AP STATISTICS REVIEW UNIT 4

Part 1: Multiple Choice

Directions: *Choose the response that best answers the question or completes the statement.*

1. If a population has a standard deviation σ, then the standard deviation of the mean of 100 randomly selected items from this population is
   a. σ
   b. 100 σ
   c. $\sigma/10$
   d. $\sigma/100$
   e. 0.1

   $\frac{\sigma}{\sqrt{n}}$

2. In a large population of adults, the mean IQ is 112 with a standard deviation of 20. Suppose 200 adults are randomly selected for a market research campaign. The distribution of the sample mean IQ is
   a. Normal, mean 112, standard deviation 20
   b. Skewed Left, mean 112, standard deviation 0.1
   c. Skewed Right, mean 112, standard deviation 1.414
   d. Normal, mean 112, standard deviation 1.414

3. You want to compute a 72% confidence interval for a population mean. Assume that the population standard deviation is known to be 10 and the sample size is 50. The value of z* to be used in this calculation is
   a. 1.080
   b. 1.645
   c. 1.7507
   d. 2.0537

   100% - 72% = 28%

4. You have measured the people with poor systolic blood pressure of a random sample of employees of a company located near you. A 95% confidence interval for the proportion of people with poor systolic blood pressure for the employees of this company is (.122, .138). Which of the following statements gives a valid interpretation of this interval?
   a. Ninety-five percent of the sample of employees has a systolic blood pressure between .122 and .138.
   b. Ninety-five percent of the population of employees has a systolic blood pressure between .122 and .138.
   c. If the procedure were repeated many times, 95% of the resulting confidence intervals would contain the population proportion of people with poor systolic blood pressure.
   d. The probability that the population proportion blood pressure is between .122 and .138 is .95.
   e. If the procedure were repeated many times, 95% of the sample proportions would be between .122 and .138

5. An analyst, using a random sample of $n = 500$ families, obtained a 90% confidence interval for mean monthly family income for a large population: ($600, $800). If the analyst had used a 99% confidence level instead, the confidence interval would be:
   a. Narrower and would involve a larger risk of being incorrect.
   b. Wider and would involve a smaller risk of being incorrect
   c. Narrower and would involve a smaller risk of being incorrect
   d. Wider and would involve a larger risk of being incorrect
   e. Wider but it cannot be determined whether the risk of being incorrect would be larger or smaller

$.098 = (1.96)\sqrt{\frac{(.5)(.5)}{100}}$

$.049 = 1.96\sqrt{\frac{(.5)(.5)}{n}}$

6. You want to estimate the proportion of Californians who want to outlaw cigarette smoking in all public places. Generally speaking, by how much must you increase the sample size to cut the margin of error in half?
   a. 2 b. ½ c. 4 d. ¼ e. 8

7. A certain population is strongly skewed to the left. We want to estimate its mean, so we 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 more clearly skewed to the left.
   II. The sampling model of the sample means will be more skewed to the left. (more normal)
   III. The variability of the sample means will be greater.
   a. I only
   b. II only
   c. III only
   d. I and III only
   e. II and III only

8. Which is true about a 99% confidence interval based on a given sample?
   I. The interval contains 99% of the population.
   II. Results from 99% of all samples will lie in this interval.
   III. The interval is wider than a 95% confidence interval would be.
      a. None
      b. I only
      c. II only
      d. III only
      e. II and III only

9. We have calculated a confidence interval based on a sample of n =180. Now we want to get a better estimate with a margin of error only one third as large. We need a new sample with n at least...
   a. 20
   b. 60
   c. 312
   d. 540
   e. 1620

   180 × 9

10. An online catalog company wants on-time delivery for at least 90% of the orders they ship. They have been shipping orders via UPS and FedEx, but will switch to a more expensive service (ShipFast) if there is evidence that this service can exceed the 90% on-time goal. As a test the company sends a random sample of orders via ShipFast, and then makes follow-up phone calls to see if these orders arrived on time. Which hypothesis should they test?
   a. $H_0: p < 0.90$
      $H_0: p = 0.90$
   b. $H_0: p > 0.90$
      $H_0: p = 0.90$
   c. $H_0: p = 0.90$
      $H_0: p < 0.90$
   d. $H_0: p = 0.90$
      $H_0: p \neq 0.90$
   e. $H_0: p = 0.90$
      $H_0: p > 0.90$

11. A researcher investigating whether j oggers are less likely to get colds than people who do not jog found a p-value of 3%. This means that:
   a. 3% of joggers get colds.
   b. Joggers get 3% fewer colds than non-joggers.
   c. There's a 3% chance that joggers get fewer colds.
   d. There's a 3% chance that joggers don't get fewer colds.
   e. None of these.

12. To plan the budget for next year a college needs to estimate what impact the current economic downturn might have on student requests for financial aid. Historically this college has provided aid to 35% of its students. Officials look at a random sample of this year's applications to see what proportion indicate a need for financial aid. Based on these data they create a 90% confidence interval of (32%, 40%). Could this confidence interval be used to test the hypothesis $H_0: p = 0.35$ versus $H_A: p \neq 0.35$ at the α=0.10 level of significance?
   a. No, because financial aid amounts may not be normally distributed.
   b. No, because they only used a sample of the applicants instead of all of them.
   c. Yes, since 35% is in the confidence interval they accept the null hypothesis, concluding that the percentage of students requiring financial aid will stay the same.
   d. Yes, since 35% is in the confidence interval they fail to reject the null hypothesis, concluding that there is not strong evidence of any change in financial aid requests.
   e. Yes, since 35% is not at the center of the confidence interval they reject the null hypothesis, concluding that the percentage of students requiring aid will increase.

13. We are about to test a hypothesis using data from a well-designed study. Which is true?
   I. A large p-value would be strong evidence against the null hypothesis.
   II. We can set a higher standard of proof by choosing α= 10% instead of 5%.
   III. If we reduce the risk of committing a Type I error, then the risk of a Type II error will also decrease.
   a. None
   b. I only
   c. II only
   d. III only
   e. I and II only

14. Suppose a device advertised to increase a car's gas mileage really does not work. We test it on a small fleet of cars (with H_0: no effective), and our data results in a p-value of 0.004. What probably happens as a result of our experiment?
   a. We correctly fail to reject H_0.
   b. We correctly reject H_0.
   c. We reject H_0, making a Type I error.
   d. We reject H_0, making a Type II error.
   e. We fail to reject H_0, committing Type II error.

15. A college alumni fund appeals for donations by phoning or emailing recent graduates. A random sample of 300 alumni shows that 40% of the 150 who were contacted by telephone actually made contributions compared to only 30% of the 150 who received email requests. Which formula calculates the 98% confidence interval for the difference in the proportions of alumni who may make donations if contacted by phone or by email?

a. $(0.40 - 0.30) \pm 2.33\sqrt{\frac{(0.35)(0.65)}{150}}$

b. $(0.40 - 0.30) \pm 2.33\sqrt{\frac{(0.35)(0.65)}{150} + \frac{(0.35)(0.65)}{150}}$

c. $(0.40 - 0.30) \pm 2.33\sqrt{\frac{(0.35)(0.65)}{300}}$

d. $(0.40 - 0.30) \pm 2.33\sqrt{\frac{(0.40)(0.60)}{150} + \frac{(0.30)(0.70)}{150}}$

e. $(0.40 - 0.30) \pm 2.33\sqrt{\frac{(0.40)(0.60)}{300} + \frac{(0.30)(0.70)}{300}}$

16. An analyst, using a random sample of $n = 800$ families, obtained a 99% confidence interval for mean monthly family income for a large population: ($900, $1000). If the analyst had used a 90% confidence level instead, the confidence interval would be:
   a. Narrower and would involve a larger risk of being incorrect.
   b. Wider and would involve a smaller risk of being incorrect
   c. Narrower and would involve a smaller risk of being incorrect
   d. Wider and would involve a larger risk of being incorrect
   e. Wider but it cannot be determined whether the risk of being incorrect would be larger or smaller

17. A company is sued for job discrimination because only 18% of the newly hired candidates were women when 48% of all applicants were women.
   a. A Type I error is deciding the company is not discriminating when it is.
   b. A Type I error is deciding the company is discriminating when, in fact, it is.
   c. A Type I error is deciding the company is not discriminating when, in fact it is discriminating.
   d. A Type I error is deciding the company is discriminating when it is not.
   e. There is no type I error in this context.

Part 2: Free Response

Directions: *Show all work. Indicate clearly the methods you use, because you will be graded on correctness of your method as well as on accuracy of your results and explanations.*

1. One researcher wants to construct a 99% confidence interval as part of a study. A colleague says such a high level isn't necessary and that a 95% confidence interval will suffice. In what ways will these intervals differ?

---

2. A survey of a random sample of 1280 student loan borrowers found that 448 had loans totaling more than $20,000 for their undergraduate education.
   a. What is the 90% confidence interval for the population proportion *p?*
   b. If we want a 95% confidence interval, what sample size would be needed to have a margin of error of 0.25?
   c. The government wants only 33% of undergrad students to use student loans and owe more than $20,000. Can the government feel confident this is happening? Explain why or why not.
   d. Would the conclusion in (c) be any different with a 99% confidence interval? Explain.

---

3. It is estimated the cats, who live in the wild or as indoor pets allowed to roam outdoors, kill an average of 10 birds a year. Assume a normal distribution with a standard deviation of 3 birds.
   a. What are the mean and standard deviation for the sampling distribution of $\bar{x}$, the mean amount of birds killed by cats per year?
   b. What is the probability an SRS of 25 cats will kill an average of between 9 and 12 birds in a year?

---

4. A research study gives a 95% confidence interval for the proportion of subjects helped by a new anti-inflammatory drug as (0.56, 0.65).
   a. Interpret this interval in the context of the problem.
   b. Explain the meaning of "95% confidence level" in this example.
   c. If we want instead to be 90% confident, would our confidence interval need to be wider or narrower?
   d. If the sample size were greater, would the margin of error be smaller or greater?

---

5. A cancer research group surveys a random sample of 500 women more than 40 years old to test the hypothesis that 28% of women in this age group have regularly scheduled mammograms. The group is concerned that fewer than 28% of women in their area have regularly scheduled mammograms. The study is conducted with the following hypothesis:

H_0: The proportion of women who have regularly scheduled mammograms is 0.28.
H_A: The proportion of women who have regularly scheduled mammograms is less than 0.28.

The P-value of the test is 0.2732.

   a. Interpret the P-value in the context of this study.
   b. What conclusions should be drawn at the 5% significance level?
   c. Given this conclusion, what possible error, Type I or Type II, might be committed, and give a possible consequence of committing this error.

---

6. An automobile manufacturer tries two distinct assembly procedures. In a random sample of 350 cars coming off the line using the first procedure there are 28 with major defects, while a random sample of 500 autos from the second line shows 32 with defects. Is the difference significant at the 10% significance level? (Be sure to conduct all steps of the hypothesis test)

Unit 4 Review FRQ solutions

① A 99% Confidence interval will be wider than a 95% Confidence interval - so we are more confident it will capture the true population proportion, however we are less precise with a 99% confidence interval.

② $n = 1280$ $\hat{p} = \frac{448}{1280} = .35$

random - random sample of 1280 student loan borrowers

10% - 1280 < 10% of all student loan borrowers

S/F - 1280(.35) = 448, 1280(.65) = 832

448 ; 832 > 10 ✓

ⓐ 90% CI → $.35 \pm (1.645) \cdot \sqrt{\frac{(.35)(.65)}{1280}} = (.3281, .3719)$

$z^* = 1.645$

Based on our sample, we are 90% confident that the true pop. proportion of student loan borrowers with loans totaling more than 20,000 for undergrad is between 32.8% and 37.2%.

ⓑ 95% CI → $z^* = 1.96$ $ME = z^* SE \rightarrow .25 = 1.96 \sqrt{\frac{(.35)(.65)}{n}}$

Sample size needs to be 14 people.

$.12755 = \sqrt{\frac{.2275}{n}}$ $n = 13.98$

$.016269 = \frac{.2275}{n}$

ⓒ According to our 90% Confidence interval, 33% is Contained in the confidence interval meaning that the government can be somewhat confident that it's happening

ⓓ The conclusion would not be different with a 99% Confidence interval since that interval would be wider than the 95% CI and still capture 33% w/i the interval.

③ ⓐ $\mu(\bar{x}) = 10$ birds $\text{stdev}(\bar{x}) = \frac{3}{\sqrt{n}} = \frac{3}{\sqrt{25}} = \frac{3}{5} = .6$ birds

ⓑ $P(9 < x < 12) \approx$ $\frac{9-10}{.6} = -1.667$ $\frac{12-10}{.6} = 3.33$

$P(-1.667 < z < 3.33) = .952$

The probability of an SRS of 25 cats will kill an average of between 9 & 12 birds in a year is 95.2%

Unit 4 review

FRQ solutions continued

4 (a) We are 95% confident that the true proportion of patients who will be helped by the new drug is between 56% and 65%.

(b) 95% of all 95% confidence intervals will capture the true population proportion of people helped by the drug

(c) To be 90% confident instead of 95% confident would mean we would have a narrower interval.

(d) If the sample size were greater, the margin of error would be smaller!

5 (a) If the proportion of women who have regularly scheduled mammograms is 28%, then we would expect to find results like ours or worse 27.3% of the time due to natural sampling variation.

(b) We would fail to reject the null at $\alpha = .05$ – we do not have sufficient evidence.

(c) If we fail to reject a false null, that's a type 2 error! We believe it's at 28% when it isn't → fail to recruit more women to get mammograms → could lead to women not being diagnosed w/ cancer.

6 $H_0: p_1 - p_2 = 0$ random - random sample of 350 cars 500 + 10% 500 + 350 < 10% of all cars made

$\hat{p}_1$ = line 1 cars

$\hat{p}_2$ = line 2 cars

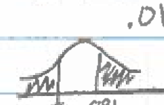
$H_A: p_1 - p_2 \neq 0$ Indep groups - Reasonable to assume independence since random samples

S/F: $350(.08) = 28 > 10$ ✓
$350(.92) = 322 > 10$ ✓
$500(.064) = 32 > 10$ ✓
$500(.936) = 468 > 10$ ✓

*Since conditions are met, we continue w/ a 2-prop z-test.

$$\hat{p}_{pooled} = \frac{28+32}{350+500} = .0706 \qquad SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{(.0706)(.9294)}{350} + \frac{(.0706)(.9294)}{500}} = .0179 \qquad z = \frac{(.08 - .064) - 0}{.01785} = .8963$$

$P(z > .8963) = .18504 \times 2 = .3701$



At the 10% significance level, we would fail to reject w/ a p-value of .3701. We do not have sufficient evidence to say they're different.