

Unit 1 Question Bank

Instructions

For the initial assignment of 10 problems please use the following parameters to select your 10:

Group	Number of Problems
1	1
2	2
3	2
4	2 or 3
5	2 or 3

Any questions done beyond those 10 count as 0.5 points on the associated unit exam.

Group 1

1. A field biologist samples 50 butterflies from a meadow. What is considered the sample, and what is considered the population?
2. A botanist counts flower colors in a meadow (red, yellow, white). What type of graph best summarizes this?
3. The heart rates (bpm) of frogs are measured as 40, 42, 42, 45, 90. Which measure of center best represents the “typical” value?
4. A dataset of lizard tail lengths has a minimum of 5 cm, a median of 10 cm, and a maximum of 15 cm. What is the range?
5. A marine biologist measures shell lengths in mussels. Which measure (mean, median, mode) changes most if one unusually large mussel is included?
6. An economist predicts a 60% chance that stock A will perform poorly and a 25% chance that stock B will perform poorly. There is also a 16% chance that both stocks will perform poorly. What is the probability that stock A performs poorly given that stock B performs poorly?

Group 2

7. In a study of 100 salmon, researchers measure the number of parasites on each fish. What measurement could they use to represent the trend in parasite counts. Would it be considered a parameter or a statistic?
8. A researcher records the weights of 40 mice. Which graph type (histogram, boxplot, bar graph) should they use to look at the frequency distribution of weights?
9. A dataset of seed counts from 12 plants has one very high outlier. Which measure of center is more resistant to that outlier?
10. An ecologist measures offspring counts in 10 bird pairs. If the minimum is 1 and the maximum is 20, what does the range tell us about variability? Which measure of variability would let us look at what happens between the minimum and maximum rather than just quantify the difference between them?
11. Which measure of position would be most helpful in identifying unusually values in a dataset of size $n = 50$?
12. An epidemiologist surveys patients with and without a rare disease, asking about diet history. What type of observational study is this?
13. A scatterplot shows fish length vs. parasite load. What correlation direction would you expect if longer fish carry more parasites?
14. If an event has probability 0.2, what is the probability that it doesn't occur? What is the term for the probability that an event does not occur?
15. A wildlife health team reports that deer using a mineral lick near a campsite show higher rates of hoof lesions than deer in remote meadows. They suggest the lick "causes" lesions. Can they assess causality from this study? If yes, explain why. If not, suggest how they could change the study to assess better causality.
16. Label the following as discrete or continuous: Number of eggs per nest, hatchling weights, nest area, bird color.

Group 3

17. Researchers sample 200 bats to study prevalence of white-nose syndrome. Define the statistic they would calculate for prevalence and the population parameter it estimates.
18. Daily milk production (liters) of 30 cows is recorded. Given the following sample statistics, construct a boxplot for the data:

\bar{x}	30
s_x	5
Min	5
Q_1	12
Med	20
Q_3	32
Max	42

19. A dataset of plant heights has $\bar{x} = 12$ cm and Median = 19 cm. Is the distribution left-skewed, right-skewed, or symmetric? Which statistic (mean or median) indicates the skew.
20. Shelter cat weights have $\bar{x} = 4.5$ kg and $s = 6.2$ kg. What does the large standard deviation relative to the mean suggest about variability and possible shape (e.g., presence of outliers or right-skew)? Give a one-sentence justification.
21. A fruit fly lifespan dataset has Min = 3, $Q_1 = 10$, Median = 15, $Q_3 = 18$, Max = 60. Use the IQR rule to determine whether 60 is an outlier.
22. A diabetes cohort study follows 1000 patients for 10 years. Explain what makes this a cohort study, one advantage of the study, and one disadvantage.
23. If a scatterplot is showing a strong correlation between two variables that cannot reasonably be linked scientifically, how could you distinguish causation from correlation?
24. In a tropical health survey, 200 children are tested for malaria parasites and screened for mosquito bites.

	Many Bites	Few Bites	Total
Parasite +	48	12	60
Parasite –	72	68	140
Total	120	80	200

Compute $P(\text{Parasite}+)$, $P(\text{Many Bites})$, and $P(\text{Parasite}+ \cap \text{Many Bites})$.

25. A retrospective case-control study investigates a suspected link between a new silage preservative and mastitis in dairy cows. Define the "cases" and "controls" in this study.
26. Roll a fair die. Let $A = \text{even}$, $B = \text{greater than 4}$. Compute $P(A \cap B)$.

Group 4

27. A conservation team samples fish only from lakes near highways. What type of sampling method is this, and what bias might it introduce?
28. Bird species counts from 50 wetlands are recorded. One researcher decides to show the counts in a boxplot, while another shows them in a histogram. How could these two graphics provide the same inference? How could they provide difference inference?
29. Egg counts per plant for two species are summarized as:
 Species A: $\bar{x} = 50$, Median = 49
 Species B: $\bar{x} = 120$, Median = 70
 For each species, state whether the *mean* or *median* is the better measure of center and briefly justify.
30. Blood glucose in rats has $\mu = 110$, $\sigma = 15$. Use the empirical rule to estimate the proportion between 80–140. Could we still use the empirical rule if Median = 40? Why or why not?
31. Which measures can be used for detecting outliers: range, standard deviation, IQR, or z -score? If there's more than one, which one is the best at detecting outliers?
32. A mask mandate study measures case counts before and after policy. Why is this observational, not experimental? What biases might arise?
33. In 20 field plots, fertilizer use (kg/ha) and plant height (cm) have correlation $r = 0.82$. Explain why this strong correlation does *not* prove that fertilizer *causes* increased height.
34. In a greenhouse, 150 tomato plants are examined for fungal infection and whiteflies.

	Whiteflies	No Whiteflies	Total
Fungus +	30	20	50
Fungus −	45	55	100
Total	75	75	150

Compute $P(\text{Fungus+} \cup \text{Whiteflies})$.

35. A hospital tracks daily ER patient arrivals for a year. Propose how to summarize the data distribution and explain what extreme outliers may mean.
36. Two events A and B have $P(A) = 0.4$, $P(B) = 0.5$, $P(A \cap B) = 0.1$. Compute $P(A \cup B)$ and discuss independence.

Group 5

Reference the data set of patient biometrics for problems 40 - 43:

Age	Sex	Chest Pain Type	Resting Blood Pressure	Serum Cholesterol
41	M	1	110	235
62	M	2	130	231
54	M	0	122	286
52	F	2	136	196
40	M	0	110	167
64	M	0	128	263
52	M	0	128	204
61	F	0	145	307
65	F	2	160	360
44	M	2	120	226

37. Calculate the mean and standard deviation of male and female patient serum cholesterol.
38. Using z -scores, determine which patient has the higher serum cholesterol, a male at 255 or a female at 280.
39. What is the correlation coefficient value for chest pain type and age?
40. What is the average of the z -scores for age?
41. A national bird survey samples only urban parks. Which type of bias does this introduce?
(a) Nonresponse (b) Undercoverage (c) Voluntary response (d) Response bias
42. Given $P(A) = 0.5$, $P(A \cap B) = 0.2$, and $P(A \cup B) = 0.8$, are A and B independent events?
43. Assume that A and B are independent. Given $P(B) = 0.8$ and $P(A^c \cup B^c) = 0.04$, solve for the following: (a) $P(A)$ (b) $P(B|A)$
44. Assume that A and B are independent. Given $P(B) = 0.3$ and $P(A^c \cup B^c) = 0.01$, solve for the following: (a) $P(A|B)$ (b) $P(A^c \cap B^c)$

45. One of these histograms represents the age at death from natural causes, (heart attack, cancer, etc.), the other represents the age at death from accidents. Which represents the age at death from accidents? What's the justification behind your answer?

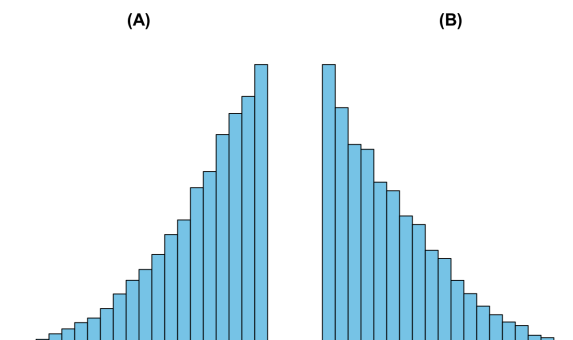


Figure 1: Histograms of Death Statistics

46. Given the data on gas prices in Kansas below:

Year	2004	2005	2006	2007	2008	2009	2010
Price	1.347	1.737	2.026	2.298	2.651	1.802	2.222

$$\bar{x} = 2.012$$

- (a) Calculate the sample variance

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)}$$

- (b) Attempt to calculate the sample variance without the squaring step

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{(n - 1)}$$

- (c) Attempt to calculate the sample variance by adding up all of the points:

$$\sum_{i=1}^n x_i$$

Then subtracting them from the sample mean (\bar{x}), squaring the result, and dividing by $n - 1$

47. The mean weight for adult male white-tailed deer is $\mu = 150$ lbs, with a standard deviation of $\sigma = 12$ lbs. The mean weight for adult female white-tailed deer is $\mu = 105$ lbs, with a standard deviation of $\sigma = 9$ lbs. Which deer is heavier relative to their sex, a male at 141 lbs or a female at 112 lbs?
48. Given an approximately symmetric histogram, which of the following are true?
- (a) The median is less than the mean.
 - (b) There is more than one mode.
 - (c) The median is equal to the mean.
 - (d) 95% of the data is within $\mu \pm 2\sigma$.
 - (e) The second quartile is approximately equal to the mode.
 - (f) The median is greater than the mean.
49. Below is a dataset from a study on the effect of mask-mandates during an outbreak of an unknown respiratory disease. Categorize the variables as quantitative or qualitative, and nominal, interval, ratio, or ordinal.

County	Cases (Pre-Mandate)	Cases (Post-Mandate)	% Change	Prop. Adhered
1	198	157	-0.207	0.873
2	232	196	-0.155	0.822
3	136	108	-0.206	0.933
4	210	170	-0.190	0.944
5	234	197	-0.158	0.873
6	913	775	-0.151	0.804
7	82	101	+0.232	0.882
8	249	210	-0.157	0.910
9	120	131	+0.092	0.820
10	227	193	-0.150	0.802

50. If you were to plot a histogram of case counts from the disease outbreak in question 49, what would the X and Y axis labels be?