Question	Points	Score
1	3	
2	4	
3	5	
4	12	
5	4	
6	5	
7	6	
8	4	
9	2	
10	6	
11	4	
12	5	
13	6	
14	5	
15	6	
16	5	
17	5	
18	3	
Total:	90	

1. For each of the following scenarios indicate what distribution is **most appropriate** for the described random variable. **No justifications are required.**

   (a) (1 point) In a manufacturing process, you're interested in the number of trials required to produce the first defective item.

   (b) (1 point) In a survey, you want to determine the number of people who prefer online shopping over in-store shopping.

   (c) (1 point) An online streaming platform wants to model the number of times a specific video is watched in an hour, given a known average view rate.

2. Consider a normal approximation to a binomial random variable. For this question it may be helpful to recall that $\Phi(1.96) = 0.975$.

   (a) (2 points) Suppose that $X \sim \text{Bin}(2, 0.0267)$. Can you use the normal approximation to the binomial in this case? Explain.

   (b) (1 point) Suppose that, regardless of your previous answer you decide to use the normal approximation. What is the approximate probability that $X > 0$?

   (c) (1 point) How does the true probability (not using the binomial approximation) differ from the result in part (b)? Does this support or contradict your result from (a)?

3. Consider two fair 6-sided dice.

   (a) (2 points) What is the probability of a 4 turning up at least once if both dice are tossed.

   (b) (2 points) What is the probability that the two dice show the same value?

   (c) (1 point) How would you interpret the probability obtained in part (b)?

4. For each of the following pairs of concepts, **briefly** describe (or define) each, indicating how they differ.

   (a) (3 points) Probability density functions versus probability mass functions.

   (b) (3 points) Independence versus mutually exclusive.

   (c) (3 points) Estimates versus estimators.

   (d) (3 points) Standard error versus standard deviation.

5. (a) (3 points) In five card poker, a flush consists of five cards of the same suit, regardless of their denominations. If we consider a standard deck of cards (denominations: ace, two, three, four, five, six, seven, eight, nine, ten, Jack, Queen, and King. Suits: hearts, diamonds, clubs, and spades), how many possible flushes are there?

   (b) (1 point) If a five card hand is selected at random, what is the probability of observing a flush?

6. Suppose that a $100(1-\alpha)\%$ confidence interval is found for a parameter, θ. For each of the following scenarios, will the new confidence interval be shorter, equal, or wider in length when compared to the original? If there is not enough information to answer, or more than one answer is possible, indicate "Not enough information".

   (a) (1 point) The new confidence interval is $100(1-\beta)\%$ where $\beta > \alpha$, with nothing else changed.

   ○ Shorter ○ Equal ○ Longer ○ Not enough Information

   (b) (1 point) The new confidence interval is computed on a larger value of n, with nothing else changed.

   ○ Shorter ○ Equal ○ Longer ○ Not enough Information

   (c) (1 point) The new confidence interval is computed on data with a smaller standard error, with nothing else changed.

   ○ Shorter ○ Equal ○ Longer ○ Not enough Information

   (d) (1 point) The new confidence interval is computed with a higher level of confidence.

   ○ Shorter ○ Equal ○ Longer ○ Not enough Information

   (e) (1 point) The new confidence interval is computed with a higher level of confidence.

   ○ Shorter ○ Equal ○ Longer ○ Not enough Information

7. Suppose that for an unknown distribution we wish to estimate the population variance, $\sigma^2 = \mathrm{var}(X)$. Suppose that the mean is known to be zero. We take a sample of n values from this population, denoted $X_1, \ldots, X_n$, independently of each other. We consider $\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ and $\widehat{\theta}_2 = X_1^2$ as estimators for σ^2.

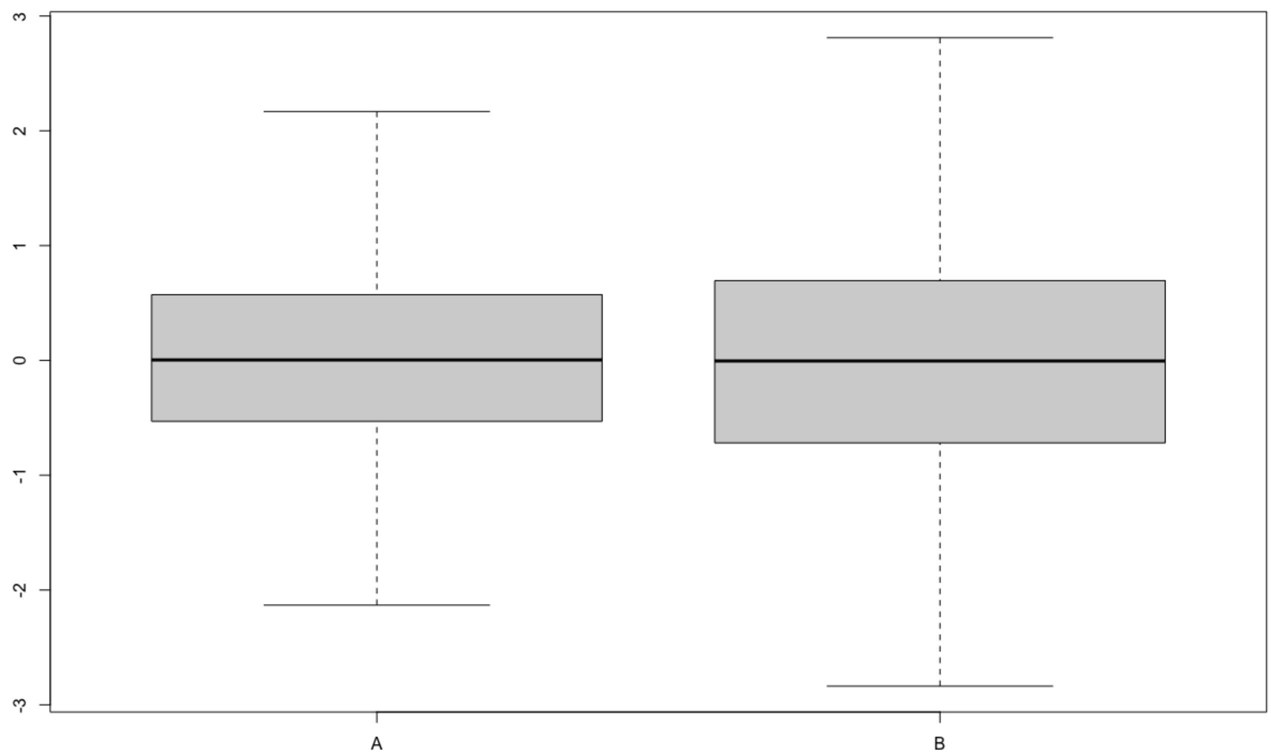
   (a) (2 points) Which estimator has a larger bias? Explain.

   (b) (2 points) Which estimator has a larger variance? Explain.

   (c) (2 points) Which estimator is better? Explain.

8. (4 points) Suppose that $X \sim \text{Unif}(-\theta, \theta)$. What is a method of moments estimator for θ based on a sample of size n

9. Consider the following box plot.



(a) (1 point) Can we determine whether A or B has a higher mean? If so, which does? If not, why?

(b) (1 point) Can we determine whether A or B has a higher $Q1$? If so, which does? If not, why?

10. Suppose that a pmf is given by

$$p(x) = \begin{cases} k(7x+3) & x = 0, 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}.$$

(a) (2 points) What is $P(X = 0)$?

(b) (1 point) What is $P(X > 1)$?

(c) (2 points) What is $E[X]$?

(d) (1 point) What is $\text{var}[X]$?

11. (4 points) A continuous random variable X, with $E[X] = 1$, has probability density function $f_X(x)$ given by

$$f_X(x) = \begin{cases} a(b-x)^2 & 0 \leq x \leq b, \\ 0 & \text{otherwise} \end{cases}.$$

Find the values of a and b.

12. Suppose that the following output was obtained from running a hypothesis test of $H_0 : \mu = 136.3$ versus $H_1 : \mu \neq 136.3$ on a sample of tensile strength observations.

Quantity	**Value**
N (Sample Size)	153
$\overline{X}$	135.39
S.E. of $\overline{X}$	0.37
t Value	–
p-value	0.014

(a) (1 point) Suppose that the hypothesis test is run at $\alpha = 0.01$. What is the conclusion of the hypothesis test?

(b) (1 point) What is the value of the test statistic used in the hypothesis test?

(c) (2 points) What is the sample variance estimate?

(d) (1 point) Suppose that a 99% confidence interval was fit to the data. What can we say about the length of this interval?

13. A sample of 14 joint specimens of a particular type gave a sample mean proportional limit stress of 8.48 MPa and a sample standard deviation of 0.79 MPa. These joints should be designed to have a proportional limit stress of at least 8.50 MPa.

   (a) (1 point) State the null and alternative hypotheses for this scenario.

   (b) (1 point) What is the value of the relevant test statistic?

   (c) (1 point) What is the sampling distribution for the test statistic, assuming the null hypothesis is correct? What assumptions are required for this to hold?

   (d) (2 points) Suppose that the p-value calculated from this test is > 0.10. Describe your conclusion in this setting.

   (e) (1 point) Given your previous response, are you more concerned with a type I or type II error? What would it refer to in this specific scenario?

14. Suppose that the true standard deviation for helium porosity of coal samples is 0.75, and the measurements are assumed to be normally distributed.

    (a) (2 points) If you need to estimate $\overline{X}$ to within 0.02 at a 95% level of confidence, how many parts should you sample? (**Note:** $Z_{0.025} = -1.96$).

    (b) (1 point) If you wish to be 99% confident, how would the required sample size change? **Do not calculate the newly required value.**

    (c) (2 points) Suppose that $\overline{X} = 4.85$ with a sample size of $n = 36$. Find and interpret the 95% confidence interval.

15. Measurements of ultimate compressive stress for green mixed oak is compared between two grades of lumber. For 11 specimens of no. 1 grade lumber, the average compressive stress was 22.1 with a standard deviation of 4.09. For 7 specimens of no. 2 grade lumber, the average was 20.4 with a standard deviation of 3.08.

    (a) (2 points) What is the sampling distribution that would be used to test hypotheses in this experiment? What assumptions are required?

    (b) (4 points) Test the hypothesis that grade 1 is better than grade 2 in terms of ultimate compressive stress. Express your p-value in terms of $F(X)$, the CDF of the sampling distribution from part (a).

16. (5 points) Cardiac function is assessed via impedance cardiography while performing the Valsalva maneuver. A set of 11 subjects perform the measure both standing and reclining, and the mean impedance ratio is reported in the following table.

Subject	Standing	Reclining
1	1.45	0.98
2	1.71	1.42
3	1.81	0.70
4	1.01	1.10
5	0.96	0.78
6	0.83	0.54
7	1.23	1.34
8	1.00	0.72
9	0.80	0.75
10	1.03	0.82
11	1.39	0.60

Describe how you would test the whether there is a difference in the mean impedance ratio between the two positions. **Note: include all of the information required for hypothesis testing, clearly defining any notation you require. Do not actually compute the required values or run the test**.

17. Consider carrying out m tests of hypotheses, independently, based on a significance level of 0.01.

   (a) (2 points) What is the probability of committing at least 1 type I error with $m = 5$? With $m = 10$?

   (b) (3 points) How many tests would need to be performed until the probability of a type I error exceeds 0.5?

18. (3 points) Describe the process and consideration for selecting α in a hypothesis test in an experiment. Why might we consider higher values of α in certain settings?