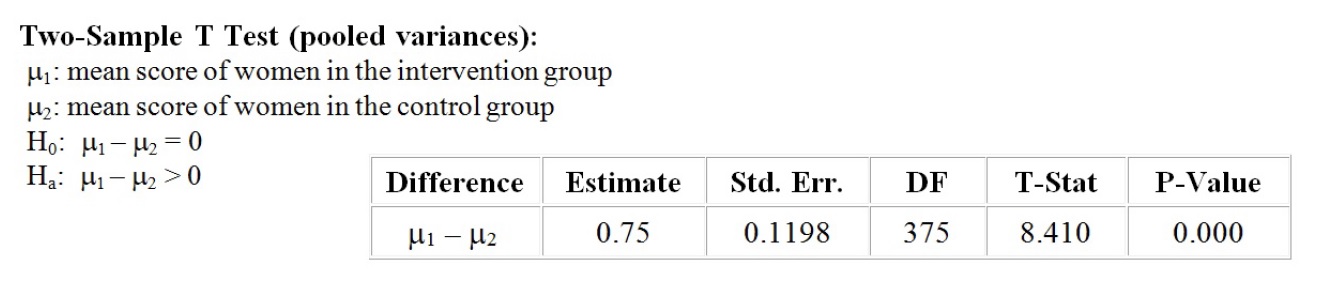
**STAT 121 Final Practice Exam**

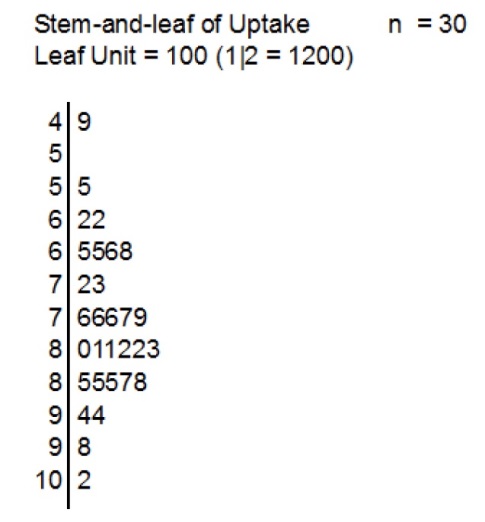
*This practice exam is not intended to duplicate the final nor is it intended to cover all possible questions that could be on the final. Hopefully, it will give you an idea of your level of preparedness and give you some practice. An answer key is at the end of the test.*

1. Smudge pots are sometimes used to protect orchards from frost. Two types of smudge pots are to be compared to determine which has the longest average burn time. Ten pots of each type are randomly selected for the study. These twenty pots will be lit and the burn time for each recorded. What is the appropriate type of analysis?
   1. One sample *t* test of significance on .
   2. Matched pairs *t* test of significance on .
   3. Two independent sample *t* test on .
   4. One sample *z* test on proportion, *p*.
   5. Two sample *z* test on
2. What is the symbol for the difference between two population means?
   1. *p*
3. 1400 bald male subjects were in a study to determine whether Rogaine is effective in stimulating hair growth. The subjects were randomly divided into two groups: one group having a placebo preparation applied containing no Rogaine and the other group having Rogaine applied. Neither the subjects nor the people recording the results knew who received the Rogaine and who did not. At the end of the treatment period, each subject was examined to determine whether he experience new hair growth. The researchers want to determine whether the proportion of those in the Rogaine treatment group who experienced new hair growth exceeded the proportion of those in the control group. What are the appropriate hypotheses?
   1. versus
   2. versus
   3. versus
   4. versus
   5. versus
   6. versus
4. Suppose you want to estimate the proportion of voters who will vote for George Smith, a candidate for state representative. How many should you sample in order to estimate *p* with a margin of error of 0.05 (5%) and 95% confidence?
   1. 1537
   2. 385
   3. 97
   4. 73
   5. 40
5. A local television network affiliate, WKSU, has recently remodeled its broadcast studio and changed in news anchor-persons. The station manager now wants to estimate the percentage of the viewing audience who currently watch its new broadcast. Which one of the following formulas should the station manager use to obtain the 99% confidence interval estimate that he desires?
6. Which one of the following represents the parameter estimated with a 99% confidence interval for difference in proportions?
   1. *p*
7. If we were to test the hypotheses versus using sample results of from a sample of size 100, what is the value we should use for the standard deviation of the sampling distribution of in the test statistic?
   1. 0.0016
   2. 0.0021
   3. 0.0400
   4. 0.0458
8. A quality inspector checks a large truckload of potatoes to determine the proportion of defective potatoes. She tests the hypotheses versus and will accept the shipment unless she has sufficient evidence that the proportion of defective potatoes significantly exceeds 6%. She plans to randomly sample 100 potatoes. Is this sample size large enough to perform a one-sample *z* test for the proportion?
   1. Yes, because *n* > 40.
   2. Yes, because there are no outliers or strong skewness in categorical data.
   3. No, because 100(.06) < 10.
   4. There is not enough information to check whether the sample size is large enough.
9. Referring to question 8, assume that the quality inspector takes a random sample of potatoes that is large enough to satisfy the conditions and finds *p*-value=0.118. Should she accept the shipment at ? Why or why not?
   1. No, because we fail to reject *H0* at .
   2. No because we reject *H0* at
   3. Yes, because we fail to reject *H0* at
   4. Yes, because we reject *H0* at
   5. There is not enough information to determine whether she should accept the shipment.
10. A company asserts that 80% of the customers who purchase its special lawn mower will have no repairs during the first two years of ownership. Your personal study of 100 randomly selected owners found 70 lasted the first two years without repair expenses. Can you use the standard Normal table to compute the probability that 70% or fewer in the sample lasted the two years without repair expenses? Why or why not?
    1. Yes, because the sample size is greater than 30 and the Central Limit Theorem applies.
    2. Yes, because both of the following are met:
    3. No, because the sample size is less than 30 (so the Central Limit Theorem does not apply).
    4. No, because both of the following are not met:
11. Referring to question 10, what is the mean and standard deviation of the sampling distribution of for samples of size *n* = 100 if *p* = 0.80?
    1. 0.70 and 0.0458
    2. 0.70 and 0.04
    3. 0.80 and 0.0458
    4. 0.80 and 0.04
    5. Not enough information given to find the mean and standard deviation of the sampling distribution of
12. Suppose we are testing versus and the test statistic from our random sample results is *z* = 1.83. What can we conclude at the 5% level of significance?
    1. The population proportion is significantly different from 0.5.
    2. The population proportion is equal to 0.5.
    3. We have insufficient evidence to conclude that the population proportion is significantly different from 0.5.
    4. Not enough information is available to answer this question.
13. 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 , what should we conclude?



* 1. The mean score of women in the intervention group is significantly greater than the mean score of women in the control group.
  2. The mean score of women in the intervention group is significantly different than the mean score of women in the control group.
  3. The mean score of women in the intervention group is significantly less than the mean score of women in the control group.
  4. 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.

1. 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?
   1. A census
   2. An observational study
   3. A double blind study
   4. An experiment
2. 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 \_\_\_\_\_\_\_\_
   1. Unbiased, since this is a recognized company.
   2. Unbiased, but not too accurate since only 20 individuals were in the sample.
   3. Biased, overstating the effectiveness of the product.
   4. Biased, understating the effectiveness of the product.
3. Which one of the following is NOT a valid sample for collecting data for inference?
   1. A simple random sample.
   2. A convenience sample.
   3. A multistage ample.
   4. A stratified sample.
4. 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?
   1. An observational study based on a convenience sample.
   2. An observational study based on a stratified sample.
   3. An observational study based on a simple random sample.
   4. An observational study based on a multistage sample.
5. 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:



What is the value of the median?

* 1. 79.5
  2. 7600
  3. 7950
  4. 8150
  5. 8500

1. For the data in question 18, how does the mean compare with the median? No computation is necessary.
   1. The mean is greater than the median.
   2. The mean is equal to the median.
   3. The mean is less than the median.
2. 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)?
   1. 5.8%
   2. 13.7%
   3. 23.9%
   4. 76.1%
3. Referring to question 20, what is the value of a difference such that 60% of the age differences are less than it?
   1. 0.96 years
   2. 0.45 years
   3. 1.78 years
   4. 2.03 years
4. 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 versus ?
   1. 0.01 < *P* < 0.02 from the t table
   2. 0.02 < *P* < 0.04 from the t table
   3. 0.0104 from the z table
   4. 0.0208 from the z table
5. A computer program gives the correlation for the following data set below as **r = 0.82**. Is this value valid?

|  |  |
| --- | --- |
| X | Y |
| 1 | 2 |
| 2 | 1 |
| 3 | 4 |
| 4 | 3 |
| 5 | 7 |
| 6 | 5 |
| 7 | 6 |

* 1. Yes, because computer programs do not make mistakes.
  2. Yes, because the relationship appears linear and both x and y variables are quantitative.
  3. No, because the relationship does not appear to be linear.
  4. No, because the correlation involves products of z-scores and each pair of x and y values do not have the same z-scores.

1. Ross, a paleontologist, found a correlation of 0.64 between X = the geographical area in square miles where fossils of the species are found and Y = the number of years that fossils are extinct species are found in the fossil record. What does r=0.64 tell us?
   1. That larger geographical areas caused longer survival of a species.
   2. About 64% of the variation in number of years a species appears can be explained by geographical area.
   3. That species with large areas for their fossils tend to have longer periods of appearance in the fossil record.
   4. About 64% of the variation in the geographical area where fossils are found can be explained by the regression on number of years a species appears.
2. A 2-way table for marital status versus TV-watching category for women in a class is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rows: Marital Status Columns: TV watching category | | | | |
|  | 0-2 hrs | 3-7 hrs | >7 hrs | Total |
| Single | 8 | 17 | 10 | 35 |
| Married | 30 | 17 | 9 | 56 |
| Total | 38 | 34 | 19 | 91 |

The P-value from the chi-square analysis of these data is 0.015. At , what can we conclude about the relationship between marital status and amount of TV watching?

* 1. There is a significant association between marital status and amount of TV watching.
  2. There is a weak linear relationship between marital status and amount of TV watching.
  3. There is no relationship between marital status and amount of TV watching.

1. A survey by the Deseret News asked a sample of 405 Utah County residents if growth over the last few years in Utah Valley had improved or deteriorated the quality of life. 54% of those surveyed said that growth had deteriorated the quality of life. The number 54% is a
   1. Parameter because it describes the percentage in a sample.
   2. Statistic because it describes the percentage in a sample.
   3. Parameter because it describes the percentage in a population.
   4. Statistic because it describes the percentage in a population.

For the next three questions consider the following table of Nobel Prize winners by field and country:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Physics | Chemistry | Medicine | **TOTAL** |
| U.S. | 49 | 29 | 61 | **139** |
| Britain | 21 | 24 | 23 | **68** |
| Germany | 15 | 25 | 12 | **52** |
| Japan | 3 | 1 | 0 | **4** |
| France | 8 | 6 | 7 | **21** |
| USSR | 7 | 1 | 2 | **10** |
| Other | 23 | 17 | 36 | **76** |
| **TOTAL** | **126** | **103** | **141** | **370** |

1. The United States received 37.57% of the prizes. This percentage is part of the \_\_\_\_\_\_\_
   1. Marginal distribution of countries.
   2. Marginal distribution of fields.
   3. Conditional distribution for the United States.
   4. Conditional distribution for physics.
2. 48.08% of Germany’s Nobel Prize winners were for chemistry. This percentage is part of the \_\_\_\_\_\_\_\_
   1. Marginal distribution of countries.
   2. Marginal distribution of fields.
   3. Conditional distribution for Germany.
   4. Conditional distribution for chemistry.
3. What proportion of all Nobel Prize winners in Physics were from the U.S.?
   1. 0.389
   2. 0.353
   3. 0.132
4. 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 values 1.66 and 1.40 are, respectively, \_\_\_\_\_\_
   1. A statistic and the parameter.
   2. The parameter and a statistic.
   3. A statistics and a statistic.
   4. A parameter and a parameter.
5. 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?
   1. 1.40
   2. 1.53
   3. 1.66
   4. It is impossible to determine from the given information.
6. Consider the random variable x, the number of days that a randomly chosen male student wears his socks before washing them. Do these probabilities satisfy the probability rules?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Probability | 0.81 | 0.12 | 0.03 | 0.01 | 0.01 | 0.01 | 0.01 |

* 1. Yes, because all probabilities are between 0.0 and 1.0 and they sum to 1.0.
  2. No, because there is no probability for x=0.
  3. No, because one or more of the probabilities exceeds 1.0.
  4. No, because the sum of all probabilities exceed 1.0.

1. The numbers in Table B (the random number table) 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 value would be greater than 5?
   1. 0.9855
   2. 0.8389
   3. 0.5073
   4. 0.4927
   5. 0.1724
   6. 0.1611
2. Why were you able to use the standard Normal table to find the probability in question 33?
   1. Because we could apply the Central Limit Theorem.
   2. Because the data were sample from a population that has a Normal distribution.
   3. Actually, we should not have used the standard Normal distribution to find the probability.
3. 13% of BYU students are from California. A Normal curve representing the sampling distribution of (the sample proportion of students who are from California in a simple random sample of 150 BYU students) is given below. First compute the mean and standard deviation of the sampling distribution of and then label the curve. Which point has a value of 0.157? Assume the points are each 1 standard deviation away from each other was *a* as the mean.



1. Referring to problem 35, what is the approximate probability that in a particular SRS of size 150 would be less than or equal to .10?
   1. 0
   2. 0.08
   3. 0.36
   4. 0.13
2. If the sample size in question 35 had been 50 instead of 150, what can we say about the probability that would be greater than or equal 0.20?
   1. The probability would be about 0.07.
   2. The probability would be about 0.93.
   3. The probability would be about 0.01.
   4. The probability should not be computed using the standard Normal table since n=50 is too small.
3. The weights of newborn babies in the US vary according to a Normal distribution with mean 7.5 pounds and standard deviation 1.25 pounds. What is the probability that a randomly chosen baby weighs more than 5.5 pounds?
   1. .1056
   2. .8944
   3. .0548
   4. .9452
4. The distribution of the age at which *all* married male BYU students got married is right skewed with years and 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?
   1. .281
   2. .438
   3. .056
   4. .562
   5. .680
   6. Can’t compute using the standard Normal table because the distribution is right skewed, not Normal.
5. Referring to question 39, what is the standard deviation of the sampling distribution of for samples of size 75?
   1. 1.5 years
   2. 0.17 years
   3. 0.02 years
   4. Cannot be determined from information given.
6. A random sample of 30 seniors from a large school district had a mean Math SAT score of and a sample standard deviation of 85. Give a 99% confidence interval for , the mean of the whole population of seniors.
   1. 450 15.5
   2. 450 30.0
   3. 450 42.8
   4. 450 219.0
7. Referring to the information in problem 41, which of the following would produce a larger margin of error keeping all else constant?
   1. Using a sample of 100 seniors.
   2. Using a confidence level of 95%.
   3. Using a sample of 10 seniors.
   4. Using a confidence level of 90%.
8. Referring to the study in problem 41, 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 .
   1. 8
   2. 61
   3. 62
   4. 87
   5. 122
9. 37% of Utah residents shop on Sunday. If a simple random sample of 300 Utah residents is taken, what is the shape of the sampling distribution of ?
   1. Right skewed
   2. Left skewed
   3. Approximately Normal
   4. No shape because the response variable is categorical.
10. 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 ?
    1. 5.6 months
    2. 2.8 months
    3. 0.56 months
    4. 0.28 months
11. You are planning a study to estimate the proportion of single BYU students that do not work during the school year. You think that the proportion is around 0.10. You want to estimate this proportion with 90% confidence and a margin of error no greater than 0.03. How large a sample do you need?
    1. 770
    2. 271
    3. 387
    4. 1323
12. ‘Statistically significant’ is synonymous with all of the following except one. Which statement is an **INCORRECT STATEMENT**?
    1. The probability of the null hypothesis being true is less than .
    2. The p-value is less than .
    3. The difference between the observed value of the statistic and the claimed parameter value is too large to attribute to chance variation.
    4. If *H0* 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 *H0*is true.
13. 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?
    1. There is less than a 5% chance that ethnocentrism is higher for attenders than for non-attenders.
    2. 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.
    3. The probability that ethnocentrism is the same for attenders as for non-attenders is 0.05.
    4. There is less than a 5% chance that if you took another sample, ethnocentrism would be higher among non-attenders than among attenders.
14. Assume that a researcher decides to test the hypotheses and with . What is the P-value if and the text statistic is *z* = -0.67.
    1. 0.6667
    2. 0.5028
    3. 0.3333
    4. 1.4972
15. Anderson Windows produces vinyl-coated aluminum windows. A thickness of 15 mm for the vinyl coating is required in order for the windows to seal properly and not jam. The standard deviation of their process for coating the windows is 1.2 mm. Their quality control procedure is to take a sample of 4 windows every 24 hours, measure the vinyl thickness of each, and compute the sample mean. The process would be considered to be out of control if the sample mean was less than \_\_\_\_\_\_\_\_.
    1. 11.4 mm
    2. 13.2 mm
    3. 13.8 mm
    4. 12.6 mm
16. 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:

*H0: mean weight loss is the same for both training regiments.*

*Ha: mean weight loss is greater for interval training than endurance training.*

What is a type II error?

* 1. To conclude that weight loss due to interval training is greater when, in fact, it is not.
  2. To conclude that weight loss due to interval training is greater when, in fact, it is.
  3. To conclude that weight loss is the same for both when, in fact, it is.
  4. To conclude that weight loss is the same for both when, in fact, it is greater for interval training.

1. A sample of 18 pigs was fed a new diet for 8 weeks. Their mean weight gain during this period was lb/day with s = .22 lb/day. When can we use the formula to estimate the mean weight gain of all pigs under this new diet?
   1. If the sample of pigs is large.
   2. If is known.
   3. If the sample of pigs is SRS and if weight gains in the sample are not outliers.
   4. If the sample pigs is SRS (no other assumptions are required due to the central limit theorem).
2. Referring to the question 52, what is the response variable?
   1. The new diet.
   2. Weight gain during the 8 week period.
   3. The mean weight gain.
   4. Final weight at the end of the 8 week period.
3. Referring to 52 and assuming conditions are met, give a 98% confidence interval for mean weight gain of the pigs.
   1. (.64 lb/day, .92 lb/day)
   2. (.65 lb/day, .91 lb/day)
   3. (.66 lb/day, .90 lb/day)
   4. (.68 lb/day, .99 lb/day)
4. Using the information in question 52, and assuming that it is appropriate to use t-procedures in this case, the p-value for the test of versus is
   1. Between .02 and .025
   2. Between .05 and .10
   3. Between .10 and .20
   4. Between .04 and .05
5. An experiment was conducted with 6 pairs of rats. Each rat in a pair came from the same litter. One rat from each pair was randomly chosen and assigned to live alone in a cage with no toys. The other rat in each pair was assigned to live with 11 other rats in a cage supplied with toys. After a month, the rats were sacrificed and their brain cortexes were weighed. The researchers were trying to show that a favorable psychological environment stimulates the growth of cortex material (grey matter) in the brain. The data are as follows:

|  |  |  |
| --- | --- | --- |
| Pair | Weight of cortex for isolated rat (mg) | Weight of cortex for non-isolated rat (mg) |
| 1 | 657 | 689 |
| 2 | 623 | 656 |
| 3 | 652 | 668 |
| 4 | 654 | 660 |
| 5 | 658 | 679 |
| 6 | 646 | 663 |

The researchers wanted to test a null hypothesis that a favorable psychological environment does not stimulate the growth of cortex material versus an alternative hypothesis that it does. What type of test should be performed in these data?

* 1. ANOVA on six means
  2. Two sample t test on means
  3. Matched pairs t test
  4. Two-sample t test on proportion

1. Using the information in the above question, what is the explanatory variable?
   1. Weight of cortex
   2. Rat litter
   3. Type of cortex material
   4. Whether rat was isolated vs. not isolated

*Use the following information in the next three 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, hours, and a standard deviation, hours. The 56 single men had a mean, hours, and a standard deviation, hours.

1. 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 \_\_\_\_\_\_\_.
   1. for matched pairs
2. To test

*H0*: The average hours of TV watching for married students equals the average TV watching for single men at BYU

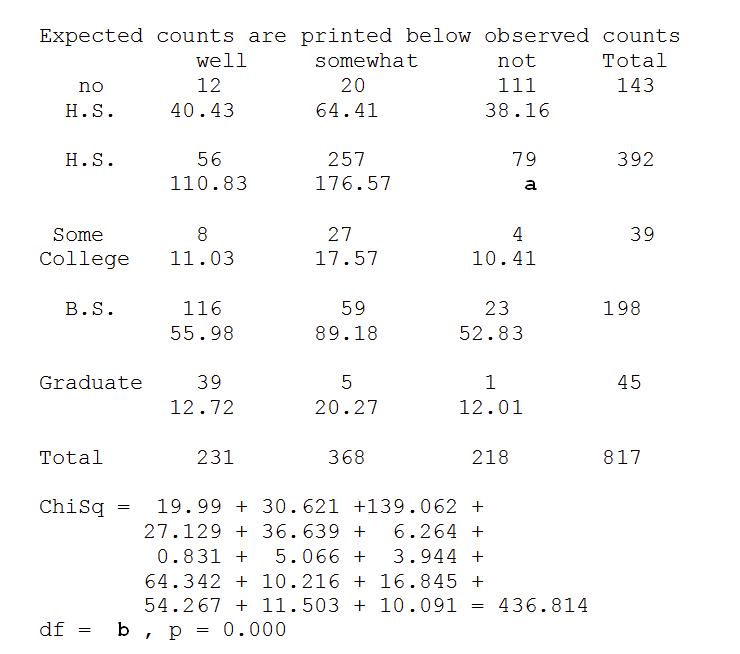
*Ha*: The average hours of TV watching for married students is greater than the average TV watching for single men at BYU

The P-value was found to be 0.018. At , what should we conclude?

* 1. 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.
  2. 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.
  3. 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.

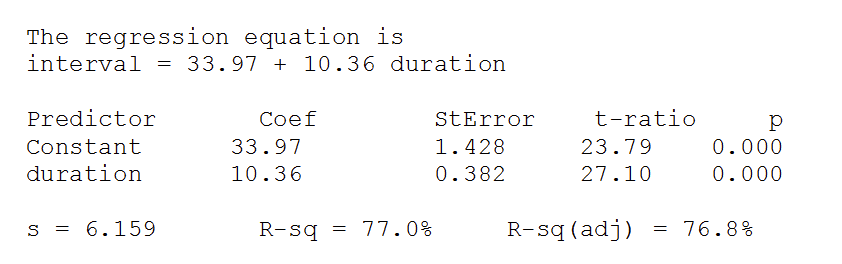
1. On the basis of the study described above, which condition needs to be met in order for the results of the approximate t-procedure to be valid?
   1. Plots of both data sets need to be bell-shaped symmetric.
   2. Students need to be selected using an SRS.
   3. The sample sizes must be equal.
2. A simple random sample of 500 17-year olds was taken. Of these, 180 knew that the author of *The Canterbury Tales* was Chaucer. What is the value of for these data?
   1. 0.034
   2. 0.180
   3. 0.360
   4. 0.500
3. Referring to the Chaucer data above, a 99% confidence interval for the percentage of all 17-year olds who know that the author of *The Canterbury Tales* was Chaucer, is (30.5%, 41.5%). Which of the following is a correct interpretation of this interval?
   1. 99% of the time, the value of will be in the interval (30.5%, 41.5%).
   2. The percentage of 17-year olds in the sample of 500 who know the author of *The Canterbury Tales* will be in the interval (30.5%, 41.5%) with 99% confidence.
   3. We are 99% confident that the percentage of all 17-year olds who know the author of *The Canterbury Tales* will be in the interval (30.5%, 41.5%).
4. In a simple random sample of American adults, 10.6% of 1067 men were left-handed and 7.9% of 1170 women were left-handed. If the test statistic for investigating whether a greater percentage of men are left-handed than left-handed women in the whole population is z = 2.22, what is the p-value?
   1. 0.9869
   2. 0.0132
   3. 0.0222
   4. 0.0264
5. A 90% confidence interval estimate of the difference between the percentage of men who are left-handed and the percentage of women who are left-handed is 2.7%­­2.0%. On the basis of this interval, can we conclude that the percentage of left-handed men differs significantly from the percentage of left-handed women?
   1. Yes, because the estimate, is positive.
   2. Yes, because the confidence interval for does not include zero.
   3. No, since confidence intervals cannot be used to compare equality of proportions.
   4. No, because the interval gives a range of values of and and does not provide information for comparing them.
6. 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?
   1. Independent two-sample t for means.
   2. Independent two-sample z for proportions.
   3. Chi-square test.
   4. Analysis of Variance.

The following information applies to questions 66-69: A simple random sample of 817 full-time workers in the U.S. aged 40-50 were asked to give their educational level (no high school, high school, some college, bachelor degree, some graduate education) and their level of financial satisfaction with their job (well satisfied, somewhat satisfied, not satisfied). The following MINITAB output test whether there is an association between education level and job satisfaction. Rows are educational levels and columns are satisfaction levels:



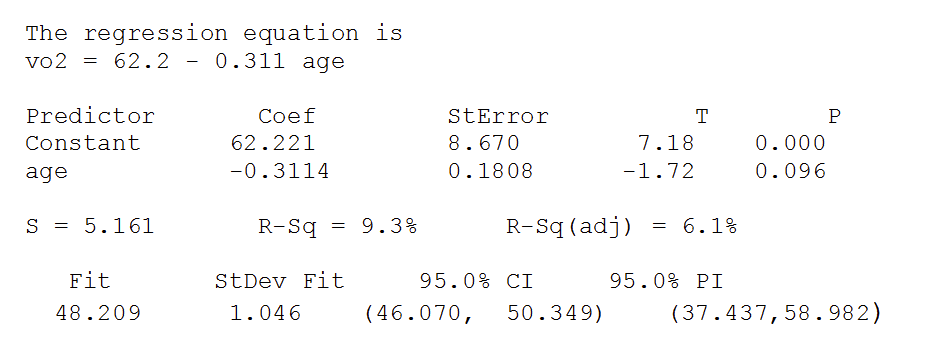
1. If education level and financial satisfaction are not associated, what is the expected number of high school graduates who are not satisfied with their work? (i.e. what is the value of ‘a’?)
   1. 104.6
   2. 79.0
   3. 21.1
   4. 6.264
2. What are the degrees of freedom? (i.e. what is the value of ‘b’?)
   1. 2
   2. 3
   3. 4
   4. 8
   5. 15
   6. 816
3. Is it appropriate to use the chi-square test for this set of data?
   1. Yes, because the sample size is over 40.
   2. Yes, because all expected counts exceed 5.
   3. No, because one row total is less than 40.
   4. No, because not all expected counts exceed 5.
4. Assuming that using the chi-square test is appropriate, the conclusion is:
   1. There is strong evidence that educational level and financial satisfaction are associated.
   2. There is strong evidence that educational level and financial satisfaction are not associated.
   3. Educational level determines (or causes) the level of financial satisfaction.
   4. There is no evidence of an association between educational level and financial satisfaction.

The following information applies to question 70 to 74: A simple random sample of 33 eruptions of the Old Faithful geyser was observed. The duration of each eruption was recorded as well as the time until the next eruption (both in minutes). The scatterplots gives a positively sloped football shape of data points. The idea was to develop an equation for predicting the time until the next eruption from duration of the current eruption. Regression analysis on MINITAB produced the following results:



1. The percent of variation in time until the next eruption explained by the least squares regression on eruption duration is
   1. 23.79%
   2. 6.159%
   3. 77.0%
   4. 87.7%
2. The predicted time until the next eruption with duration of 4.5 minutes is
   1. 114.6 min
   2. 163.4 min
   3. 46.6 min
   4. 80.6 min
3. A measure of variability of the time until the next eruption (y) about the regression line is
   1. 27.10
   2. 6.159
   3. 0.3822
   4. 1.428
4. For each one minute increase in duration of eruption, the average increase in the time until the next eruption is estimated to be
   1. 10.36
   2. 27.10
   3. 0.382
   4. 33.97
5. Using the results of a test on slope given in the regression output, what can we conclude?
   1. Duration of eruption can be used to predict time until the next eruption of Old Faithful.
   2. Duration of eruption cannot be used to predict time until the next eruption of Old Faithful.
   3. There is a steep relationship between duration and time until the next eruption of Old Faithful.
   4. There is strong evidence of no relationship between duration and time until the next eruption of Old Faithful.

For problems 75-79: A regression analysis was run to investigate the relationship between VO2 (oxygen consumption in ml/kg) and age (years) for a sample of middle-aged men. The following MINITAB output was obtained, based on a sample of 31 men. The last line of the output was printed in response to specifying that 95% confidence/prediction intervals were desired for age=45 years.



1. A 95% interval for the VO2 of Bill, who is 45 years old, is
   1. (46/070, 50.349)
   2. (37.437, 58.982)
   3. (8.670, 62.221)
   4. Not given
2. A 95% interval for the mean V02 of men who are 45 years old is
   1. (46.070, 50.349)
   2. (37.437, 58.982)
   3. (8.670, 62.221)
   4. Not given
3. What is the P-value for the test of versus ?
   1. 0.096
   2. 0.048
   3. 0.311
   4. 0.000
4. What percent of the variation in VO2 can be explained by age?
   1. 62.2%
   2. 48.2%
   3. 31.1%
   4. 9.3%
   5. 5.2%
5. To do regression inference, the data must satisfy all of the following **except**
   1. The response variable (y) has a Normal distribution at each value of x.
   2. The values of the explanatory variable (x) must follow a Normal distribution.
   3. The true relationship must be linear.
   4. The standard deviation of the y’s about the true line is the same everywhere.
6. Which of the following set of hypotheses is appropriate for testing equality of three means in ANOVA?
   1. versus
   2. versus
   3. versus At least one differs from the others
   4. versus At least one differs from the others
7. 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
   1. there is strong evidence that hair dryers and black-and-white televisions cause leukemia.
   2. 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.
   3. while these results may be statistically significant, they are not necessarily practically significant.
   4. because multiple tests were performed, the results are only suggestive, not conclusive.

**True or False Questions**

1. When no information is available about the value of p and we need to determine sample size needed to estimate proportion, we can safely use in the sample size formula.
2. We check n and n before obtaining the p-value to test .
3. We compute the standard error of using the formula:
4. The mean of the sampling distribution of the sample proportion equals .
5. , the difference in two population means, represents the parameter used to compare the means of two populations.
6. Margin of error for an approximate confidence interval for p is
7. Correlation ignores the distinction between explanatory and response variables.
8. Changing the unit of measurement in the X or Y variable changes the value of r.
9. Correlation is a valid measure of strength of relationship whenever the relationship between two different quantitative measurements on each individual appears linear in the scatterplot.
10. When many tests of significance are performed on one set of data, the researcher is guilty of performing multiple analyses and inflating the overall α.

**Answers**

1. C
2. D
3. B
4. B
5. F
6. G
7. D
8. C
9. C
10. B
11. D
12. C
13. A
14. B
15. C
16. B
17. A
18. C
19. C
20. C
21. D
22. B
23. B
24. C
25. A
26. B
27. A
28. C
29. A
30. B
31. C
32. A
33. F
34. A
35. B
36. D
37. D
38. D
39. B
40. B
41. C
42. C
43. C
44. C
45. D
46. B
47. A
48. B
49. B
50. B
51. D
52. C
53. B
54. B
55. C
56. C
57. D
58. B
59. A
60. B
61. C
62. C
63. B
64. B
65. D
66. A
67. D
68. B
69. A
70. C
71. D
72. B
73. A
74. A
75. B
76. A
77. A
78. D
79. B
80. C
81. D
82. T
83. F
84. T
85. F
86. T
87. F
88. T
89. F
90. T
91. T