TT TEXTBOOK TESTBANKS

Solutions Manual for Mathematical Statistics with Applications 8th Wackerly

SOLUTIONS MANUAL SAMPLE PREVIEW

PUBLISHER

Cengage

COPYRIGHT

2026

RESOURCE TYPE

Solutions Manual

ISBN

9798214013435

Solutions Manual for Mathematical Statistics with Applications 8th Edition by Wackerly, Chen

ISBN: 9798214013435

Solution and Answer Guide

DENNIS D. WACKERLY, WACKERLY, 8E, ©2027, 9798214013435;
CHAPTER 1: WHAT IS STATISTICS?

TABLE OF CONTENTS

END OF SECTION EXERCISE SOLUTIONS

1.1 Answers may vary

1.

a. Population of interest: all U.S. citizens who are members of Generation X (as defined by the study—commonly birth years ~1965–1980)
Inferential objective: Estimate a population proportion: the proportion of Gen X U.S. citizens interested in starting their own business.
How to collect a sample: Use a national survey with a probability-based sampling method (e.g., stratified random sampling by region, gender, etc.). Screen respondents to confirm Gen X and U.S. citizenship, then ask the interest question.

b. Population of interest: all healthy adults in the United States
Inferential objective: Estimate the population mean body temperature (and compare to 98.6°F informally or via a confidence interval).
How to collect a sample: Select a random sample of adults (possibly stratified by age/sex/region). Measure temperature using a standardized method (same type of thermometer, same body location, similar conditions/time of day), and exclude those who are ill.

c. Population of interest: all single-family dwelling units in the city
Inferential objective: Estimate the population's mean weekly water consumption.
How to collect a sample: Take a simple random sample (or stratified by neighborhood/size) of single-family addresses from utility records and use meter data for a defined week (or average across several weeks to reduce week-to-week variation).

d. Population of interest: all automobile tires manufactured by that company during the current production year
Inferential objective: Estimate the population proportion with unsafe tread.
How to collect a sample: Use quality-control sampling from production lots: randomly sample tires throughout the year (stratify by plant, line, tire model, month). Measure tread depth and classify "unsafe" by a defined standard.

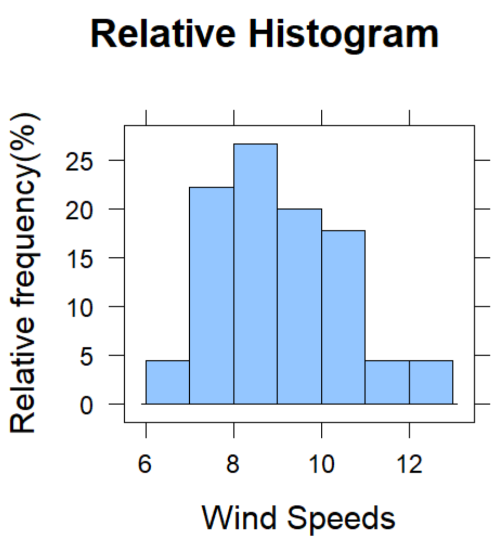
e. Population of interest: all adult residents of the state
Inferential objective: Determine whether the population proportion favoring unicameral legislature is greater than 0.50 (a majority). (This is a hypothesis test about a proportion.)
How to collect a sample: Conduct a statewide random poll (stratified by region/urban-rural, etc.). Use multiple contact methods (phone/text/online) and weight appropriately to match known demographics.

f. Population of interest: patients with the disease who become recurrence-free after treatment (often defined more precisely as patients meeting eligibility criteria in a target health system or in the U.S.).
Inferential objective: Estimate the mean (or typical) time to recurrence; in practice this is often a survival/time-to-event objective.
How to collect a sample: Use a cohort study: enroll a sample of eligible patients and follow them over time, recording recurrence times. This usually requires handling censoring (patients lost to follow-up or recurrence not observed by study end).

g. Population of interest: all transistors of that specific type (from the relevant production process/time period).
Inferential objective: Test whether the population mean lifetime is greater than 500 hours (hypothesis test on a mean).
How to collect a sample: Randomly sample transistors from production (possibly stratified by batch/line), then run a life test under standardized operating conditions until failure (or use accelerated life testing, with appropriate statistical adjustment).

2.

a.

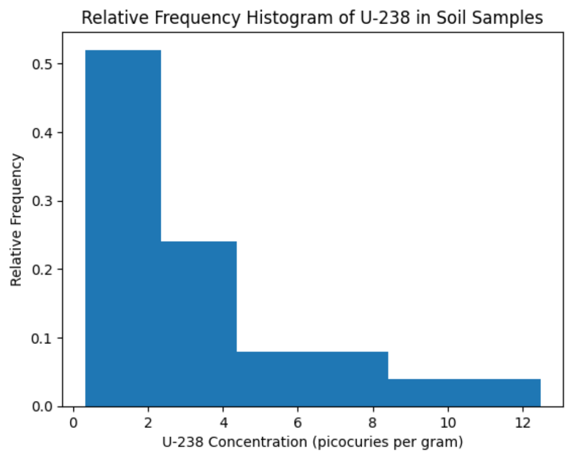


b. Yes, it is quite windy there.

c. 9/45, or approx. 20%

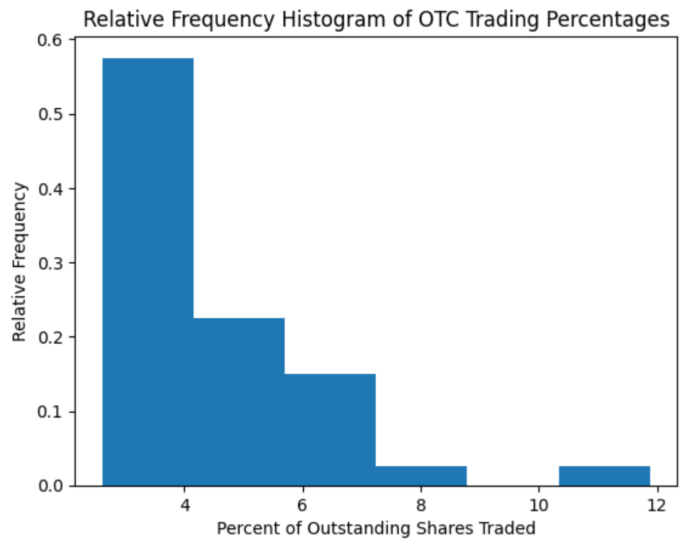
d. It is not especially windy in the overall sample.

3.



4.

a.



b. $18/40 = 45\%$

c. $29/40 = 72.5\%$

5.

a. The $2.45 - 2.65$ and $2.65 - 2.85$ GPA categories have the tallest bars, so these intervals are associated with the largest proportion of students.

b. The proportion in each category is: $6/30 = 0.20$. So, 20% of the students fall in each of these two categories.

c. The proportion of students that has GPAs less than 2.65 is $3/30 + 3/30 + 3/30 + 6/30 = 0.50$ or 50%.

6.

a. The modal category is 2 (quarts of milk). About 36% (9 people) of the 25 are in this category.

b. $0.2 + 0.12 + 0.04 = 0.36$

c. Note that 8% purchased 0 while 4% purchased 5. Thus, $1 - .08 - .04 = .88$ purchased between 1 and 4 quarts.

7.

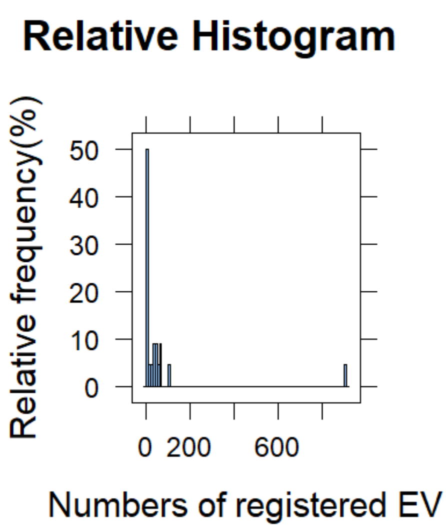
a. The histogram is bimodal.

b. Yes. The distribution has two peaks (2 modes).

c. The two peaks likely result from combining male and female height data, which have different average heights.

8.

a.



b. The data appears to have outliers. Two states in the West region have sample values much higher than the rest.

9.

a. $9.7 = 12 - 1\sigma$, $14.3 = 12 + 1\sigma$. About 68% lie within $\mu \pm 1\sigma$.

b. $7.4 = 12 - 2\sigma$, $16.6 = 12 + 2\sigma$. About 95% lie within $\mu \pm 2\sigma$.

c. From the empirical rule, 68% within $\mu \pm 1\sigma$ and an additional 13.5% fall between $\mu + 1\sigma$ and $\mu + 2\sigma$. So, $0.68 + 0.135 = 0.815$ or 81.5%

d. $5.1 = 12 - 3\sigma$, $18.9 = 12 + 3\sigma$. About 99.7% lie within $\mu \pm 3\sigma$. So, $1 - 0.997 = 0.003$ or 0.3%.

10.

a. $14 - 17 = -3$.

b. Since 68% lie within 1 standard deviation of the mean, 32% should lie outside. By symmetry, 16% should lie below 1 standard deviation from the mean.

c. If normally distributed, approximately 16% of people would spend less than –3 hours on the internet. Since this doesn't make sense, the population is not normal.

11.

$\sum(y_i - \bar{y})^2 = \sum(y_i^2 - 2\bar{y}y_i + \bar{y}^2) = \sum y_i^2 - 2\bar{y}\sum y_i + \sum \bar{y}^2$

So that

$$\sum(y_i - \bar{y})^2 = \sum y_i^2 - 2\bar{y}(n\bar{y}) + n\bar{y}^2 = \sum y_i^2 - n\bar{y}^2$$

Writing $\bar{y} = \frac{1}{n}\sum y_i$, we get

$$n\bar{y}^2 = n\left(\frac{1}{n}\sum y_i\right)^2 = \frac{1}{n}\left(\sum y_i\right)^2$$

So that

$$\sum(y_i - \bar{y})^2 = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2$$

Dividing both sides by $n-1$ and summing from 1 to n, we get

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]$$

12.

Using the data, $\sum_{i=1}^{6} y_i = 14$ and $\sum_{i=1}^{6} y_i^2 = 40$. So, $s^2 = (40 - 14^2/6)/5 = 1.47$, $s = 1.21$.

13.

a. $\sum y_i = 407.0$, $\sum y_i^2 = 3773.28$, $n = 45$. So, $\bar{y} = \frac{407.0}{45} = 9.004$. $s^2 = \frac{1}{44}\left[3773.28 - \frac{(407.0)^2}{45}\right]$

$= 2.095 \Rightarrow s = \sqrt{2.095} = 1.447$.

b.

k	interval	frequency	Exp. frequency
1	(7.597, 10.492)	31	30.6
2	(6.149, 11.939)	42	42.75
3	(4.702, 13.387)	45	44.87

14.

a. $\sum_{i=1}^{25} y_i = 80.63$ and $\sum_{i=1}^{25} y_i^2 = 500.7459$, we have $\bar{y} = 3.23$ and $s = 3.17$

b.

k	interval	frequency	Exp. frequency
1	0.063, 6.397	21	17
2	−3.104, 9.564	23	23.75
3	−6.271, 12.731	25	25

15.

a. $\sum y_i = 175.48$, $\sum y_i^2 = 906.4118$, $n = 40$. So, $\bar{y} = \dfrac{175.48}{40} = 4.387$. s^2

$$= \frac{1}{39}\left[906.4118 - \frac{(175.48)^2}{40}\right] = 3.504 \Rightarrow s = \sqrt{3.504} = 1.871.$$

b.

k	interval	frequency	Exp. frequency
1	(2.516, 6.258)	35	27.2
2	(0.644, 8.130)	39	38
3	(−1.227, 10.001)	39	39.88

16.

a. Without the extreme value, $y = 4.19$ and $s = 1.44$.

b. These counts compare more favorably:

k	interval	frequency	Exp. frequency
1	2.75, 5.63	25	26.52
2	1.31, 7.07	36	37.05
3	−0.13, 8.51	39	39

17.

For 1.2: min = 6.0, max = 12.3 range = 6.3. So, range/4 = 6.3/4 = 1.575. $s = 1.447$, which is 0.128 lower than the estimate.
For 1.3: min = 0.32, max = 12.48 range = 12.16. So, range/4 = 12.16/4 = 3.04. $s = 3.167$, which is 0.127 higher than the estimate.
For 1.4: min = 2.61, max = 11.88 range = 9.27. So, range/4 = 9.27/4 = 2.3175. $s = 1.871$, which is 0.446 lower than the estimate.

18.

min = 200, max = 800 $\Rightarrow$ range = 600. From 17, $s \approx \dfrac{range}{4} = \dfrac{600}{4} = 150$.

19.

If chloroform amounts were normally distributed, then by the empirical rule, about 95% of the observations would fall within

$$\mu \pm 2\sigma = 34 \pm 2(53) = 34 \pm 106 = (-72,140).$$

Since chloroform concentration cannot be negative, the distribution of chloroform amounts cannot be normal.

20.

$15,500 = 14,200 + 1\sigma = \mu + 1\sigma$. About 68% of values lie within $\mu \pm 1\sigma$. So, 32% lie outside this interval. By symmetry, half of that or 16% lie above $\mu + 1\sigma = 15,500$.

21.

$20 = 22 - 1\sigma = \mu - 1\sigma$. About 68% of values lie within $\mu \pm 1\sigma$. So, 32% lie outside. By symmetry, 16% lie below $\mu - 1\sigma$. Thus, $1 - 0.16 = 0.84$ or 84% should have weight gains in excess of 20 pounds. Hence, the sample information does not contradict the manufacturer's claim.

22.

Probability is the measurement of uncertainty for statistical inference.

23.

No. You can only say that 40% is an estimate of the true population of smokers, since this value comes from a sample of 50 patients, and is subject to sampling error. Different random samples would likely produce different results.

24.

$$\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y} = \sum_{i=1}^{n} y_i - n\bar{y} = \sum_{i=1}^{n} y_i - n \cdot \frac{1}{n}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i = 0$$

25.

a. $30 = 25 + 1\sigma = \mu + 1\sigma$. So, by the empirical rule, 16% lie above $\mu + 1\sigma$.

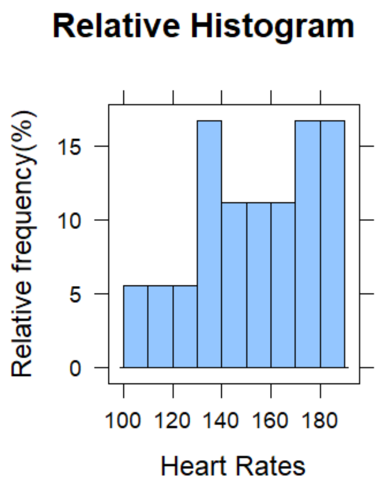
b. $15 = \mu - 2\sigma, 45 = \mu + 4\sigma$. By the empirical rule, 95% of the data fall within $\mu + 2\sigma$. So, by symmetry, 2.5% fall below $\mu - 2\sigma$. Also, about 0.15% falls beyond $\mu + 3\sigma$. So, beyond $\mu + 4\sigma$ is approximately 0.

c. $90 = 25 + 13\sigma$. That's extremely far from the mean. So, the probability is approximately 0. Thus, it is very unlikely that a commercial will last longer than 1.5 minutes or 90 seconds.

26.

a. $(188.22 - 105.6)/4 = 20.66$

b.



c. Using R, we have that $\bar{y} = 152.76$ and $s = 25.16$.

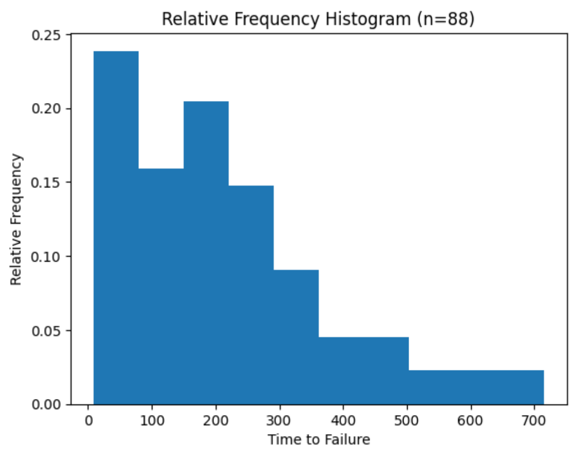
d.

k	interval	frequency	Exp. frequency
1	127.6 177.92	12	12.24
2	102.44 203.08	16	17.1
3	77.28 228.24	18	18

27.

a. Range = max – min = 716 – 8 = 708. So, $s \approx \frac{Range}{4} = \frac{708}{4} = 177$

b.



c. $\bar{y} = 210.80$, $s = 162.17$

d.

k	interval	frequency	Exp. frequency
1	(48.62, 372.97)	63	59.84
2	(-113.55, 535.14)	82	83.6
3	(-275.72, 697.31)	87	87.74

28.

For Ex. $1.12, 3/1.21 = 2.48$. For Ex. 1.26, $82.62/25.16 = 3.28$. For Ex. 1.27, $708/162.17 = 4.37$. The ratio increases as the sample size increases.

29.

$\bar{y} = 72, s = 8, n = 340$

$64 = 72 - 1s, 80 = 72 + 1s$. So, by the empirical rule, $0.68(340) = 231$ in the interval 64 to 80.

$56 = 72 - 2s, 88 = 72 + 2s$. So, by the empirical rule, $0.95(340) = 323$ in the interval 56 to 88.

30.

13mg/L is 1 standard deviation below the mean, so 16%.

31.

a. $2.98 = 3.00 - 2(0.01) = \mu - 2\sigma$, $3.02 = 3.00 + 2(0.01) = \mu + 2\sigma$
So, bearings that fail are those outside the interval: $\mu \pm 2\sigma$. By the empirical rule, about 95% of observations fall within $\mu + 2\sigma$. So, 5% fall outside the interval.

b. The bearing diameters are approximately normally distributed.

32.

If $\mu = 10$ and ▯ $= 1.2$, we expect 34% to be between 10 and $10 + 1.2 = 11.2$. Also, approximately $95\%/2 = 47.5\%$ will lie between 10 and 12.4. So, approximately $47.5\% - 34\% = 13.5\%$ should lie between 11 and 13.

33.

$40 = 60 - 2(10) = \mu - 2\sigma$. By the empirical rule, about 95% of values are within $\mu \pm 2\sigma$. This leaves 5% outside this interval. By symmetry, the probability of a yield less than 40 is approximately 0.025 or 2.5%. This could raise suspicion of an abnormality. The assumptions are that daily yield is approximately normally distributed and the mean and variance are stable over time.

34.

Let m be the number of i such that $|y_i - \bar{y}| \geq ks$.

For each of the m observations outside the interval, we have

$(y_i - \bar{y})^2 \geq (ks)^2 = k^2s^2$. For the remaining $n-m$ observations, $(y_i - \bar{y})^2 \geq 0$. Therefore,

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 \geq m(k^2s^2) \Rightarrow s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \geq \frac{1}{n-1}m(k^2s^2)$$

Thus,

$$1 \geq \frac{mk^2}{n-1} \Rightarrow m \leq \frac{n-1}{k^2}$$

So, the number of measurements inside the interval is at least

$$n-m \geq n - \frac{n-1}{k^2} \Rightarrow \frac{n-m}{n} \geq 1 - \frac{n-1}{nk^2}$$

Now, since $\frac{n-1}{n} \leq 1$,

$$1 - \frac{n-1}{nk^2} \geq 1 - \frac{1}{k^2}$$

Thus, the fraction of measurements inside the interval $\bar{y} - ks$ to $\bar{y} + ks$ is at least

$$1 - \frac{1}{k^2}$$

35.

Using Chebyshev's inequality, we have that

$1 - \frac{1}{k^2} \geq 0.75 \Rightarrow k \geq 2$.

Then, for $k = 2$, with $\bar{y} = 5.5$ and $s = 2.5$, we have

$$\bar{y} \pm 2s = 5.5 \pm 2(2.5) = 5.5 \pm 5$$

So, at least 75% of the days absent count must fall in the interval (0.5, 10.5).

36.

The point 13 is $13 - 5.5 = 7.5$ units above the mean, or $7.5/2.5 = 3$ standard deviations above the mean. By Chebyshev's Inequality, at least $1 - 1/32 = 8/9$ will lie within 3 standard deviations of the mean. Thus, at most 1/9 of the values will exceed 13.

37.

a. $s \approx \frac{Range}{4} = \frac{172-108}{4} = 16$

b. $\bar{y} = 136.07, s = 17.10$

c. At least $1 - \frac{1}{k^2} \geq 0.75 \Rightarrow k = 2$.

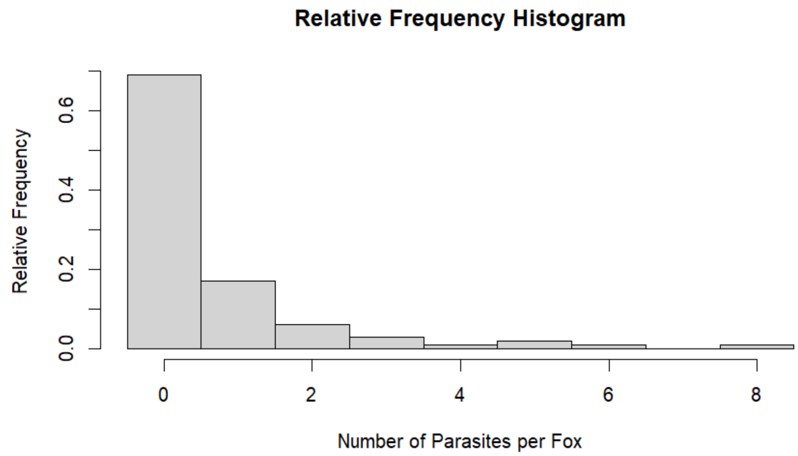
So,

$a = \bar{y} - 2s = 136.07 - 2(17.10) = 101.86, b = \bar{y} + 2s = 136.07 + 2(17.10) = 170.27$

d. The number of readings in the interval (101.86, 170.27) is 14. And $\frac{14}{15} = 0.9333$ or 93.3%. So, yes, the inequality worked.

38.

a.



b. With $\sum_{i=1}^{100} y_i = 66$ and $\sum_{i=1}^{100} y_i^2 = 234$ we have that $\bar{y} = 0.66$ and $s = 1.39$.

c. Within 2 standard deviations: 95, within 3 standard deviations: 96. The calculations agree with Chebyshev's Inequality.

39.

If the lead content readings were normally distributed, then by the empirical rule about 95% of the values would lie within the interval
$\mu \pm 2\sigma = 0.027 \pm 2(0.12) = (-0.213, 0.267)$.
Since this interval includes negative numbers and lead concentrations must be nonnegative, the distribution of lead readings cannot be normal.

40.

By Chebyshev's inequality, at least ¾=75% lie between $(0,140)$, at least $8/9$ lie between $(0,193)$, and at least 15/16 lie between $(0,246)$. Note: the lower bounds are all truncated to 0 since the measurement cannot be negative.