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. *[Ch 1.1]* A field biologist samples 50 butterflies from a meadow. Which is the sample, and which is the population?

2. *[Ch 1.2]* A botanist counts flower colors in a meadow (red, yellow, white). What type of graph best summarizes this?

3. *[Ch 1.3]* The heart rates (bpm) of frogs are measured as 40, 42, 42, 45, 90. Which measure of center best represents the "typical" value?

4. *[Ch 1.4]* A dataset of lizard tail lengths has a minimum of 5 cm and a maximum of 15 cm. What is the range?

5. *[Ch 1.5]* A marine biologist measures shell lengths in mussels. Which measure (mean, median, mode) changes most if one unusually large mussel is included?

Group 2

6. *[Ch 1.1]* In a study of 100 salmon, researchers measure the number of parasites on each fish. Identify whether the parameter of interest is a population mean or a sample mean.

7. *[Ch 1.2]* A researcher records the weights of 40 mice. Which graph type (histogram, boxplot, bar graph) should they use?

8. *[Ch 1.3]* A dataset of seed counts from 12 plants has one very high outlier. Which measure of center (mean or median) is more resistant?

9. *[Ch 1.4]* An ecologist measures offspring counts in 10 bird pairs. What does the range tell us about variability?

10. *[Ch 1.5]* Which measure of position would be most helpful in identifying unusually long mussel shells in a dataset: quartiles, IQR, or mean?

11. *[Ch 2.1]* An epidemiologist surveys patients with and without a rare disease, asking about diet history. What type of observational study is this?

12. *[Ch 2.2]* A scatterplot shows fish length vs. parasite load. What correlation direction would you expect if longer fish carry more parasites?

13. *[Ch 2.3]* If an event has probability 0.2, what is the probability of its complement?

14. *[Ch 2.4]* 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. List two possible confounders and describe how you'd redesign the study to better assess causality.

15. *[Ch 3.1]* Which of these is discrete: number of eggs per nest, or hatchling weights?

Group 3

17. *[Ch 1.1]* Researchers sample 200 bats to study prevalence of white-nose syndrome. Define the statistic they calculate and the population parameter it estimates.

18. *[Ch 1.2]* Weekly milk production (liters) of 30 cows is recorded. Construct a frequency table with 5 classes and explain what patterns you might look for.

19. *[Ch 1.3]* A lab records heights of 20 plants. Discuss how the mean and median together describe the distribution, especially if skewed.

20. *[Ch 1.4]* Shelter cat weights have a very large standard deviation compared to the mean. What does this suggest about spread and distribution shape?

21. *[Ch 1.5]* Fruit fly lifespan data are summarized with a five-number summary. How can this help detect outliers?

22. *[Ch 2.1]* A diabetes cohort study follows 1,000 patients for 10 years. Explain what makes this a cohort study and one advantage.

23. *[Ch 2.2]* Fish length and parasite count are measured. How would you distinguish causation from correlation here?

24. *[Ch 2.3]* 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

(a) Compute $P(\text{Parasite+})$, $P(\text{Many Bites})$, and $P(\text{Parasite+} \cap \text{Many Bites})$. (b) Test whether "Parasite+" and "Many Bites" appear independent.

25. *[Ch 2.4]* A retrospective case–control study investigates a suspected link between a new silage preservative and mastitis in dairy cows. Define cases and controls, identify one bias risk and mitigation, and suggest one matching variable.

26. *[Ch 3.2]* Roll a fair die. Let A = even, B = greater than 4. Compute $P(A \cap B)$.

Group 4

28. *[Ch 1.1]* A conservation team samples fish from several lakes to estimate parasite prevalence. Discuss potential biases and suggest a better design.

29. *[Ch 1.2]* Bird species counts from 50 wetlands are recorded. Explain how using both a histogram and a boxplot together gives different insights.

30. *[Ch 1.3]* Egg counts for two insect species are compared; one distribution is symmetric, the other skewed. Which measure of center should be reported for each?

31. *[Ch 1.4]* Blood glucose in rats has $\mu = 110$, $\sigma = 15$. Use the empirical rule to estimate the proportion between 80–140.

32. *[Ch 1.5]* A cholesterol study presents both boxplots and IQR. Why are these more robust than range for detecting outliers?

33. *[Ch 2.1]* A mask mandate study measures case counts before and after policy. Why is this observational, not experimental? What biases might arise?

34. *[Ch 2.2]* Data show a strong positive correlation between plant height and fertilizer. Explain why this doesn't prove causation.

35. *[Ch 2.3]* 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(Fungus+ OR Whiteflies).

36. *[Ch 2.4]* A hospital tracks daily ER patient arrivals for a year. Propose how to summarize the data distribution and explain what extreme outliers may mean.

37. *[Ch 3.2]* 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

40. *[Ch 2.1]* A national bird survey selects only urban parks. Discuss why this introduces undercoverage bias and propose a corrected design.

41. *[Ch 1.5]* A dataset of insect counts per plot is extremely skewed by one plot with 500 individuals. Propose two numerical and one graphical summary that best communicate the data.

42. *[Ch 1.5]* A study of turtle egg survival follows nests across 10 years. Discuss how both median and IQR provide more robust insight than mean and range, especially in variable years.

43. *[Ch 2.3]* A microbiology lab cultures 120 mice for bacterial strains A and B.

	Strain B+	**Strain B–**	**Total**
Strain A+	25	15	40
Strain A–	35	45	80
Total	60	60	120

Find $P(\text{A+ OR B+})$.

44. *[Ch 2.4]* A longitudinal study measures white blood cell counts monthly for 5 years. Propose two challenges with missing data and how they affect interpretation.

45. *[Ch 1.5]* A marine ecologist measures mercury concentration in fish: mean=0.4 ppm, SD=0.05 ppm. Explain how z-scores can be used to classify "high-risk" fish, and how cutoffs might be set.

46. *[Ch 2.1]* In a cohort study, 300 patients are followed for incidence of diabetes. Discuss potential confounding variables and how they can be controlled in analysis.

47. *[Ch 1.5]* A data set of whale lengths has min = 15 m, Q_1 = 18 m, Median = 20 m, Q_3 = 21 m, max = 35 m. Interpret the spread, identify outliers, and explain how skewness affects conclusions.

48. *[Ch 1.4]* A dataset on bacterial colony counts has extreme variability. Discuss why variance/SD are more informative than range, and how transformations (like log) could improve analysis.

49. *[Ch 3.2]* In a deer study, 200 animals are sampled: 60 are CWD+, 120 have ticks, 42 have both. (a) Compute $P(\text{CWD+})$, $P(\text{Ticks})$, and $P(\text{CWD+} \cap \text{Ticks})$. (b) Compute $P(\text{CWD+} \mid \text{Ticks})$ and interpret in context. (c) Do CWD and ticks appear independent? Justify.

50. *[Ch 1.3]* 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 $n-1$