AP STATISTICS

Exam Prep Workbook

FRQ + MCQ Practice

90 Multiple-Choice Questions & 30 Free-Response Investigative Tasks

Grades 11 – 12

Aligned to College Board AP Statistics Course Framework

Three Thematic Units • Three Cumulative Practice Tests

Complete Worked Solutions • Scoring Rubrics • TI-84 Guide

Unit 1	Exploring Data: Distributions, Bivariate Data, Regression & Transformations
Unit 2	Sampling, Experimentation & Probability: Design, Rules, Distributions
Unit 3	Statistical Inference: Confidence Intervals & Significance Tests

UM Printables

Premium AP Resources for Educators & Students

Table of Contents

AP Statistics Formula & Table Reference Sheet

I. Descriptive Statistics

Sample Mean	$\bar{x} = (1/n)\ \Sigma x_i$
Sample Standard Deviation	$s = \text{sqrt}[\ (1/(n-1))\ \Sigma(x_i - \bar{x})^2\ ]$
Interquartile Range	$IQR = Q_3 - Q_1$
Outlier Boundaries	Lower = $Q_1 - 1.5 \times IQR$; Upper = $Q_3 + 1.5 \times IQR$
Z-Score	$z = (x - \mu) / \sigma$
Correlation Coefficient	$r = [1/(n-1)]\ \Sigma[(x_i-\bar{x})/s_x][(y_i-\bar{y})/s_y]$
Coefficient of Determination	$r^2 = SS_{Reg} / SS_{Total}$
Slope of LSRL	$b_1 = r(s_y / s_x)$
Y-Intercept of LSRL	$b_0 = \bar{y} - b_1\bar{x}$
LSRL Equation	$\hat{y} = b_0 + b_1 x$
Residual	residual = $y_i - \hat{y}_i$
Pooled Sample Variance	$s_p^2 = [(n_1-1)s_1^2 + (n_2-1)s_2^2] / (n_1+n_2-2)$

II. Probability & Distributions

Addition Rule	$P(A \chi B) = P(A) + P(B) - P(A \chi B)$
Multiplication Rule	$P(A \chi B) = P(A) \chi P(B \mid A)$
Conditional Probability	$P(A \mid B) = P(A \chi B) / P(B)$
Binomial Probability	$P(X=k) = (nCk)\ p^k\ (1-p)^{n-k}$
Geometric Probability	$P(X=k) = (1-p)^{k-1}\ p$
Mean of Discrete RV	$\mu_x = \Sigma\ x_i p_i$
Variance of Discrete RV	$\sigma_x^2 = \Sigma\ (x_i - \mu_x)^2\ p_i$
Normal Distribution	$z = (x - \mu) / \sigma$

III. Statistical Inference

General CI Formula	statistic ± (critical value) × (standard error)
1-Proportion z-Interval	$\hat{p} \pm z^* \text{ sqrt}[\ \hat{p}(1-\hat{p}) / n\ ]$
2-Proportion z-Interval	$(\hat{p}_1-\hat{p}_2) \pm z^* \text{ sqrt}[\ \hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2\ ]$
1-Mean t-Interval	$\bar{x} \pm t^* (s / \text{sqrt}(n))$
2-Mean t-Interval	$(\bar{x}_1-\bar{x}_2) \pm t^* \text{ sqrt}[\ s_1^2/n_1 + s_2^2/n_2\ ]$
Chi-Square Test	$\chi^2 = \Sigma\ (O-E)^2 / E$
Slope CI	$b_1 \pm t^* \ SE_{b1}$

UNIT 1: Exploring Data

Section 1A: No-Calculator Multiple-Choice

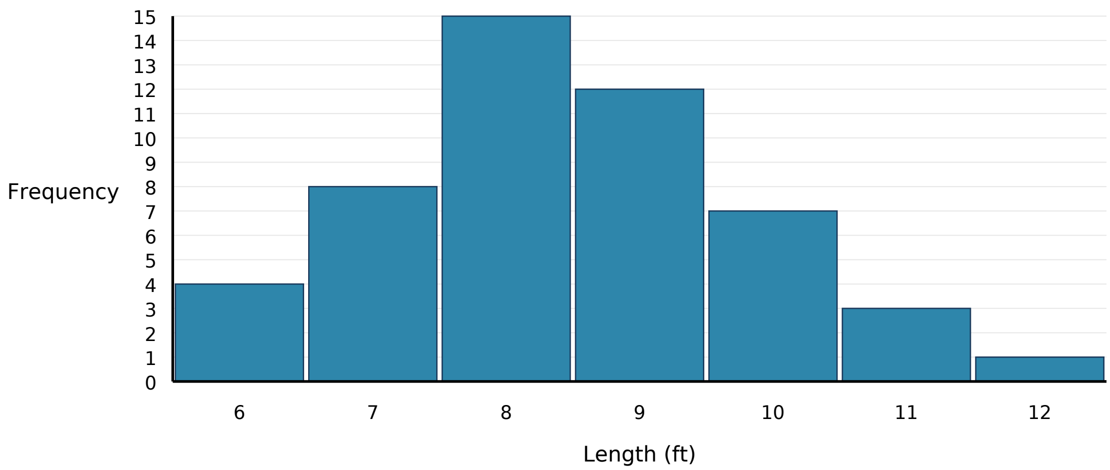
Questions 1 – 15 | Approximate Time: 25 minutes

Shape, Center, & Spread

Question 1

A marine biologist recorded the lengths of 50 adult bottlenose dolphins. The histogram displays the distribution of lengths. Which of the following statements correctly compares the mean and median of the distribution?

Figure 1: Distribution of Dolphin Lengths (in feet)



(A) The mean is equal to the median.

(B) The mean is greater than the median.

(C) The mean is less than the median.

(D) The relationship between the mean and median cannot be determined from a histogram.

Solution **Answer: (B)**

The histogram shows a distribution with a tail extending to the right (towards the larger values like 11 and 12 feet), indicating a **right-skewed** shape. In a right-skewed distribution, the mean is pulled in the direction of the tail. Therefore, the mean is greater than the median.

IQR & 5-Number Summary

Question 2

A set of 12 data values is shown: 3, 5, 6, 8, 9, 10, 11, 12, 14, 15, 18, 20. What is the interquartile range (IQR) of the dataset?

(A) 7.0
(B) 7.5
(C) 8.0
(D) 9.5

Solution **Answer: (B)**

To find the IQR, we first find the median. Since there are 12 values, the median is the average of the 6th and 7th values: (10 + 11) / 2 = 10.5.

The first quartile (Q1) is the median of the lower half (first 6 values): 3, 5, 6, 8, 9, 10. Q1 = (6 + 8) / 2 = 7.

The third quartile (Q3) is the median of the upper half (last 6 values): 11, 12, 14, 15, 18, 20. Q3 = (14 + 15) / 2 = 14.5.

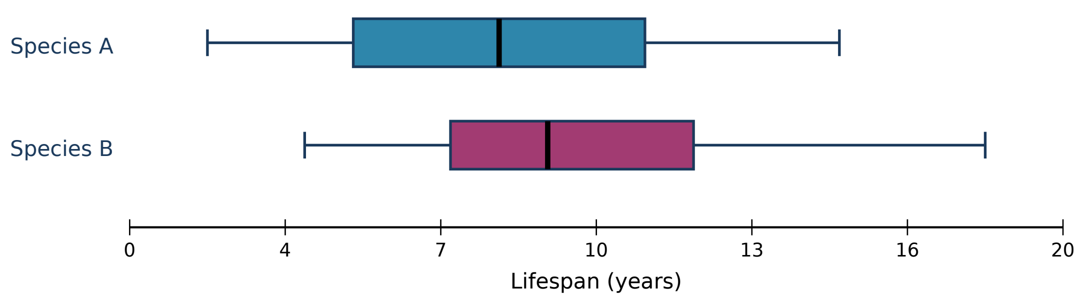
IQR = Q3 - Q1 = 14.5 - 7 = 7.5.

Comparing Distributions

Question 3

The boxplots above display the distribution of lifespans for two species of plants. Based on the boxplots, which of the following statements is true?

Figure 2: Lifespan of Two Plant Species (in years)



(A) Species B has a larger interquartile range (IQR).
(B) Species A has a higher median lifespan.
(C) Both distributions appear to be skewed left.
(D) Species B has a greater range of lifespans.

Solution **Answer: (D)**

Let us evaluate the statistics from the boxplots:

Species A: Min=2, Q1=5, Median=8, Q3=11, Max=15. IQR = 11 - 5 = 6; Range = 15 - 2 = 13.
Species B: Min=4, Q1=7, Median=9, Q3=12, Max=18. IQR = 12 - 7 = 5; Range = 18 - 4 = 14.

Evaluating the options:
A) False. Species A (6) > Species B (5).
B) False. Species A median (8) < Species B median (9).
C) False. Neither distribution appears primarily skewed left based on the boxplots.
D) True. Species B range (14) > Species A range (13).

Residuals

Question 4

A least-squares regression line is computed for a dataset relating the number of hours studied and exam scores. If the residual for a given data point is positive, which of the following must be true?

(A) The actual exam score is higher than the predicted exam score.
(B) The actual exam score is lower than the predicted exam score.
(C) The actual exam score is equal to the predicted exam score.
(D) The correlation coefficient for the regression line is positive.

Solution **Answer: (A)**

A residual is calculated as: Residual = Actual y - Predicted y ($y - \hat{y}$).

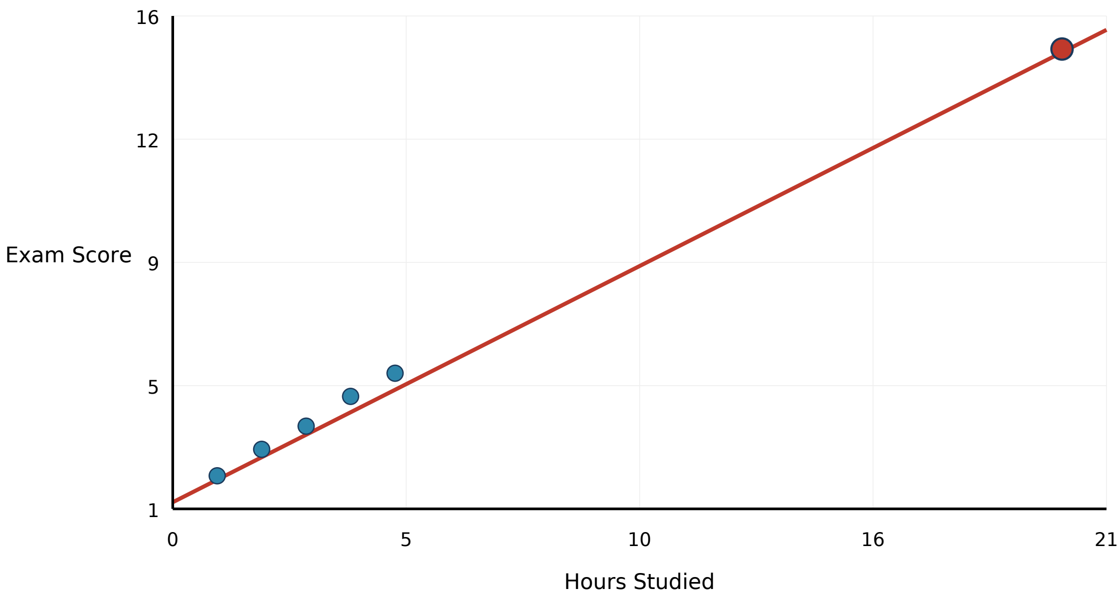
If the residual is positive, then $y - \hat{y} > 0$, which means $y > \hat{y}$. This indicates that the actual data point lies **above** the least-squares regression line. In context, the actual exam score is higher than the predicted exam score.

Influential Points

Question 5

The scatterplot displays the relationship between hours studied and exam scores. What effect does the highlighted point (20, 15) have on the correlation coefficient (r) and the slope of the least-squares regression line compared to a model without this point?

Figure 3: Study Hours vs. Exam Score



(A) It decreases the correlation and decreases the slope.
(B) It increases the correlation and increases the slope.
(C) It increases the correlation but decreases the slope.
(D) It decreases the correlation but increases the slope.

Solution **Answer: (B)**

The highlighted point (20, 15) is located in the upper-right corner of the scatterplot, far from the main cluster of data. It follows the positive trend established by the other points.

An influential point that aligns with the general trend will **strengthen** the linear association, thereby increasing the correlation coefficient (r). Because it has an unusually large x-value and fits the trend, it pulls the regression line toward itself, thereby increasing the slope.

Z-Scores

Question 6

The heights of adult males are approximately normally distributed with a mean of 69 inches and a standard deviation of 2.8 inches. What is the z-score for an adult male who is 74 inches tall?

(A) -1.79
(B) 1.79
(C) -5.0
(D) 5.0

Solution **Answer: (B)**

The z-score formula is $z = (x - \mu) / \sigma$.

Given $x = 74$, $\mu = 69$, and $\sigma = 2.8$:
$z = (74 - 69) / 2.8 = 5 / 2.8 \approx 1.7857$.

Rounding to two decimal places, the z-score is 1.79. A positive z-score indicates the value is above the mean.

Effects of Transformations

Question 7

A set of exam scores has a mean of 72 and a standard deviation of 8. If the teacher adds 5 bonus points to every student's score, what are the new mean and standard deviation?

(A) Mean = 72, Standard Deviation = 13
(B) Mean = 77, Standard Deviation = 13
(C) Mean = 72, Standard Deviation = 8
(D) Mean = 77, Standard Deviation = 8

Solution **Answer: (D)**

When a constant is added to every value in a dataset:
- The **mean** increases by that constant: 72 + 5 = 77.
- The **standard deviation** is unaffected because adding a constant does not change the spread of the data; it merely shifts the entire distribution: 8.

Therefore, the new mean is 77 and the new standard deviation is 8.

Coefficient of Determination

Question 8

A linear regression model is fit to data relating the age of a used car (in years) to its selling price (in dollars). The model yields an r^2 value of 0.81. Which of the following is the best interpretation of this value?

(A) 81% of the data points lie on the regression line.
(B) 81% of the variation in selling price can be explained by the linear relationship with age.
(C) There is an 81% correlation between the age and selling price of the car.
(D) The slope of the regression line is 0.81.

Solution **Answer: (B)**

The coefficient of determination, r^2, represents the proportion of the variance in the dependent variable (selling price) that is predictable from the independent variable (age).

An r^2 of 0.81 means that 81% of the variation in selling price can be explained by the linear relationship with the car's age. The remaining 19% is due to other factors or random variation.

Outlier Rule

Question 9

A data set has Q1 = 45 and Q3 = 65. Using the 1.5 × IQR rule, which of the following values would be classified as an outlier?

(A) 16
(B) 94
(C) 14
(D) 90

Solution **Answer: (C)**

First, calculate the IQR: IQR = Q3 - Q1 = 65 - 45 = 20.

Next, calculate the outlier boundaries using the 1.5 × IQR rule:
Lower Bound = Q1 - 1.5(IQR) = 45 - 1.5(20) = 45 - 30 = 15.
Upper Bound = Q3 + 1.5(IQR) = 65 + 1.5(20) = 65 + 30 = 95.

A value is an outlier if it is strictly less than 15 or strictly greater than 95. The value 14 is less than 15, so it is an outlier. (16 and 94 are within the bounds).

Transformations to Linearity

Question 10

A scatterplot of y versus x shows a strong curved pattern. When the natural logarithm of y is plotted against x, the relationship appears linear. Which of the following models is most appropriate for the original data?

(A) $y = a + bx$
(B) $y = ab^x$
(C) $y = ax^b$
(D) $y = a + b \ln(x)$

Solution **Answer: (B)**

We are told that plotting ln(y) against x produces a linear relationship. This means:
$\ln(y) = a + bx$

To find the model for the original data, we exponentiate both sides to solve for y:
$e^{\ln(y)} = e^{a + bx}$
$y = e^a \times e^{bx}$

Let $A = e^a$ and $B = e^b$. Then the model becomes:
$y = A \times B^x$

This is the form of an exponential model, commonly written as $y = ab^x$.

Resistant Statistics

Question 11

Which of the following is most resistant to the effects of extreme outliers?

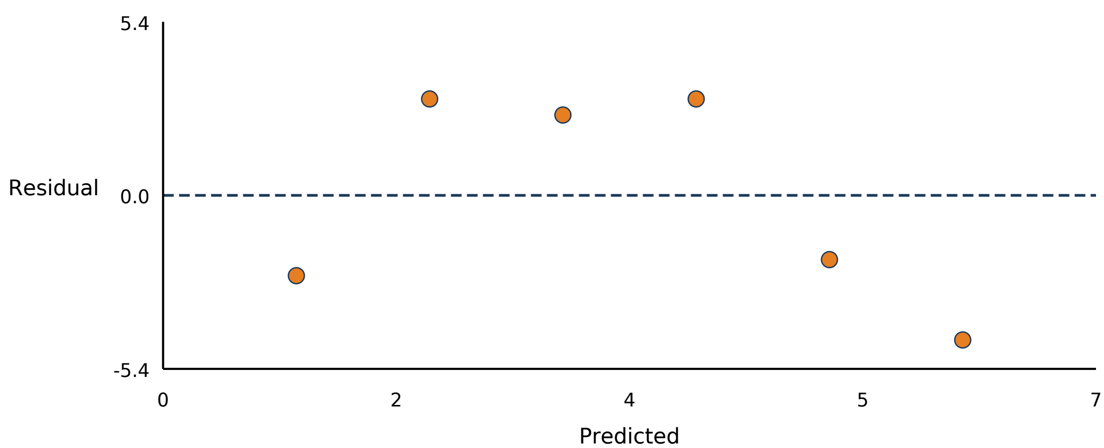
Investigative Task: FRQ 4

FRQ 4

A statistician fits a linear model to predict the fuel efficiency (mpg) of a car based on its weight. The residual plot is shown above.

(a) Is a linear model appropriate for these data? Justify your answer.
(b) What does the pattern in the residual plot suggest about the true relationship between weight and fuel efficiency?

Figure 11: Residual Plot



AP Scoring Rubric (4 Points Total)
Part (a) [2 Points]: States that the linear model is NOT appropriate (1 pt); Justifies by noting the distinct curved pattern (U-shape) in the residual plot (1 pt).
Part (b) [2 Points]: Explains that the curved pattern suggests the true relationship is non-linear (e.g., quadratic or exponential decay) (2 pts).

Annotated Sample Response

Part (a): *No, the linear model is not appropriate. A good linear model should produce a residual plot with a random scatter of points around the horizontal line at zero. The residual plot here shows a distinct U-shaped curve (residuals go from negative, to positive, back to negative).*

Part (b): *The U-shaped pattern in the residual plot indicates that the true relationship between weight and fuel efficiency is non-linear. As weight increases, fuel efficiency likely decreases at a decreasing rate (a curve), meaning a linear model underestimates the efficiency of very light and very heavy cars, while overestimating the efficiency of mid-weight cars.*

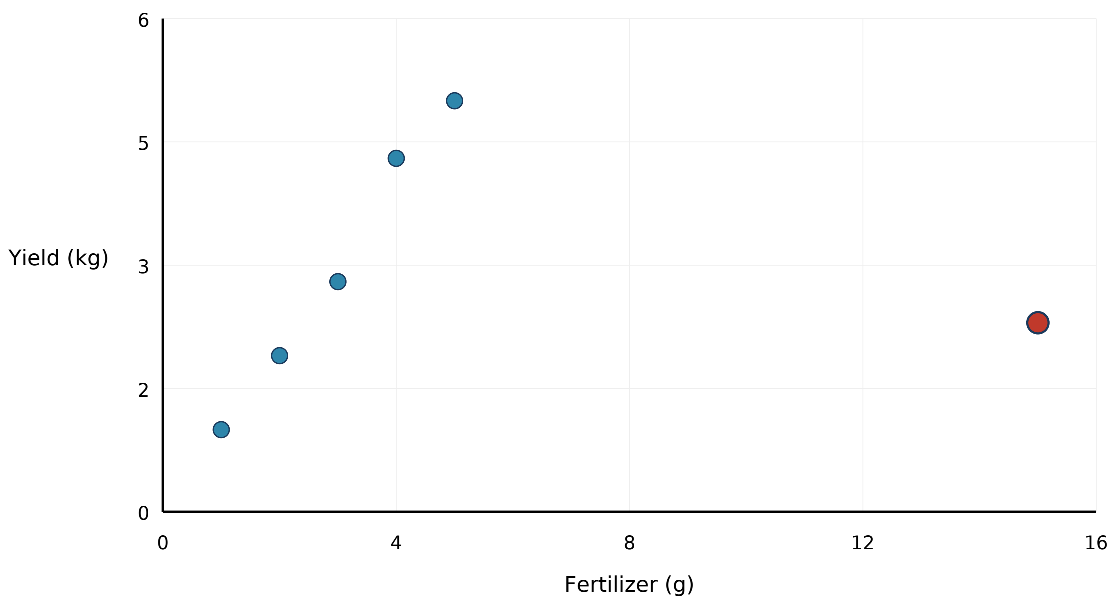
Investigative Task: FRQ 5

FRQ 5

An agronomist tests the effect of fertilizer amount (g) on crop yield (kg). The scatterplot shows the data. The highlighted point corresponds to a plot that was accidentally over-fertilized (15g) but suffered from poor drainage, resulting in a low yield.

(a) Classify the point (15, 2.5) in terms of its influence on the slope of the regression line.
(b) If the point (15, 2.5) is removed, will the correlation coefficient (r) increase, decrease, or stay the same? Justify your answer.

Figure 12: Fertilizer Amount vs Crop Yield



AP Scoring Rubric (4 Points Total)

Part (a) [2 Points]: Identifies the point as an influential point (1 pt); Explains it is an outlier in the x-direction and pulls the regression line down, flattening the slope (1 pt).

Part (b) [2 Points]: States r will increase (towards a stronger positive value) (1 pt); Explains that removing this point removes a data value that deviates from the strong positive linear pattern of the rest, thus strengthening the linear association (1 pt).

Annotated Sample Response

Investigative Task: FRQ 13

FRQ 13

Sarah is a 70% free-throw shooter in basketball. Assume each shot is independent.

(a) If Sarah takes 10 free throws, what is the probability she makes exactly 7 of them?
(b) If Sarah takes 10 free throws, what is the probability she makes at least 7 of them?
(c) If Sarah shoots free throws until she makes her first one, what is the probability it takes her exactly 3 shots?

AP Scoring Rubric (4 Points Total)

Part (a) [1 Point]: Uses binompdf(10, 0.7, 7) to get 0.2668.

Part (b) [2 Points]: Calculates P(X≥7) = 1 - binomcdf(10, 0.7, 6) = 1 - 0.3504 = 0.6496.

Part (c) [1 Point]: Recognizes geometric distribution and calculates (0.3)^2(0.7) = 0.063.

Annotated Sample Response

Part (a): *This is a binomial distribution with n=10 and p=0.70. We want P(X=7).*
$P(X=7) = (10C7)(0.7)^7(0.3)^3$
Using TI-84: ***binompdf(10, 0.7, 7)*** *= 0.2668.*

Part (b): *We want the probability of making at least 7, so P(X ≥ 7) = P(X=7) + P(X=8) + P(X=9) + P(X=10). It is easier to use the complement: P(X ≥ 7) = 1 - P(X ≤ 6).*
Using TI-84: 1 - ***binomcdf(10, 0.7, 6)*** *= 1 - 0.3504 = 0.6496.*

Part (c): *This is a geometric distribution. The probability that the first success occurs on the 3rd trial means she must miss the first two and make the third.*
$P(X=3) = (1 - p)^2(p) = (0.3)^2(0.7) = 0.09 \times 0.7 = 0.063.$

Investigative Task: FRQ 14

FRQ 14

A city council wants to estimate the proportion of residents who support building a new stadium. They email a survey to 5000 randomly selected residents from the city's voter registration list. Only 800 surveys are returned.

(a) Identify the type of bias most likely to affect this survey and explain why it is a concern.
(b) Suggest one practical way the city council could reduce the bias identified in part (a).

AP Scoring Rubric (4 Points Total)

Part (a) [2 Points]: Identifies Nonresponse bias (1 pt); Explains that the 4200 who didn't respond may have different views than the 800 who did, so the sample may not represent the population (1 pt).

Part (b) [2 Points]: Provides a reasonable method to reduce nonresponse, such as offering incentives, sending reminders, or conducting follow-up phone calls (2 pts).

Annotated Sample Response

Part (a): *The most likely bias affecting this survey is* ***nonresponse bias****. Because only 800 out of 5000 people responded (a 16% response rate), there is a high likelihood that the 4200 people who did not respond have different opinions about the stadium than those who took the time to respond. For example, perhaps only those strongly in favor or strongly against bothered to respond. This makes the sample unrepresentative of the whole population.*

Part (b): *To reduce nonresponse bias, the city council could send follow-up reminder emails to those who haven't responded, or they could offer a small incentive (like a gift card raffle) for completing the survey. Another option would be to follow up with phone calls to a random subset of the non-respondents to gather their opinions and see if they differ significantly from the initial respondents.*

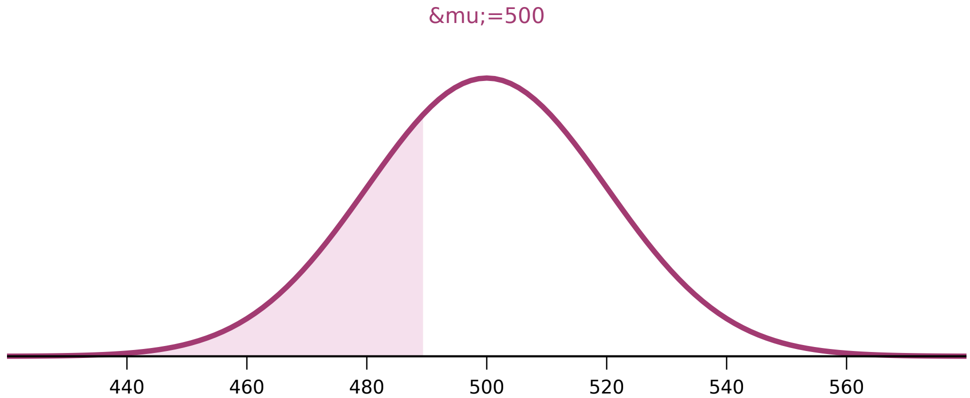
Investigative Task: FRQ 20

FRQ 20

The lifetime of a certain brand of lightbulb is normally distributed with a mean of 500 hours and a standard deviation of 40 hours. A consumer advocacy group tests 4 randomly selected bulbs. Let $\bar{x}$ represent the average lifetime of the 4 bulbs.

(a) Describe the shape, center, and spread of the sampling distribution of $\bar{x}$.
(b) What is the probability that the average lifetime of the 4 bulbs is less than 490 hours?

Figure 17: Sampling Distribution of x-bar (n=4)



AP Scoring Rubric (4 Points Total)

Part (a) [2 Points]: States shape is normal (since pop is normal), center is {MU}_xbar = 500, and spread is {SIGMA}_xbar = 40/sqrt(4) = 20 (2 pts).

Part (b) [2 Points]: Calculates z-score: (490 - 500)/20 = -0.5 (1 pt); Finds probability: normalcdf(-999, -0.5) = 0.3085 (1 pt).

Annotated Sample Response

Part (a): *Because the population distribution is normally distributed, the sampling distribution of $\bar{x}$ will also be exactly normal. Its center (mean) will be the same as the population mean: {MU}$_{\bar{x}}$ = 500 hours. Its spread (standard deviation) is the standard error: {SIGMA}$_{\bar{x}}$ = {SIGMA}/sqrt(n) = 40/sqrt(4) = 20 hours.*

Part (b): *We want P($\bar{x}$ < 490) for a distribution with {MU}=500 and {SIGMA}=20.*
First, find the z-score: z = (490 - 500) / 20 = -10 / 20 = -0.50.
*Using Table A or TI-84 (**normalcdf(-9999, -0.5)**):*
P(Z < -0.50) = 0.3085.
There is a 30.85% chance that the average lifetime of 4 bulbs is less than 490 hours.

CUMULATIVE MINI PRACTICE TESTS

Practice Test 1: Units 1 & 2

10 Questions | Time: 20 Minutes

Transformations

Question 1

A dataset has a mean of 50 and a standard deviation of 5. If every value is multiplied by 2, what is the new standard deviation?

(A) 5
(B) 10
(C) 25
(D) 100

Solution **Answer: (B)**

Multiplying every value by a constant 'a' multiplies the standard deviation by |a|. 5 * 2 = 10.

Resistant Stats

Question 2

Which of the following is a measure of spread that is resistant to outliers?

(A) Standard deviation
(B) Range
(C) Mean
(D) IQR

Solution **Answer: (D)**

The Interquartile Range (IQR) measures the spread of the middle 50% of the data, ignoring extreme high and low values, making it resistant to outliers.

Residuals

Question 3

A residual plot shows a random scatter of points above and below zero. What does this indicate?

(A) A non-linear model is better.
(B) A linear model is appropriate.
(C) There are influential points.
(D) The correlation is 0.

TI-84 / GDC Technology Command Guide

A quick reference for the most frequently used statistical operations on the TI-83/84 graphing calculator.

1-Sample Z-Interval / Z-Test	Press **STAT**, scroll right to **TESTS**. Select **7:ZInterval** or **1:Z-Test**. Choose "Stats" if given summary data, "Data" if using a list. Input: {MU}0 (for test), {SIGMA}, x̄, n, and C-Level (for interval).
1-Sample T-Interval / T-Test	Press **STAT** > **TESTS**. Select **8:TInterval** or **2:T-Test**. Input: {MU}0, x̄, Sx, n, C-Level.
2-Sample Z-Interval / Z-Test	Press **STAT** > **TESTS**. Select **9:2-SampZInt** or **3:2-SampZTest**. Input: {SIGMA}1, {SIGMA}2, x̄1, n1, x̄2, n2.
2-Sample T-Interval / T-Test	Press **STAT** > **TESTS**. Select **0:2-SampTInt** or **4:2-SampTTest**. Input: x̄1, Sx1, n1, x̄2, Sx2, n2. Choose "Pooled: No" unless specifically told variances are equal.
1-Proportion Z-Interval / Z-Test	Press **STAT** > **TESTS**. Select **A:1-PropZInt** or **5:1-PropZTest**. Input: p0 (for test), x (number of successes, NOT a proportion), n, C-Level.
2-Proportion Z-Interval / Z-Test	Press **STAT** > **TESTS**. Select **B:2-PropZInt** or **6:2-PropZTest**. Input: x1, n1, x2, n2.
Chi-Square Test for Independence	Press **STAT** > **TESTS**. Select **C:{CHI}-Test**. Input: [A] for Observed, [B] for Expected. (Go to **2nd** + **MATRX** to edit matrices).
Chi-Square Goodness-of-Fit	Press **STAT** > **TESTS**. Select **D:{CHI}-GOF-Test**. (Available on TI-84 Plus CE and later). Input observed list (L1) and expected list (L2), df.

Thank You for Choosing UM Printables!

We are dedicated to providing premium, rigorously tested educational resources that empower educators and students alike. Your success is our mission!

If you found this AP Statistics Exam Prep Workbook helpful, please consider leaving a review on Teachers Pay Teachers (TPT). Your feedback not only helps us improve but also guides other educators in finding high-quality materials for their classrooms.

How to Leave a Review:
1. Go to your "My Purchases" page on TPT.
2. Find this resource and click the "Provide Feedback" button.
3. Share your honest experience and earn TPT credits!

Happy Studying and Good Luck on the AP Exam!

— The UM Printables Team