Name _____ Date _____ Per_____ #_____

AP REVIEW PACKET #4 INSTRUCTIONS – Due by Thursday 3/28 7:50 – 30 points

1 Read Topics 6 & 7: Overview of Methods of Data Collection and Planning and Conducting Surveys, pp 161-185, in the Barron's book. Yay, new theme! As you read, answer the guided reading questions. Look over the practice problems at the end of each topic.

2) Answer the multiple choice questions in this packet, these will be graded for accuracy. Show all work and then copy your answers into the blanks below.

26. _____ 27. _____ 28. _____ 29. _____ 30. _____

3) Answer the FRAPPY Questions.

Optional Bonus Activity: Go to http://learner.org/resources/series65.html. Watch programs 4& 5 & complete the worksheet.

Guided Reading Questions for Packet #5:

1) In studying statistics we learn how to estimate _____ characteristics by considering a _____.

2) Analysis of _____ gathered sample data is usually a _____ exercise.

3) A simple random sample is one in which _____ _____sample of the desired size has an

_____ _____ of being selected.

4) What are the 9 sources of bias in surveys? (Note some of the language is different than the textbook, that's OK.)

5) Why don't the alternative procedures result in a simple random sample?

Multiple Choice Questions. 26. A baseball pitcher has a history of throwing only 40% of his pitches for strikes. The coach has suspended the player until he improves his strike throwing percentage to at least 55%. If a practice session of 40 pitches is scheduled, and the player has not actually improved, what is the probability that the player will throw at least 55% of his pitches for strikes?

(A) $P\left(z > \frac{0.55-0.40}{\sqrt{\frac{0.40(0.60)}{40}}}\right)$ (B) $P\left(z > \frac{0.55-0.40}{\sqrt{\frac{0.50(0.50)}{40}}}\right)$ (C) $P\left(z > \frac{0.55-0.40}{\sqrt{\frac{0.55(0.45)}{40}}}\right)$ (D) $\binom{40}{22}(0.40)^{22}(0.60)^{18}$ (E) $\binom{40}{22}(0.55)^{22}(0.45)^{18}$

27. In a recent high school basketball tournament where over 750 games were played, the mean team score was 68 points and the standard deviation was 13 points. The scores were approximately normally distributed. A coach was overheard saying that his team scored 95 points in one game. About what proportion of teams' scores during the tournament were more than 95 points?

(A) 0.0035 (B) 0.0190 (C) 0.05 (D) 0.9810 (E) 2.07

28. The traffic safety officer of a local police force was trying to see if there was an association between the number of cars that did not use a main intersection in town because of the traffic light and the number of tickets written for speeding on the alternate route. The correlation between these two variables was found to be 0.58. Which of the following statements is true?

(A) About 58% of the variation in the number of speeding tickets can be explained by the linear relationship between the number of speeding tickets issued and the number of cars that did not use the main intersection in town.
(B) Any potential linear relationship between the number of cars not using the main intersection in town and the number of speeding tickets written on an alternate route would be positive.
(C) If one uses the main intersection, one is 58% more likely to receive a ticket than using the alternate route.
(D) Since the correlation is not close to 1, there cannot be a linear relationship between the number of cars not using the main intersection in town and the number of speeding tickets written on an alternate route.
(E) Getting a speeding ticket is a direct cause of taking the alternate route.

29. A major automobile manufacturer is trying to improve its customer service at its dealerships across the United States. A survey of 200 customers in Arizona who recently purchased a vehicle from this manufacturer were asked if they were satisfied with the customer service at the dealership. Is it reasonable to generalize the conclusion to the population of all customers in the United States that purchased from this manufacturer?

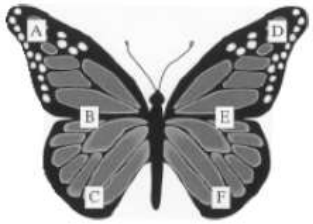
(A) No, because customers were only sampled in one state.
(B) No, because 200 is not a large enough sample.
(C) No, because only one sample was taken.
(D) Yes, because the sample size is more than 30.
(E) Yes, because the population of all new vehicle owners by this manufacturer is more than 2,000.

30. A politician is considering running for public office. He wants to measure his name recognition by doing a survey of voters in his district. Which of the following survey methods would produce unbiased results?

(A) The politician stands in front of a grocery store in his district on Saturday morning and asking each person entering the store if he or she recognizes the politician's name.
(B) Placing pollsters in front of every grocery store in his district on Saturday morning and asking each person entering the store if he or she recognizes the politician's name.
(C) Sending a survey card to all registered voters in the district asking them to call a phone number to state whether or not they recognize the politician's name.
(D) Calling people from his district listed in the phone book and asking each if he or she recognizes the politician's name.
(E) None of these methods would produce unbiased results.

FRAPPY Questions:

1) Researchers often mark wildlife in order to identify particular individuals across time or space. A study of buttery migration is designed to determine which location on the butteries' wings is best for marking. The six possible locations are those shown as A through F in the figure to the right. The buttery in the figure is a monarch (Danaus plexippus). Because marks in certain locations may be more likely to attract predators or cause problems than marks in other locations, the goal is to determine whether the six marking locations result in equivalent chances of successful migration. To test this, researchers plan to mark 3,600 butteries and release them, then count how many arrive displaying each marking location at the end of the migratory path.



(a) Briefly describe a method you could use to assign the marking locations if you wanted to ensure that exactly 600 butteries were marked in each location.

(b) Briefly describe a method you could use to assign the marking locations if you wanted to be independent from one buttery to the next, and wanted each location assigned with a probability 1/6 each time.

(c) Using your method of assignment from part (b), explain how you would analyze the data collected from this study.

(d) If butteries are marked using your method of assignment from part (a), would you change your method of analysis? Explain your reasoning.

2) The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and in people who eat less than on apple a week. They are going to conduct a study with 50 people in each group.

Fifty people who report that they routinely eat an apple a day and 50 clinic patients who report that they eat less than one apple a week will be identified. The dentists will examine the patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups.

a) Why is this an observational study and not an experiment?

b) Explain the concept of confounding in the context of this study. Include an example of a possible confounding variable.

c) If the mean number of new cavities for those who ate an apple a day was statistically significantly smaller than the mean number of new cavities for those who ate less than one apple a week, could one conclude that the lower number of new cavities can be attributed to eating an apple a day? Explain.

3)Sunshine Farms wants to know whether there is a difference in consumer preference for two new juice products - Citrus Fresh and Tropical Taste. In an initial blind taste test, 8 randomly selected consumers were given unmarked samples of the two juices. The product that each consumer tasted first was randomly decided by the flip of a coin. After tasting the two juices, each consumer was asked to choose which juice he or she preferred and the results were recorded.

(a) Let p represent the population proportion of consumers who prefer Citrus Fresh. In terms of p, state the hypotheses that Sunshine Farms in interested in testing.

(b) One might consider using a one-proportion z-test to test the hypotheses in part (a). Explain why this would <u>not</u> be a reasonable procedure for this sample.

(c) Let X represent the number of consumers in the sample who prefer Citrus Fresh. Assuming there is no difference in consumer preference; find the probability for each possible value of X. Record the x-values and the corresponding probabilities in the table to the right.

x	p(x)

(d) When testing the hypotheses in part (a), Sunshine Farms will conclude that there is a consumer preference if too many or too few individuals prefer Citrus Fresh. Based on your probabilities in part (c), Is it possible for the significance level (probability of rejecting the null hypothesis when it is true) for this test to be exactly 0.05? Justify you're answer.

(e) The preference data for the 8 randomly selected consumers are given in the table to the right. Based on these preferences and your previous work, test the hypotheses in part (a).(on the back or another page)

Individual	Juice Preference
1	Tropical Taste
2	Citrus Fresh
3	Tropical Taste
4	Tropical Taste
5	Tropical Taste
6	Citrus Fresh
7	Tropical Taste
8	Tropical Taste

f) Sunshine Farms plans to add one of these two new juices - Citrus Fresh or Tropical Taste - to its production schedule. A follow-up study will be conducted to decide which of the two juices to produce. Make one recommendation for the follow-up study that would make it better than the initial study. Provide a statistical justification for your recommendation in the context of the problem.

Bonus Video: TIME SERIES

1. When we measure something at regular intervals over time we end up with a _____.

2. If a time series is stable we say that it is _____.

3. What is a smaller pattern that repeats through a time series called? _____

4. What is the length of time each cycle takes? _____

5. What refers to anything that varies with a yearly cycle? _____

6. What is the most important example of this variation with a yearly cycle? _____

7. What is the overall tendency to increase or decrease in a time series? _____

8. What is a method for smoothing time series data illustrated with the Boston Marathon data?

   _____

9. What two important issues must be examined when dealing with time series? _____

   _____