Q SCI 381: Introduction to Probability and Statistics
S. Scherba, Jr. Assignment #8: Review Exercises

Shorter answers and a problem (*30 points*).

1. (*1 point*) What is the objective of simple linear regression?
2. (*1 point*) In simple linear regression, we seek to estimate two parameters the equation of a straight-line function: $Y = a + bx$, where a is the intercept and b is the slope. We want to find the estimates for those parameters that provide the "best fit" to our data. What is meant by "best fit"?
3. (*3 points*) In estimating these parameters in this manner, we apply the calculus to minimize the residual sum of squares. In so doing we get a pair of equations. What are they called? What are the solutions to these two equations?
4. (*6 points*) What are the statistical assumptions explicitly assumed in simple linear regression?
5. (*1 point*) With regard to using the regression model what we develop (or fit), what warning must always be remembered?
6. (*2 points*) The usefulness of an estimated regression equation is based on two important measures. What are these two measures and what do they measure?
7. (*2 points*) It can be shown that the equation for R^2 (the coefficient of determination) is identical to that for r^2 (the square of the correlation coefficient). Why do these two measures not measure that same thing?
8. (*14 points*) A physiologist once proposed that the oxygen consumed by an animal should be proportional to its surface area, and since surface area is proportional to the weight raised to the 2/3 power, then oxygen consumption should be proportional to the 2/3 power of the animal's weight. A random sample of 10 animals from certain specie was collected. The physiologist's theory results in a power function. To test this hypothesis, each animal's oxygen consumption was measured and its weight recorded.

Let $C =$ oxygen consumption in l/hr
$W =$ weight of the animal in kg:

$C = a * (W)^{\frac{2}{3}}$, which we can **linearize** as follows:

$$
\begin{aligned}
&C = a*(W)^{\frac{2}{3}} \\
&\Leftrightarrow \ln(c) = ln(a) + \frac{2}{3}\ln(W) \\
&\Leftrightarrow \ln(c) = ln(a) + 0.66\ln(W) \\
&\Leftrightarrow Y = \alpha + \beta x
\end{aligned}
$$

Hence, for each observation we have: $Y_i = \alpha + \beta x_i$

The physiologist obtains the following data:

Animal (i)	1	2	3	4	5	6	7	8	9	10
Y (ln((l/hr))	5.25	6.04	6.08	6.57	6.52	7.25	7.40	8.25	8.30	8.50
x (ln (kg))	1.25	1.75	2.25	2.55	3.08	3.59	4.07	4.52	5.05	5.25

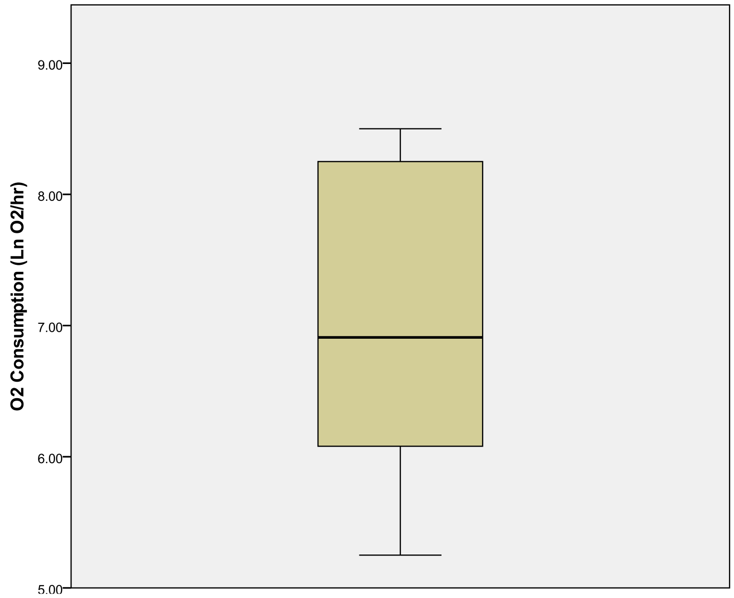
Assuming the probability of making a Type I error $= \alpha = 0.05$ (α^* in the class notes), the physiologist **wishes to test**:

$$
\begin{aligned}
&H_0 : \beta = 0.66 \\
&H_A : \beta \neq 0.66
\end{aligned}
$$

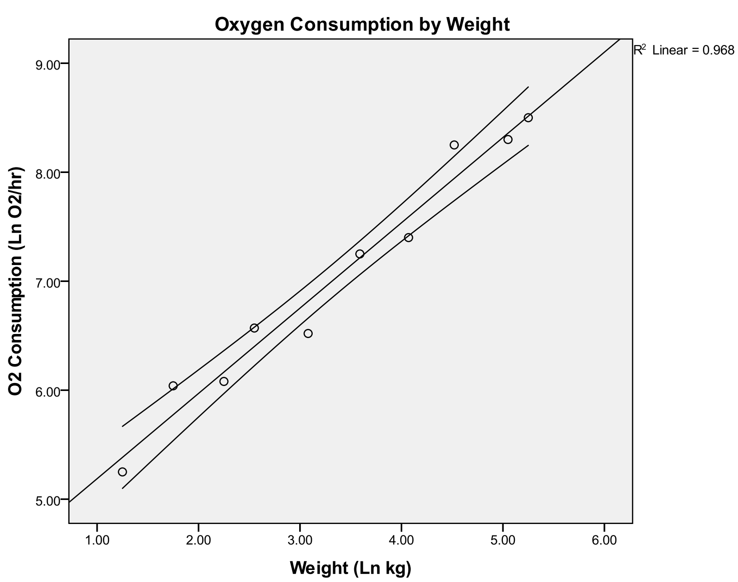
The physiologist comes to you for help in interpreting what the output means with regard to her regression analysis. She used a statistical software package to do the analysis, and **she provides you with the output of that analysis**. **What she provides you is all you have to use**. You do not have the opportunity to do your own analysis. **NOTE:** The questions come after the computer output, which is given on the following pages.

Using SPSS statistical software, the physiologist obtains the following output:

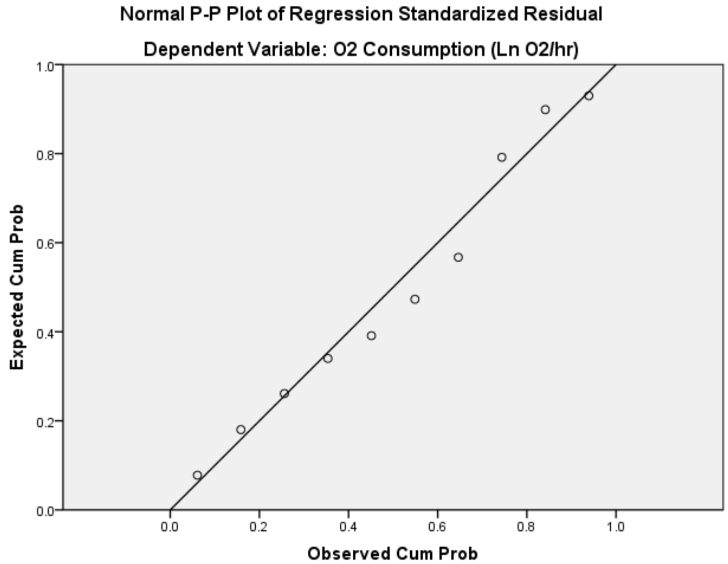
Box Plot

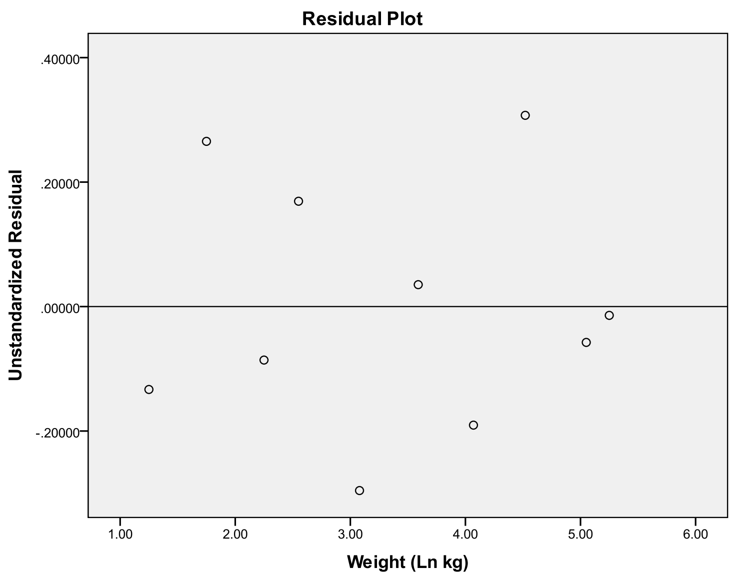


Scatter Plot



Note: The P-P plot is the Q-Q plot





Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.984[a]	**.968**	.964	.20829

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.984[a]	**.968**	.964	.20829

a. Predictors: (Constant), Weight (Ln kg)

b. Dependent Variable: O2 Consumption (Ln O2/hr)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	**(Constant)**	**4.405**	.179		24.551	.000	3.991	4.819
	Weight (Ln kg)	**.783**	**.050**	.984	15.647	.000	.667	.898

a. **Dependent Variable: O2 Consumption (Ln O2/hr)**

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.621	1	10.621	244.823	.000[a]
	Residual	.347	8	.043		
	Total	10.968	9			

a. **Predictors: (Constant), Weight (Ln kg)**

b. **Dependent Variable: O2 Consumption (Ln O2/hr)**

From the regression output, the **test statistic can be calculated as**:

$$t = \frac{\hat{\beta} - b_0}{\frac{s_e}{\sqrt{S_{xx}}}} = \frac{0.783 - 0.66}{0.050} = 2.46$$

<u>Questions to be answered</u>:

a. (*4 points*) Using the symbols in the problem, **clearly and specifically state** what the <u>linear model</u> coefficients are estimated to be. Using the values of these coefficients, **write the estimated regression equation**.

b. (*4 points*) Do the assumptions of linear regression appear to hold? In your answer, state exactly what output from the computer software you would look at to answer that question, what you would check for in looking at it, and what you conclude after looking at it.

c. (*2 points*) What percentage of the variability in our data can be explained by the regression model? What do we call this measure?
$R^2 = 0.968$.

d. (*2 points*) What is the critical value that is appropriate for the physiologist to use to test the hypothesis (that is, using the appropriate table, give the numerical value of the critical value)? Using the value of the test statistic given in the problem, state if the physiologist should reject or fail to reject the null hypothesis that the slope is approximately 0.66.

e. (*2 points*) State your conclusion in (d), above, in the context of the problem.