AP Statistics
Unit 11 Multiple Choice Practice

Name Key
Date ______

1. The Excellent Drug Company claims its aspirin tablets will relieve headaches faster than any other aspirin on the market. To determine whether Excellent's claim is valid, random samples of size 15 are chosen from aspirins made by Excellent and the Simple Drug Company. An aspirin is given to each of the 30 randomly selected persons suffering from headaches and the number of minutes required for each to recover from the headache is recorded. The sample results are:

	$\bar{x}$	s^2
Excellent (E)	8.4	4.2
Simple (S)	8.9	4.6

C

A 5% significance level test is performed to determine whether Excellent's aspirin cures headaches significantly faster than Simple's aspirin. The appropriate hypotheses to be tested are

(a) H_0: $\mu_E - \mu_S = 0$; H_a: $\mu_E - \mu_S > 0$
(b) H_0: $\mu_E - \mu_S = 0$; H_a: $\mu_E - \mu_S \neq 0$
(c) H_0: $\mu_E - \mu_S = 0$; H_a: $\mu_E - \mu_S < 0$
(d) H_0: $\mu_E - \mu_S < 0$; H_a: $\mu_E - \mu_S = 0$
(e) H_0: $\mu_E - \mu_S > 0$; H_a: $\mu_E - \mu_S = 0$

2. In a large midwestern university (the class of entering freshmen being on the order of 6000 or more students), an SRS of 100 entering freshmen in 1993 found that 20 finished in the bottom third of their high school class. Admission standards at the university were tightened in 1995. In 1997 an SRS of 100 entering freshmen found that 10 finished in the bottom third of their high school class. Let p_1 and p_2 be the proportion of all entering freshmen in 1993 and 1997, respectively, who graduated in the bottom third of their high school class. What conclusion should we draw?

D

(a) We are 95% confident that the admissions standards have been tightened.
(b) Reject H_0 at the $\alpha = 0.01$ significance level.
(c) Fail to reject H_0 at the $\alpha = 0.05$ significance level.
(d) There is significant evidence at the 5% level of a decrease in the proportion of freshmen who graduated in the bottom third of their high school class that were admitted by the university.
(e) If we reject H_0 at the $\alpha = 0.05$ significance level based on these results, we have a 5% chance of being wrong.

3. 42 of 65 randomly selected people at a baseball game report owning an iPod. 34 of 52 randomly selected people at a rock concert occurring at the same time across town reported owning an iPod. A researcher wants to test the claim that the proportion of iPod owners at the two venues is not the same. A 90% confidence interval for the difference in population proportions is $(-0.154, 0.138)$.
Which of the following gives the correct outcome of the researchers' test of the claim?

B

(a) Since the confidence interval includes 0, the researcher can conclude that the proportion of iPod owners at the two venues is the same.
(b) Since the confidence interval includes 0, the researcher can conclude that the proportion of iPod owners at the two venues may be the same.
(c) Since the confidence interval includes 0, the researcher can conclude that the proportion of iPod owners at the two venues is different.
(d) Since the confidence interval includes more positive than negative values, we can conclude that a higher proportion of people at the baseball game own iPods than at the rock concert.
(e) We cannot draw a conclusion about a claim without performing a significance test.

The next two (4 and 5) refer to this scenario.

Different varieties of fruits and vegetables have different amounts of nutrients. These differences are important when these products are used to make baby food. We wish to compare the carbohydrate content of two varieties of peaches. The data were analyzed with SAS, and the following output was obtained:

VARIETY	N	MEAN	STD DEV	STD ERROR	MIN	MAX
1	5	33.6	3.781	1.691	29.000	38.000
2	7	25.0	10.392	3.927	2.000	33.000

VARIANCES	T	DF	PROB > |T|
UNEQUAL	2.0110	8.0	0.0791
EQUAL	1.7490	10.0	0.1109

4. We wish to test if the two varieties are significantly different in their mean carbohydrate content. The null and alternative hypotheses are

C

(a) $H_0 : \mu_1 = \mu_2; H_a : \mu_1 < \mu_2$
(b) $H_0 : \mu_1 = \mu_2; H_a : \mu_1 > \mu_2$
(c) $H_0 : \mu_1 = \mu_2; H_a : \mu_1 \neq \mu_2$
(d) $H_0 : \bar{x}_1 = \bar{x}_2; H_a : \bar{x}_1 < \bar{x}_2$
(e) $H_0 : \bar{x}_1 = \bar{x}_2; H_a : \bar{x}_1 \neq \bar{x}_2$

5. The test statistic and *P*-value are

D

(a) 1.7490; 0.0318
(b) 1.7490; 0.0554
(c) 2.0110; 0.1582
(d) 2.0110; 0.0791
(e) 2.0110; 0.0396

6. Some researchers have conjectured that stem-pitting disease in peach tree seedlings might be controlled with weed and soil treatment. An experiment was conducted to compare peach tree seedling growth with soil and weeds treated with one of two herbicides. In a field containing 20 seedlings, 10 were randomly selected from throughout the field and assigned to receive Herbicide A. The remaining 10 seedlings were to receive Herbicide B. Soil and weeds for each seedling were treated with the appropriate herbicide, and at the end of the study period, the height (in centimeters) was recorded for each seedling. A box plot of each data set showed no indication of non-normality. The following results were obtained:

Herbicide A:	$\bar{X}_1 = 94.5$ cm	$s_1 = 10$ cm
Herbicide B:	$\bar{X}_2 = 109.1$ cm	$s_2 = 9$ cm

C

Referring to the information above, a 95% confidence interval for $\mu_2 - \mu_1$ (using the conservative value for the degrees of freedom) is

A) 14.6 ± 7.36. B) 14.6 ± 7.80. C) 14.6 ± 9.62. D) 14.6 ± 13.93.

E) 14.6 ± 33.18.

7. Thirty-five people from a random sample of 125 workers from Company A admitted to using sick leave when they weren't really ill. Seventeen employees from a random sample of 68 workers from Company B admitted that they had used sick leave when they weren't ill. A 95% confidence interval for the difference in the proportions of workers at the two companies who would admit to using sick leave when they weren't ill is

(a) $0.03 \pm \sqrt{\frac{(0.28)(0.72)}{125} + \frac{(0.25)(0.75)}{68}}$

(b) $0.03 \pm 1.96\sqrt{\frac{(0.28)(0.72)}{125} + \frac{(0.25)(0.75)}{68}}$

B

(c) $0.03 \pm 1.645\sqrt{\frac{(0.28)(0.72)}{125} + \frac{(0.25)(0.75)}{68}}$

(d) $0.03 \pm 1.96\sqrt{\left(\frac{1}{125} + \frac{1}{68}\right)(0.269)(0.731)}$

(e) $0.03 \pm 1.645\sqrt{\left(\frac{1}{125} + \frac{1}{68}\right)(0.269)(0.731)}$

Use the following to answer questions **8 and 9**:

A sociologist is studying the effect of having children within the first three years of marriage on the divorce rate. From city marriage records, she selects a random sample of 400 couples that were married between 1985 and 1990 for the first time, with both members of the couple being between the ages of 20 and 25. Of the 400 couples, 220 had at least one child within the first three years of marriage. Of the couples that had children, 83 were divorced within five years, while of the couples that didn't have children, only 52 were divorced within three years. Suppose p_1 is the proportion of couples married in this time frame that had a child within the first three years and were divorced within five years and p_2 is the proportion of couples married in this time frame that did not have a child within the first two years and were divorced within five years.

B

8. Referring to the information above, the estimate of $p_1 - p_2$ is
A) 0.0775. B) 0.0884. C) 0.3100. D) 0.3375. E) 0.3773.

9. Referring to the information above, the sociologist had hypothesized that having children early would increase the divorce rate. She tested the one-sided alternative and obtained a P-value of 0.0314. The correct conclusion is that
A) if you want to decrease your chances of getting divorced, it is best to wait several years before having children.
B) having more children increases the risk of divorce during the first 5 years of marriage.
C) if you want to decrease your chances of getting divorced, it is best not to marry when you are closer to 30 years old.
D) it is best not to have children.
E) there is evidence of an association between divorce rate and having children early in a marriage.

E

Use the following to answer questions **10 through 12:**

Some researchers have conjectured that stem-pitting disease in peach tree seedlings might be controlled with weed and soil treatment. An experiment was conducted to compare peach tree seedling growth with soil and weeds treated with one of two herbicides. In a field containing 20 seedlings, 10 were randomly selected from throughout the field and assigned to receive Herbicide A. The remaining 10 seedlings were to receive Herbicide B. Soil and weeds for each seedling were treated with the appropriate herbicide, and at the end of the study period, the height (in centimeters) was recorded for each seedling. A box plot of each data set showed no indication of non-normality. The following results were obtained:

Herbicide A:	$\bar{X}_1 = 94.5$ cm	$s_1 = 10$ cm
Herbicide B:	$\bar{X}_2 = 109.1$ cm	$s_2 = 9$ cm

10. Referring to the information above, a 95% confidence interval for $\mu_2 - \mu_1$ (using the conservative value for the degrees of freedom) is

C

A) 14.6 ± 7.36. B) 14.6 ± 7.80. C) 14.6 ± 9.62. D) 14.6 ± 13.93.
E) 14.6 ± 33.18.

11. Referring to the information above, suppose we wished to determine if there tended to be a significant difference in mean height for the seedlings treated with the different herbicides. To answer this question, we decide to test the hypotheses H_0: $\mu_2 - \mu_1 = 0$, H_a: $\mu_2 - \mu_1 \neq 0$. Based on our data, the value of the two-sample t test statistic is

C

A) 14.60. B) 7.80. C) 3.43. D) 2.54. E) 1.14.

12. Referring to the information above, suppose we wished to determine if there tended to be a significant difference in mean height for the seedlings treated with the different herbicides. To answer this question, we decide to test the hypotheses H_0: $\mu_2 - \mu_1 = 0$, H_a: $\mu_2 - \mu_1 \neq 0$. The 90% confidence interval is 14.6 ± 7.80 cm. Based on this confidence interval,

A) we would not reject the null hypothesis of no difference at the 0.10 level.
B) we would reject the null hypothesis of no difference at the 0.10 level.
C) we would reject the null hypothesis of no difference at the 0.05 level.
D) the P-value is less than 0.10.
E) both C) and D) are correct.

Both

B & D say the same thing!

Statistics X/H/AP
Unit 13 Packet Problems

Name ____________________
Date ____________________

Complete all ~~4 problems~~. These problems are previous AP problems and will ~~be graded.~~
Show all work and explain using full sentences. Review Free Response

1. Sleep researchers know that some people are early birds (E), preferring to go to bed by 10 P.M. and arise by 7 A.M., while others are night owls (N), preferring to go to bed after 11 P.M. and arise after 8 A.M. A study was done to compare dream recall for early birds and night owls. One hundred people of each of the two types were selected at random and asked to record their dreams for one week. Some of the results are presented below.

	Number of Dreams Recalled During the Week			Proportion Who Recalled	
Group	Mean	Median	Standard Deviation	No dreams	5 or more dreams
Early birds	7.26	6.0	6.94	0.24	0.55
Night owls	9.55	9.5	5.88	0.11	0.69

(a) The researchers believe that night owls may have better dream recall than do early birds. One parameter of interest to the researchers is the mean number of dreams recalled per week with μ_E representing this mean for early birds and μ_N representing this mean for night owls. The appropriate hypotheses would then be $H_0: \mu_E - \mu_N = 0$ and $H_a: \mu_E - \mu_N < 0$. State two other pairs of hypotheses that might be used to test the researchers' belief. Be sure to define the parameter of interest in each case.

(b) Use the data provided to carry out a test of the hypotheses about the mean number of dreams recalled per week given in the statement of part (a). Do the data support the researchers' belief ?

a)

1. p_E = proportion of early birds who recall dreams
p_N = proportion of night owls who recall dreams

$H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N < 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E < p_N$

OR

p_E = proportion of early birds who do not recall dreams
p_N = proportion of night owls who do not recall dreams

$H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N > 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E > p_N$

NOTE: Either of these, BUT NOT BOTH, can be used as one of the possibilities for part (a).

2. p_E = proportion of early birds who recall 5 or more dreams
p_N = proportion of night owls who recall 5 or more dreams

$H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N < 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E < p_N$

OR

p_E = proportion of early birds who do not recall 5 or more dreams
p_N = proportion of night owls who do not recall 5 or more dreams

$H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N > 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E > p_N$

b) →

b)

Part 1: States a correct pair of hypotheses, identifies a correct test (by name or by formula) and checks appropriate requirements.

μ_E = mean number of dreams early birds recall
μ_N = mean number of dreams night owls recall

$$H_0: \mu_E = \mu_N \qquad \text{OR} \qquad H_0: \mu_E - \mu_N = 0$$
$$H_a: \mu_E < \mu_N \qquad\qquad\quad H_a: \mu_E - \mu_N < 0$$

Two-sample t-test $\qquad t = \dfrac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\dfrac{s_E^2}{n_E} + \dfrac{s_N^2}{n_N}}}$

Requirements:

1. Problem states that independent random samples were taken.
2. Normal population distributions or large samples. Since these are not normal, we need to note that $n_E = 100$ and $n_N = 100$ are both large in order to perform the t-test.

Part 2: Correct mechanics, including the value of the test statistic, df, and P-value (or rejection region)

$$t = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\dfrac{s_E^2}{n_E} + \dfrac{s_N^2}{n_N}}} = \frac{7.26 - 9.55}{\sqrt{\dfrac{(6.94)^2}{100} + \dfrac{(5.88)^2}{100}}} = -2.52$$

So, $\quad t = -2.52 \qquad \text{df} = 192 \qquad \text{P-value} = 0.006$

- It is OK to use conservative df of 99.
- Using t-tables: P-value < 0.01
- Using calculator: $t = -2.517578$, P-value $= 0.006304$, df $= 192.799$

Because the P-value is small (or less than an α selected and stated by the student), reject H_0. There is convincing evidence that the mean number of dreams night owls recall is greater for than the mean number of dreams early birds recall.

2. The principal at Crest Middle School, which enrolls only sixth-grade students and seventh-grade students, is interested in determining how much time students at that school spend on homework each night. The table below shows the mean and standard deviation of the amount of time spent on homework each night (in minutes) for a random sample of 20 sixth-grade students and a separate random sample of 20 seventh-grade students at this school.

	Mean	Standard Deviation
Sixth-grade students	27.3	10.8
Seventh-grade students	47.0	12.4

Based on dotplots of these data, it is not unreasonable to assume that the distribution of times for each grade were approximately normally distributed.

(a) Estimate the difference in mean times spent on homework for all sixth- and seventh-grade students in this school using an interval. Be sure to interpret your interval.

(b) An assistant principal reasoned that a much narrower confidence interval could be obtained if the students were paired based on their responses; for example, pairing the sixth-grade student and the seventh-grade student with the highest number of minutes spent on homework, the sixth-grade student and seventh-grade student with the next highest number of minutes spent on homework, and so on. Is the assistant principal correct in thinking that matching students in this way and then computing a matched-pairs confidence interval for the mean difference in time spent on homework is a better procedure than the one used in part (a) ? Explain why or why not.

Part (a):

Solution

Step 1: Identifies appropriate confidence interval by name or by formula and checks appropriate assumptions.

Two-sample confidence interval for the difference in means, $(\mu_1 - \mu_2)$ is

$$(\bar{X}_1 - \bar{X}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where
μ_1 = the mean time spent on homework by all 6th graders at Crest Middle School
μ_2 = the mean time spent on homework by all 7th graders at Crest Middle School

Assumptions: (1) Independent random samples,
(2) Large samples or normal population distributions

The problem states that samples were independent random samples and that dotplots of the data showed it was not unreasonable to assume that the population distributions are approximately normal. So, it is okay to proceed.

Step 2: Correct mechanics

A 95% confidence interval for $(\mu_1 - \mu_2)$ is computed as

$$(47.0 - 27.3) \pm t^* \sqrt{\frac{12.4^2}{20} + \frac{10.8^2}{20}} \quad \text{which gives} \quad 19.7 \pm t^*(3.68)$$

Most calculators use $t^* = 2.026$ based on 37.297 degree of freedom. Then the confidence interval is (12.25, 27.15).

Solutions to Step 2				
Procedure	**Degrees of freedom**	**Confidence level**	t^*	**Confidence Interval**
Unequal variances	37.297	0.90 0.95 0.99	1.687 2.026 2.715	(13.50, 25.90) (12.25, 27.15) (9.72, 29.68)
Conservative approach	19	0.90 0.95 0.99	1.729 2.093 2.861	(13.34, 26.06) (11.99, 27.40) (9.17, 30.23)

Step 3: Interpretation

(For 95% confidence level) Based on these samples, we can be 95% confident that the difference in <u>mean</u> times spent studying for seventh graders and sixth graders is between 12.25 minutes and 27.15 minutes.

Part (b):

Matching students in this way, using the measured responses, is inappropriate because it creates artificial association between responses from independent samples. (1) Pairs should be created on the basis of one or more variables that might be related to the response, not on the response itself. (2) The matching plan must be established before the data are collected.

Part (b) is scored as either essentially correct or incorrect.

Part (b) is essentially correct if it

1. says that it is not reasonable to pair in the proposed way
   AND
2. gives an explanation that says either that pairing must be done prior to collecting the data or that pairing must be done using variables other than the response, or proposes an acceptable scheme of matching using some variable other than the response.

Note: The assistant principal's suggestion to match sixth and seventh graders based on their responses would create an artificial positive correlation between the responses from two independent random samples and result in an interval estimate of the difference in the means that tends to be too narrow to provide the required level of confidence.

3. A large company has two shifts—a day shift and a night shift. Parts produced by the two shifts must meet the same specifications. The manager of the company believes that there is a difference in the proportions of parts produced within specifications by the two shifts. To investigate this belief, random samples of parts that were produced on each of these shifts were selected. For the day shift, 188 of its 200 selected parts met specifications. For the night shift, 180 of its 200 selected parts met specifications.

(a) Use a 96 percent confidence interval to estimate the difference in the proportions of parts produced within specifications by the two shifts.

(b) Based only on this confidence interval, do you think that the difference in the proportions of parts produced within specifications by the two shifts is significantly different from 0 ? Justify your answer.

Part (a):

Step 1: Identifies the appropriate confidence interval by name or formula and checks appropriate conditions.

Two sample z confidence interval for $p_D - p_N$, the difference in the proportions of parts meeting specifications for the two shifts OR $(\hat{p}_D - \hat{p}_N) \pm z^* \sqrt{\dfrac{\hat{p}_D(1-\hat{p}_D)}{n_D} + \dfrac{\hat{p}_N(1-\hat{p}_N)}{n_N}}$.

Conditions:
1. Independent random samples from two separate populations
2. Large samples, so normal approximation can be used

The problem states that random samples of parts were selected from the two different shifts. We need to assume that these parts were produced independently. That is, each employee works the day shift or night shift, but not both, and the machine quality does not vary over time. Since the sample sizes are both 200 and the number of successes (188 and 180) and the number of failures (12 and 20) for each shift are larger than 10, it is reasonable to use the large sample procedures.

Step 2: Correct mechanics

$$\hat{p}_D = \frac{188}{200} = 0.94 \text{ and } \hat{p}_N = \frac{180}{200} = 0.90$$

$$(0.94 - 0.9) \pm 2.0537\sqrt{\frac{0.94 \times 0.06}{200} + \frac{0.9 \times 0.1}{200}}$$

$$0.04 \pm 2.0537 \times 0.0271$$

$$0.04 \pm 0.0556$$

$$(-0.0156, 0.0956)$$

Step 3: Interpretation

Based on these samples, we can be 96 percent confident that the difference in the proportions of parts meeting specifications for the two shifts is between -0.0156 and 0.0956.

Part (b):

Since zero is in the 96 percent confidence interval, zero is a plausible value for the difference $p_D - p_N$, and we do not have evidence to support the manager's belief. In other words, there does not appear to be a statistically significant difference between the proportions of parts meeting specifications for the two shifts at the $\alpha = 0.04$ level.