

Name _____ Date _____ Per _____

AP REVIEW PACKET #2 INSTRUCTIONS (30 points) – Due by Thursday 3/14 7:50am

1) Read Topic 3: Comparing Distributions, pp 89-103, in the Barron's book. As you read, answer the guided reading questions. It is highly suggested that you also do the questions at the end of the Topic Review.

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.

11. _____ 12. _____ 13. _____ 14. _____ 15. _____

3) Answer the Free-Response Questions in this packet

Optional Bonus Activity: Go to <http://learner.org/resources/series65.html>. Watch program 3 answer the questions on the worksheet.

Guided Reading Questions for Packet #2:

1) When asked for a comparison, making two separate lists is _____ and will be _____.

2) How do the shapes compare for Drug A and B?

3) Parallel boxplots are useful in presenting a _____ of the comparison of several distributions.

4) The interquartile range for U.S. population is _____ in 1860 and _____ in 1980.

5) To visually compare two or more distributions use:

- _____, either one above the other or _____ by _____.
- Double _____.
- _____, either one above the other or _____ by _____.
- _____ to _____ stemplots.
- _____ boxplots.
- _____ frequency plots on the same grid.
- For all of the above, make note of any _____ and _____ in _____, _____, and _____.

Packet Multiple Choice Questions:

11. A 95% confidence interval is to be calculated in order to find an estimate for a population proportion. What is the smallest sample size that will guarantee a margin of error of at most 5%?
- (A) 225 (C) 400
(B) 350 (D) 575
(E) 800
12. Some AP Statistics students were interested in finding out if there was a relationship between the number of hours of study for a chapter test and the score on that test. On the basis of the number of hours their classmates studied for the chapter 3 test and the scores on the test (out of 100%), the least-squares regression line calculated was $\hat{y} = 72.53 + 5.88x$, where x is the number of hours studied and \hat{y} is the predicted score on the test. Which statement correctly interprets the meaning of the slope of this regression line?
- (A) For each additional hour studied, the predicted score on the test increases by approximately 73%.
- (B) For each additional hour studied, the predicted score on the test increases by approximately 6%.
- (C) For each additional percent of increase on the test, the predicted score on the test increases approximately 73%.
- (D) For each additional percent of increase on the test, the predicted score on the test increases approximately 6%.
- (E) We cannot use this regression equation, since cause-effect has not been proven.
13. A group of statistics students want to know if there is a difference in how well their classmates like two of a beverage producer's latest juice blends. Ten students were randomly selected from the class to be taste testers. Each taste tester was randomly given one of the two juices, recorded how well it was liked on a scale of 1 to 10, and then given the other juice to evaluate. The scores given by the tasters are shown below.

Taste tester No.	1	2	3	4	5	6	7	8	9	10
Lemon-Grape	7.8	8.6	9.5	9.6	9.8	9.9	8.5	9.3	8.0	9.8
Cran-Pineapple	6.6	9.9	7.2	9.0	9.7	8.8	8.9	7.5	6.2	8.6

The statistics students want to use confidence intervals to determine if there is a difference in taste preference between the two juices. They think of three possible ways to do this.

- I. Do a two-sample 95% t -confidence interval for the true difference between the mean ratings of Lemon-Grape and Cran-Pineapple juices. The result is

$$-0.18 < \mu_{\text{LemonGrape}} - \mu_{\text{CranPineapple}} < 1.86.$$
- II. Do a paired 95% t -confidence interval for the true mean difference between the ratings of Lemon-Grape and Cran-Pineapple juices. The result is

$$0.05 < \mu_{\text{LemonGrape} - \text{CranPineapple}} < 1.63.$$
- III. Do a one-sample 95% t -confidence interval for the true mean rating of each juice and see if the intervals overlap. The results are

$$8.52 < \mu_{\text{LemonGrape}} < 9.64 \text{ and } 7.32 < \mu_{\text{CranPineapple}} < 9.16.$$

Which of the methods listed above is/are appropriate to answer the students' original question?

- (A) I only
 - (B) II only
 - (C) III only
 - (D) I and III only
 - (E) None of the methods are appropriate. A hypothesis test is required to draw any sort of conclusion from these data.
14. A doll-making company has been able to streamline their production process by having three jobs: assembling the doll, putting clothing on the doll, and putting the doll in the box. The table below illustrates the mean times (in seconds) and standard deviations for each task.

	Mean	Standard Deviation
Assembly	36	2.5
Clothing	22	1.8
Boxing	8	0.75

- (A) $\mu_x = 60.6$ $\sigma_x = 3.17$
- (B) $\mu_x = 60.6$ $\sigma_x = 5.05$
- (C) $\mu_x = 66$ $\sigma_x = 0.88$
- (D) $\mu_x = 66$ $\sigma_x = 3.17$
- (E) $\mu_x = 66$ $\sigma_x = 5.05$

The distributions of times for each step are approximately normal and independent. What are the mean and standard deviation of the total time (in seconds) to complete all three tasks?

15. When is a linear regression t -test used?
- (A) To find the value of a residual
 - (B) To find the value of the slope of the true regression line
 - (C) To find the confidence interval for the intercept of the true regression line
 - (D) To find the value of the intercept of the true regression line
 - (E) To find out if there is a meaningful linear relationship between two variables

Packet Free Response Questions:

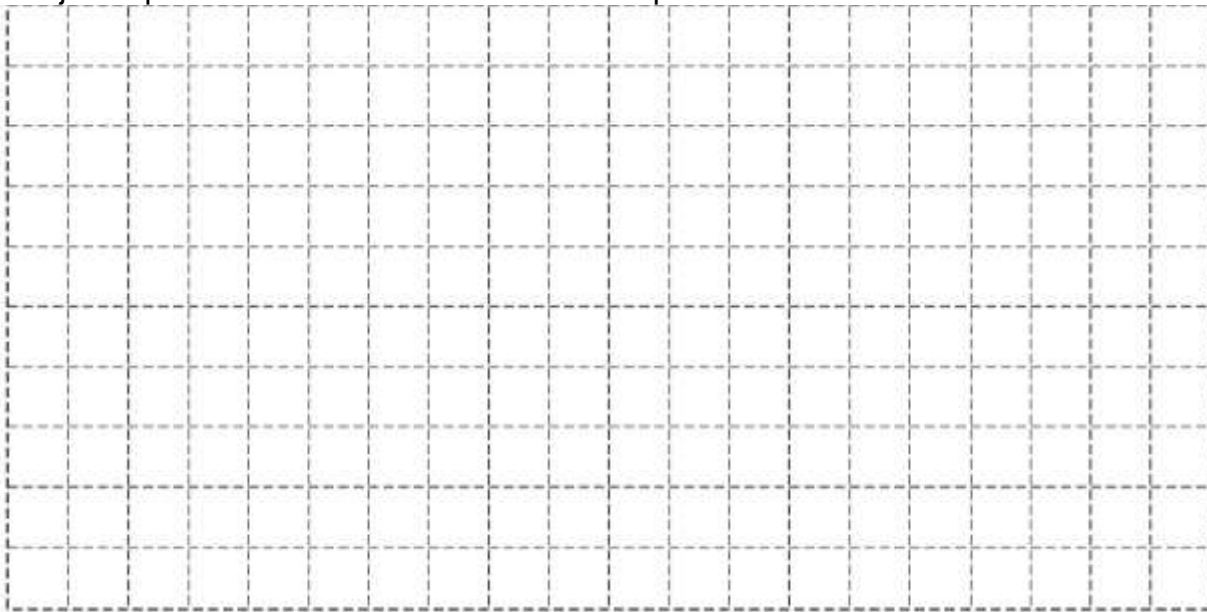
1. A simple random sample of 100 high school seniors was selected from a large school district. The gender of each student was recorded, and each student was asked the following questions.

1. Have you ever had a part-time job?
2. If you answered yes to the previous question, was your part-time job in the summer only?

The responses are summarized in the table below.

Job Experience	Gender		Total
	Male	Female	
Never had a part-time job	21	31	52
Had a part-time job during summer only	15	13	28
Had a part-time job but not only during summer	12	8	20
Total	48	52	100

(a) On the grid below, construct a graphical display that represents the association between gender and job experience for the students in the sample.



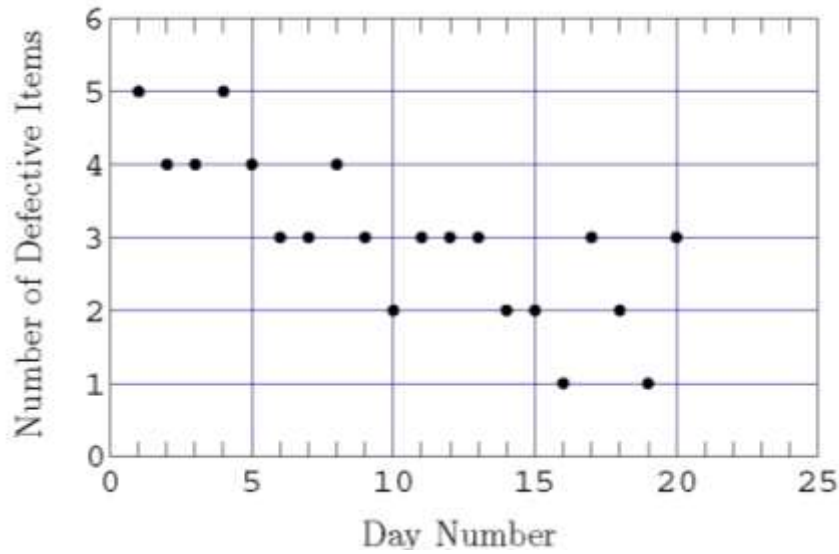
(b) Write a few sentences summarizing what the display in part (a) reveals about the association between gender and job experience for the students in the sample.

(c) Which test of significance should be used to test if there is an association between gender and job experience for the population of high school seniors in the district?

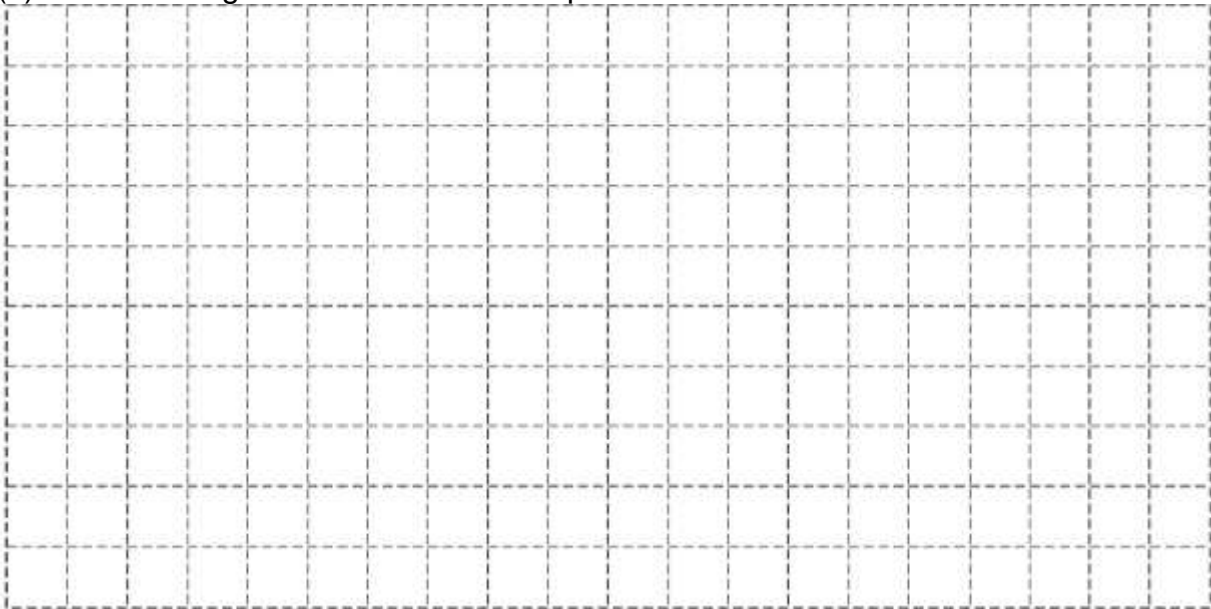
State the null and alternative hypotheses for the test, but do not perform the test.

Name _____ Date _____ Per _____

2. A plot of the number of defective items produced during 20 consecutive days at a factory is shown below.



(a) Draw a histogram that shows the frequencies of the number of defective items.



(b) Give one fact that is obvious from the histogram but is not obvious from the scatterplot.

(c) Give one fact that is obvious from the scatterplot but is not obvious from the histogram.

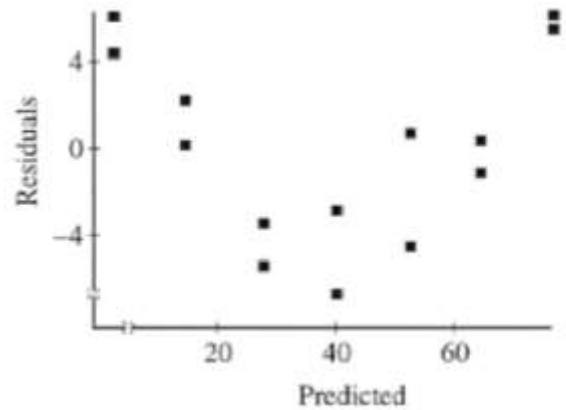
Name _____ Date _____ Per _____

3. In a study of the application of a certain type of weed killer, 14 fields containing large numbers of weeds were treated. The weed killer was prepared at seven different strengths by adding 1, 1.5, 2, 2.5, 3, 3.5, or 4 teaspoons to a gallon of water. Two randomly selected fields were treated with each strength of weed killer. After a few days, the percentage of weeds killed on each field was measured. The computer output obtained from fitting a least squares regression line to the data is shown below. A plot of the residuals is provided as well.

Dependent variable is: percent killed

R squared = 97.2% R squared (adjusted) = 96.9%

s = 4.505 with 14-2 = 12 degrees of freedom



Source	Sum of Squares	df	Mean Square	F-
Regression	8330.16	1	8330.16	410
Residual	243.589	12	20.2990	

Variable	Coefficient	s.e. of Coefficient	t-ratio	Prob
Constant	-20.5893	3.242	-6.35	≤ 0:0001
No. Teaspoons	24.3929	1.204	20.3	≤ 0:0001

(a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation.

(b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of weed killer are used, which of the following would you expect?

- The prediction will be too large.
- The prediction will be too small.
- A prediction cannot be made based on the information given on the computer output.

Explain your reasoning.

Optional Assignment: Video 3 - PICTURING DISTRIBUTIONS

1. What shape does the "Weekly Earnings" distribution have? _____

2. Discuss the findings of the Colorado Springs study of "Comparable Worth."

3. Which measures of location, the mean or median, is more resistant to the influence of extreme observations? _____

4. What numbers make up the five-number summary of a distribution? _____

5. What study is used to illustrate the use of five-number summaries and boxplots to compare distributions? Describe the results of this study.

6. What is the distance between the first and third quartiles called? _____

7. In a normal distribution, what one number can give the most information about the spread of the data?

8. Which study in the video illustrates the use of standard deviation to measure the spread about the mean as center? Describe the results of this study.

