AP Stats
Review for the AP Exam - Ch 10 & 11

Name: Key
Hour: ______

1. What does 98% confidence mean?

Approximately 98 of every 100 confidence intervals created will capture the true population parameter.

2. What is a test statistic?

$$\frac{\text{estimate} - \text{hypothesized}}{\text{std. dev. of estimate}}$$

3. What is a p-value?

Probability computed assuming Ho is true and the observed outcome is at least as extreme as the actual.

4. What is a Type I Error? A Type II Error?

I: Rejecting Ho when it is true
II: Failing to reject Ho, when it is false

5. What is the probability of making a Type I Error?

same as α significance level

6. What is the power of a test? How do you calculate power?

$1-\beta$ (probability of Type II)
Power = Probability a test will reject Ho when Ho is false

7. How do you increase power?

Increase significance level α
Increase sample size

8. A city is interested in building a waste management facility in a certain area. One hundred randomly selected residents from this area were asked, "Do you support the city's decision to build a waste management facility in your area?" Of the 100 residents interviewed, 54 said no, 4 said yes, and 42 had no opinion. A large sample z-confidence interval, $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, was constructed from these data to estimate the proportion of this area's residents who support building a waste management facility in their area. Which of the following statements is correct for this confidence interval?

(A) This confidence interval is valid because a sample size of more than 30 was used.
(B) This confidence interval is valid because each area resident was asked the same question.
(C) This confidence interval is valid because no conditions are required for constructing a large sample confidence interval for a proportion.
(D) This confidence interval is not valid because the quantity $n\hat{p}$ is too small. $4 \ngeq 10$
(E) This confidence interval is valid because "no opinion" was included as a response category for the question.

9. A researcher has conducted a survey using a simple random sample of 50 registered voters to create a confidence interval to estimate the proportion of registered voters favoring the election of a certain candidate for mayor. Assume that the sample proportion does not change. Which of the following best describes the anticipated effect on the width of the confidence interval if the researcher were to survey a random sample of 200, rather than 50, registered voters?

(A) The width of the new interval would be about one-fourth the width of the original interval.
(B) The width of the new interval would be about one-half the width of the original interval.
(C) The width of the new interval would be about the same as the width of the original interval.
(D) The width of the new interval would be about twice the width of the original interval.
(E) The width of the new interval would be about four times the width of the original interval.

10. A planning board in Elm County is interested in estimating the proportion of its residents that are in favor of offering incentives to high-tech industries to build plants in that county. A random sample of Elm County residents was selected. All of the selected residents were asked, "Are you in favor of offering incentives to high-tech industries to build plants in your county?" A 95 percent confidence interval for the proportion of residents in favor of offering incentives was calculated to be 0.54 ± 0.05. Which of the following statements is correct?

(A) At the 95% confidence level, the estimate of 0.54 is within 0.05 of the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county.
(B) At the 95% confidence level, the majority of area residents are in favor of offering incentives to high-tech industries to build plants in the county.
(C) In repeated sampling, 95% of sample proportions will fall in the interval (0.49, 0.59).
(D) In repeated sampling, the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county will fall in the interval (0.49, 0.59).
(E) In repeated sampling, 95% of the time the true proportion county residents in favor of offering incentives to high-tech industries to build plants in the county will be equal to 0.54.

11. A polling organization asks a random sample of 1,000 registered voters which of two candidates they plan to vote for in an upcoming election. Candidate A is preferred by 400 respondents, Candidate B is preferred by 500 respondents, and 100 respondents undecided. George uses a large sample confidence interval for two proportions to estimate the difference in the population proportions favoring the two candidates. This procedure is not appropriate because

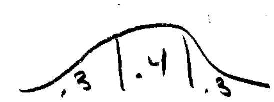
(A) the two sample proportions were not computed from independent samples.
(B) the sample size was too small.
(C) the third category, undecided, make the procedure invalid.
(D) the sample proportions are all different; therefore the variances are probably different as well.
(E) George should have taken the difference $\frac{500-400}{1,000}$, and then used a large sample confidence interval for a single proportion instead.

12. The lengths of individual shellfish in a population of 10,000 shellfish are approximately normally distributed with mean 10 centimeters and standard deviation 0.2 centimeter. Which of the following is the shortest interval that contains approximately 4,000 shellfish lengths?

(A) 0 cm to 9.949 cm
(B) 9.744 cm to 10 cm
(C) 9.744 cm to 10.256 cm
(D) 9.895 cm to 10.105 cm
(E) 9.9280 cm to 10.080 cm

$\frac{4000}{10000} = .4$

$z^* = -.524$



$10 \pm .524(.2)$

$= (9.895, 10.105)$

13. A survey was conducted to determine what percentage of college seniors would have chosen to attend a different college if they had known what they know now. In a random sample of 100 seniors, 34 percent indicated that they would have attended a different college. A 90 percent confidence interval for the percentage of all seniors who would have attended a different college is

$.34 \pm 1.645\left(\sqrt{\frac{.34(.66)}{100}}\right)$ invNorm (.05)

(A) 24.7% to 43.3%. (B) 25.8% to 42.2%. (C) 26.2% to 41.8%. (D) 30.6% to 37.4%. (E) 31.2% to 36.8%.

14. Courtney has constructed a cricket out of paper and rubber bands. According to the instructions for making the cricket, when it jumps it will land on its feet half of the time and on its back the other half of the time. In the first 50 jumps, Courtney's cricket landed on its feet 35 times. In the next 10 jumps, it landed on its feet only twice. Based on this experience, Courtney can conclude that

(A) the cricket was due to land on its feet less than half the time during the final 10 jumps, since it had landed too often on its feet during the first 50 jumps.
(B) a confidence interval for estimating the cricket's true probability of landing on its feet is wider after the final 10 jumps than it was before the final 10 jumps.
(C) a confidence interval for estimating the cricket's true probability of landing on its feet after the final 10 jumps is exactly the same as it was before the final 10 jumps.
(D) a confidence interval for estimating the cricket's true probability of landing on its feet is more narrow after the final 10 jumps than it was before the final 10 jumps.
(E) a confidence interval for estimating the cricket's true probability of landing on its feet based on the initial 50 jumps does not include 0.2, so there must be a defect in the cricket's construction to account for the poor showing in the final 10 jumps.

15. A consulting statistician reported the results from a learning experiment to a psychologist. The report stated that on one particular phase of the experiment a statistical test yielded a *p*-value of 0.24. Based on this *p*-value, which of the following conclusions should the psychologist make?

(A) The test was statistically significant because a *p*-value of 0.24 is greater than a significance level of 0.05.
(B) The test was statistically significant because $p = 1 - 0.24 = 0.76$ and this is greater than a significance level of 0.05.
(C) The test was not statistically significant because 2 times 0.24 = 0.48 and that is less than 0.5.
(D) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 24% of the time.
(E) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 76% of the time.

16. The analysis of a random sample of 500 households is a suburb of a large city indicates that a 98 percent confidence interval for the mean family income is ($41,300, $58,630). Could this information be used to conduct a test of the null hypothesis H_0: $\mu = 40{,}000$ against the alternative hypothesis H_a: $\mu \neq 40{,}000$ at the $\alpha = 0.02$ level of significance?

(A) No, because the value of σ is not known.
(B) No, because it is not known whether the data are normally distributed.
(C) No, because the entire data set is needed to do this test.
(D) Yes, since $40,000 is not contained in the 98 percent confidence interval, the null hypothesis would be rejected in favor of the alternative, and it could be concluded that the mean family income is significantly different from $40,000 at the $\alpha = 0.02$ level.
(E) Yes, since $40,000 is not contained in the 98 percent confidence interval, the null hypothesis would not be rejected, and it could be concluded that the mean family income is not significantly different from $40,000 at the $\alpha = 0.02$ level.

17. A t-statistic was used to conduct a test of the null hypothesis H_0: $\mu = 0$ against the alternative H_a: $\mu \neq 0$. The p-value was 0.056. A two-sided confidence interval for μ is to be constructed. Of the following, which is the largest level of confidence for which the confidence interval will NOT contain 0?

(A) 90% confidence
(B) 93% confidence
(C) 95% confidence
(D) 98% confidence
(E) 99% confidence

18. In a test of the hypotheses H_0: $\mu = 100$ versus H_a: $\mu > 100$, the power of the test when $\mu = 101$ would be greatest for which of the following choices of sample size n and significance level α?

(A) $n = 10$, $\alpha = 0.05$
(B) $n = 10$, $\alpha = 0.01$
(C) $n = 20$, $\alpha = 0.05$
(D) $n = 20$, $\alpha = 0.01$
(E) It cannot be determined form the information given.

19. Ten students were randomly selected from a high school to take part in a program designed to raise their reading comprehension. Each student took a test before and after completing the program. The mean of the differences between the score after the program and the score before the program is 16. It was decided that all students in the school would take part in this program during the next school year. Let μ_A denote the mean score after the program and μ_B denote the mean score before the program for all students in the school. The 95 percent confidence interval estimate of the true mean difference for all students is (9, 23). Which of the following statements is a correct interpretation of this confidence interval?

(A) $\mu_A > \mu_B$ with probability 0.95.
(B) $\mu_A < \mu_B$ with probability 0.95.
(C) μ_A is around 23 and μ_B is around 9.
(D) For any μ_A and μ_B with $(\mu_A - \mu_B) \geq 14$, the sample result is quite likely.
(E) For any μ_A and μ_B with $9 < (\mu_A - \mu_B) < 23$, the sample result is quite likely.

20. You attend a large university with approximately 15,000 students. You want to construct a 90% confidence interval estimate, within 5%, for the proportion of students who favor outlawing country music. How large a sample do you need?

(use .5(.5) if p and $\hat{p}$ are unknown)

$$.05 = 1.645\sqrt{\frac{.5(.5)}{n}}$$

$n \approx 270.603$

need at least 271 students

(2003 #2)

24. When a law firm represents a group of people in a class action lawsuit and wins that lawsuit, the firm receives a percentage of the group's monetary settlement. That settlement amount is based on the total number of people in the group – the larger the groups and the larger the settlement, the more money the firm will receive.

A law firm is trying to decide whether to represent car owners in a class action lawsuit against the manufacturer of a certain make and model for a particular defect. If 5 percent or less of the cars of this make and model have the defect, the firm will not recover its expenses. Therefore, the firm will handle the lawsuit only if it is convinced that more than 5 percent of cars of this make and model have the defect. The firm plans to take a random sample of 1,000 people who bought this car and ask them if they experienced this defect in their cars.

(a) Define the parameter of interest and state the null and alternative hypotheses that the law firm should test.

p = true proportion of all cars of this make and model with the defect.

$H_0: p = .05$

$H_a: p > .05$

(b) In the context of this situation, describe Type I and Type II errors and describe the consequences of each of these for the law firm.

I) The law firm rejects H_0 believing more than 5% of the cars have the defect when it is actually less than 5%. A consequence may be a waste of money and resources because the law firm takes the case when they should not.

II) The law firm fails to reject H_0 believing at most 5% of the cars have the defect when actually more than 5% do. A consequence could be the firm loses an opportunity to make money on a case they could have won.

(2002 #6)

25. A survey given to a random sample of students at a university included a question about which of two well-known comedy shows, S or F, students preferred. The students were asked the question, "Do you prefer S or F ?" The responses are shown below.

Preference		
S	F	Total
185	139	324

(a) Based on the results of this survey, construct and interpret a 95% confidence interval for the proportion of students in the population who would respond S to the question, "Do you prefer S or F ?"

① One-sample z Confidence interval for p, the proportion of all students at This university who would respond "S" to the question.

② SRS is stated

$N \geq 10n = 10(324) = 3240$. It is likely there are more than 3240 students at this university, so independence is satisfied.

$n\hat{p} = 185$ and $n(1-\hat{p}) = 139$. Since both are ≥ 10, we can use normal approximation.

③ $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = \dfrac{185}{324} \pm 1.96 \sqrt{\dfrac{\frac{185}{324}\left(\frac{139}{324}\right)}{324}} = (.517, .625)$

④ Based on this sample, I am 95% confident the true proportion of all students at this university who would respond "S" when asked "Do you prefer S or F?" is between .517 and .625.

(b) What is the meaning of "95% confidence" in part (a) ?

About 95 of every 100 intervals created with this method will capture the true population proportion of all students at this University who choose "S" when asked the question.

(c) In a follow-up survey, a separate group of randomly selected students was asked "Do you prefer F or S ?" The responses are shown below.

Preference		
S	F	Total
68	88	156

Based on these two surveys, is there evidence that the stated preference depends on the order in which the comedy shows were listed in the survey question? Justify your answer.

① $H_0: p_1 - p_2 = 0 \quad H_a: p_1 - p_2 \neq 0$

② Two-sample z-test for the true difference in the proportion of all students at this university who would answer 'S'.

(original − new)

③ SRS is stated

$N_1 \geq 10(324) = 3240 \quad N_2 \geq 10(156) = 1560$. We assume there are more than 3240 students at this university and independence is satisfied.

$\hat{p}_c = \frac{185+68}{324+156} = \frac{253}{480} \approx .527$

$n_1\hat{p}_c = 170.775$
$n_1(1-\hat{p}_c) = 153.225$
$n_2\hat{p}_c = 82.225$
$n_2(1-\hat{p}_c) = 73.775$

All are ≥ 5, so we can use normal approximation

④ $z = \dfrac{\left(\frac{185}{324} - \frac{68}{156}\right) - 0}{\left(\sqrt{\frac{.527(4.73)}{324}}\right)^2 + \left(\sqrt{\frac{.527(4.73)}{156}}\right)^2} = 2.777$ P-value $= 2P(z > 2.777) = .005$

⑤ Since $p = .005 < \alpha = .05$, we reject H_0. There is evidence that the stated preference depends on the order in which the shows are stated in the question.

(d) Suppose the test in part (c) indicates that the order in which the shows were listed does make a difference.

Is the pooled value $\frac{185 + 68}{324 + 156} = 0.527$ a reasonable estimate for the proportion of students at the university who would respond S ? If so, justify your answer. If not, what would be a more reasonable estimate? Explain why.

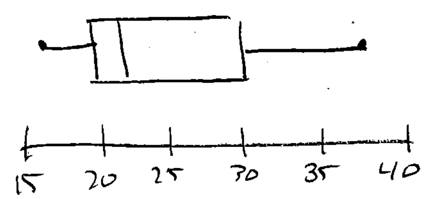
No since the sample sizes are not the same. Since the order of the questions matter, the pooled value is not a reasonable estimate. (Do not Pool) It would be better to weight them equally and find the mean: $\frac{\frac{185}{324} + \frac{68}{156}}{2} = .503$

(2000 #2)

26. Anthropologists have discovered a prehistoric cave dwelling that contains a large number of adult human footprints. To study the size of the adults who used the cave dwelling, they randomly selected 20 of the footprints from the population of all footprints in the cave and measured the length of those footprints. Some statistics resulting from this random sample are as follows.

Skewed Right? ~N

Sample size	20	Minimum	15.2 cm
Mean	24.8 cm	First quartile	18.7 cm
Standard Deviation	7.5 cm	Median	21.5 cm
		Third quartile	30.0 cm
		Maximum	37.0 cm



The anthropologists would like to construct a 95 percent confidence interval for the mean foot length of the adults who used the cave dwelling.

(a) What conditions are necessary in order for this confidence interval to be appropriate?

1. Must be a random sample of foot lengths of all adults who used the cave dwelling.
2. Foot lengths must be independent or sample size: $N > 10n$
3. Distribution of foot lengths is ~N or $n > 30$

(b) Discuss whether each of the conditions listed in your response to (a) appears to be satisfied in this situation.

1. The footprints may have come from the same person or children. Therefore, not a random sample.
2. If they came from the same person, the samples are not independent. Do we believe there were more than $10(20) = 200$ people in the caves.
3. Sample size is < 30. Boxplot is skewed slightly right