Unit 5 Practice on Inferencing with Proportions Name______________________________

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

1) We have calculated a confidence interval based upon a sample of $n = 200$. Now we want to get a better estimate with a margin of error only one fifth as large. We need a new sample with n at least…

A) 450 B) 1000 C) 40 D) 5000 E) 240

2) A certain population is strongly skewed to the right. We want to estimate its mean, so we will collect a sample. Which should be true if we use a large sample rather than a small one?

I. The distribution of our sample data will be closer to normal.
II. The sampling model of the sample means will be closer to normal.
III. The variability of the sample means will be greater.

A) II only
B) I only
C) III only
D) II and III only
E) I and III only

3) A relief fund is set up to collect donations for the families affected by recent storms. A random sample of 400 people shows that 28% of those 200 who were contacted by telephone actually made contributions compared to only 18% of the 200 who received first class mail requests. Which formula calculates the 95% confidence interval for the difference in the proportions of people who make donations if contacted by telephone or first class mail?

A) $(0.28 - 0.18) \pm 1.96\sqrt{\dfrac{(0.23)(0.77)}{400}}$

B) $(0.28 - 0.18) \pm 1.96\sqrt{\dfrac{(0.23)(0.77)}{200}}$

C) $(0.28 - 0.18) \pm 1.96\sqrt{\dfrac{(0.23)(0.77)}{200} + \dfrac{(0.23)(0.77)}{200}}$

D) $(0.28 - 0.18) \pm 1.96\sqrt{\dfrac{(0.28)(0.72)}{400} + \dfrac{(0.18)(0.82)}{400}}$

E) $(0.28 - 0.18) \pm 1.96\sqrt{\dfrac{(0.28)(0.72)}{200} + \dfrac{(0.18)(0.82)}{200}}$

4) Which is true about a 95% confidence interval based on a given sample?

I. The interval contains 95% of the population.
II. Results from 95% of all samples will lie in the interval.
III. The interval is narrower than a 98% confidence interval would be.

A) II only
B) II and III only
C) None
D) III only
E) I only

5) A truck company wants on-time delivery for 98% of the parts they order from a metal manufacturing plant. They have been ordering from Hudson Manufacturing but will switch to a new, cheaper manufacturer (Steel-R-Us) unless there is evidence that this new manufacturer cannot meet the 98% on-time goal. As a test the truck company purchases a random sample of metal parts from Steel-R-Us, and then determines if these parts were delivered on-time. Which hypothesis should they test?

A) $H_0: p = 0.98$, $H_A: p > 0.98$
B) $H_0: p = 0.98$, $H_A: p < 0.98$
C) $H_0: p = 0.98$, $H_A: p \neq 0.98$
D) $H_0: p < 0.98$, $H_A: p > 0.98$
E) $H_0: p > 0.98$, $H_A: p = 0.98$

6) We are about to test a hypothesis using data from a well-designed study. Which is true?
I. A small P-value would be strong evidence against the null hypothesis.
II. We can set a higher standard of proof by choosing $\alpha = 10\%$ instead of 5%.
III. If we reduce the alpha level, we reduce the power of the test.
A) None B) I and III only C) II only D) III only E) I only

7) A pharmaceutical company investigating whether drug stores are less likely than food stores to remove over-the-counter drugs from the shelves when the drugs are past the expiration date found a P-value of 2.8%. This means that:
A) 2.8% more drug stores remove over-the-counter drugs from the shelves when the drugs are past the expiration date.
B) None of these.
C) There is a 97.2% chance the drug stores remove more expired over-the-counter drugs.
D) 97.2% more drug stores remove over-the-counter drugs from the shelves when the drugs are past the expiration date than food stores.
E) There is a 2.8% chance the drug stores remove more expired over-the-counter drugs.

8) To plan the course offerings for the next year a university department dean needs to estimate what impact the "No Child Left Behind" legislation might have on the teacher credentialing program. Historically, 40% of this university's pre-service teachers have qualified for paid internship positions each year. The Dean of Education looks at a random sample of internship applications to see what proportion indicate the applicant has achieved the content-mastery that is required for the internship. Based on these data he creates a 90% confidence interval of (33%, 41%). Could this confidence interval be used to test the hypothesis H_0: $p = 0.40$ versus H_A: $p < 0.40$ at the $\alpha = 0.05$ level of significance?
A) No, because the dean only reviewed a sample of the applicants instead of all of them.
B) Yes, since 40% is not the center of the confidence interval he rejects the null hypothesis, concluding that the percentage of qualified applicants will decrease.
C) Yes, since 40% is in the confidence interval he fails to reject the null hypothesis, concluding that there is not strong enough evidence of any change in the percent of qualified applicants.
D) No, because the dean should have used a 95% confidence interval.
E) Yes, since 40% is in the confidence interval he accepts the null hypothesis, concluding that the percentage of applicants qualified for paid internship positions will stay the same.

9) Suppose that a conveyor used to sort packages by size does not work properly. We test the conveyor on several packages (with H_0: incorrect sort) and our data results in a P-value of 0.016. What probably happens as a result of our testing?
A) We reject H_0, making a Type II error.
B) We correctly fail to reject H_0.
C) We correctly reject H_0.
D) We fail to reject H_0, committing a Type II error.
E) We reject H_0, making a Type I error.

10) We test the hypothesis that $p = 35\%$ versus $p < 35\%$. We don't know it but actually $p = 26\%$. With which sample size and significance level will our test have the greatest power?
A) $\alpha = 0.03$, $n = 250$
B) $\alpha = 0.03$, $n = 400$
C) The power will be the same as long as the true proportion p remains 26%
D) $\alpha = 0.01$, $n = 250$
E) $\alpha = 0.01$, $n = 400$

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

11) Dolphin births A state has two aquariums that have dolphins, with more births recorded at the larger aquarium than at the smaller one. Records indicate that in general babies are equally likely to be male or female, but the gender ratio varies from season to season. Which aquarium is more likely to report a season when over two-thirds of the dolphins born were males? Explain.

12) Approval rating A newspaper article reported that a poll based on a sample of 1150 residents of a state showed that the state's Governor's job approval rating stood at 58%. They claimed a margin of error of ±3%. What level of confidence were the pollsters using?

13) Pumpkin pie A can of pumpkin pie mix contains a mean of 30 ounces and a standard deviation of 2 ounces. The contents of the cans are normally distributed. What is the probability that four randomly selected cans of pumpkin pie mix contain a total of more than 126 ounces?

14) Depression A recent psychiatric study from the University of Southampton observed a higher incidence of depression among women whose birth weight was less than 6.6 pounds than in women whose birth weight was over 6.6 pounds. Based on a *P*-value of 0.0248 the researchers concluded there was evidence that low birth weights may be a risk factor for susceptibility to depression. Explain in context what the reported P-value means.

15) Truckers On many highways state police officers conduct inspections of driving logbooks from large trucks to see if the trucker has driven too many hours in a day. At one truck inspection station they issued citations to 49 of 348 truckers that they reviewed.

a. Based on the results of this inspection station, construct and interpret a 95% confidence interval for the proportion of truck drivers that have driven too many hours in a day.

b. Explain the meaning of "95% confidence" in part A.

16) Tax advice Each year people who have income file income tax reports with the government. In some instances people seek advice from accountants and financial advisors regarding their income tax situations. This advice is meant to lower the percentage of taxes paid to the government each year. A random sample of people who filed tax reports resulted in the data in the table below. Does this data indicate that people should seek tax advice from an accountant or financial advisor? Test an appropriate hypothesis and state your conclusion.

	Had Tax Advice?	
Paid Lower % of taxes	No	Yes
No	48	19
Yes	24	86

Answer Key
Testname: UNIT 5 PRACTICE

1) D
2) A
3) E
4) D
5) B
6) B
7) B
8) C
9) E
10) B
11) The smaller aquarium would experience more variability in the season percentage of male births. We would expect the larger aquarium to stay more consistent and closer to the 50-50 ratio for gender births.
12) Since $ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$, we have $0.03 = z^*\sqrt{\frac{(0.58)(0.42)}{1150}}$ or $z^* \approx 2.06$. Confidence level is 96%.
13) Two methods are shown below to solve this problem:

Method 1:
Let P = one can of pumpkin pie mix and T = four cans of pumpkin pie mix.
We are told that the contents of the cans are normally distributed, and can assume that the content amounts are independent from can to can.
$E(T) = E(P_1 + P_2 + P_3 + P_4) = E(P_1) + E(P_2) + E(P_3) + E(P_4) = 120$ ounces
Since the content amounts are independent,
$Var(T) = Var(P_1 + P_2 + P_3 + P_4) = Var(P_1) + Var(P_2) + Var(P_3) + Var(P_4) = 16$
$SD(T) = \sqrt{Var(T)} = \sqrt{16} = 4$ ounces
We model T with $N(120, 4)$ $\quad z = \frac{126 - 120}{4} = 1.5 \quad P = P(T > 126) = P(z > 1.5) = 0.067$
There is a 6.7% chance that four randomly selected cans of pumpkin pie mix contain more than 126 ounces.

Method 2:
Using the Central Limit Theorem approach, let $\bar{y}$ = average content of cans in sample
Since the contents are Normally distributed, $\bar{y}$ is modeled by $N\left(30, \frac{2}{\sqrt{4}}\right)$.
$P\left(\bar{y} > \frac{126}{4}\right) = P(\bar{y} > 31.5) = P\left(z > \frac{31.5 - 30}{1}\right) = P(z > 1.5) = 0.067$
There is about a 6.7% chance that 4 randomly selected cans will contain a total of over 126 ounces.
14) If birth weight was not a risk factor for susceptibility to depression, an observed difference in incidence of depression this large (or larger) would occur in only 2.48% of such samples.

15) a. Conditions:

* Independence: We assume that one trucker's driving times do not influence other trucker's driving times.
* Random Condition: We assume that trucks are stopped at random.
* 10% Condition: This sample of 348 truckers is less than 10% of all truckers.
* Success/Failure: 49 tickets and 299 tickets are both at least 10, so our sample is large enough.

Under these conditions the sampling distribution of the proportion can be modeled by a Normal model. We will find a one-proportion z-interval.

We know $n = 348$ and $\hat{p} = 0.14$, so $SE(\hat{p}) = \sqrt{\dfrac{\hat{p}\hat{q}}{n}} = \sqrt{\dfrac{(0.14)(0.86)}{348}} = 0.0186$

The sampling model is Normal, for a 95% confidence interval the critical value is $z^* = 1.96$.

The margin of error is $ME = z^* \times SE(\hat{p}) = 1.96(0.0186) = 0.0365$.

The 95% confidence interval is 0.14 ± 0.0365 or (0.1035, 0.1765).

We are 95% confident that between 10.4% and 17.7% of truck drivers have driven too many hours in a day.

b. If we repeated the sampling and created new confidence intervals many times we would expect about 95% of those intervals to contain the actual proportion of truck drivers that have driven too many hours in a day.

16) We want to know whether the percentage of people who paid lower taxes was different based on whether they sought tax advice or not.

H_0: $p_{advice} - p_{no\ advice} = 0$, H_A: $p_{advice} - p_{no\ advice} > 0$

Conditions:

* Independence: The people who filed tax reports were randomly selected and do not influence each other.
* Random Condition: The people who filed tax reports were randomly selected.
* 10% Condition: 72 is less than 10% of people who didn't get tax advice, and 105 is less than 10% of people who did get tax advice.
* Success/Failure: All observed counts (48, 19, 24, and 86) are at least 10.

Because all conditions have been satisfied, we can model the sampling distribution of the difference in proportions with a Normal model. We can perform a two-proportion z-test.

Let 'Advice' group be the people who sought tax advice and the 'No advice' group be the people who did not seek tax advice.

We know: $n_{advice} = 105$, $n_{no\ advice} = 72$, $\hat{p}_{advice} = 0.819$, $\hat{p}_{no\ advice} = 0.333$.

$$\hat{p}_{pooled} = \frac{24 + 86}{72 + 105} = 0.621$$

$$SE_{pooled}(\hat{p}_{advice} - \hat{p}_{no\ advice}) = \sqrt{\frac{(0.62)(0.38)}{72} + \frac{(0.62)(0.38)}{105}} = 0.074$$

The observed difference in sample proportions is $\hat{p}_{advice} - \hat{p}_{no\ advice} = 0.819 - 0.333 = 0.486$

$$z = \frac{(\hat{p}_{advice} - \hat{p}_{no\ advice}) - 0}{SE_{pooled}(\hat{p}_{advice} - \hat{p}_{no\ advice})} = \frac{0.49 - 0}{0.074} = 6.62$$

$P = P(z > 6.62) < 0.0001$

The P-value is small, so we reject the null hypothesis. There is strong evidence of a difference in the tax percentages paid between the group who has tax advice and the group who did not have tax advice. It appears that if you have tax advice you are more likely to pay a lower percentage of taxes to the government.