STAT 230 HW1 (Introductory Statistics Review)

1. What is the symbol for the difference between two population means?
   a. \( \bar{x} \)
   b. \( \mu \)
   c. \( x_1 - x_2 \)
   d. \( \mu_1 - \mu_2 \)
   e. \( p \)
   f. \( \hat{p} \)
   g. \( p_1 - p_2 \)
   h. \( \hat{p}_1 - \hat{p}_2 \)

2. A multimedia program designed to improve dietary behavior among low-income women was evaluated by comparing women who were randomly assigned to intervention and control groups. The intervention was a 30-minute session in a computer kiosk in the Food Stamp office. One of the outcomes was the score on a knowledge test taken about two months after the program. In testing whether the women in the intervention group scored higher than the women in the control group on average, the following results were obtained using Stat Crunch. At \( \alpha = 0.5 \), what should we conclude?

   Two-Sample T Test (pooled variances):
   \( \mu_1 \): mean score of women in the intervention group
   \( \mu_2 \): mean score of women in the control group
   \( H_0: \mu_1 - \mu_2 = 0 \)
   \( H_a: \mu_1 - \mu_2 > 0 \)

<table>
<thead>
<tr>
<th>Difference</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>DF</th>
<th>T-Stat</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_1 - \mu_2 )</td>
<td>0.75</td>
<td>0.1198</td>
<td>375</td>
<td>8.410</td>
<td>0.000</td>
</tr>
</tbody>
</table>

   a. The mean score of women in the intervention group is significantly greater than the mean score of women in the control group.
   b. The mean score of women in the intervention group is significantly different than the mean score of women in the control group.
   c. The mean score of women in the intervention group is significantly less than the mean score of women in the control group.
   d. There is insufficient evidence to conclude that the mean score of women in the intervention group is significantly greater than the mean score of women in the control group.
3. A study was carried out to investigate possible effects of ultrasound on the birth weight of babies. The average birth weight of babies born at Johns Hopkins Hospital in a certain year whose mothers had received an ultrasound examination during pregnancy was compared to the average birth weight of babies born in the same hospital in the same year whose mothers had not received an ultrasound examination. What type of study is this?
   a. A census
   b. An observational study
   c. A double blind study
   d. An experiment

4. A company’s news release for a diet product reported a study showing that most people who lose weight using their product keep it off. The study was based on a sample of 20 graduates of the company’s program, who endorse it in commercials. The results of the sample are probably ________
   a. Unbiased, since this is a recognized company.
   b. Unbiased, but not too accurate since only 20 individuals were in the sample.
   c. Biased, overstating the effectiveness of the product.
   d. Biased, understating the effectiveness of the product.

5. Which one of the following is NOT a valid sample for collecting data for inference?
   a. A simple random sample.
   b. A convenience sample.
   c. A multistage sample.
   d. A stratified sample.

6. In order to assess the opinions of BYU students about a proposed library remodel, a reporter interviews the first ten people entering the library who are willing to express an opinion. What type of study is this?
   a. An observational study based on a convenience sample.
   b. An observational study based on a stratified sample.
   c. An observational study based on a simple random sample.
   d. An observational study based on a multistage sample.
7. In a study of the effect of nitrites on bacteria, researchers measured the rate of uptake of an amino acid for 30 cultures of bacteria growing in a solution. Here is the resulting stem plot:

```
Stem-and-leaf of Uptake, n = 30
Leaf Unit = 100 (12 = 1200)

4    9
5
5
6    22
6    55568
7    23
7    66679
8    011223
8    55578
9    44
9    8
10   2
```

What is the value of the median?

a. 79.5  
   b. 7600  
   c. 7950  
   d. 8150  
   e. 8500

8. For the data in question 7, how does the mean compare with the median? No computation is necessary.

   a. The mean is greater than the median.
   b. The mean is equal to the median.
   c. The mean is less than the median.

9. Difference in age between husbands and wives (husband age minus wife age) of BYU married couples has an approximate Normal distribution with a mean of 1.5 years and a standard deviation of 2.1 years. Negative differences indicate that the wife is older. Approximately what percentage of BYU married couples are “wife-older” marriages (i.e. what percent of BYU married couples have a difference less than 0)?

   a. 5.8%  
   b. 13.7%  
   c. 23.9%  
   d. 76.1%

10. Referring to question 9, what is the value of a difference such that 60% of the age differences are less than it?

   a. 0.96 years  
   b. 0.45 years  
   c. 1.78 years  
   d. 2.03 years
11. A school district wondered if the average score of their students in the Math SAT exam differed from the national average of 450. A random sample of 30 seniors from the district had a mean of 486 with a sample standard deviation of 85. What is the P-value for the test of $H_0: \mu = 450$ versus $H_1: \mu \neq 450$?
   a. $0.01 < P < 0.02$ from the t table  
   b. $0.02 < P < 0.04$ from the t table  
   c. $0.0104$ from the z table  
   d. $0.0208$ from the z table

12. Consider 109 students in a college class as the population of interest. The mean ‘number of different people dated in the last month’ for these 109 students is 1.66. A simple random sample of 10 students was selected, and their mean was 1.40. The numbers 1.66 and 1.40 are, respectively, values for ______
   a. a statistic and the parameter.  
   b. a parameter and a statistic.  
   c. a statistic and a statistic.  
   d. a parameter and a parameter.

13. If all possible simple random samples of size 10 were taken from the 109 values described in the above question, and the mean calculated for each of these samples, what would be the mean of all these sample means?
   a. 1.40  
   b. 1.53  
   c. 1.66  
   d. It is impossible to determine from the given information.

14. The numbers in a random number table (a long random sequence of the numbers 0-9) have a mean of 4.5 and a standard deviation of 2.87. If you were to take a simple random sample of 32 numbers in table B by placing your finger at a random spot and taking the next 32 numbers, what is the probability that your $\bar{x}$ value would be greater than 5?
   a. 0.9855  
   b. 0.8389  
   c. 0.5073  
   d. 0.4927  
   e. 0.1724  
   f. 0.1611

15. Why were you able to use the standard Normal table to find the probability in the previous question?
   a. Because we could apply the Central Limit Theorem.  
   b. Because the data were sample from a population that has a Normal distribution.  
   c. Actually, we should not have used the standard Normal distribution to find the probability.

16. The weights of newborn babies in the US vary according to a Normal distribution with mean 7.5 lbs and std deviation 1.25 lbs. What is the probability that a randomly chosen baby weighs more than 5.5 lbs?
   a. .1056  
   b. .8944  
   c. .0548  
   d. .9452
17. The distribution of the age at which all married male BYU students got married is right skewed with \( \mu = 22.9 \) years and \( \sigma = 1.5 \) years. What is the probability that an SRS of 75 married male BYU students would have a sample mean between 22.8 and 23.0?
   a. 0.281
   b. 0.438
   c. 0.056
   d. 0.562
   e. 0.680
   f. Can’t compute using the standard Normal table because the distribution is right skewed, not Normal.

18. Referring to question 17, what is the standard deviation of the sampling distribution of \( \bar{x} \) for samples of size 75?
   a. 1.5 years
   b. 0.17 years
   c. 0.02 years
   d. Cannot be determined from information given.

19. A random sample of 30 seniors from a large school district had a mean Math SAT score of \( \bar{x} = 450 \) and a sample standard deviation of 85. Give a 99% confidence interval for \( \mu \), the mean of the whole population of seniors.
   a. 450 \( \pm \) 15.5
   b. 450 \( \pm \) 30.0
   c. 450 \( \pm \) 42.8
   d. 450 \( \pm \) 219.0

20. Referring to the information in problem 19, which of the following would produce a larger margin of error keeping all else constant?
   a. Using a sample of 100 seniors.
   b. Using a confidence level of 95%.
   c. Using a sample of 10 seniors.
   d. Using a confidence level of 90%.

21. Referring to the study in problem 19, suppose we wanted to plan a similar study with just female seniors. We want to have a margin of error of 25 with 95% confidence. What is the sample size needed to achieve this margin of error? Assume that the standard deviation for all female seniors is \( \sigma = 100 \).
   a. 8
   b. 61
   c. 62
   d. 87
   e. 122
22. Consider a simple random sample of 100 married BYU students taken by a family scientist to estimate the average length of engagement of BYU students. The sample standard deviation of length of engagement is 2.8 months. What is the value of the standard error of \( \bar{x} \)?
   a. 5.6 months   c. 0.56 months
   b. 2.8 months   d. 0.28 months

23. ‘Statistically significant’ is synonymous with all of the following except one. Which statement is an INCORRECT STATEMENT?
   a. The probability of the null hypothesis being true is less than \( \alpha \).
   b. The p-value is less than \( \alpha \).
   c. The difference between the observed value of the statistic and the claimed parameter value is too large to attribute to chance variation.
   d. If \( H_0 \) were true, the probability of obtaining a test statistic as or more extreme than what we actually observed is too small for us to believe that \( H_0 \) is true.

24. A sociologist reported that “Ethnocentrism was significantly higher (P<0.05) among church-attenders than among non-attenders.” Which is a correct interpretation of this p-value?
   a. There is less than a 5% chance that ethnocentrism is higher for attenders than for non-attenders.
   b. If ethnocentrism were really the same for attenders as for non-attenders, the chance of obtaining a difference as large or larger than the observed difference is less than 0.05.
   c. The probability that ethnocentrism is the same for attenders as for non-attenders is 0.05.
   d. There is less than a 5% chance that if you took another sample, ethnocentrism would be higher among non-attenders than among attenders.

25. A researcher in exercise physiology believes that interval training leads to greater weight loss than endurance training; she plans an experiment to investigate her belief. Her hypotheses are:

   \( H_0: \) mean weight loss is the same for both training regiments.

   \( H_a: \) mean weight loss is greater for interval training than endurance training.

   What is a type II error?
   a. To conclude that weight loss due to interval training is greater when, in fact, it is not.
   b. To conclude that weight loss due to interval training is greater when, in fact, it is.
   c. To conclude that weight loss is the same for both when, in fact, it is.
   d. To conclude that weight loss is the same for both when, in fact, it is greater for interval training.
26. A sample of 18 pigs was fed a new diet for 8 weeks. Their mean weight gain during this period was \( \bar{x} = 0.78 \) lb/day with \( s = 0.22 \) lb/day. When can we use the formula \( \bar{x} \pm t^{*} \frac{s}{\sqrt{n}} \) to estimate the mean weight gain of all pigs under this new diet?
   a. If the sample of pigs is large.
   b. If \( \sigma \) is known.
   c. If the sample of pigs is SRS and if weight gains in the sample are not outliers.
   d. If the sample pigs is SRS (no other assumptions are required due to the central limit theorem).

27. Referring to the question 26, what is the response variable?
   a. The new diet.
   b. Weight gain during the 8 week period.
   c. The mean weight gain.
   d. Final weight at the end of the 8 week period.

28. Referring to 26 and assuming conditions are met, give a 98% confidence interval for mean weight gain of the pigs.
   a. (.64 lb/day, .92 lb/day)
   b. (.65 lb/day, .91 lb/day)
   c. (.66 lb/day, .90 lb/day)
   d. (.68 lb/day, .99 lb/day)

29. Using the information in question 26, and assuming that it is appropriate to use t-procedures in this case, the p-value for the test of \( H_0: \mu = 0.7 \) versus \( H_a: \mu \neq 0.7 \) is
   a. Between .02 and .025
   b. Between .05 and .10
   c. Between .10 and .20
   d. Between .04 and .05

---

Use the following information in the next two questions. Individuals in a sample of male BYU students were asked how many hours of TV they usually watched per week. They were also asked if they were married or single. The 35 married men had a mean, \( \bar{x}_1 = 6.21 \) hours, and a standard deviation, \( s_1 = 5.05 \) hours. The 56 single men had a mean, \( \bar{x}_2 = 3.66 \) hours, and a standard deviation, \( s_2 = 4.13 \) hours.

30. In order to estimate with 90% confidence the difference between the average time married men at BYU watch TV and the average time single men at BYU watch TV, we should compute a confidence interval for ________.
   a. \( \mu \)
   b. \( \mu_1 - \mu_2 \)
   c. \( \mu_d \) for matched pairs
   d. \( p_1 - p_2 \)
31. To test

\[ H_0: \text{The average hours of TV watching for married students equals the average TV watching for single men at BYU} \]

\[ H_a: \text{The average hours of TV watching for married students is greater than the average TV watching for single men at BYU} \]

The P-value was found to be 0.018. At \( \alpha = 0.05 \), what should we conclude?

a. The average hours of TV watching for married men at BYU is significantly greater than the average hours of TV watching for single men at BYU.

b. The average hours of TV watching for married men at BYU is the same as the average hours of TV watching for single men at BYU.

c. The average hours of TV watching for married men at BYU is not significantly greater than the average hours of TV watching for single men at BYU.

32. In a study on storage of wheat, calcium content was compared at harvest time, at one month, at two months and at four months. Calcium content was measured on six samples of wheat at each storage time. What statistical procedure should be used to determine whether length of storage affects calcium content?

a. Independent two-sample t for means.

b. Independent two-sample z for proportions.

c. Chi-square test.

d. Analysis of Variance.

33. Which of the following set of hypotheses is appropriate for testing equality of three means in ANOVA?

a. \( H_0: \mu_1 = \mu_2 = \mu_3 \) versus \( H_a: \mu_1 \neq \mu_2 \neq \mu_3 \)

b. \( H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3 \) versus \( H_a: \bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3 \)

c. \( H_0: \mu_1 = \mu_2 = \mu_3 \) versus \( H_a: \text{At least one } \mu_i \text{ differs from the others} \)

d. \( H_0: \mu_1 = \mu_2 = \mu_3 \) versus \( H_a: \text{At least one } \bar{x}_i \text{ differs from the others} \)

34. A study was conducted on all small electrical appliances to determine whether any link could be found between leukemia and appliance use. Statistically significant links were found between only hair dryers and black-and-white televisions even though over 50 appliances were tested. Wise consumers of statistical information would conclude that

a. there is strong evidence that hair dryers and black-and-white televisions cause leukemia.

b. a difference between incidence of leukemia for those using either of the two appliances and those not using the appliances as large or larger than that observed is unlikely to be due to chance.

c. while these results may be statistically significant, they are not necessarily practically significant.

d. because multiple tests were performed, the results are only suggestive, not conclusive.
True or False Questions

35. $\mu_1 - \mu_2$, the difference in two population means, represents the parameter used to compare the means of two populations.

36. When many tests of significance are performed on one set of data, the researcher is guilty of performing multiple analyses and inflating the overall $\alpha$. 