

# STUDENT SOLUTIONS MANUAL

Lindsay Packer

# STATISTICS

Informed Decisions  
Using Data

Michael Sullivan, III

# STUDENT SOLUTIONS MANUAL

**Lindsay Packer**

*College of Charleston*

# STATISTICS

**Informed Decisions  
Using Data**

**Michael Sullivan, III**



Editor-in-Chief: Sally Yagan  
Acquisitions Editor: Petra Recter  
Supplement Editor: Joanne Wendelken  
Assistant Managing Editor: John Matthews  
Production Editor: Allyson Kloss  
Supplement Cover Manager: Paul Gourhan  
Supplement Cover Designer: Joanne Alexandris  
Manufacturing Buyer: Ilene Kahn



© 2004 Pearson Education, Inc.  
Pearson Prentice Hall  
Pearson Education, Inc.  
Upper Saddle River, NJ 07458

All rights reserved. No part of this book may be reproduced in any form or by any means, without permission in writing from the publisher.

Pearson Prentice Hall® is a trademark of Pearson Education, Inc.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Printed in the United States of America

10 9 8 7 6 5 4 3

**ISBN 0-13-046492-9**

Pearson Education Ltd., *London*  
Pearson Education Australia Pty. Ltd., *Sydney*  
Pearson Education Singapore, Pte. Ltd.  
Pearson Education North Asia Ltd., *Hong Kong*  
Pearson Education Canada, Inc., *Toronto*  
Pearson Educación de México, S.A. de C.V.  
Pearson Education—Japan, *Tokyo*  
Pearson Education Malaysia, Pte. Ltd.  
Pearson Education, *Upper Saddle River, New Jersey*

## Table of Contents

Chapter 1	Data Collection	1
Chapter 2	Organizing and Summarizing Data	11
Chapter 3	Numerically Summarizing Data	45
Chapter 4	Describing the Relation Between Two Variables	65
Chapter 5	Probability	89
Chapter 6	Discrete Probability Distributions	105
Chapter 7	The Normal Probability Distribution	123
Chapter 8	Confidence Intervals about a Single Parameter	155
Chapter 9	Hypothesis Testing	177
Chapter 10	Inference on Two Samples	201
Chapter 11	Chi-Square Procedures	219
Chapter 12	Inference on the Least-Squares Regression Model; ANOVA	239
Chapter 13	Nonparametric Statistics	259

## Chapter 1. Data Collection

### 1.1 Introduction to the Practice of Statistics

#### 1.1 Concepts and Vocabulary

1. Statistics is the science of collecting, organizing, summarizing and analyzing information in order to draw conclusions.
3. Descriptive statistics consists of organizing and summarizing information. Inferential statistics consists of generalizing results from a sample to the population, and measuring the reliability of those results.
5. A qualitative variable classifies individuals based on some attribute or characteristic. Some examples are gender, zip codes, class (freshman, sophomore etc) and ethnicity. A quantitative variable is a numerical variable on which arithmetic operations can be sensibly performed. Some examples are temperature, height, blood pressure and life expectancy.

#### 1.1 Exercises: Basic Skills

1. Qualitative.
3. Quantitative.
5. Quantitative.
7. Quantitative.
9. Qualitative.
11. Discrete.
13. Discrete.
15. Continuous.
17. Discrete.
19. Continuous.
21. The population consists of all adult residents of the United States aged 18 years or older. The members of this population are the individuals. The sample consists of the 119 adult residents contacted by the Gallup Organization.
23. The population consists of all soybean plants grown by this farmer. The individuals are the individual soybean plants. The sample consists of the 100 plants randomly selected by the farmer.

#### 1.1 Exercises: Applying the Concepts

25. (a) To determine the genetic and non-genetic factors to structural brain abnormalities on schizophrenia.  
(b) The sample consisted of 58 pairs of twins (29 with schizophrenia, and 29 healthy).  
(c) The study would have calculated average brain volumes in the two groups of subjects. They found that brain volumes were 2.2% smaller in the schizophrenic patients.  
(d) An increased genetic risk to develop schizophrenia is related to reduced brain growth early in life.

27. (a) To determine whether there is a relationship between music cognition and cognitions pertaining to abstract operations such as mathematical or spatial reasoning.  
 (b) The sample consisted of 36 college students.  
 (c) The researchers calculated the mean scores of these students on a spatial reasoning test, taken both after the students had listened to Mozart, and after the students had sat in silence. The mean scores were 119, after listening to Mozart, and 110 after listening to silence.  
 (d) Subjects perform better on abstract/spatial reasoning tests after listening to Mozart.
29. (a) To determine what percentage of households had been victimized by crime in the preceding year.  
 (b) The sample consisted of the households of 1012 adults, aged 18 years or older.  
 (c) The researchers calculated the percentage of these households that had been victimized by crime in the preceding 12 months. It was 24%.  
 (d) Gallup News Service concluded that 24% of all households had been victimized by crime in the preceding year.
31. The individuals are the five students in Michael Sullivan's business calculus course. The variables are gender (qualitative), age (quantitative, continuous) and number of siblings (quantitative, discrete). The data for these variables are: F, M, F etc for gender; 19, 19, 19 etc for age; and 1, 1, 2 etc for number of siblings.
33. The individuals are the five states. The variables are age for driver's license (quantitative, continuous), blood alcohol concentration limit (quantitative, continuous), mandatory belt-use law seating positions (qualitative), and maximum allowable speed limit in 1999 (quantitative, continuous). The data are: 16, 18 etc for age for driver's license; 0.08, 0.08 etc for blood alcohol concentration limit; front, front etc for mandatory belt-use law seating positions; and 70, 65 etc for maximum allowable speed limit in 1999.

## 1.2 Observational Studies; Simple Random Sampling

### 1.2 Concepts and Vocabulary

1. An observational study uses data obtained by studying individuals in a sample without trying to manipulate or influence the variable(s) of interest. In a designed experiment, a treatment is applied to the individuals in a sample in order to isolate the effects of the treatment on a response variable. Observational studies are appropriate where the control of certain variables is either impossible or unethical. Designed experiments are appropriate when it is possible to control certain variables and this is necessary for the study (for example to establish causation).
3. Sampling is used in statistics because it can be prohibitively expensive or impossible to obtain census data.

**1.2 Exercises: Applying the Concepts**

1. Observational study.      3. Experiment.      5. Observational study.  
 7. Observational study.      9. Experiment.      11. Observational study

13. (a) All possible samples of size two, selected without replacement, are:

Graham, Murkowski	Graham, Kyl	Graham, Baucus	Graham, Conrad	Murkowski, Kyl
Murkowski, Baucus	Murkowski, Conrad	Kyl, Baucus	Kyl, Conrad	Baucus, Conrad

(b) If the members are selected at random, then there is a one in ten (or 10%) chance of these two being selected.

15. (a) Answers will vary.      (b) Answers will vary.

17. (a) Number each student in the list of registered students, from 1 to 19,935. Generate 25 random numbers, without repetition, between 1 and 19,935 using a random number generator or table. Select the 25 students with these numbers.

(b) Answers will vary.

**1.3 Other Types of Sampling****1.3 Concepts and Vocabulary**

- Stratified random sampling may be appropriate if the population of interest can be divided into groups or strata that are homogeneous (or similar) in some way.
- Convenience samples are typically selected in a nonrandom manner (such as self-selection) and this makes it impossible to quantify the likely margin of error in statistics calculated from those samples. Convenience samples may also be self-selected, which will frequently result in sample bias.

**1.3 Exercises: Basic Skills**

1. Systematic.      3. Cluster.      5. Simple random.      7. Cluster.  
 9. Convenience.

### 1.3 Exercises: Applying the Concepts

11. Answers will vary. One design would be a stratified random sample, with two strata being rail commuters and those who do not commute by rail, as these two groups each might be fairly homogeneous in their reactions to the proposal.
13. Answers will vary. One design would be a cluster sample, with the clusters being city blocks. Randomly select city blocks and survey every household in the selected blocks.
15. Answers will vary. Since the company already has a list (frame) of 6,600 individuals with high cholesterol, a simple random sample would be an appropriate design.
17. (a)  $N = 4502$ ,  $n = 50$ ,  $4502/50 = 90.04 \therefore k = 90$ .  
 (b) Randomly select a number between 1 and 90. Suppose that we select 15. Then the individuals to be surveyed will be the 15th, 105th, 195th, 285th and so on up to the 4425th employee on the company list.

## 1.4 Sources of Errors in Sampling

### 1.4 Concepts and Vocabulary

1. The population may be very large (for example the population of the United States), making it difficult to obtain a complete frame, and the population may change frequently (for example the voter roll), making it difficult to keep the frame up to date.
3. A closed question is one in which the respondent must choose one from a list of prescribed responses. An open question is one in which the respondent is free to choose his or her own response.
5. A talented interviewer will be able to elicit truthful responses even to sensitive questions.
7. A pro is that the interviewer is more likely to find the individual at home at this time. A con is that many individuals will be irritated at having their dinner interrupted and will refuse to respond.
9. Open questions allow respondents to give answers that more accurately reflect their true opinions. However, if there are too many distinct answers, then the results may be difficult to use. Closed questions limit the responses and so the results are easier to analyze. However, the choices do not always allow the respondent to answer as he or she would wish.

### 1.4 Exercises: Basic Skills

1. (a) Flawed sampling method.  
 (b) A simple random sample would be a good choice of sampling method as the vice president would have access to a list of all students.



3. (a) Flawed survey due to a poorly worded question.  
 (b) The survey should inform the respondent of the current penalty for selling a gun illegally and the question should be worded as: "Do you approve or disapprove of harsher penalties for individuals who sell guns illegally?" The order of "approve" and "disapprove" should be switched from one individual to the next.
5. (a) Flawed survey. If a foreign language is the primary language in a household, then the head of the household might not be able to respond to the question.  
 (b) A better design would be to use personal or telephone interviews.
7. (a) Flawed survey because the students are unlikely to give honest answers if their teacher is administering the survey.  
 (b) An impartial party should administer the survey in order to increase the rate of truthful responses.
9. It is very likely that the order of these two questions will affect the response to B. Either B could be asked first, or the order of the two questions could be switched at random.

## 1.5 The Design of Experiments

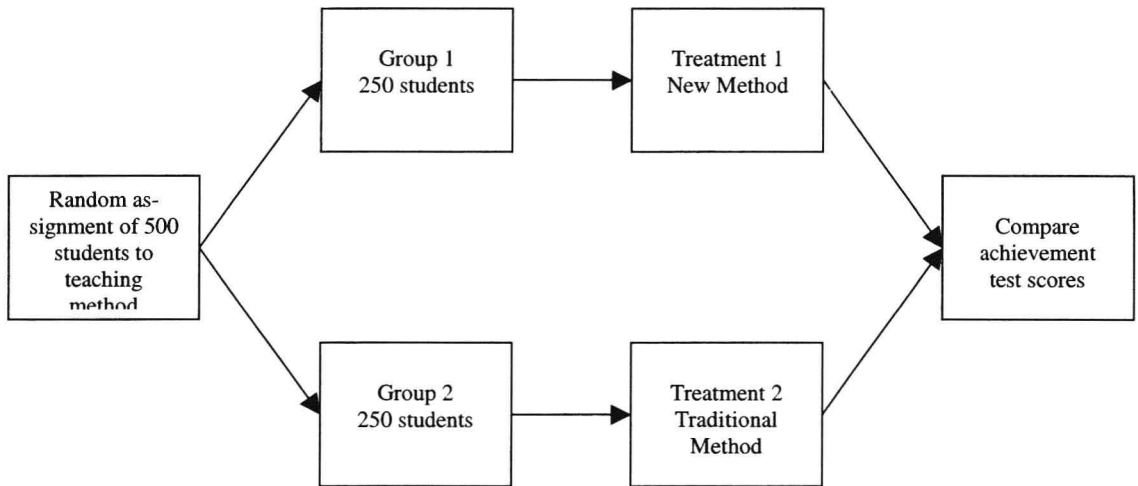
### 1.5 Concepts and Vocabulary

1. (a) An experimental unit is a person or object to which a treatment is applied.  
 (b) A treatment is a condition applied to an experimental unit.  
 (c) A response variable is a variable that measures a response of interest to the experimenter.  
 (d) A predictor variable measures a factor that might affect a response variable.  
 (e) A double-blind experiment is one in which neither the subject nor the experimenter knows what treatment is being administered to the subject.  
 (f) A placebo is an innocuous treatment, such as a sugar pill, administered to a subject in a manner indistinguishable from an actual treatment.  
 (g) Confounding occurs when the effect of two predictor variables on a response variable cannot be distinguished.

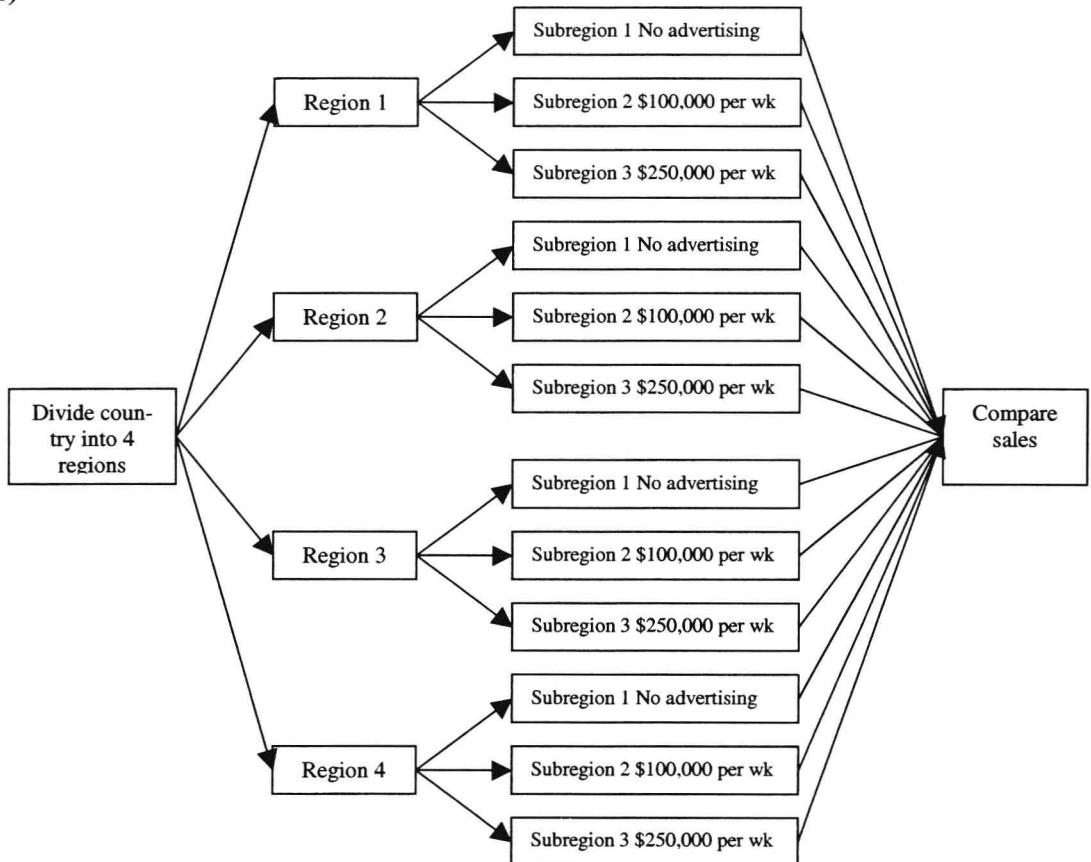
### 1.5 Exercises: Applying the Concepts

1. (a) Achievement test score.  
 (b) Method of teaching. Two levels.  
 (c) School district, grade level and teacher.  
 (d) By selecting students from a single school district.  
 (e) Completely randomized design.  
 (f) The 500 students.

(g)

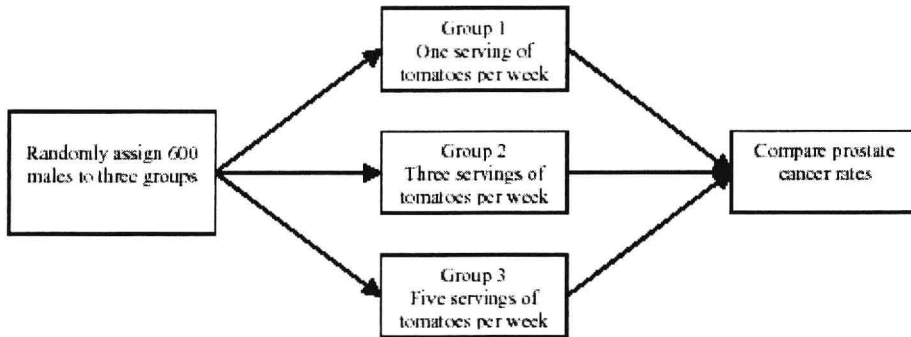


3. (a) Product sales.  
 (b) Advertising (\$ per week). Three levels.  
 (c) Region of the country.  
 (d) Randomized block design.  
 (e)

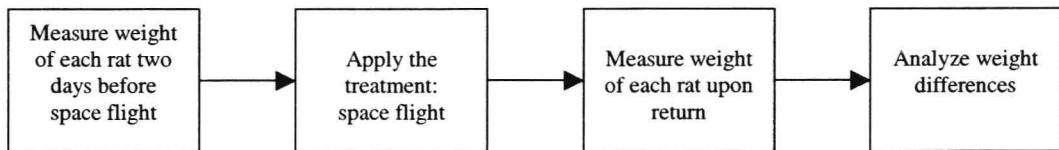


5. (a) Occurrence of prostate cancer.  
 (b) Weekly consumption of tomatoes. Three levels.

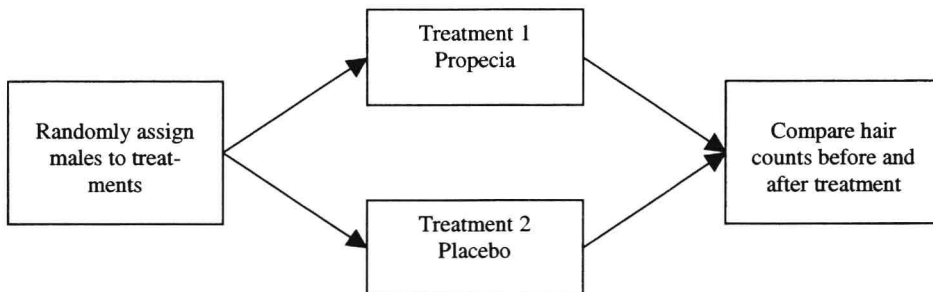
- (c) Age, eating habits and other lifestyle variables.
- (d) Completely randomized design.
- (e) The 600 males aged 30 years.
- (f)



7. (a) Weight difference before and after space flight.  
 (b) Space flight.  
 (c) Matched pairs.  
 (d)



9. (a) Hair counts.  
 (b) Medication. Two levels (Propecia or placebo).  
 (c) Completely randomized design.  
 (d)

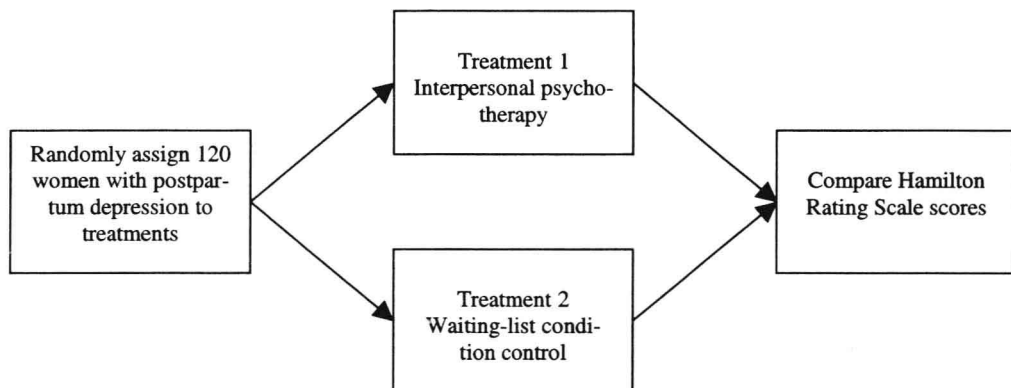


11. Answers will vary. The experimental units would be people who work out. Some factors to take into account might be age, gender, and physical condition prior to exercising (such as weight, blood pressure etc). A matched pairs design would be the best way to control for these factors. The response variable would be the change in lung capacity from before to after the exercise regimen.

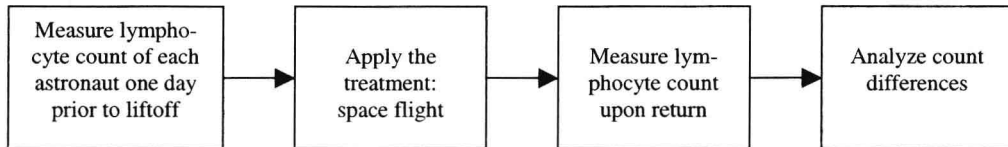
13. (a) Completely randomized. (b) Answers will vary.

## Chapter 1 Review Exercises

1. Statistics is the science of collecting, organizing, summarizing and analyzing information in order to draw conclusions.
3. A sample is a subset of the population.
5. In a designed experiment, a treatment is applied to the individuals in a sample in order to isolate the effects of the treatment on the response variable.
7. Errors in sampling consist of sampling error (the error resulting from using sample data to estimate a characteristic of an entire population) and non-sampling error (from poor sampling design). Sampling error can be quantified by using the theory of probability. Some common non-sampling errors are non-response, which can be ameliorated by call-back or incentives, and poorly worded questions, which can be avoided by careful survey design.
9. Qualitative.
11. Quantitative, continuous.
13. Quantitative, discrete.
15. Observational study.
17. Observational study.
19. Systematic.
21. Stratified.
23. Answers will vary.
25. Answers will vary.
27. (a) Score on the Hamilton Rating Scale for Depression.  
 (b) Group. Two levels (psychotherapy or no treatment).  
 (c) Age and medical history are two, neither of which is controlled.  
 (d) Completely randomized design.  
 (e) The 120 women with postpartum depression.  
 (f)



29. (a) Lymphocyte count.  
 (b) Space flight.  
 (c) Matched pairs.  
 (d) The four members of Skylab.  
 (e)



- 31.** In a matched pairs design, experimental units are paired based on some common characteristic. One of each pair is assigned at random to each of two treatments, and the difference between the pairs is analyzed. In a completely randomized design, each experimental unit is randomly assigned to one of the treatments, and no attempt is made to compare individual units receiving different treatments—only the two (or more) treatment groups as a whole are compared.



## Chapter 2. Organizing and Summarizing Data

### 2.1 Organizing Qualitative Data

#### 2.1 Concepts and Vocabulary

1. Raw data are the data as originally collected, before they have been organized.
3. A Pareto chart is a bar chart with bars drawn in order of decreasing frequency or relative frequency.
5. Pareto charts emphasize those observations with the higher frequencies or relative frequencies.

#### 2.1 Exercises: Basic Skills

1. (a) 34–52. (b) 18–33. (c) 27%.
3. (a) United States. (b)  $\approx 18$  million.
5. (a) 447,927  
(b) Total votes cast =  $49,922,623 + 49,659,871 + \dots + 98,226 = 103,268,793$   
Gore:  $49,922,623 / 103,268,793 = 48.3\%$   
Bush:  $49,659,871 / 103,268,793 = 48.1\%$   
(c) Nader:  $2,756,008 / 103,268,793 = 2.7\%$ . No.
7. (a) 16%. (b) Natural gas. (c) LPG.  
(d) The shift of population towards the southern states.  
(e) The oil crises that drove up the price of heating oil in the past.

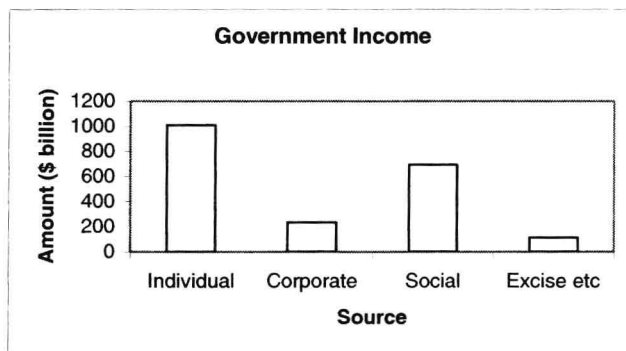
#### 2.1 Exercises: Applying the Concepts

9. (a) Total receipts =  $1009.5 + 234.7 + 691.5 + 111.2 = 2046.9$   
Relative frequency of individual income tax =  $1009.5 / 2046.9 = 0.4932$  and so on.

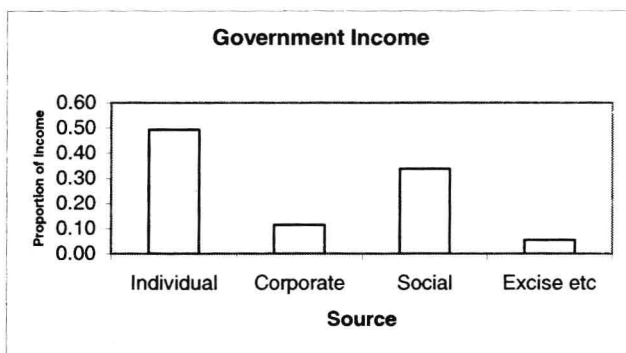
Source of income	Relative Frequency
Individual Income Tax and Tax Withholdings	0.4932
Corporate Income Taxes	0.1147
Social Insurance and Retirement Receipts	0.3378
Excise, Estate and Gift Taxes, Customs, and Miscellaneous Receipts	0.0543

- (b) 49.32%.

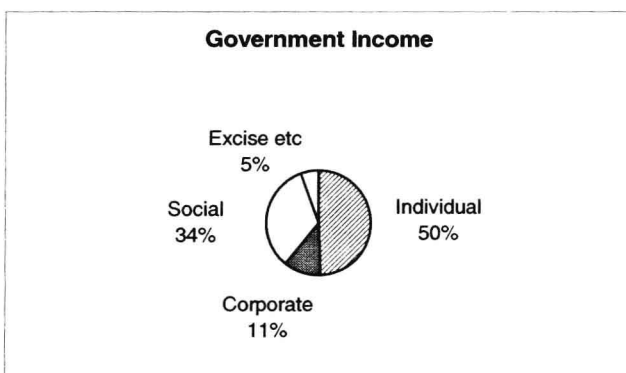
(c)



(d)



(e)



11. (a) Total students surveyed =  $125 + 324 + 552 + 1257 + 2518 = 4776$   
Relative frequency of "never" =  $125 / 4776 = 0.0262$  and so on.

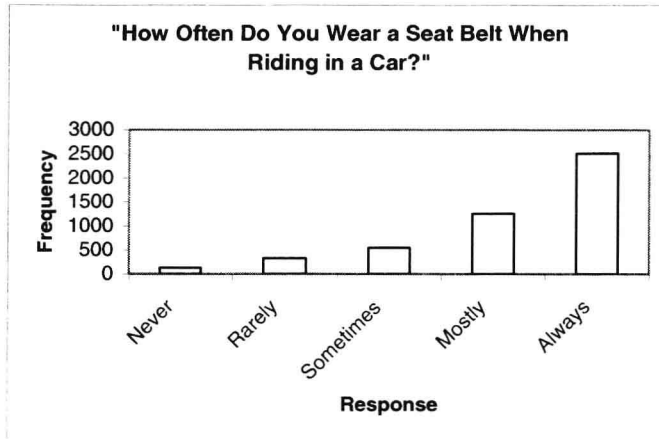
Response	Relative Frequency
Never	0.0262
Rarely	0.0678
Sometimes	0.1156
Most of the time	0.2632
Always	0.5272

(b) 52.72%

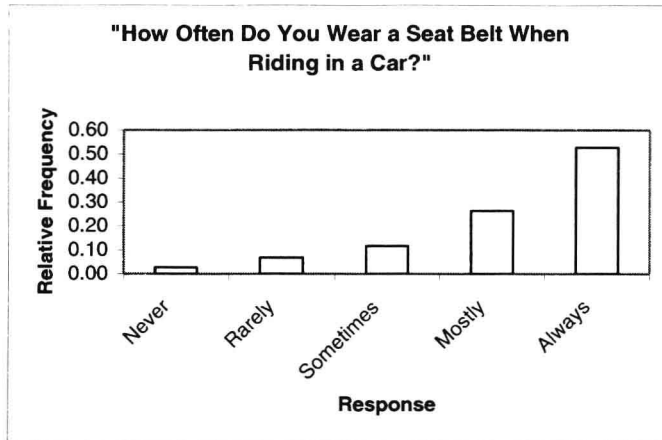
(c)  $2.62 + 6.78 = 9.40\%$



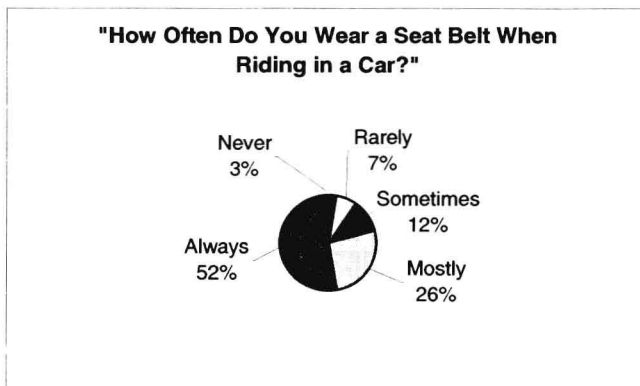
(d)



(e)



(f)



13. (a) Total foreign =  $4772 + 8364 + 840 + 180 + 15472 + 836 = 30464$  (thousand)  
 Relative frequency for Europe =  $4772 / 30464 = 0.1566$  and so on.

Region	Relative Frequency
Europe	0.1566
Asia	0.2746
Africa	0.0276