Prob/Stats
Practice Final Exam

Choose the *best* answer.

1) A local support group has 1050 members. A study of 372 people with anxiety disorders, and 75% of them reported a link between energy drinks and the inability to focus. Which statement is true?
   a) The population is 1050, the sample is 372, and 75% is a statistic.
   b) The population is 1050, the sample is 372, and 75% is a parameter.
   c) The population is 372, the sample is 1050, and 75% is a statistic.
   d) The population is 372, the sample is 1050, and 75% is a parameter.

The population is the total group from which the sample is taken. Here it is **1,050**.

The sample is the subset taken from the population. Here it is **372**.

A statistic is a descriptive measure (using a numerical value) of a sample, e.g., $\bar{x}, s$. A parameter is a descriptive measure (using a numerical value) of a population, e.g., μ, σ. In this problem, the 75% value relates to the sample, so it is a **statistic**.

Answer A

Determine whether the data are qualitative or quantitative.
2) the colors of automobiles on a used car lot
   A) quantitative   B) qualitative

A **quantitative variable** is one that takes on **numerical values**. Examples: age, weight, salary, GPA, temperature, etc.

A **categorical variable** is one that is not quantitative. Examples: dog breed, hair color, high school attended, political affiliation, etc.

Color is not described with a number, so it is **qualitative**.   **Answer B**

3) the number of complaint letters received by the United States Postal Service in a given day
   A) qualitative   B) quantitative

The **number** of complaint letters is described with a number, so it is **quantitative**.
          Answer B

Here are the **definitions of levels of measurement** that you will need for Questions 4 to 6. They are listed here from lowest to highest level.

- **N**ominal: qualitative only. Characterized by names, labels, qualities. No mathematical computations can be made at this level. Examples: TV stations, colors.
- **O**rdinal: qualitative or quantitative. Can be arranged in order. Differences between data elements are not meaningful. Example: BCS college football rankings.
- **I**nterval: can be ordered and you can calculate meaningful differences between entries. At the interval level, zero represents a location on a scale and does not imply "none." Examples: years, temperatures in ˚F or ˚C.
- **R**atio: similar to the interval level, but a zero implies "none," i.e., a true zero as we typically think of it in mathematics. Also, we can form ratios of data values so that one is a multiple of another. Examples: ages, counts of things.

Increasing from lowest to highest level of measurement, these form the acronym **NOIR**, which is a French word meaning dark.

Identify the data set's level of measurement.

4) ages of students in a statistic class

    A) ratio   B) nominal   C) ordinal   D) interval

We can take ratios of ages and the ratios will mean something, so they are at the **ratio** level of measurement.

<div align="right">

Answer A

</div>

5) temperatures of 22 selected refrigerators

    A) interval   B) nominal   C) ordinal   D) ratio

Temperatures can be ordered and there are meaningful differences between them. Ratios of temperatures do not mean anything as long as we are using the Fahrenheit or Celsius scales. So, they would be at the **interval** level of measurement.

<div align="right">

Answer A

</div>

Note (for Chemistry and Physics students): If we were using the Kelvin Scale (˚C + 273), the temperatures would have meaningful ratios to each other, and so would be at the ratio level of measurement.

6) marriage status (married, single, or divorced) of the faculty at the University of Colorado

    A) nominal   B) ordinal   C) ratio   D) interval

Marital status uses names or labels. We cannot perform mathematical computations with martial statuses. Therefore, they are at the **nominal** level of measurement.

<div align="right">

Answer A

</div>

Here are the **definitions of sampling techniques** that you will need for Questions 7 to 9.

Types of Samples: Note that all statistical sampling approaches have the common goal that chance, rather than human choice, is used to select the sample.

- **Cluster Sample:** a sampling approach in which entire groups (i.e., clusters) are chosen at random; a census is taken of each cluster. *Each cluster should be representative of the entire population.* All clusters should be heterogeneous and similar to each other. The problem with cluster samples is that the clusters are often not homogeneous and representative.

- **Convenience Sample:** a sample of individuals who are conveniently available. Convenience samples often fail to be representative.

- **Multistage Sample:** a sampling approach that combines several sampling methods. Example: stratify the country by geographic region; randomly select cities from each region; interview a cluster of residents from each city. Care should be taken at each step not to introduce bias.

- **Simple Random Sample (SRS):** a sample of size n in which each set of n elements has an equal chance of being selected.   This is the standard against which other sampling methods are measured.

- **Stratified Random Sample:** the population is divided into subgroups (i.e., strata), and random samples are taken from each subgroup. This is better than a simple random sample if the strata are relatively homogeneous and different from each other. It results in reduced sampling variability, and can point out differences in responses among groups.

- **Systematic Sample:** individuals are selected systematically (e.g., every 10th person). Systematic samples can be representative if there is no relationship between the order of the sampling frame and the variables of interest.

Identify the sampling technique used.

7) Every fifth person boarding a plane is searched thoroughly.
   A) systematic
   B) cluster
   C) convenience
   D) random
   E) stratified

Selecting every n-th item in a population is **Systematic Sample**.
Answer A

8) A researcher for an airline interviews all of the passengers on five randomly selected flights.
   A) systematic
   B) cluster
   C) random
   D) convenience
   E) stratified

When everyone in randomly selected groups are surveyed, we have a **Cluster Sample**.
Answer B

9) A community college student interviews everyone in a statistics class to
determine the percentage of students that own a car.
   A) cluster
   B) convenience
   C) random
   D) stratified
   E) systematic

The statistics class is a **convenient group** to survey; the survey is not random over the population of interest. This is a **Convenience Sample**.

Answer B

Decide which method of data collection you would use to collect data for the study

10) A study where the effects of a plague hitting Los Angles is observed

A) simulation   B) observational study   C) experiment   D) survey

What are each of these?

A) **Simulation:** repeated trials of an event, from which data are collected and analyzed to determine the answer to a question.

B) **Observational Study:** a study on data in which there is no manipulation of factors (i.e., variables). The variables and results are observed and studied.

C) **Experiment:** manipulates factor levels to create treatments, randomly assigns subjects to the treatments, and compares the responses across treatment levels.

D) **Sample Survey:** examination of information about a representative subset of a population. Used when performing a census is impractical.

E) **Census (not an answer to this problem):** examination of information about every member of a population. This is the best approach when the population is small and accessible.

If no plague is hitting Los Angeles, and we want to see what would happen if one did, we would **simulate** the impact of the plague.

Answer A

On the other hand, if a plague were actually hitting Los Angeles and we wanted to observe what was happening, we would use an Observational Study.

Decide which type of experimental design is evident in the experiment described.

11) An experiment is conducted to see whether the reflex response time of men is increased with a new nutritional supplement . 40 men participate in the experiment, and each man is partnered with another subject who has the most similar response time. One man from each pair will take the new nutritional supplement, and the other man will receive a placebo as a control.

A) completely randomized design   B) randomized block design
C) stratified random sample   D) matched pairs design

What are each of these?

A) Completely Randomized Design: subjects are assigned to different treatment groups purely through random selection.

B) Randomized Block Design: similar subjects are grouped together into blocks. The results of the experiment are recorded for each block.

C) Stratified Random Sample: the population is divided into subgroups (i.e., strata), and random samples are taken from each subgroup.

D) Matched Pairs Design: subjects are paired up according to a similarity. Within each pair, the treatment each subject receives is randomly chosen.

The magenta oval in the paragraph above identifies the key word to the solution of this problem. "Each man is **partnered** with another subject who has the most similar response time." This is a **Matched Pairs Design**.   **Answer D**

12) A candy company is changing its formula for its most popular chocolate bar. Before committing to the new candy, the company decides to do an experiment to see which formula is preferred by customers. 40 subjects are randomly chosen to try both candy bars, and then to choose which formula they prefer. The conductors of the experiment know which formula the subjects will receive (the order of the medicine given is randomly chosen). Which of the following is a true statement?

a) A placebo is used.
b) The experiment is blind.   See the definitions below. The customers do not know which
c) The experiment is double-blind.   chocolate bars they are eating, but the experimenter does.
d) None of these statements is true.   This is a **Single Blind Experiment**.   **Answer B**

Key Terms

Blinding: Keeping individuals from knowing how the subjects are assigned to treatment groups. Anyone who is unaware of the assignments is said to be **blinded**. Blinding is used to reduce bias.

Class A: those who can *influence* the results (e.g., subjects, administrators, technicians).
Class B: those who *evaluate* the results (e.g., judges, treating physicians, statisticians).

- **Single Blind:** When one of the above groups, but not both, is blinded.
- **Double Blind:** When both of the above groups are blinded.

Placebo: A "fake" treatment which is known to have no effect.

13) There are 4750 students who are currently enrolled at your school. You wish to form a sample of 9 students to answer some survey questions. Select the first three students who will belong to the survey.

31524 49587 76612 39789 13537 48086 59483 60680

a) 315, 244, 958
b) 31, 52, 44
c) 3152, 4495, 8776
d) 3152, 4495, 1353

Using this list of random numbers, we take the first three sets of four digits that produce numbers less than or equal to 4750, the size of the population we are dealing with.

Answers **A** and **B** do not have 4 digits, and answer **C** has a number that exceeds 4,750. Answer **D** meets all of our criteria. See the sequence rewritten below, four digits at a time. We cannot use 8776, 6123 or 9789. So, after 3152 and 4495, the next good number is 1353.

3152 4495 8776 6123 9789 1353   **Answer D**

Use the diagram below for questions 14 and 15.
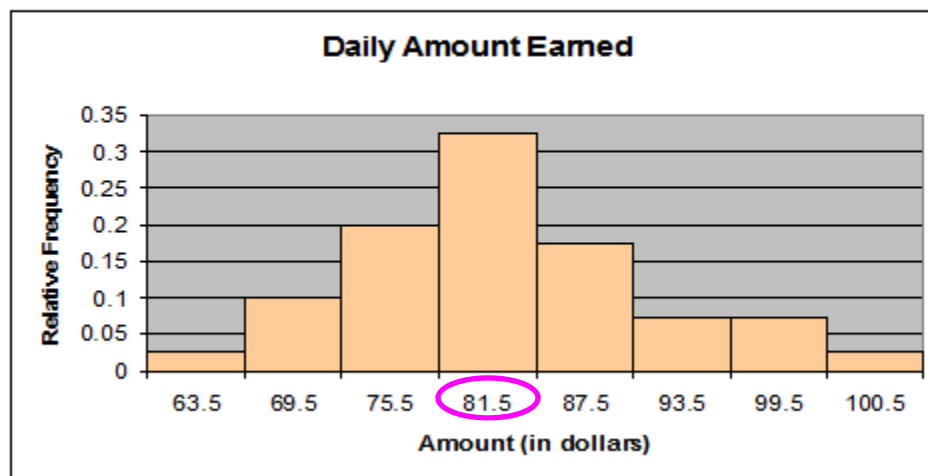
Note that the diagram is labeled with class midpoints. To find the endpoints of each class, divide the class width by 2 and subtract it from, or add it, to each midpoint. For example, the class width (also called the bin width) is: $69.5 - 63.5 = 75.5 - 69.5 = 81.5 - 75.5 = \cdots = 6$. (By the way, this answers Problem 15.)

So the class labeled 81.5 has a lower boundary of $81.5 - \frac{6}{2} = 78.5$, and an upper boundary of $81.5 + \frac{6}{2} = 84.5$.

Now, let's look at problems 14 and 15.

14) According to the relative frequency distribution shown below, what percentage of people earned an average of more than $78.50 in one day?

a) 10%   b) 25.5%   c) 32.5%   d) 68.5%

We need to add the values associated with the five bars to the right of the 78.5 endpoint to answer this question.

$$0.325 + 0.175 + 0.075 + 0.075 + 0.025 = 0.675 = 67.5\% \sim 68.5\%$$

Answer D

Alternatively, the five rightmost bars (with earnings above $78.50) must be what's left of 100% when you subtract out the three leftmost bars (with earnings below $78.50). So,

$$1 - 0.025 - 0.100 - 0.200 = 0.675 = 67.5\% \sim 68.5\%$$

15) What is the class width for the diagram above?

a) 3   b) 6   c) 12   d) 37

The class width is the distance between lower (or upper, or midpoint) limits of consecutive classes. It is calculated in the final paragraph on the last page to be **6**.   **Answer B**

16) A city in the Pacific Northwest recorded its highest temperature of 74 degrees Fahrenheit, and its lowest temperature at 23 degrees Fahrenheit for a particular year. Find the upper and lower limits of the first class if you wish to construct a frequency distribution with 10 classes.

a) 23-27   b) 18-28   c) 23-29   d) 23-28

Steps:
1. Find the range of the distribution: $\text{Range} = 74 - 23 = 51$.
2. Obtain an initial estimated class width by dividing the range by the number of classes: $51 \div 10 = 5.1$.
3. Round the result in Step 2 **up** to the next highest value that you find acceptable for a class width. I would round 5.1 up to 6 since all of the answers are given in integers.
4. The first class, then starts at the lowest value and the next class starts at the lower value plus the class width.

The answer key says the next class begins with 29, so the first one ends with 28. I am not a big fan of this since I wonder where that would put the temperature for 28.6°. An answer to this may be if the class boundaries extend beyond the definition of the class, so that the true upper and lower limits of the range 23-28 were 22.5-28.5 (textbook page 44). We must assume that 22.5-28.5 is the true definition of the class 23-28, which seems odd to me. If you are confused by this, you have company. This is not how things work in the real world.

I like answer C, which I think is the best answer, but the textbook requires: **Answer D**

Provide an appropriate response.

17) Use the ogive below to approximate the cumulative frequency for 24 hours.

Leisure Time of College Students
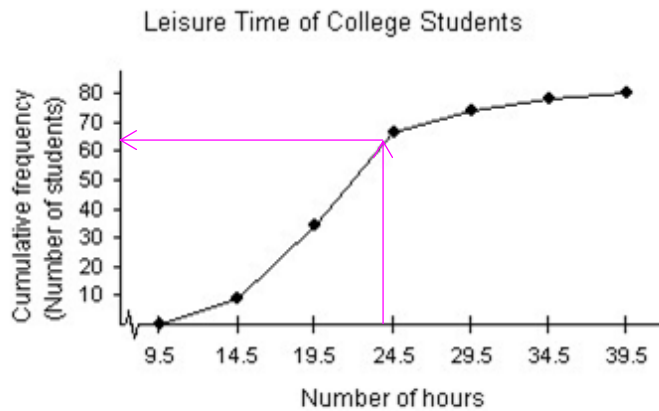
a) 17   **b) 63**   c) 27   d) 80

To find the cumulative frequency for 24 hours, follow the magenta arrows in the above figure.

Steps:

1. Start on the x-axis at 24 hours, which is a little to the left of 24.5 hours.
2. Move vertically until you intersect the ogive.
3. Move left to the y-axis and read the value off the y-axis scale.

Solution: the cumulative frequency for 24 hours is **a little over 60 students**.

Answer B

Motor Vehicle Accidents for Teenage Drivers

Frequency

Number of Accidents
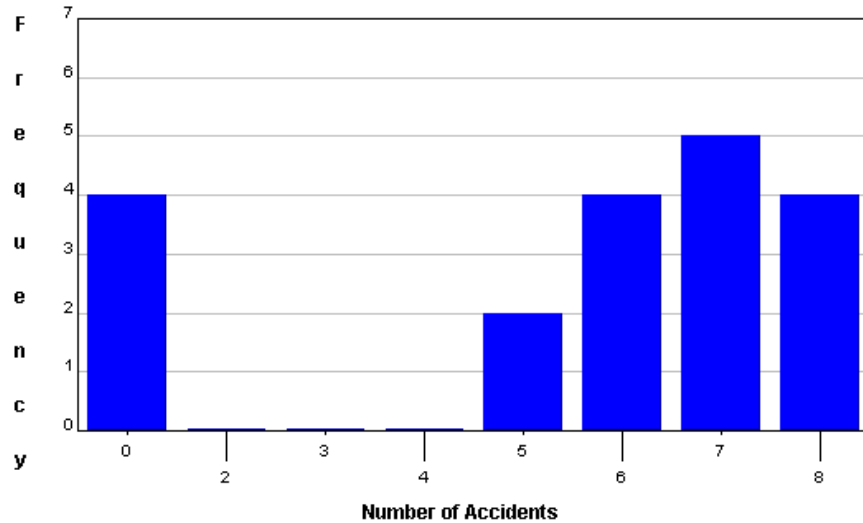
For questions 18 to 19, I would first calculate frequencies and cumulative frequencies for each class.

Number of Accidents	0	2	3	4	5	6	7	8
Frequency	4	0	0	0	2	4	5	4
Cumulative Frequency	4	4	4	4	6	10	15	19

18) What is the median for the above diagram?

a) 2   b) 4.5   c) 6   d) 7.5

To find the median, construct the above table, then find the midpoint. The total frequency is 19. To find the midpoint, add 1 and divide the result by 2.

Our midpoint is $\frac{19+1}{2} = 10$, meaning the median is the 10th item in the ordered data set.

Note: using this process, if you get a fraction, you would average the ordered items on each side of the fraction. For example, if the frequency were 18 you would get $\frac{18+1}{2} = 9.5$, you would average the 9th item and 10th item in the ordered set to get the median.

Look at the cumulative frequency row to see which interval contains the 10th item. It is the interval for **6 accidents**.   **Answer C**

19) What is the mode for the above diagram?

a) 3   b) 6   c) 7   d) 8

The mode is the highest bar, i.e., the one with the highest frequency. In this data set, "**7 accidents" has the highest frequency, which is 5 teenage drivers.**   **Answer C**

20) What is the mean for the above diagram?

a) 2.5   b) 3.46   c) 4.71   d) 5.32

Accidents (x)	0	2	3	4	5	6	7	8	Total ($\sum$)
Frequency (w)	4	0	0	0	2	4	5	4	19
Product ($x \cdot w$)	0	0	0	0	10	24	35	32	101

The mean is calculated from totals in the above table. The formula for the mean is:

$$\bar{x} = \frac{\sum(x \cdot w)}{\sum w},$$ where x is the number of accidents and w is the frequency.

Using the totals in the above chart, we get $\bar{x} = \frac{\sum(x \cdot w)}{\sum w} = \frac{101}{19} = 5.32$

Answer D

21) Which measure of central tendency would best represent the above data?

a) mean   b) median   c) mode   d) range

Consider each answer:

A) The mean is the average of all of the values in a distribution. It can be affected by very low or very high values, so it may not be a good measure of central tendency for some distributions.

B) The median is not affected by very low or very high values, so it is generally considered the best measure of central tendency for most distributions. 50% of the population is above the median and 50% of the population is below the median.

C) The mode is the value that occurs most frequently. It is useful in describing shape (e.g., a bimodal distribution), but is generally not considered to be a good measure of central tendency for many distributions.

D) The range is not a measure of central tendency. It is the difference between the minimum and maximum values of a distribution, but does not describe what goes on between those two values.

Answer: B

Use the information below for questions 22 and 23.

The heights (in inches) of 20 adult males are listed below.

70 72 71 70 69 73 69 68 70 71
67 71 70 74 69 68 71 71 71 72

22) Find the population standard deviation.

   a) 1.682   b) 1.725   c) 1.834   d) 1.973

I like to calculate the variance first, so I will end up answering Problem 23 before I answer Problem 22.

I like to calculate the **variance** and **standard deviation** using a table.

$\sigma^2 = \dfrac{\Sigma\,(x-\mu)^2 \cdot f}{n}$, is the **population** variance, where μ is the mean, f is the set of frequencies for the given x-values and n is the total frequency (i.e., $n = \Sigma f$).

Note that if we were asked about the **sample variance** instead of the population variance, we would use the same formula, but with $(n-1)$ in the denominator instead of n.

$\sigma = \sqrt{\sigma^2}$, is the population standard deviation

You can do this problem with your calculator, but I am going to show you the calculations using the table to the right (created with Microsoft Excel) so you can see what is going on under the surface. I start by tabulating the frequencies of each height and enter the heights and the frequencies in the first two columns.

Height in inches (x)	Frequency (f)	Product (f · x)	Deviation from Mean (x-μ)	Squared Deviations (x-μ)²	Product (x-μ)² · f
67	1	67	-3.35	11.2225	11.2225
68	2	136	-2.35	5.5225	11.0450
69	3	207	-1.35	1.8225	5.4675
70	4	280	-0.35	0.1225	0.4900
71	6	426	0.65	0.4225	2.5350
72	2	144	1.65	2.7225	5.4450
73	1	73	2.65	7.0225	7.0225
74	1	74	3.65	13.3225	13.3225
Total	20	1,407			56.5500
		÷			÷
n:	20			n:	20
		=			=
Mean:	70.35			Variance:	2.828
				Standard Deviation:	1.682

Results: $s^2 = 2.828$

$s = \sqrt{s^2} = 1.682$

Answer: A

23) Find the population variance.

   a) 2.209   b) 2.541   c) 2.828   d) 2.976

From the calculations above, $s^2 = 2.828$ **Answer: C**

24) Test scores for an anatomy and physiology class had a mean of 79 with a standard deviation of 4.5. Test scores for the AP Government class had a mean of 69 with a standard deviation of 3.7. Suppose a student get a 65 on the anatomy and physiology test and an 64 on the AP Government test. Calculate the z-score for each test. On which test did the student perform better?

a) Anatomy and Physiology test   b) AP Government test   c) Equally as well on both

$\mu_{anatomy} = 79$ $\sigma_{anatomy} = 4.5$ $x_{anatomy} = 65$

$$z_{anatomy} = \frac{x - \mu}{\sigma} = \frac{65 - 79}{4.5} = -3.11$$

$\mu_{government} = 69$ $\sigma_{government} = 3.7$ $x_{government} = 64$

$$z_{government} = \frac{x - \mu}{\sigma} = \frac{64 - 69}{3.7} = -1.35$$

So, relative to his peers, the student did better on the Government test because his z-score is higher.   **Answer: B**

25) The mean length of the first 20 space shuttle flights was about 7 days, and the standard deviation was about 2 days. On the basis of Chebychev's Theorem, at least how many of the flights lasted between 3 days and 11 days?

a) 4   b) 10   c) 15   d) 75

Chebychev's Theorem:

If $k > 1$, the portion of any data set lying within k σ's of μ is at least $1 - \frac{1}{k^2}$.

Now, on to solving this problem. Here are the steps:

1. Calculate how many days there are between the mean (7 days) and the top or bottom of the desired range. The range "3 to 11 days" is within 4 days of the mean of 7 days. That is, $[3, 11] = \{7 \pm 4\}$.

2. 4 days divided by a standard deviation of 2 days is 2 standard deviations, so we have $k = 2$.

3. Next we calculate the fraction in **Chebychev's Theorem:** $1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = \frac{3}{4}$.

4. So, we multiply $\frac{3}{4}$ by the number of flights to get $\frac{3}{4} \cdot 20 = 15$ flights between 3 and 11 days.

      Answer: C

I will do the next set of problems using the TI-84 calculator and will show you which functions I use. **You must be able to use the calculator to do these problems.** For those who need help using the calculator, there are several TI-84 resources available at:

http://www.mathguy.us/BySubject/Statistics.php

For normal curve probabilities, use: normalcdf(lower bound, upperbound, μ, σ)

26) Find $P(z < -2.10)$

a) 0.0179
b) 0.9821
c) 0.9778
d) 0.0222

Using the TI-84 or similar calculator, we calculate:

normalcdf(-10000, -2.10, 0, 1) = **0.0179**   **Answer: A**

27) In a recent year, the sophomore class taking the Nevada High School Proficiency Mathematics exam had a mean score of 330 with a standard deviation of 34. Assume that the scores are normally distributed. Find the probability that a student had a score higher than 303.

a) 0.2136
b) 0.1567
c) 0.7864
d) 0.8433

Using the TI-84 or similar calculator, we calculate:

normalcdf(303, 10000, 330, 34) = **0.7864**   **Answer: C**

28) A normal distribution has a mean of 65.1 and a standard deviation of 12.3. Find the probability of an x value less than 35.5.

a) 0.9542
b) 0.0458
c) 0.9919
d) 0.0081

Using the TI-84 or similar calculator, we calculate:

normalcdf(-10000, 35.5, 65.1, 12.3) = **0.0081**   **Answer: D**

29) The average number of hours per week a high school student watches television is 12.5 hours, with a standard deviation of 2.4 hours. Find the probability of a student watching between 10.6 and 13.8 hours of television.

a) 0.3257
b) 0.6743
c) 0.5083
d) 0.4917

Using the TI-84 or similar calculator, we calculate:

normalcdf(10.6, 13.8, 12.5, 2.4) = **0.4917**   **Answer: D**

30) Find $P(z > -0.74)$

a) 0.7704
b) 0.2296
c) 0.1251
d) 0.8749

Using the TI-84 or similar calculator, we calculate:

normalcdf$(-0.74, \ 10000, \ 0, \ 1) = \mathbf{0.7704}$ **Answer: A**

Next set of problems

Given a normal curve probability, find a z-score with: invNorm(area, μ, σ)

31) The area to the right of a z-score is 0.5871. What is the z-score?

a) 0.78
b) 0.22
c) –0.78
d) –0.22

The area that is input into the calculator is the area to the left of the z-score. So first, we must calculate the area on the left: $1 - 0.5871 = 0.4129$.

Using the TI-84 or similar calculator, we calculate:

$z = $ invNorm$(0.4129, \ 0, \ 1) = \mathbf{-0.22}$ **Answer: D**

32) The probability of being lower than a specific z-score is 4.9%. What is the value of the z-score?

a) –1.65
b) 1.65
c) –0.03
d) 0.03

The area that is input into the calculator is the area to the left of the z-score. This is what we are given in this problem. Recall that $4.9\% = 0.049$.

Using the TI-84 or similar calculator, we calculate:

$z = $ invNorm$(0.049, \ 0, \ 1) = \mathbf{-1.65}$ **Answer: A**

33) What z-score is at the 60th percentile?

a) 0.25
b) –0.25
c) 0.75
d) –0.75

The 60th percentile is the point with an area of 0.60 to the left of the point. Recall that $60\% = 0.60$.

Using the TI-84 or similar calculator, we calculate:

$z = $ invNorm$(0.60, \ 0, \ 1) = \mathbf{+0.25}$ **Answer: A**

34) SAT English scores are normally distributed with a mean score of 690 and a standard deviation of 75. What score is at the 80th percentile?

a) Around 753
b) Around 754
c) Around 761
d) Around 765

The 80th percentile is the point with an area of 0.80 to the left of the point. Recall that $80\% = 0.80$.

Using the TI-84 or similar calculator, we calculate:

$$z = \text{invNorm}(0.80, \ 0, \ 1) = +0.8416$$

We must find the value of x associated with this z-score.

We are given: $\mu = 690$, $\sigma = 75$; and we just calculated $z = +0.8416$. Write the equation for z, substitute in the appropriate values and use some Algebra to calculate x.

$$z = \frac{x - \mu}{\sigma} \quad \Rightarrow \quad +0.8416 = \frac{x - 690}{75} \quad \Rightarrow \quad x = 753.1$$

Answer: A

35) The mean home price in Reno is $354,889, with a standard deviation of $45,600. 45% of the homes in Reno are more expensive than John's home. What is the price of John's home?

a) $299,414.51
b) $355,481.97
c) $360,619.16
d) $375,577.23

45% of homes are **more** expensive. This is the area to the **right of the mean**. We want the area to the left of the mean. $100\% - 45\% = 55\% = 0.55$.

Using the TI-84 or similar calculator, we calculate:

$$z = \text{invNorm}(0.55, \ 0, \ 1) = +0.125661$$

We must find the value of x associated with this z-score.

We are given: $\mu = 354889$, $\sigma = 45600$; and we just calculated $z = +0.125661$. Write the equation for z, substitute in the appropriate values and use some Algebra to calculate x.

$$z = \frac{x - \mu}{\sigma} \quad \Rightarrow \quad +0.125661 = \frac{x - 354889}{45600} \quad \Rightarrow \quad x = \$360,619$$

Answer: C

36) The average age of males entering the military is 18.4 years old, with a standard deviation of 0.7. random samples of 175 recent enlistees are taken. What is the standard error of the mean of the sampling distribution?

a) 0.053
b) 0.647
c) 0.755
d) 0.080

The formula for the standard error of the mean is: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

We are given $\sigma = 0.7$ and $n = 175$.

So, we calculate: $\sigma_{\bar{x}} = \dfrac{0.7}{\sqrt{175}} = 0.0529.$ **Answer: A**

37) Electric bills for residents of Reno have a mean of $120 and a standard deviation of $12. Random samples of 45 electric bills are drawn from this population and the mean of each sample is determined. Find the mean of the sampling distribution.

a) $12.8
b) $45.2
c) $75
d) $120

The mean of the sampling distribution is the same as the mean of the population. That is, $\mu_{\bar{x}} = \mu$. So, $\mu_{\bar{x}} = \$120$.   **Answer: D**

38) A random sample of 150 students has a grade point average with a standard deviation of 0.78. Find the margin of error if $c = 0.98$.

a) 0.15   b) 0.08   c) 0.11   d) 0.12

Since the sample is size 30 or over, we use a z-statistic. If it were under size 30, we would use a t-statistic. The formula for the margin of error is:

$$E = z_c \cdot \frac{\sigma}{\sqrt{n}}$$

We are given: $c = .98$, $\sigma = 0.78$, $n = 150$.   Then,

$$E = z_c \cdot \frac{\sigma}{\sqrt{n}} = z_{.98} \cdot \frac{0.78}{\sqrt{150}} = 2.326 \cdot \frac{0.78}{\sqrt{150}} = \mathbf{0.1482}$$

Answer: A

Note: I calculated z_c on the TI-84 using the formula: $z_c = \text{invNorm}\left(0.5 + \frac{c}{2}, \ 0, \ 1\right)$ since the margin of error implies a two-tail test.

So, $z_{.98} = \text{invNorm}\left(0.5 + \frac{0.98}{2}, \ 0, \ 1\right) = \text{invNorm}(0.99, \ 0, \ 1) = 2.326$

CALCULATOR TRICK: You can calculate the margin of error directly on your calculator by assuming a mean of zero. On the TI-84:

1.   STAT – RIGHT – RIGHT (to TESTS) – 7:ZInterval
2. Inpt: STATS
3. Enter your values for σ, n, and C-$Level$, and set $\bar{x} = 0$
4. Move the cursor down to CALCULATE and hit ENTER

The margin of error will show up in parentheses as a negative value and a positive value. Use the positive value. I was able to do that for this problem without using the formula for E.

39) Estimate the standard deviation of the normal curve shown below.
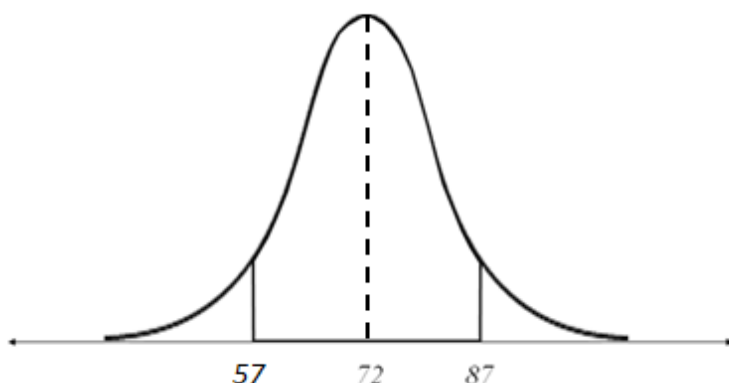
a) 72   b) 15   c) 30   d) 57

Plots like this show the mean in the middle and vertical lines at intervals of one standard deviation to the left and right of the mean. So, $\sigma = 87 - 72 = 72 - 57 = 15$.

Answer: B

40) The speeds of vehicles along a stretch of highway are normally distributed, with a mean of 72 miles per hour and a standard deviation of 7 miles per hour. Find the speed x corresponding to a z-score of 1.96.

a) 79.37

b) 84.26

c) 78.83

d) 85.72

We can algebraically manipulate the formula for z: $z = \dfrac{x - \mu}{\sigma}$ to obtain the following formula for x: $x = \mu + z\sigma$.

We are given: $\mu = 72$, $\sigma = 7$, $z = +1.96$. Then,

$$x = \mu + z\sigma = 72 + 1.96 \cdot 7 = \mathbf{85.72}$$ **Answer: D**

41) Scores for the ASVAB (Armed Services Vocational Aptitude Battery) test are normally distributed, with a mean of 88 and a standard deviation of 4.9. To be eligible to join the military, you must score in the top 25%. What is the lowest score you can earn and still be eligible to enlist?

a) 88

b) 91

c) 92

d) 95

We want to score in the top 25%, so our z-score would have an area of 75% to its left. We need: $z = \text{invNorm}(0.75,\ 0,\ 1) = +0.6745$

Once again, we will use the formula: $x = \mu + z\sigma$.

We have: $\mu = 88$, $\sigma = 4.9$, $z = +0.6745$. Then,

$$x = \mu + z\sigma = 88 + 0.6745 \cdot 4.9 = \mathbf{91.30}$$ **Answer: C**

Note that a score of $\mathbf{91}$ is below $\mathbf{91.3}$, so you need a $\mathbf{92}$ to make the cut.

42) In order to set rates, an insurance company is trying to estimate the number of sick days that full time workers at an auto repair shop take per year. A previous study indicated that the standard deviation was 2.8 days. How large a sample must be selected if the company wants to be 95% confident that the true mean differs from the sample mean by no more than 1 day?

a) 141 b) 1024 c) 31 d) 512

Let's assume the sample is size 30 or over so we can use a z-statistic. We can revisit this if we are incorrect. The formula for the margin of error is:

$$E = z_c \cdot \frac{\sigma}{\sqrt{n}} \quad \text{which we can manipulate to get:} \quad n = \left(z_c \cdot \frac{\sigma}{E}\right)^2$$

We are given: $c = .95$, $\sigma = 2.8$, $E = 1$. Then,

$$z_{.95} = \text{invNorm}\left(0.5 + \frac{0.95}{2}, \ 0, \ 1\right) = \text{invNorm}(0.975, \ 0, \ 1) = 1.960$$

$$n = \left(z_c \cdot \frac{\sigma}{E}\right)^2 = \left(z_{.95} \cdot \frac{2.8}{1}\right)^2 = \left(1.960 \cdot \frac{2.8}{1}\right)^2 = 30.1$$

A sample size of 30 is too small, and we must have an integer sample size. So, we round up to get $n = 31$. Note that the sample size is ≥ 30, so we are okay using a z-statistic.

Answer: C

43) Assume that the salaries of elementary teachers in the United States are normally distributed with a mean of $32,000, and a standard deviation of $3000. If 100 teachers are randomly selected, find the probability that their mean salary is greater than $32,500.

a) 0.9525
b) 0.1312
c) 0.0478
d) 0.3312

We are given $x = 32500$, $\mu = 32000$, $\sigma = 3000$, $n = 100$.

Since $n \geq 30$, we can use the z-statistic. We can also simplify things by using the TI-84's normalcdf function, entering $\sigma_{\bar{x}}$ as the standard deviation.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3000}{\sqrt{100}} = 100$$

The probability of salaries exceeding $32,500, then, is:

$$\text{normalcdf}(32500, \ 100000, \ 32000, \ 300) = \mathbf{0.0478}$$

Answer: C

44) A random sample of 80 students has a test score with $\bar{x} = 79.5$ and $s = 12.5$. Construct the confidence interval for the population mean, μ if $c = 0.90$.

a) $(74.9, 86.3)$ b) $(77.2, 81.8)$
c) $(76.4, 82.5)$ d) $(75.3, 80.6)$

First, note that $n \geq 30$, so we can use a z-statistic.

The formula for the confidence interval is: $\bar{x} \pm z_c \cdot \dfrac{\sigma}{\sqrt{n}}$

We are given: $\bar{x} = 79.5$, $c = .90$, $\sigma = 12.5$, $n = 80$.   Then,

CI: $\bar{x} \pm z_c \cdot \dfrac{\sigma}{\sqrt{n}} = \bar{x} \pm z_{.90} \cdot \dfrac{12.5}{\sqrt{80}} = 79.5 \pm 1.645 \cdot \dfrac{12.5}{\sqrt{80}} = \mathbf{79.5 \pm 2.3}$

$79.5 - 2.3 = \mathbf{77.2}$ $79.5 + 2.3 = \mathbf{81.8}$

Answer: B

Note: I calculated z_c on the TI-84 using the formula: $z_c = \text{invNorm}\left(0.5 + \frac{c}{2},\ 0,\ 1\right)$ since this problem requires two tails.

So, $z_{.90} = \text{invNorm}\left(0.5 + \frac{0.90}{2},\ 0,\ 1\right) = \text{invNorm}(0.95,\ 0,\ 1) = 1.645$

45) A group of 35 bowlers showed that their average score was 225 with a standard deviation of 12. Find the 95% confidence interval of the mean score of all bowlers.

a) $(219, 227)$ b) $(220, 224)$
c) $(226, 234)$ d) $(221, 229)$

First, note that $n \geq 30$, so we can use a z-statistic.

The formula for the confidence interval is: $\bar{x} \pm z_c \cdot \dfrac{\sigma}{\sqrt{n}}$

We are given: $\bar{x} = 225$, $c = .95$, $\sigma = 12$, $n = 35$.   Then,

CI:: $\bar{x} \pm z_c \cdot \dfrac{\sigma}{\sqrt{n}} = \bar{x} \pm z_{.95} \cdot \dfrac{12}{\sqrt{35}} = 225 \pm 1.96 \cdot \dfrac{12}{\sqrt{35}} = \mathbf{225 \pm 4}$

$225 - 4 = \mathbf{221}$ $225 + 4 = \mathbf{229}$

Answer: D

Note: I calculated z_c on the TI-84 using the formula: $z_c = \text{invNorm}\left(0.5 + \frac{c}{2},\ 0,\ 1\right)$ since this problem requires two tails.

So, $z_{.95} = \text{invNorm}\left(0.5 + \frac{0.95}{2},\ 0,\ 1\right) = \text{invNorm}(0.975,\ 0,\ 1) = 1.96$

46) Construct a 95% confidence interval for the population mean, μ. Assume the population has a normal distribution. In a recent study of 21 seniors, the mean number of hours per week that they spent doing homework was 17.1, with a standard deviation of 6.7 hours.

a) (14.6, 19.6)   b) (13.6, 19.8)   c) (14.1, 20.2)   d) (12.9, 21.3)

We are told the distribution is **Normal**, but the population σ is not known. The sample is small and we must substitute the sample standard deviation, s, for σ, so we must use the t-distribution.

The formula for the confidence interval is: $\bar{x} \pm t_c \cdot \dfrac{\sigma}{\sqrt{n}}$

We are given: $\bar{x} = 17.1,\quad c = .95,\quad \sigma = 6.7,\quad n = 21,\quad df = 21 - 1 = 20$

Calculated t_c on the TI-84 using the formula: $t_c = \text{invT}\left(0.5 + \dfrac{c}{2},\ 20\right)$ since this problem requires two tails.

$t_{.95} = \text{invT}\left(0.5 + \dfrac{0.95}{2},\ 20\right) = \text{invT}(0.975,\ 20) = 2.086$

Then,

CI: $\bar{x} \pm t_c \cdot \dfrac{\sigma}{\sqrt{n}} = \bar{x} \pm t_{.95} \cdot \dfrac{6.7}{\sqrt{21}} = 17.1 \pm 2.086 \cdot \dfrac{6.7}{\sqrt{21}} = \mathbf{17.1 \pm 3.0}$

$17.1 - 3.0 = \mathbf{14.1}$ $17.1 + 3.0 = \mathbf{20.1}$ **Answer: C**

47) The standard IQ has a mean of 100 and a standard deviation of 7. We want to be 95% certain that we are within 3 IQ points of the true mean. Determine the required sample size.

a) 153   b) 3   c) 21   d) 97

Let's assume that IQ's are Normally distributed so we can use a z-statistic. The formula for the margin of error is:

$E = z_c \cdot \dfrac{\sigma}{\sqrt{n}}$ which we can manipulate to get: $n = \left(z_c \cdot \dfrac{\sigma}{E}\right)^2$

We are given: $c = .95,\quad \sigma = 7,\quad E = 3.\quad z_{.95} = 1.96$ from Problem 45.   Then,

$n = \left(z_c \cdot \dfrac{\sigma}{E}\right)^2 = \left(z_{.95} \cdot \dfrac{7}{3}\right)^2 = \left(1.96 \cdot \dfrac{7}{3}\right)^2 = 21$

                                        Answer: C

48) The mean life for a Toyota Camry is at most 17.9 years. State this claim mathematically. Write the null and alternative hypotheses. Identify which is the claim.

a) $H_0 : \mu \le 17.9$ (claim)   b) $H_0 : \mu \ge 17.9$
    $H_a : \mu > 17.9$ $H_a : \mu < 17.9$ (claim)

c) $H_0 : \mu > 17.9$ (claim)   d) $H_0 : \mu < 17.9$ (claim)
    $H_a : \mu \le 17.9$ $H_a : \mu \ge 17.9$

The **null hypothesis, H_0,** is a statistical hypothesis that contains a statement of equality (i.e., it contains an equal sign, such as $=, \le$ or $\ge$).

The **alternative hypothesis, H_a or H_1,** is the complement of the null hypothesis. It contains a statement of strict inequality (using $<$ or $>$).

"At most 17.9 years" means "≤ 17.9 years." So, $H_0: \mu \le 17.9$

Answer: A

49) A researcher claims that 62% of voters favor universal health care. Determine whether the hypotheses test for this claim is left-tailed, right-tailed, or two-tailed.
a) left-tailed   b) right-tailed   c) two-tailed

We are interested in whether the claim of 62% is either too high or too low. Therefore, we would use a **two-tailed test**. If we were concerned about only too high or too low, instead of both, we would use a one-tailed tests.

Answer: C

50) The mean score for all NFL games during a particular season was less than 30 points per game. Identify the type II error for the hypothesis test of this claim.
a) Reject H_0, when $\mu \geq 30$
b) Fail to reject H_0, when $\mu < 30$.
c) Not enough sufficient evidence to decide.

Type I Error: Reject the null hypothesis when it is true.

Type II Error: Do not (i.e., fail to) reject the null hypothesis when it is false.

For this problem, we have:

H_0: $\mu \geq 30$

H_a: $\mu < 30$

So a type II error would be failing to reject H_0 when H_0 is false (i.e., $\mu < 30$).   **Answer: B**

Type I and Type II Errors		
Action Taken	H_0 is TRUE	H_0 is FALSE
Do Not Reject H_0	Correct action	Type II Error
Reject H_0	Type I Error	Correct action

51) Given H_0: $\sigma \geq 0.05$ and H_a: $\sigma < 0.05$, determine whether the hypotheses test is left-tailed, right-tailed, or two-tailed.

a) two-tailed   b) left-tailed   c) right-tailed

The rejection region is where the tail of the test is; we look to H_a to see where that tail is. Since a less than sign indicates a location on the left, this is a **left-tailed test.**

Answer: B

52) For a hypothesis test, the P-value is 0.0641 and the level of significance is $\alpha = 0.05$. What decision should be made?

a) Reject H0   b) Fail to reject H0   c) No conclusion can be made

About α and the P-value

α is the significance level of the test.

The **P-value** is the probability of obtaining a test statistic at least as extreme as the one observed from the sample data, assuming the H_0 is true. It is the probability (area under the curve) that you can calculate when you apply your z-test or t-test to the sample.

Rules for rejecting H_0.
- If $P < \alpha$, reject H_0.
- If $P \geq \alpha$, do not reject H_0.

For this problem, $(P = 0.0641) \geq (\alpha = 0.05)$, so we would **fail to reject H_0.**

Answer: B

Use the following information for #53-55.

A tea drinker's society estimates that the mean consumption of tea by a person in the United States is more than 8 gallons per year. In a random sample of 100 people, you find that the mean consumption is 7.9 gallons per year with a standard deviation of 2.67 gallons.

53) What is the critical value at $\alpha = 0.07$?

a) -1.45   (b) 1.45   c) -1.81   d) 1.81

Let's set his up:

H_0: $\mu \leq 8$ gallons (note: $\mu = 8$)

H_a: $\mu > 8$ gallons

$n = 100$, $\bar{x} = 7.9$, $\sigma = 2.67$

Since H_a contains a " > " sign, this would be a right tailed test with $\alpha = .07$.

We use a z-statistic because $n \geq 30$. We would use a t-statistic if $n < 30$.

To find the critical value for a right-hand test, we look up:

$z = \text{invNorm}(1 - \alpha, \ 0, \ 1) = \text{invNorm}(0.93, \ 0, \ 1) = +1.476$

This is roughly the answer in B, so let's take that one.   **Answer: B**

54) What is the P-value?

a) 0.3229   b) 0.3745   c) 1.2918   (d) 0.6460

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{7.9 - 8}{2.67 / \sqrt{100}} = -0.3745$$

$P(z > -0.374) = \text{normalcdf}(-0.3745, \ 10{,}000, \ 0, \ 1) = 0.6460$ **Answer: D**

⌐ Note: use a " > " sign, consistent with H_a, because we are using a right-tailed test.

Alternatively, you could calculate the standard error of the mean, then calculate the P-value as follows:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.67}{\sqrt{100}} = 0.267$$

$P(z > -0.374) = \text{normalcdf}(7.9, \ 10{,}000, \ 8, \ 0.267) = 0.6460$

55) What is the correct interpretation of the hypothesis test?

    a) Reject the claim that the mean consumption is more than 8 gallons.
    b) There is not enough evidence to reject the claim that the mean consumption is more than 8 gallons.
    c) Support the claim that the mean consumption is more than 8 gallons.
    d) There is not enough evidence to support the claim that the mean consumption is more than 8 gallons.

When we do a hypothesis test, we determine whether we can reject H_0. Remember that H_a is our claim. When we fail to reject H_0, we do not have enough evidence to support H_a.

In the information presented above, the claim is that $\mu > 8$ gallons, as expressed in H_a. In order to reject H_0, we need a z-value above the critical value of 1.476 (from problem 53). Our z-value (from problem 54) is -0.374.

For this problem,
- The claim is H_a: $\mu > 8$ gallons (i.e., consumption exceeds 8 gallons).
- We cannot reject H_0; therefore, there is not enough evidence to support H_a.

Answer: D

For #56-59: As part of your work for an environmental awareness group, you want to test a claim that the mean waste generated by adults in the United States is more than 4 pounds per day. In a random sample of 10 adults in the United States, you find that the mean waste generated per person per day is 4.54 pounds with a standard deviation of 1.21 pounds. The level of significance is $\alpha = 0.05$.

56) What type of test should be done to test the hypothesis?
    a) A right-tailed t-test   b) A right-tailed z-test
    c) A two-tailed t-test   d) A two-tailed z-test

Let's set his up:
    H_0: $\mu \leq 4$ pounds per day
    H_a: $\mu > 4$ pounds per day

    $n = 10,\ \ \bar{x} = 4.54,\ \ \sigma = 1.21$

Since H_a contains a " $>$ " sign, this would be a right tailed test with $\alpha = .05$.

Answer: A

57) Find the critical value(s).
a) -1.645, 1.645   b) 1.645   c) 1.833   d) -1.833, 1.833

We use a t-statistic because $n < 30$. We would use a z-statistic if $n \geq 30$.

We have $n - 1 = 10 - 1 = 9$ degrees of freedom.

To find the critical value for a right-hand test, we look up:

$$z = \text{invT}(1 - \alpha, \ df) = \text{invT}(0.95, \ 9) = +1.833 \qquad \textbf{Answer: C}$$

58) Which answer shows the correct rejection region(s)?
a) t < -1.645 and t > 1.645   b) t < -1.833 and t > 1.833
c) t > 1.645   d) t > 1.833

We are using a right-tailed test, so the rejection region is to the right of the critical region. So, we reject when $t > 1.833$.

Answer: D

Remember that this is one time in life when rejection is good. When we reject H_0, we accept H_a. So the rejection region is where the result lie that allow us to conclude something (H_a is "true").

59) Interpret the results of the hypothesis test.
a) Reject the claim that the mean waste is more than 4 pounds per day
b) There is not enough evidence to reject the claim that the mean waste is more than 4 pounds per day.
c) Support the claim that the mean waste is more than 4 pounds per day.
d) There is not enough evidence to support the claim that the mean waste is more than 4 pounds per day.

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{4.54 - 4}{1.21 / \sqrt{10}} = 1.411$$

In order to reject H_0, we need a t-value above the critical value of 1.833 (calculated in Problem 57). We do not have this. Therefore, there is not enough evidence to support this claim. So …

For this problem,
- The claim is H_a: $\mu > 4$ pounds per day.
- We cannot reject H_0; therefore, there is not enough evidence to support H_a.

Answer: D

Use the data for home prices for # 60 – 66.

Price (Thousands of $)	$140	$170	$190	$220	$240	$260	$280
Sales of New Homes This Year	136	114	94	75	82	40	20

60) Find the equation of the regression line.

a) $\hat{y} = -0.79x + 249.86$ b) $\hat{y} = 246.16x + 0.77$

c) $\hat{y} = -0.68x + 232.57$ d) $\hat{y} = -0.77x + 246.16$

You do this problem with your TI-84 using the following steps:

1. Enter the Price (x-variable) as List 1 (STAT – EDIT – enter data in L1)
2. Enter the Sales (y-variable) as List 2 (STAT – EDIT – enter data in L2)
3. Perform the linear regression on List 1 and List 2 (STAT – CALC – 8:Linreg(a+bx))

My TI-84 provides the following output:
$y = a + bx$
$a = 246.1638418$
$b = -.7747645951$
$r^2 = .9372170732$
$r = -.9680997228$

We can use this information to answer Problems 60 to 65

For problem 60, the regression line is: $\hat{y} = a + bx$ or $\hat{y} = 246.16 - 0.7748x$

Answer: D

61) What is the value of the correlation coefficient, r?
a) 0.94   b) -0.97   c) -0.95   d) 0.82

This value is give above in magenta. $r = -0.97$ **Answer: B**

62) Predict the number of new homes sold at a price of $300,000.
a) around 15   b) around 14   c) around 13   d) around 12

Using the regression line: $\hat{y} = 246.16 - 0.7748x$, we substitute in a value of $x = 300$ to get:

$\hat{y} = 246.16 - 0.7748(300) = 13.72$ **Answer: B**

Note that this problem requires you to extrapolate beyond the range of the data provided, which is generally a bad idea.

63) 200 homes are sold this year at a certain price. Use the regression line to estimate the price.
a) around $81,860   b) around $74,502   c) around $62,394   (d) around $59,948

y is the number of homes sold, and x is the price (in 000's) of the homes sold.

Using the regression line: $\hat{y} = 246.16 - 0.7748x$, we substitute in a value of $y = 200$ to get:

$200 = 246.16 - 0.7748(x)$

Solve for x to get $x = \mathbf{59.6\ thousand}$ **Answer: D**

Note that this problem requires you to extrapolate beyond the range of the data provided, which is generally a bad idea.

64) Find the residual value for a home priced at $240,000.
a) 21.74   b) -21.74   c) 20.64   d) -20.64

The residual is the difference between the actual value and its corresponding predicted value. The formula is: $e = y - \hat{y}$, where e is the value of the residual, y is the actual value and $\hat{y}$ is the predicted value.

For a home priced at $240,000, $x = 240$. $\hat{y} = 246.16 - 0.7748(240) = 60.21$

From the chart in Problem 60, we see that $y = 82$. Then,

$e = y - \hat{y} = 82 - 60.21 = \mathbf{21.79}$ **Answer: A**

65) What percentage of the data can be explained by the regression line?
a) 94%   b) 97%   c) 79%   d) 83%

This question should ask: "What percentage of the **variation** in the data can be explained by the regression line?" As stated, the question is unanswerable. As modified by adding the word "variation," we want the value of R^2, expressed as a percentage.

We can read this value from Problem 60. $R^2 = 0.9372170732 \sim \mathbf{0.94} = \mathbf{94\%}$

Answer: A

66) A collection of a set of data (x) has a mean 14 with a standard deviation of 2.4. Another variable (y) has a mean of 28 with a standard deviation of 5. The correlation coefficient is -0.94. Find the equation of the linear regression line.

a) $\hat{y} = -1.96x + 55.44$

b) $\hat{y} = 1.96x + 0.56$

c) $\hat{y} = -0.45x + 34.3$

d) $\hat{y} = 0.45x + 21.7$

The equation of the regression line is:

$$\hat{y} = a + bx \quad \text{where:} \quad b = r\frac{s_y}{s_x} \quad \text{and} \quad a = \bar{y} - b\bar{x}$$

We are given: $\bar{x} = 14, \quad s_x = 2.4, \quad \bar{y} = 28, \quad s_y = 5, \quad r = -0.94$

Then, $b = r\frac{s_y}{s_x} = -0.94\frac{5}{2.4} = -1.96$

Since the only answer with this coefficient of x is **A**, that must be the solution: **Answer: A**

Let's calculate a anyway.

$a = \bar{y} - b\bar{x} = 28 - (-1.96)14 = 55.44$