CHAPTER 2

A REVIEW OF BASIC STATISTICAL CONCEPTS

ANSWERS TO PROBLEMS AND CASES

1. Descriptive Statistics

Variable	N	Mean	Median	StDev	SE Mean
Orders	28	21.32	17.00	13.37	2.53

Variable	Min	Max	Q1	Q3
Orders	5.00	54.00	11.25	28.75

a. $\overline{X} = 21.32$

b. S = 13.37

c. $S^2 = 178.76$

d. If the policy is successful, smaller orders will be eliminated and the mean will increase.

e. If the change causes all customers to consolidate a number of small orders into large orders, the standard deviation will probably decrease. Otherwise, it is very difficult to tell how the standard deviation will be affected.

f. The best forecast over the long-term is the mean of 21.32.

2. Descriptive Statistics

Variable	N	Mean	Median	StDev	SE Mean
Prices	12	176654	180000	39440	11385

Variable	Min	Max	Q1	Q3
Prices	121450	253000	138325	205625

$\overline{X} = 176{,}654$ and S = 39,440

3. a. Point estimate: $\bar{X} = 10.76\%$

b. $1-\alpha = .95 \Rightarrow Z = 1.96,\ n = 30,\ \bar{X} = 10.76,\ S = 13.71$

$\bar{X} \pm 1.96(S/\sqrt{n}) = 10.76 \pm 1.96(13.71/\sqrt{30}) = 10.76 \pm 4.91$

(5.85%, 15.67%)

c. df = 30–1 = 29, t = 2.045

$\bar{X} \pm 2.045\left(S/\sqrt{n}\right) = 10.76 \pm 2.045\left(13.71/\sqrt{30}\right) = 10.76 \pm 5.12$

(5.64%, 15.88%)

d. We see that the 95% confidence intervals in b and c are not much different because the multipliers 1.96 and 2.045 are nearly the same magnitude. This explains why a sample of size n = 30 is often taken as the cutoff between large and small samples.

4. a. Point estimate: $\bar{X} = \frac{23.41 + 102.59}{2} = 63$

95% error margin: (102.59 – 23.41)/2 = 39.59

b. 1–α = .90 ⇒ Z = 1.645, $\bar{X} = 63,\ S/\sqrt{n} = 39.59/1.96 = 20.2$

$\bar{X} \pm 1.645\left(S/\sqrt{n}\right) = 63 \pm 1.645\left(20.2\right) = 63 \pm 33.23$

(29.77, 96.23)

5. H_0: μ = 12.1 n = 100 α = .05

H_1: μ > 12.1 S = 1.7 $\bar{X}$ = 13.5

Reject H_0 if Z > 1.645

$$Z = \frac{13.5 - 12.1}{1.7/\sqrt{100}} = 8.235$$

Reject H_0 since the computed Z (8.235) is greater than the critical Z (1.645). The mean has increased.

6. point estimate: 8.1 seats

interval estimate: $8.1 \pm 1.96 \frac{5.7}{\sqrt{49}}$ ⇒ 6.5 to 9.7 seats

Forecast 8.1 empty seats per flight; very likely the mean number of empty seats will lie between 6.5 and 9.7.

7. n = 60, $\bar{X} = 5.60,\ S = .87$

$H_0: \mu = 5.9$

$H_1: \mu \neq 5.9$ two-sided test, α = .05, critical value: |Z|= 1.96

Test statistic: $Z = \frac{\bar{X} - 5.9}{S/\sqrt{n}} = \frac{5.60 - 5.9}{.87/\sqrt{60}} = -2.67$

Since |–2.67| = 2.67 > 1.96, reject H_0 at the 5% level. The mean satisfaction rating is

different from 5.9.
p-value: P($Z < -2.67$ or $Z > 2.67$) = 2 P($Z > 2.67$) = 2(.0038) = .0076, very strong evidence against H_0.

8. df = $n - 1 = 14 - 1 = 13$, $\bar{X} = 4.31,\ S = .52$

$H_0: \mu = 4$

$H_1: \mu > 4$ one-sided test, α = .05, critical value: $t = 1.771$

Test statistic: $t = \dfrac{\bar{X} - 4}{S/\sqrt{n}} = \dfrac{4.31 - 4}{.52/\sqrt{14}} = 2.23$

Since 2.23 > 1.771, reject H_0 at the 5% level. The medium-size serving contains an average of more than 4 ounces of yogurt.

p-value: P($t > 2.23$) = .022, strong evidence against H_0

9. H_0: $\mu = 700$ n = 50 α = .05
H_1: $\mu \neq 700$ S = 50 $\bar{X} = 715$

Reject H_0 if Z < -1.96 or Z > 1.96

$Z = \dfrac{715 - 700}{50/\sqrt{50}} = 2.12$

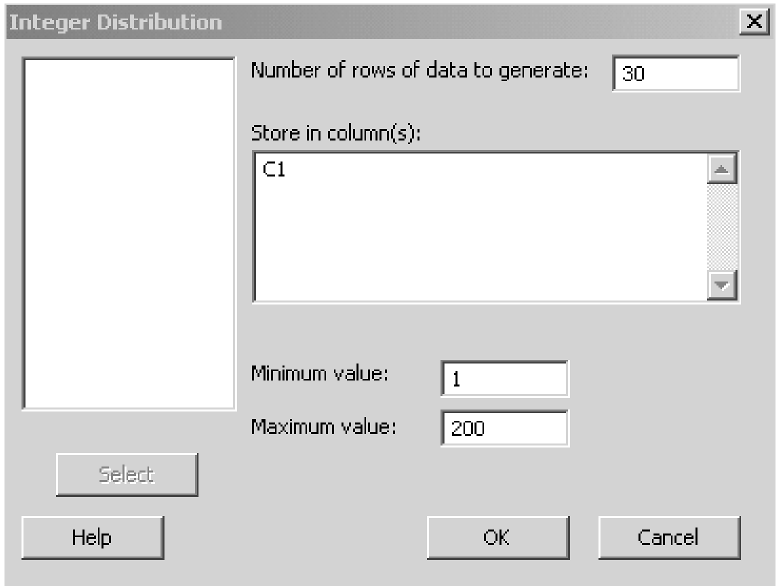
Since the calculated Z is greater than the critical Z (2.12 > 1.96), reject the null hypothesis. The forecast does not appear to be reasonable.

p-value: P($Z < -2.12$ or $Z > 2.12$) = 2 P($Z > 2.12$) = 2(.017) = .034, strong evidence against H_0

10. This problem can be used to illustrate how a random sample is selected with Minitab. In order to generate 30 random numbers from a population of 200 click the following menus:

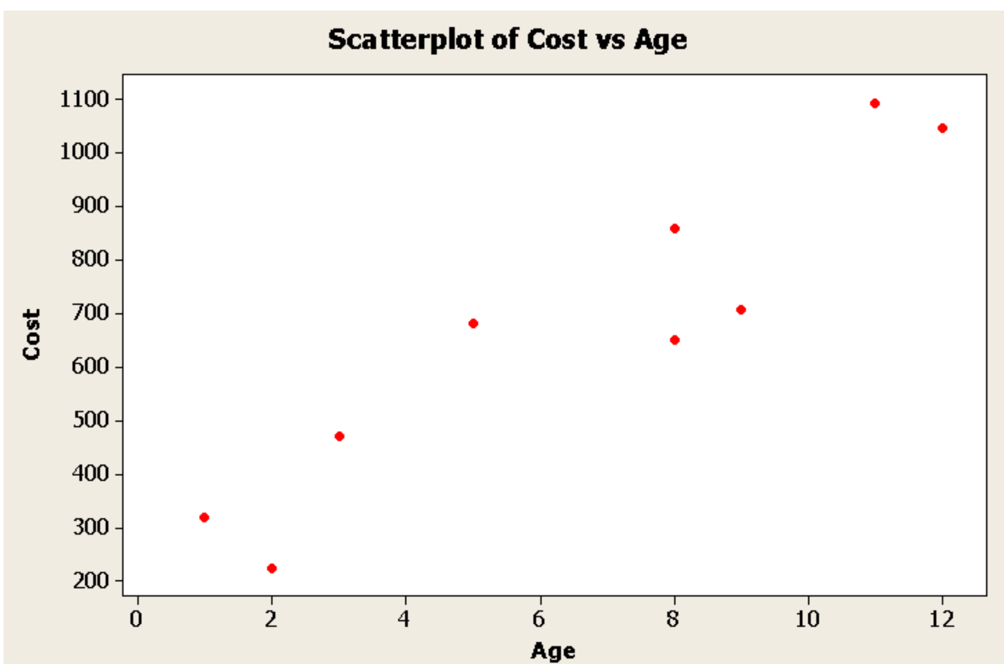
Calc>Random Data>Integer

The Integer Distribution dialog box shown in the figure below appears. The number of random digits desired, 30, is entered in the Number of rows of data to generate space. C1 is entered for Store in column(s) and 1 and 200 are entered as the Minimum and Maximum values. OK is clicked and the 30 random numbers appear in Column 1 of the worksheet.



The null hypothesis that the mean is still 2.9 is true since the actual mean of the population of data is 2.91 with a standard deviation of 1.608; however, a few students may reject the null hypothesis, committing a Type I error.

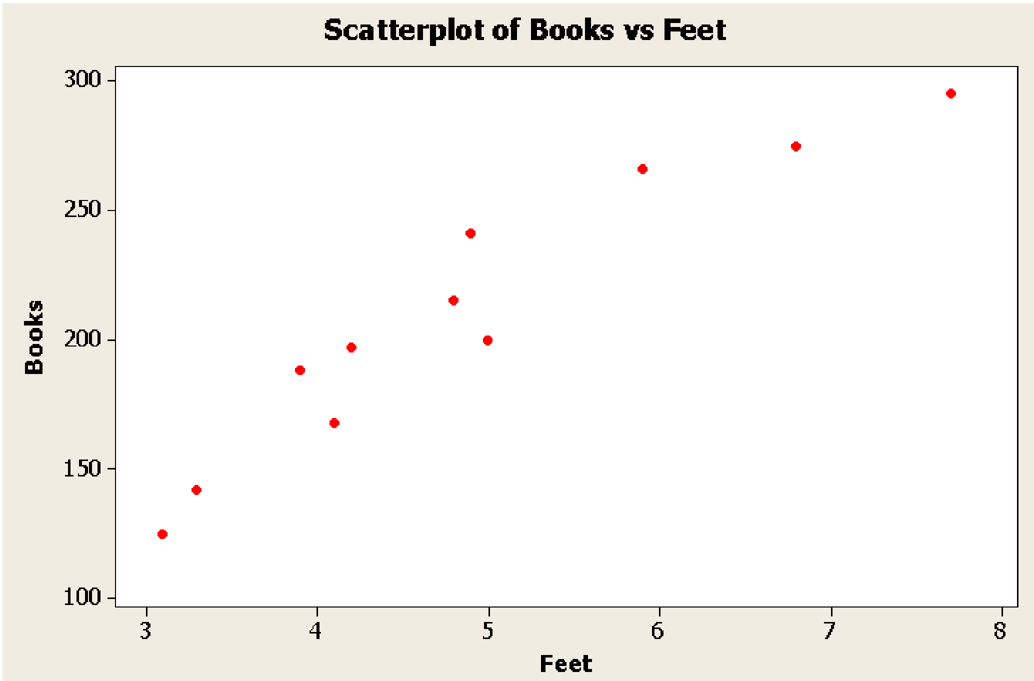
11. a.



b. Positive linear relationship

c. $\sum Y = 6058$ $\sum Y^2 = 4{,}799{,}724$ $\sum X = 59$
$\sum X^2 = 513$ $\sum XY = 48{,}665$ $r = .938$

12. a.



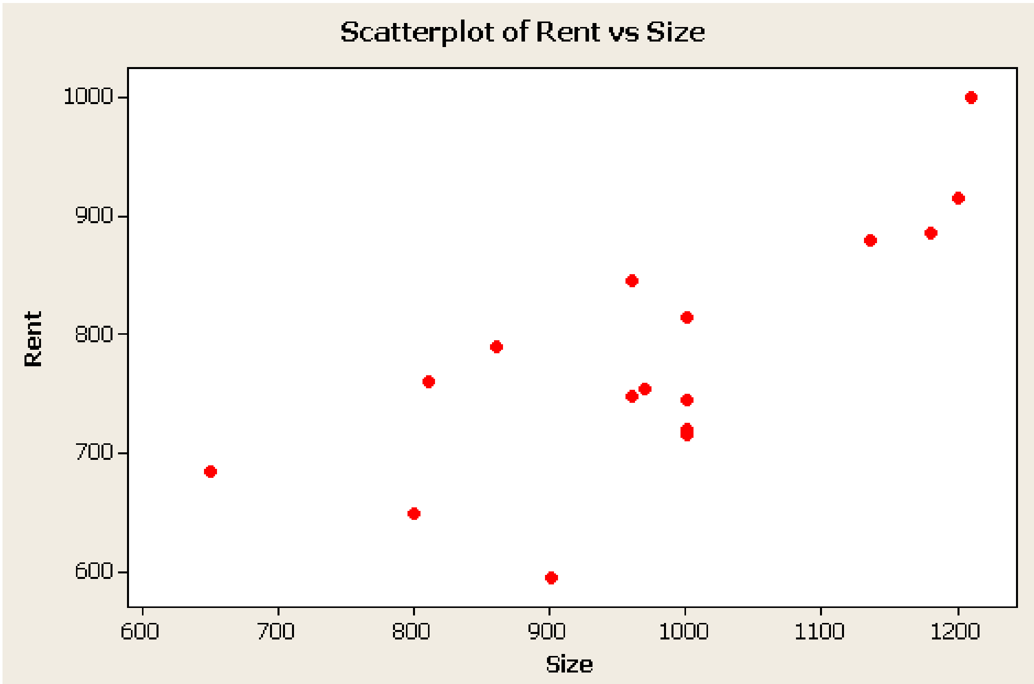
b. Positive linear relationship

c. $\sum Y = 2312$ $\sum Y^2 = 515{,}878$ $\sum X = 53.7$
$\sum X^2 = 282.55$ $\sum XY = 12{,}029.3$ $r = .95$

$\hat{Y} = 32.5 + 36.4X$
$\hat{Y} = 32.5 + 36.4(5.2) = 222$

13. This is a good population for showing how random samples are taken. If three-digit random numbers are generated from Minitab as demonstrated in Problem 10, the selected items for the sample can be easily found. In this population, $\rho = 0.06$ so most students will get a sample correlation coefficient *r* close to 0. The least squares line will, in most cases, have a slope coefficient close to 0, and students will not be able to reject the null hypothesis H_0: $\beta_1 = 0$ (or, equivalently, $\rho = 0$) if they carry out the hypothesis test.

14. a.



b. Rent = 275.5 + .518 Size

c. Slope coefficient = .518 $\Rightarrow$ Increase of $.518/month for each additional square foot of space.

d. Size = 750 $\Rightarrow$ Rent = 275.5 + .518(750) = $664/month

15. $n = 175$, $\bar{X}=45.2,\ S=10.3$

Point estimate: $\bar{X}=45.2$

98% confidence interval: $1-\alpha = .98 \Rightarrow Z = 2.33$

$\bar{X}\pm 2.33\left(S/\sqrt{n}\right)=45.2\pm 2.33\left(10.3/\sqrt{175}\right)=45.2\pm 1.8$ $\Rightarrow$ (43.4, 47.0)

Hypothesis test:

$H_0: \mu=44$

$H_1: \mu\neq 44$ two-sided test, $\alpha = .02$, critical value: $|Z|= 2.33$

Test statistic: $Z=\dfrac{\bar{X}-44}{S/\sqrt{n}}=\dfrac{45.2-44}{10.3/\sqrt{175}}=1.54$

Since $|Z| = 1.54 < 2.33$, do not reject H_0 at the 2% level.

As expected, the results of the hypothesis test are consistent with the confidence interval for μ; μ = 44 is not ruled out by either procedure.

16. a. $H_0: \mu = 63{,}700$
$H_1: \mu > 63{,}700$

b. $H_0: \mu = 4.3$
$H_1: \mu \neq 4.3$

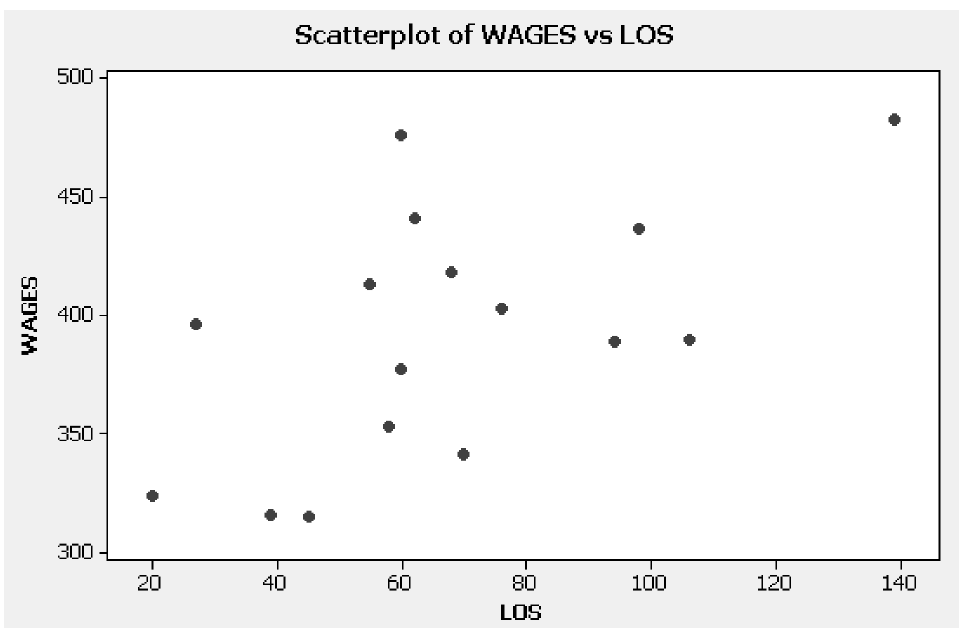
c. $H_0: \mu = 1300$
$H_1: \mu < 1300$

17. Large sample 95% confidence interval for mean monthly return μ:

$$-1.10 \pm 1.96 \frac{5.99}{\sqrt{39}} = -1.10 \pm 1.88 \Rightarrow (-2.98, .78)$$

μ = .94 (%) is not a realistic value for mean monthly return of client's account since it falls outside the 95% confidence interval. Client may have a case.

18. a.



b. r = .581, positive linear association between wages and length of service. Other variables affecting wages may be size of bank and previous experience.

c. $WAGES = 324.3 + 1.006\ LOS$
$WAGES = 324.3 + 1.006\ (80) = 405$

CASE 2-1: ALCAM ELECTRONICS

In our consulting work, business people sometimes tell us that business schools teach a risk-taking attitude that is too conservative. This is often reflected, we are told, in students choosing too low a significance level: such a choice requires extreme evidence to move one from the status quo. This case can be used to generate a discussion on this point as David chooses $\alpha = .01$ and ends up "accepting" the null hypothesis that the mean lifetime is 5000 hours.

Alice's point is valid: the company may be put in a bad position if it insists on very dramatic evidence before abandoning the notion that its components last 5000 hours. In fact, the indifference α (*p*-value) is about .0375; at any higher level the null hypothesis of 5000 hours is rejected.

CASE 2-2: MR. TUX

In this case, John Mosby tries some primitive ways of forecasting his monthly sales. The things he tries make some sort of sense, at least for a first cut, given that he has had no formal training in forecasting methods. Students should have no trouble finding flaws in his efforts, such as:

1. The mean value for each year, if projected into the future, is of little value since month-to-month variability is missing.
2. His free-hand method of fitting a regression line through his data can be improved upon using the least squares method, a technique now found on inexpensive hand calculators. The large standard deviation for his monthly data suggests considerable month-to-month variability and, perhaps, a strong seasonal effect, a factor not accounted for when the values for a year are averaged.

Both the hand-fit regression line and John's interest in dealing with the monthly seasonal factor suggest techniques to be studied in later chapters. His efforts also point out the value of learning about well-established formal forecasting methods rather than relying on intuition and very simple methods in the absence of knowledge about forecasting. We hope students will begin to appreciate the value of formal forecasting methods after learning about John's initial efforts.

CASE 2-3: ALOMEGA FOOD STORES

Julie’s initial look at her data using regression analysis is a good start. She found that the r-squared value of 36% is not very high. Using more predictor variables, along with examining their significance in the equation, seems like a good next step. The case suggests that other techniques may prove even more valuable, techniques to be discussed in the chapters that follow.

Examining the residuals of her equation might prove useful. About how large are these errors? Are forecast errors in this range acceptable to her? Do the residuals seem to remain in the same range over time, or do they increase over time? Are a string of negative residuals followed by a string of positive residuals or vice versa? These questions involve a deeper understanding of forecasting using historical values and these matters will be discussed more fully in later chapters.