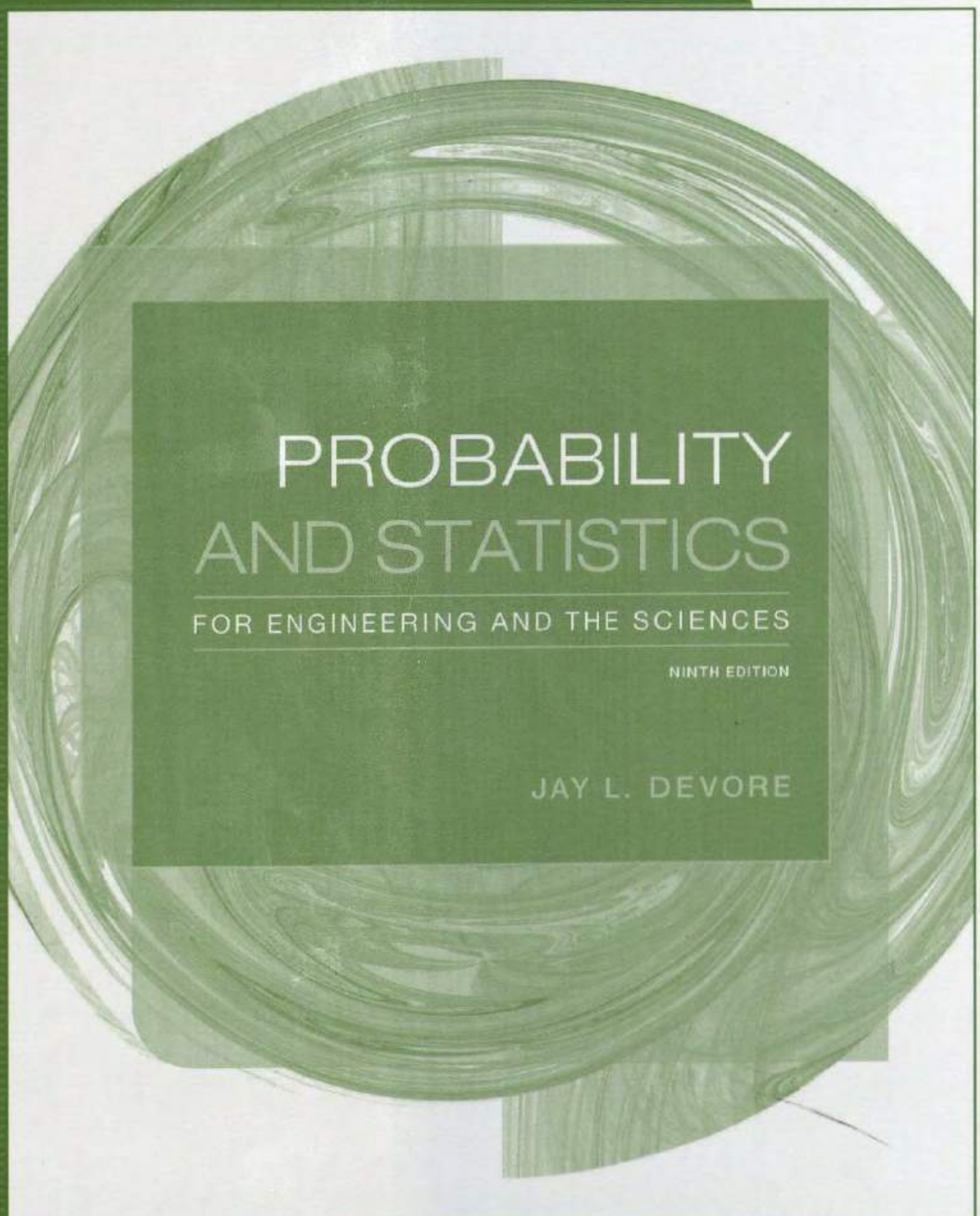
Student Solutions Manual



MATT CARLTON

Student Solutions Manual

Probability and Statistics for Engineering and the Sciences

NINTH EDITION

Jay L. Devore
California Polytechnic State University

Prepared by

Matt Carlton
California Polytechnic State University

CENGAGE Learning

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

CENGAGE Learning

For product information and technology assistance, contact us at **Cengage Learning Customer & Sales Support, 1-800-354-9706**.

For permission to use material from this text or product, submit all requests online at **www.cengage.com/permissions**
Further permissions questions can be emailed to **permissionrequest@cengage.com**.

ISBN: 978-1-305-26059-7

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at: **www.cengage.com/global**.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**.

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**.

Printed in the United States of America
Print Number: 01 Print Year: 2014

CONTENTS

CENGAGE
Learning

Printed in the United States of America
Print Number: 01 Print Year: 2014

CHAPTER 1

Section 1.1

1.

a. *Los Angeles Times, Oberlin Tribune, Gainesville Sun, Washington Post*

b. Duke Energy, Clorox, Seagate, Neiman Marcus

c. Vince Correa, Catherine Miller, Michael Cutler, Ken Lee

d. 2.97, 3.56, 2.20, 2.97

3.

a. How likely is it that more than half of the sampled computers will need or have needed warranty service? What is the expected number among the 100 that need warranty service? How likely is it that the number needing warranty service will exceed the expected number by more than 10?

b. Suppose that 15 of the 100 sampled needed warranty service. How confident can we be that the proportion of *all* such computers needing warranty service is between .08 and .22? Does the sample provide compelling evidence for concluding that more than 10% of all such computers need warranty service?

5.

a. No. All students taking a large statistics course who participate in an SI program of this sort.

b. The advantage to randomly allocating students to the two groups is that the two groups should then be fairly comparable before the study. If the two groups perform differently in the class, we might attribute this to the treatments (SI and control). If it were left to students to choose, stronger or more dedicated students might gravitate toward SI, confounding the results.

c. If all students were put in the treatment group, there would be no firm basis for assessing the effectiveness of SI (nothing to which the SI scores could reasonably be compared).

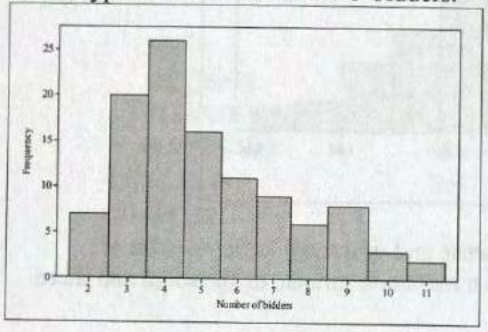
7. One could generate a simple random sample of all single-family homes in the city, or a stratified random sample by taking a simple random sample from each of the 10 district neighborhoods. From each of the selected homes, values of all desired variables would be determined. This would be an enumerative study because there exists a finite, identifiable population of objects from which to sample.

17. The sample size for this data set is $n = 7 + 20 + 26 + \ldots + 3 + 2 = 108$.

a. "At most five bidders" means 2, 3, 4, or 5 bidders. The proportion of contracts that involved at most 5 bidders is $(7 + 20 + 26 + 16)/108 = 69/108 = .639$.
Similarly, the proportion of contracts that involved at least 5 bidders (5 through 11) is equal to $(16 + 11 + 9 + 6 + 8 + 3 + 2)/108 = 55/108 = .509$.

b. The number of contracts with between 5 and 10 bidders, inclusive, is $16 + 11 + 9 + 6 + 8 + 3 = 53$, so the proportion is $53/108 = .491$. "Strictly" between 5 and 10 means 6, 7, 8, or 9 bidders, for a proportion equal to $(11 + 9 + 6 + 8)/108 = 34/108 = .315$.

c. The distribution of number of bidders is positively skewed, ranging from 2 to 11 bidders, with a typical value of around 4-5 bidders.

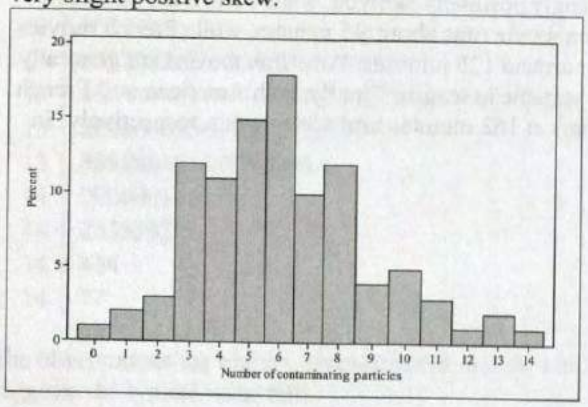


19.

a. From this frequency distribution, the proportion of wafers that contained at least one particle is $(100-1)/100 = .99$, or 99%. Note that it is much easier to subtract 1 (which is the number of wafers that contain 0 particles) from 100 than it would be to add all the frequencies for 1, 2, 3,… particles. In a similar fashion, the proportion containing at least 5 particles is $(100 - 1-2-3-12-11)/100 = 71/100 = .71$, or, 71%.

b. The proportion containing between 5 and 10 particles is $(15+18+10+12+4+5)/100 = 64/100 = .64$, or 64%. The proportion that contain strictly between 5 and 10 (meaning strictly *more* than 5 and strictly *less* than 10) is $(18+10+12+4)/100 = 44/100 = .44$, or 44%.

c. The following histogram was constructed using Minitab. The histogram is *almost* symmetric and unimodal; however, the distribution has a few smaller modes and has a very slight positive skew.



CHAPTER 6

Section 6.1

1.

a. We use the sample mean, $\overline{x}$, to estimate the population mean μ. $\hat{\mu} = \overline{x} = \frac{\Sigma x_i}{n} = \frac{219.80}{27} = 8.1407$.

b. We use the sample median, $\tilde{x} = 7.7$ (the middle observation when arranged in ascending order).

c. We use the sample standard deviation, $s = \sqrt{s^2} = \sqrt{\frac{1860.94 - \frac{(219.8)^2}{27}}{26}} = 1.660$.

d. With "success" = observation greater than 10, x = # of successes = 4, and $\hat{p} = \frac{x}{n} = \frac{4}{27} = .1481$.

e. We use the sample (std dev)/(mean), or $\frac{s}{\overline{x}} = \frac{1.660}{8.1407} = .2039$.

3.

a. We use the sample mean, $\overline{x} = 1.3481$.

b. Because we assume normality, the mean = median, so we also use the sample mean $\overline{x} = 1.3481$. We could also easily use the sample median.

c. We use the 90th percentile of the sample: $\hat{\mu} + (1.28)\hat{\sigma} = \overline{x} + 1.28s = 1.3481 + (1.28)(.3385) = 1.7814$.

d. Since we can assume normality,
$$P(X < 1.5) \approx P\left(Z < \frac{1.5 - \overline{x}}{s}\right) = P\left(Z < \frac{1.5 - 1.3481}{.3385}\right) = P(Z < .45) = .6736.$$

e. The estimated standard error of $\overline{x} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{.3385}{\sqrt{16}} = .0846$.

5. Let θ = the total audited value. Three potential estimators of θ are $\hat{\theta}_1 = N\overline{X}$, $\hat{\theta}_2 = T - N\overline{D}$, and $\hat{\theta}_3 = T \cdot \frac{\overline{X}}{\overline{Y}}$.

From the data, $\overline{y} = 374.6$, $\overline{x} = 340.6$, and $\overline{d} = 34.0$. Knowing N = 5,000 and T = 1,761,300, the three corresponding estimates are $\hat{\theta}_1 = (5{,}000)(340.6) = 1{,}703{,}000$, $\hat{\theta}_2 = 1{,}761{,}300 - (5{,}000)(34.0) = 1{,}591{,}300$, and $\hat{\theta}_3 = 1{,}761{,}300\left(\frac{340.6}{374.6}\right) = 1{,}601{,}438.281$.

e. First, we need to figure out s^2 based on the information we have been given: s^2 = MSE = SSE/df = .29/5 = .058. Then, the 95% PI is $21.36 \pm 2.571\sqrt{.058+(.1141)^2} = 21.36 \pm 0.685 = (20.675, 22.045)$.

75.

a. To test H_0: $\beta_1 = \beta_2 = 0$ vs. H_a: either β_1 or $\beta_2 \neq 0$, first find R^2: SST = $\Sigma y^2 - (\Sigma y)^2 / n = 264.5 \Rightarrow R^2 =$ 1 – SSE/SST = 1 – 26.98/264.5 = .898. Next, $f = \dfrac{.898/2}{(1-.898)/(10-2-1)} = 30.8$, which at df = (2,7) corresponds to a *P*-value of ≈ 0. Thus, H_0 is rejected at significance level .01 and the quadratic model is judged useful.

b. The hypotheses are H_0: $\beta_2 = 0$ vs. H_a: $\beta_2 \neq 0$. The test statistic value is t = (–2.3621 – 0)/.3073 = –7.69, and at 7 df the *P*-value is $2P(T \geq |-7.69|) \approx 0$. So, H_0 is rejected at level .001. The quadratic predictor should not be eliminated.

c. x = 1 here, $\hat{\mu}_{Y\cdot 1} = \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(1)^2 = 45.96$, and $t_{.025,7}$ = 1.895, giving the CI $45.96 \pm (1.895)(1.031) = (44.01, 47.91)$.

77.

a. The hypotheses are H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ versus H_a: at least one $\beta_i \neq 0$. From the output, the *F*-statistic is f = 4.06 with a *P*-value of .029. Thus, at the .05 level we reject H_0 and conclude that at least one of the explanatory variables is a significant predictor of power.

b. Yes, a model with R^2 = .834 would appear to be useful. A formal model utility test can be performed: $f = \dfrac{R^2/k}{(1-R^2)/[n-(k+1)]} = \dfrac{.834/3}{(1-.834)/[16-4]} = 20.1$, which is much greater than $F_{.05,3,12}$ = 3.49. Thus, the mode including $\{x_3, x_4, x_3x_4\}$ is useful.

We <u>cannot</u> use an *F* test to compare this model with the first-order model in (a), because neither model is a "subset" of the other. Compare $\{x_1, x_2, x_3, x_4\}$ to $\{x_3, x_4, x_3x_4\}$.

c. The hypotheses are H_0: $\beta_5 = \ldots = \beta_{10} = 0$ versus H_a: at least one of these $\beta_i \neq 0$, where β_5 through β_{10} are the coefficients for the six interaction terms. The "partial *F* test" statistic is $f = \dfrac{(\text{SSE}_l - \text{SSE}_k)/(k-l)}{\text{SSE}_k/[n-(k+1)]} = \dfrac{(R_k^2 - R_l^2)/(k-l)}{(1-R_k^2)/[n-(k+1)]} = \dfrac{(.960-.596)/(10-4)}{(1-.960)/[16-(10+1)]} = 7.58$, which is greater than $F_{.05,6,5}$ = 4.95. Hence, we reject H_0 at the .05 level and conclude that at least one of the interaction terms is a statistically significant predictor of power, in the presence of the first-order terms.

79. There are obviously several reasonable choices in each case. In **a**, the model with 6 carriers is a defensible choice on all three grounds, as are those with 7 and 8 carriers. The models with 7, 8, or 9 carriers in **b** merit serious consideration. These models merit consideration because R_k^2, MSE_k, and C_k meet the variable selection criteria given in Section 13.5.

81.

a. The relevant hypotheses are $H_0: \beta_1 = \ldots = \beta_5 = 0$ vs. H_a: at least one among $\beta_1, \ldots, \beta_5 \neq 0$. $f = \dfrac{.827/5}{.173/11} = 106.1 \geq F_{.05,5,111} \approx 2.29$, so *P*-value < .05. Hence, H_0 is rejected in favor of the conclusion that there is a useful linear relationship between *Y* and at least one of the predictors.