26 packet

STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Administrators in a large school district wanted to determine whether students who attended a new magnet school for one year achieved greater improvement in science test performance than students who did not attend the magnet school. Knowing that more parents would want to enroll their children in the magnet school than there was space available for those children, the district administrators decided to conduct a lottery of all families who expressed interest in participating. In their data analysis, the administrators would then compare the change in test scores of those children who were selected to attend the magnet school with the change in test scores of those who applied to attend the magnet school but who were not selected.

GO ON TO THE NEXT PAGE.

The tables below show the scores on the same science pretest and the same science posttest for 20 students. Of the 20 students, 8 were randomly selected from the magnet school and 12 were randomly selected from those who applied to attend the magnet school but who were not selected and then attended their original school.

Magnet School		
Pretest Score	Posttest Score	Posttest − Pretest
80	97	17
78	98	20
86	84	−2
78	79	1
64	89	25
71	77	6
71	83	12
73	88	15
$\bar{x} = 75.125$	$\bar{x} = 86.875$	$\bar{x} = 11.750$
$s = 6.770$	$s = 7.699$	$s = 9.407$

Original School		
Pretest Score	Posttest Score	Posttest − Pretest
83	80	−3
80	89	9
63	65	2
79	78	−1
83	93	10
77	79	2
66	70	4
80	84	4
73	80	7
90	90	0
77	78	1
90	91	1
$\bar{x} = 78.417$	$\bar{x} = 81.417$	$\bar{x} = 3.000$
$s = 8.207$	$s = 8.512$	$s = 3.977$

GO ON TO THE NEXT PAGE.

(a) Perform a test to determine whether students who attend the magnet school demonstrate a significantly higher mean difference in test scores (Posttest − Pretest) than students who applied to attend the magnet school but who were not selected and then attended their original school.

Administrators were also interested in using pretest scores on this test as a predictor of posttest scores on the test. The following computer output contains the results from separate regression analyses on the magnet school scores and on the original school scores. The accompanying graph displays the data and separate regression lines for the magnet and original schools.

Regression Analysis: Post_Magnet versus Pre_Magnet

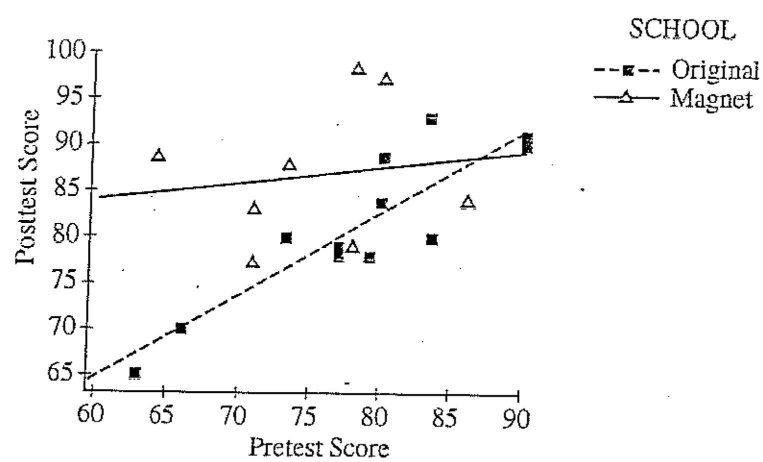
Predictor	Coef	SE Coef	T	P
Constant	73.27	34.55	2.12	0.078
Pre_Magnet	0.1811	0.4583	0.40	0.706

S = 8.20920 R-Sq = 2.5% R-Sq(adj) = 0.0%

Regression Analysis: Post_Original versus Pre_Original

Predictor	Coef	SE Coef	T	P
Constant	9.24	11.91	0.78	0.456
Pre_Original	0.9204	0.1512	6.09	0.000

S = 4.11463 R-Sq = 78.8% R-Sq(adj) = 76.6%



(b) (i) State the equation of the regression line for the magnet school and interpret its slope in the context of the question.

(ii) State the equation of the regression line for the original school and interpret its slope in the context of the question.

(c) To determine whether there is a significant correlation between pretest score and posttest score, a test of the following hypotheses will be performed.

H_0 : There is no correlation between pretest score and posttest score (true slope $= 0$)

versus

H_a : There is a correlation between pretest score and posttest score (true slope $\neq 0$)

(i) Using the regression output, state the p-value and conclusion for this test at the magnet school. Assume the conditions for inference have been met.

(ii) Using the regression output, state the p-value and conclusion for this test at the original school. Assume the conditions for inference have been met.

(d) What additional information do the regression analyses give you about student performance on the science test at the two schools beyond the comparison of mean differences in part (a) ?

STOP

END OF EXAM

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanations.

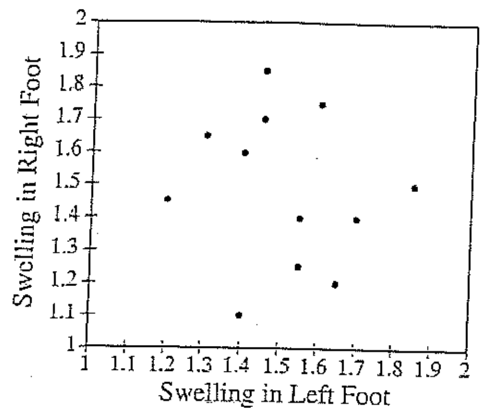
6. The nerves that supply sensation to the front portion of a person's foot run between the long bones of the foot. Tight-fitting shoes can squeeze these nerves between the bones, causing pain when the nerves swell. This condition is called Morton's neuroma. Because most people have a dominant foot, muscular development is not the same in both feet. People who have Morton's neuroma may have the condition in only one foot or they may have it in both feet.

   Investigators selected a random sample of 12 adult female patients with Morton's neuroma to study this disease further. The data below are measurements of nerve swelling as recorded by a physician. A value of 1.0 is considered "normal," and 2.0 is considered extreme swelling. The population distribution of the swelling measurements is approximately normal for adult females who have Morton's neuroma.

Dominant Foot	Swelling in Dominant Foot	Swelling in Nondominant Foot	Foot with Neuroma
Left	1.40	1.10	Left
Left	1.55	1.25	Left
Left	1.65	1.20	Left
Left	1.55	1.40	Both
Left	1.70	1.40	Left
Left	1.85	1.50	Both
Right	1.45	1.20	Right
Right	1.65	1.30	Right
Right	1.60	1.40	Right
Right	1.70	1.45	Both
Right	1.85	1.45	Both
Right	1.75	1.60	Both

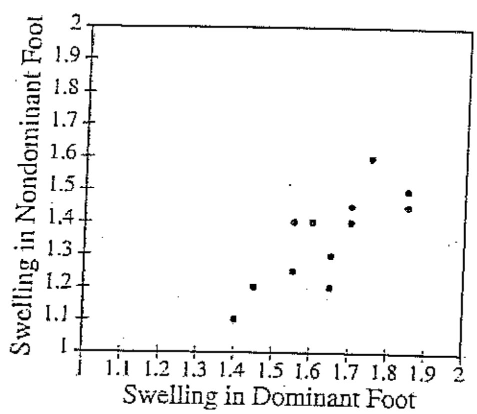
GO ON TO THE NEXT PAGE.

(a) A scatterplot of the ordered pairs (swelling in left foot, swelling in right foot), is shown below.



The scatterplot suggests there are two distinct groups of patients. Patients within each group share a common trait. Use the scatterplot above and the table on page 10 to determine the common trait and explain how this trait differs for the two groups.

(b) A scatterplot of the ordered pairs (swelling in dominant foot, swelling in nondominant foot), is shown below.



What conclusion can be drawn from this scatterplot that is not apparent from the scatterplot in part (a) ?

(c) Can you conclude that there is a difference between the mean swelling in the dominant foot and the mean swelling in the nondominant foot for adult females who have Morton's neuroma in at least one foot? Give a statistical justification to support your answer.

(For easy reference, the table of data from page 10 also appears at the bottom of page 12.)

(d) The nerve swelling measurement is used to indicate whether a foot has Morton's neuroma. Use the 24 measurements of nerve swelling to suggest a criterion for diagnosing Morton's neuroma. Justify your suggestion graphically.

(For easy reference, the table of data from page 10 also appears below.)

Dominant Foot	Swelling in Dominant Foot	Swelling in Nondominant Foot	Foot with Neuroma
Left	1.40	1.10	Left
Left	1.55	1.25	Left
Left	1.65	1.20	Left
Left	1.55	1.40	Both
Left	1.70	1.40	Left
Left	1.85	1.50	Both
Right	1.45	1.20	Right
Right	1.65	1.30	Right
Right	1.60	1.40	Right
Right	1.70	1.45	Both
Right	1.85	1.45	Both
Right	1.75	1.60	Both

STOP

END OF EXAM

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.

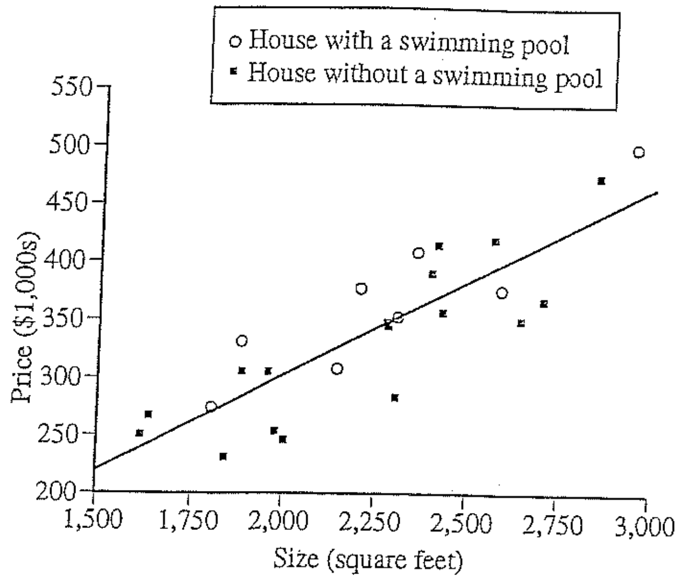
Percent of Section II score—25

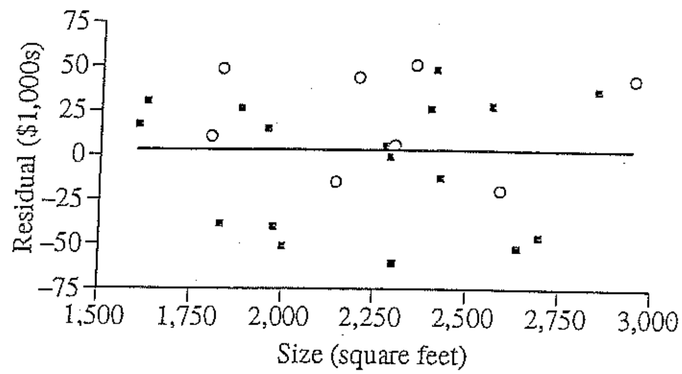
Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A real estate agent is interested in developing a model to estimate the prices of houses in a particular part of a large city. She takes a random sample of 25 recent sales and, for each house, records the price (in thousands of dollars), the size of the house (in square feet), and whether or not the house has a swimming pool. This information, along with regression output for a linear model using size to predict price, is shown below and on the next page.

Price ($1,000s)	Size (square feet)	Pool	Residual ($1,000s)
274	1,799	yes	6
330	1,875	yes	49
307	2,145	yes	−18
376	2,200	yes	42
352	2,300	yes	1
409	2,350	yes	50
375	2,589	yes	−23
498	2,943	yes	42
248	1,600	no	13
265	1,623	no	26
228	1,829	no	−45
303	1,875	no	22
303	1,950	no	10
251	1,975	no	−46
244	2,000	no	−57
347	2,274	no	1
345	2,279	no	−2
282	2,300	no	−69
389	2,392	no	23
413	2,410	no	44
353	2,428	no	−19
419	2,560	no	26
348	2,639	no	−58
365	2,701	no	−52
474	2,849	no	33

GO ON TO THE NEXT PAGE.





Linear Fit

Price = −28.144 + 0.165 Size

Summary of Fit

RSquare 0.722

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	−28.144	48.259	−0.58	0.5654
Size	0.165	0.0213	7.72	<.0001

(a) Interpret the slope of the least squares regression line in the context of the study.

GO ON TO THE NEXT PAGE.

(b) The second house in the table has a residual of 49. Interpret this residual value in the context of the study.

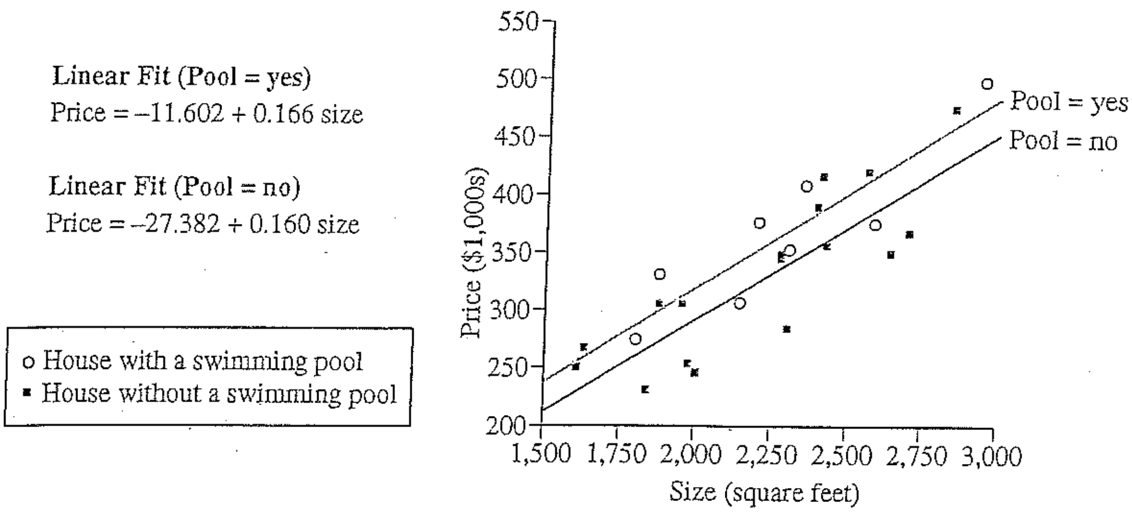
The real estate agent is interested in investigating the effect of having a swimming pool on the price of a house.

(c) Use the residuals from all 25 houses to estimate how much greater the price for a house with a swimming pool would be, on average, than the price for a house of the same size without a swimming pool.

To further investigate the effect of having a swimming pool on the price of a house, the real estate agent creates two regression models, one for houses with a swimming pool and one for houses without a swimming pool. Regression output for these two models is shown below.

Linear Fit (Pool = yes)
Price = −11.602 + 0.166 size

Linear Fit (Pool = no)
Price = −27.382 + 0.160 size



(d) The conditions for inference have been checked and verified, and a 95 percent confidence interval for the true difference in the two slopes is $(-0.099, 0.110)$. Based on this interval, is there a significant difference in the two slopes? Explain your answer.

(e) Use the regression model for houses with a swimming pool and the regression model for houses without a swimming pool to estimate how much greater the price for a house with a swimming pool would be than the price for a house of the same size without a swimming pool. How does this estimate compare with your result from part (c) ?

STOP

END OF EXAM

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Two treatments, A and B, showed promise for treating a potentially fatal disease. A randomized experiment was conducted to determine whether there is a significant difference in the survival rate between patients who receive treatment A and those who receive treatment B. Of 154 patients who received treatment A, 38 survived for at least 15 years, whereas 16 of the 164 patients who received treatment B survived at least 15 years.

(a) Treatment A can be administered only as a pill, and treatment B can be administered only as an injection. Can this randomized experiment be performed as a double-blind experiment? Why or why not?

(b) The conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the difference between the proportion of the population who would survive at least 15 years if given treatment A and the proportion of the population who would survive at least 15 years if given treatment B.

In many of these types of studies, physicians are interested in the ratio of survival probabilities, $\frac{p_A}{p_B}$, where p_A represents the true 15-year survival rate for all patients who receive treatment A and p_B represents the true 15-year survival rate for all patients who receive treatment B. This ratio is usually referred to as the relative risk of the two treatments.

For example, a relative risk of 1 indicates the survival rates for patients receiving the two treatments are equal, whereas a relative risk of 1.5 indicates that the survival rate for patients receiving treatment A is 50 percent higher than the survival rate for patients receiving treatment B. An estimator of the relative risk is the ratio of estimated probabilities, $\frac{\hat{p}_A}{\hat{p}_B}$.

(c) Using the data from the randomized experiment described above, compute the estimate of the relative risk.

The sampling distribution of $\frac{\hat{p}_A}{\hat{p}_B}$ is skewed. However, when both sample sizes n_A and n_B are relatively large, the distribution of $\ln\left(\frac{\hat{p}_A}{\hat{p}_B}\right)$ — the natural logarithm of relative risk — is approximately normal with a mean of $\ln\left(\frac{p_A}{p_B}\right)$ and a standard deviation of $\sqrt{\frac{1-p_A}{n_A p_A}+\frac{1-p_B}{n_B p_B}}$, where p_A and p_B can be estimated by using $\hat{p}_A$ and $\hat{p}_B$.

GO ON TO THE NEXT PAGE.

When a 95 percent confidence interval for $\ln\left(\frac{p_A}{p_B}\right)$ is known, an approximate 95 percent confidence interval for $\frac{p_A}{p_B}$ — the relative risk of the two treatments — can be constructed by applying the inverse of the natural logarithm to the endpoints of the confidence interval for $\ln\left(\frac{p_A}{p_B}\right)$.

(d) The conditions for inference are met for the data in the experiment above, and a 95 percent confidence interval for $\ln\left(\frac{p_A}{p_B}\right)$ is (0.3868, 1.4690). Construct and interpret a 95 percent confidence interval for the relative risk, $\frac{p_A}{p_B}$, of the two treatments.

(e) What is an advantage of using the interval in part (d) over using the interval in part (b) ?

STOP

END OF EXAM

STATISTICS

Section II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

6. A pharmaceutical company has developed a new drug to reduce cholesterol. A regulatory agency will recommend the new drug for use if there is convincing evidence that the mean reduction in cholesterol level after one month of use is more than 20 milligrams/deciliter (mg/dl), because a mean reduction of this magnitude would be greater than the mean reduction for the current most widely used drug.

   The pharmaceutical company collected data by giving the new drug to a random sample of 50 people from the population of people with high cholesterol. The reduction in cholesterol level after one month of use was recorded for each individual in the sample, resulting in a sample mean reduction and standard deviation of 24 mg/dl and 15 mg/dl, respectively.

   (a) The regulatory agency decides to use an interval estimate for the population mean reduction in cholesterol level for the new drug. Provide this 95 percent confidence interval. Be sure to interpret this interval.

   (b) Because the 95 percent confidence interval includes 20, the regulatory agency is not convinced that the new drug is better than the current best-seller. The pharmaceutical company tested the following hypotheses.

   $$H_0: \mu = 20 \text{ versus } H_a: \mu > 20,$$

   where μ represents the population mean reduction in cholesterol level for the new drug.

   The test procedure resulted in a t-value of 1.89 and a p-value of 0.033. Because the p-value was less than 0.05, the company believes that there is convincing evidence that the mean reduction in cholesterol level for the new drug is more than 20. Explain why the confidence interval and the hypothesis test led to different conclusions.

   (c) The company would like to determine a value L that would allow them to make the following statement.

   We are 95 percent confident that the true mean reduction in cholesterol level is greater than L.

   A statement of this form is called a one-sided confidence interval. The value of L can be found using the following formula.

   $$L = \bar{x} - t^* \frac{s}{\sqrt{n}}$$

   This has the same form as the lower endpoint of the confidence interval in part (a), but requires a different critical value, t^*. What value should be used for t^*?

   Recall that the sample mean reduction in cholesterol level and standard deviation are 24 mg/dl and 15 mg/dl, respectively. Compute the value of L.

   (d) If the regulatory agency had used the one-sided confidence interval in part (c) rather than the interval constructed in part (a), would it have reached a different conclusion? Explain.

END OF EXAMINATION

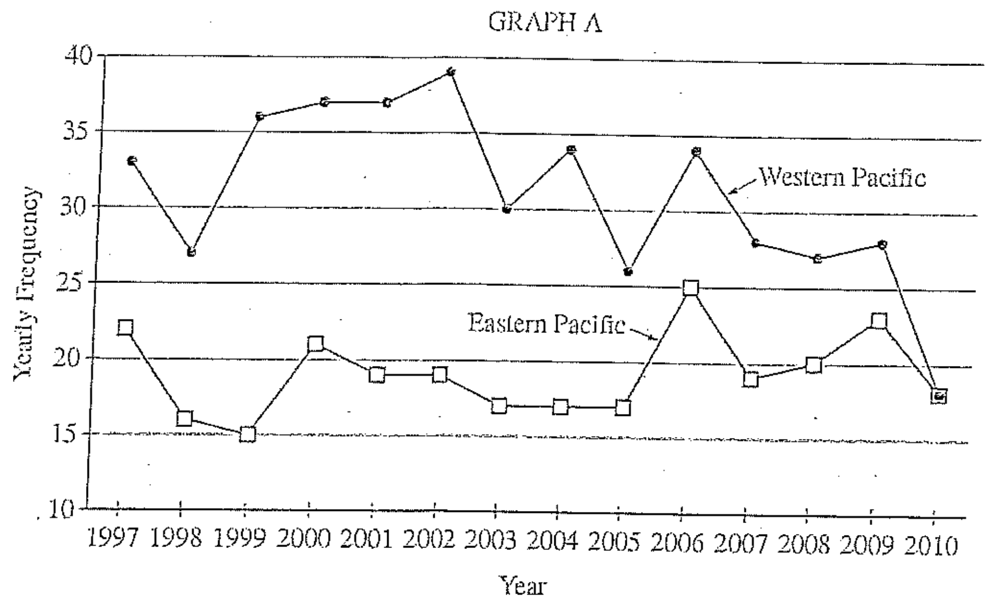
STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Tropical storms in the Pacific Ocean with sustained winds that exceed 74 miles per hour are called typhoons. Graph A below displays the number of recorded typhoons in two regions of the Pacific Ocean—the Eastern Pacific and the Western Pacific—for the years from 1997 to 2010.



(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

(b) For each region, describe how the yearly frequencies changed over the time period from 1997 to 2010.

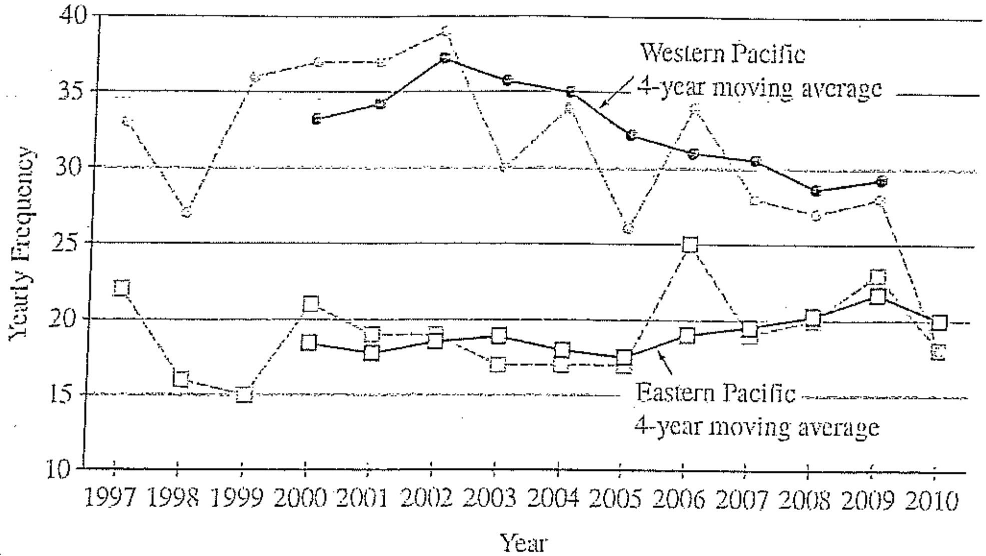
A moving average for data collected at regular time increments is the average of data values for two or more consecutive increments. The 4-year moving averages for the typhoon data are provided in the table below. For example, the Eastern Pacific 4-year moving average for 2000 is the average of 22, 16, 15, and 21, which is equal to 18.50.

Year	Number of Typhoons in the Eastern Pacific	Eastern Pacific 4-year moving average	Number of Typhoons in the Western Pacific	Western Pacific 4-year moving average
1997	22		33	
1998	16		27	
1999	15		36	
2000	21	18.50	37	33.25
2001	19	17.75	37	34.25
2002	19	18.50	39	37.25
2003	17	19.00	30	35.75
2004	17	18.00	34	35.00
2005	17	17.50	26	32.25
2006	25	19.00	34	31.00
2007	19	19.50	28	30.50
2008	20	20.25	27	28.75
2009	23	21.75	28	29.25
2010	18	20.00	18	

(c) Show how to calculate the 4-year moving average for the year 2010 in the Western Pacific. Write your value in the appropriate place in the table.

(d) Graph B below shows both yearly frequencies (connected by dashed lines) and the respective 4-year moving averages (connected by solid lines). Use your answer in part (c) to complete the graph.

GRAPH B



(e) Consider graph B.

i) What information is more apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

ii) What information is less apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

STOP

END OF EXAM