Problem 3

Rosner textbook 8^{th} edition, 9.7-9.8.

Health Services Administration

Suppose we want to compare the length of hospital stay for patients with the same diagnosis at two different hospitals. The results are shown in Table 1.

Table 2: Comparison of length of stay in 2 hospitals

First hospital	21, 10, 32, 60, 8, 44, 29, 5, 13, 26, 33
Second hospital	86, 27, 10, 68, 87, 76, 125, 60, 35, 73, 96, 44, 238

9.7 Why might a t test not be very useful in this case?

The underlying distribution don't follow normal distribution.

9.8 Carry out a nonparametric procedure for testing the hypothesis that lengths of stay are comparable in the two hospitals.

wilcoxon rank sum

```
        Wilcoxon rank sum test with continuity correction

data:  hos1 and hos2
W = 17.5, p-value = 0.001925
alternative hypothesis: true location shift is not equal to 0
```

2.

Rosner textbook 8^{th} edition, 9.17-9.21.

Hypertension

Polyunsaturated fatty acids in the diet favorably affect several risk factors for cardiovascular disease. The principal dietary polyunsaturated fat is linoleic acid. To test the effects of dietary supplementation with linoleic acid on blood pressure, 17 adults consumed 23 g/day of safflower oil, high in linoleic acid, for 4 weeks. Systolic blood pressure (SBP) measurements were taken at baseline (before ingestion of oil) and 1 month later, with the mean values over several readings at each visit given in Table 9.12.

TABLE 9.12 Effect of linoleic acid on SBP

Subject	Baseline SBP	1-month SBP	Baseline – 1-month SBP
1	119.67	117.33	2.34
2	100.00	98.78	1.22
3	123.56	123.83	−0.27
4	109.89	107.67	2.22
5	96.22	95.67	0.55
6	133.33	128.89	4.44
7	115.78	113.22	2.56
8	126.39	121.56	4.83
9	122.78	126.33	−3.55
10	117.44	110.39	7.05
11	111.33	107.00	4.33
12	117.33	108.44	8.89
13	120.67	117.00	3.67
14	131.67	126.89	4.78
15	92.39	93.06	−0.67
16	134.44	126.67	7.77
17	108.67	108.67	0.00

9.17 What parametric test could be used to test for the effect of linoleic acid on SBP?
The paired t test can be applied.

9.18 Perform the test in Problem 9.17, and report a p-value.

9.19 What nonparametric test could be used to test for the effect of linoleic acid on SBP?
We can use Wilcoxon signed-rank test.

9.20 Perform the test in Problem 9.19, and report a p-value.

9.21 Compare your results in Problems 9.18 and 9.20, and discuss which method you feel is more appropriate here.

Problem 1

Kutner textbook 5^{th} edition, 1.5, 1.16, 1.18.

1.5 When asked to state the simple linear regression model, a student wrote it as follows $E(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$. Do you agree?
No. $E(Y_i)$ does not have the error term ε_i. The correct model is: $E(Y_i) = \beta_0 + \beta_1 X_i$

1.16 Evaluate the following statement: "For the least squares method to be fully valid, it is required that the distribution of Y be normal.
No. Least squares method is a mathematical approach to obtain the estimates of the coefficients, therefore it does not require any assumption.

1.18 According to equation (1.17), $\sum e_i = 0$ when regression model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ is fitted to a set of n cases by the method of least squares. Is it also true that $\sum \varepsilon_i = 0$? Comment.
No. ε is an unknown random variable, it is the difference between Y_i and $E(Y_i)$, and $E(Y_i)$ is determined by the true regression model.
For example, suppose we have 5 observations in one data set. It would not be unusual for all 5 observations to fall above the TRUE regression line, and in that case $\sum \varepsilon_i$ would be strictly greater than 0.

Problem 2

Kutner textbook 5^{th} edition, 1.21, 1.25.

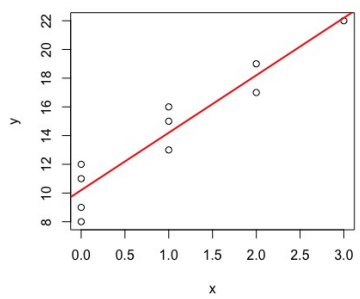
1.21 *Airfreight breakage*
A substance used in biological and medical research is shipped by air-freight to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y). Assume that first-order regression model (1.1) is appropriate.

i	1	2	3	4	5	6	7	8	9	10
X_i	1	0	2	0	3	1	0	1	2	0
Y_i	16	9	17	12	22	13	8	15	19	11

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here?

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2000     0.6633  15.377 3.18e-07 ***
x             4.0000     0.4690   8.528 2.75e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.483 on 8 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.8885
F-statistic: 72.73 on 1 and 8 DF,  p-value: 2.749e-05
```



b. Obtain a point estimate of the expected number of broken ampules when $X = 1$ transfer is made.

14.2 $\hat{Y}_1 = 10.2 + 4X_i = 10.2 + 4(1) = 14.2$

c. Estimate the increase in the expected number of ampules broken when there are 2 transfers as compared to 1 transfer.

4*2=8

1.25 *Airfreight breakage*

a. Obtain the residual for the first case. What is its relation to ε_1?

1.8 $e_1 = Y_1 - \hat{Y}_1 = 16 - (10.2 + 4 \times 1) = 1.8$

$E(e_1) = \varepsilon_1$

b. Compute $\sum e_i^2$ and MSE. What is estimated by MSE?

17.6 MSE = sum/n−2 = 2.2