of Correlation representing the strength of the relationship between Age and FEV.

(6) When we estimate the "Mean Response" of a sub-population with a common value of the predictor's value X = x_h, the variance is given by:

$$s^2(\hat{Y}_h) = MSE\left\{\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right\}$$

(7) Calculate this *variance* when x_h = 30; you *can use* all the above computer output plus the following descriptive statistics:

n=25	**Age (X)**	**FEV (Y)**
Minimum	26	1.86
Maximum	54	4.91
Mean	39.8	3.66
Variance	58.56	0.79
St Deviation	7.65	0.89

(8) Suppose you have a numerical result for # (1f), [the following number is not necessarily correct; it is given just in case you could not answer or skipped #II1f)]

$$s^2(\hat{Y}_h) = .0786$$

(9) However, we like to we treat the "Predicted/fitted value" as an estimate of a new individual observation, calculate its *standard error*.

Problem 10:

In the last decade prostate specific antigen doubling time (PSA-DT; the time required for serum PSA to double its value) has been extensively researched in the prostate cancer literature; it is a reliable predictor of many major clinical endpoints. The use of PSA-DT is based on the finding that serum PSA in patients with prostate cancer follows an exponential growth curve model:

Model 1a: $PSA(t) = PSA(t_0)\exp[\beta(t - t_0)]; \quad t > t_0$

Where t_0 is the time origin at which the exponential growth stage starts and β>0 is parameter representing disease severity.

(1) In 1992 a retrospective study of banked serum samples of patients with prostate cancer showed that the exponential increase in serum PSA begins 7 to 9 years before the tumor is detected clinically. In other words, the time of disease inception t_0 cannot be determined. Prove that if $t_0 < t_1$, we still have the same exponential model:

Model 1b: $PSA(t) = PSA(t_1)\exp[\beta(t - t_1)]; \quad t > t_1$

(That means we can use any time t_1 in the exponential growth stage as "time origin" instead of the unknown time of disease inception t_0).

(2) Show how to express Model 1b as a simple linear regression model:

Model 2: $\mathbf{Y_t = \alpha + \beta t + \varepsilon}$

(3) where Y_t = ln[PSA(t)]. What is the meaning of the intercept α? Why the slope β can be used as a parameter representing disease severity? Assume that the error term ε satisfies assumptions of the normal error regression model; what are these assumptions? Do we really need to know a specific value t_1 (in Model 1b) to perform data analysis or just any sample in the exponential growth stage? What if we include data points before time t_0?

(4) Find a possible link or relationship between the regression coefficients α and β (either one or both) and the PSA-DT.

(5) What is the link or relationship between the coefficient of correlation $r(Y_t,t)$ and coefficient of correlation $r(Y_t,t-t_0)$? Can we calculate or approximate the coefficient of correlation r[PSA(t), $t-t_0$] if we know $r(Y_t,t)$ or $r(Y_t,t-t_0)$ or both?

(6) Use the relationship found in (3), show how to calculate or approximate the standard error SE(PSA-DT) using the variance-covariance matrix for the regression coefficients (variances of estimates of α and β and their covariance; numbers are often provided by computer out put such as SAS).

(7) Given the following small data set:

Time, t	PSA	Y=ln(PSA)
100	1.2	0.182
280	1.6	0.470
640	4	1.386
1000	8.4	2.128

a) Calculate the coefficient of correlation r(Y,t), the estimated slope b, the estimated intercept a, the PSA-DT, the standard error SE(b), the test statistic for the test of independence and its degree of freedom, the standard error SE(PSA-DT)

b) Set up the ANOVA table, include up to the F ratio statistics; p-value is not required. Can we get the value of F ratio statistic without the ANOVA table?

c) Suppose we change the data by changing the time origin:

Time, t	PSA	Y=ln(PSA)
0	1.2	0.182
180	1.6	0.470
540	4	1.386
900	8.4	2.128

Do we still get the same results in (a)?

d) Suppose we change the sampling time to the followings:

Time, t	PSA	Y=ln(PSA)
100		
280		
820		
1000		

How does the new SE(PSA-DT) compare to the result in (a)? explain your answer.

Problem 11:

The following data were collected during an experiment in which 10 laboratory animals were inoculated with a pathogen. The variables are Time after inoculation (X, in minutes) and Temperature (Y, in Celsius degrees).

	X, Time Minutes)	Y, Temperature (^{0}C)	x^2	y^2	xy
	24	38.8	576.00	1505.44	931.20
	28	39.5	784.00	1560.25	1106.00
	32	40.3	1024.00	1624.09	1289.60
	36	40.7	1296.00	1656.49	1465.20
	40	41.0	1600.00	1681.00	1640.00
	44	41.1	1936.00	1689.21	1808.40
	48	41.4	2304.00	1713.96	1987.20
	52	41.6	2704.00	1730.56	2163.20
	56	41.8	3136.00	1747.24	2340.80
	60	41.9	3600.00	1755.61	2514.00
Total	**420**	408.1	18960.00	16663.85	17245.60

(6) Draw a Scatter Diagram to show the association, if any, between these two variables (correct scale is not very important); can you draw any conclusion/observation without doing any calculation?
(7) Calculate the Coefficient of Correlation and its 95% Confidence Interval.
(8) Form the regression line of Y on X by calculating the estimate Intercept and Slope; if the model holds, what would be the temperature for an animal, chosen at random, after 30 minutes? Explain, in the context of this problem, why it is risky to predict the response outside the range of values of the independent variable represented in the sample – say at 5 hours.
(9) What fraction of the total variability of Y is explained by its relationship to X? Form the ANOVA Table.
(10) Test for H_0: Slope = 0 at the .05 level of significance and state your conclusion in term of this problem description.

Problem 22:

The Pearson Correlation Coefficient (r) between two variables X and Y can be expressed in several equivalent forms; one of which is

$$r(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

(1) If a and c are two positive constants and b and d are any two constants, prove that:
$r(aX+b, cY+d) = r(X,Y)$
Is this result in (1) still true if we do not assume that a and c are positive?

(2) What is the value of that Correlation Coefficient in question 2 of Problem 11 if Temperature is measured in Fahrenheit scale (^{0}F)? Explain your answer.

Problem 13:

Let X and Y be two variables in a study; the regression line that can be used to predict Y from X values is:

$$\text{Predicted } y = b_0 + b_1 x$$

The estimated intercept and slope can be expressed in several equivalent forms; one of which is

$$b_1 = r\frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

Where x-bar is sample mean and s_x is the sample standard deviation of X.

(1) If a and c are two positive constants and b and d are any two constants, consider the data transformation:

$U = aX + b$

$V = cY + d$

And let denote the estimated intercept and slope of the regression line predicting V from U as B_0 and B_1. Express B_0 and B_1 as function of a, b, c, d, and b_0 and b_1

(2) What would be the results of (1) in the special case that a=1 and b=0? What are the values of the Intercept and Slope in question 3 of Problem 11 if Temperature is measured in Fahrenheit scale (^{0}F)? Explain your answer.

Problem 14:

For an experiment like the one in Problem 1, Investigator #1 may be interested in predicting Y from X, and fits and computes a regression line for this purpose. Investigator #2, however, may be interested in predicting X from Y, and computes his regression line for that purpose.

(1) Are these two regression lines the same? If so, why? If not, compute the ratio and the product of the two slopes as function of standard statistics.
(2) Calculate the product of those two slopes for data in Problem 11.