Problems: T-TESTS

1. Two sets of 60 randomly selected 5th grade children were taught spelling according to one of two methods. One method was the old one that had been used for the last twenty years. The other was a new, computer assisted technique. You wish to know if the new method is an improvement over the old one. At the end of the school year, 40 of the 60 children who used the new method passed a vocabulary test, whereas only 24 of the 60 children who used the old method passed the test. (Note: Use a method other than chi-square to answer this question.)

a. State the appropriate null and alternative hypotheses.

b. What is your rejection rule? (Use the .01 significance level.)

c. State your conclusions in words.

2. The Women’s Youth Crisis Centers (WYCCs) in Minneapolis are run entirely by women volunteers from local religious groups. These women founded the WYCCs in 1985 with the express goal of “reducing the pregnancy rate among teenagers in our high schools.” (Based on high school health records in 1985, the pregnancy rate was such that one out of every twenty teenage women in Minneapolis high schools had been pregnant previously or became pregnant during that year.) At the founding of the WYCCs there was considerable disagreement among the "founding mothers" as to whether religious terminology (e.g., references to God, prayer, etc.) should be used by counselors during their "pregnancy-preventative counseling." Unable to resolve this disagreement, they set up two counseling centers: the Religious Center (RC), where religious terminology was used during counseling, and the Secular Center (SC), where no such terminology was used. To keep track of the centers’ performance, high school health records of teenaged visitors to either center were obtained after they graduated from high school. In 1992 1000 teenage women had been counseled at the WYCCs (300 at the Religious Center and 700 at the Secular Center) and had subsequently graduated from high school. Of the 300 teenage visitors to the Religious Center, 12 became pregnant after their initial visit to the RC; of the 700 teenage visitors to the Secular Center, 21 became pregnant after their initial visit to the SC.

a. Do the data suggest that religious terminology contributes to the effectiveness of pregnancy-preventative counseling? (Use the .05 level of significance and show your work.)

b. Using the 1985 pregnancy rate among teenage women in Minneapolis high schools as your point of comparison, do your data provide evidence that women visitors to the WYCCs are less likely than this to become pregnant after their initial visit to one of the centers? (Hints: This question does not require that you distinguish between the two centers in any way. As a result, you will want to combine your data on the Religious and Secular Centers. And once again, use α = .05 and show your work.)

c. Calculate the P-value of your finding in part b. State in words what this P-value means. (Hint: This statement should be phrased something like, "The P-value of ______ is the probability that . . .")
d. A colleague challenges your conclusion in part b, by showing you that five percent of teenage women in Minneapolis high schools had been pregnant previously or became pregnant during 1992. (I.e., the pregnancy rate in 1992 is the same as it was in 1985.) In a sentence or two, how might you explain the seeming inconsistency between the P-value (calculated in part c) that suggests a decrease in the pregnancy rate and the truth (just made known to you by your colleague) that there has been no change in the pregnancy rate?

3. In the 1984 NORC survey, respondents were asked if they were "afraid to walk at night in their neighborhood." (Note: Do not use chi-square to answer this question.)

a. Construct a 95% confidence interval for the difference in proportions between males and females who are afraid to walk in their neighborhoods at night. Interpret the interval.

b. Use the same data to test the null hypothesis that the two proportions are equal.

c. You have now used two techniques (confidence interval and hypothesis test) to compare the proportions of males vs. females who are afraid to walk in their neighborhoods at night. Compare these two techniques by explaining why the null hypothesis is only rejected at the .05 level of significance when the 95% confidence interval does not contain the value of zero.

You will need the following one-line SPSS program to get the data to do this problem:

crosstabs tables = fear by sex.

4. During the last 35 years Isaacs Counseling Services (ICS) has trained thousands of social workers for employment in welfare bureaus throughout the United States. The board of directors at ICS has recently demanded that the company evaluate the effectiveness with which their training instills altruistic values in those whom they train. You are in charge of making this evaluation. After trying various methods of evaluating altruism, you decide that the best altruism measure requires that the respondent choose between altruistic and nonaltruistic alternatives in a real-life situation. Here is the situation that you set up: Each person-being-evaluated is left in a waiting room with another person. After a few minutes of waiting, the "other person" slumps in his chair, then slowly slides down off the chair and onto the floor (presumably) unconscious.

When faced with this situation, 35 out of a random sample of 42 ICS graduates took the altruistic alternative, and tried to revive (or otherwise help) the other person. In contrast, 26 out of a random sample of 39 social workers who were not graduates of the ICS program took this same altruistic alternative. (Hint: Do not make the mistake of trying to use a chi-square test in answering this question.)
a. Give the null and alternative hypotheses that you would use to test whether ICS graduates are more altruistic than other social workers?

b. What rejection rule would you use in a test of the hypotheses in part a? (Use the .05 significance level.)

c. Do you have evidence at the .05 significance level that ICS graduates are more altruistic than other social workers? Explain your answer.

d. Assume that with your altruism measure, precisely 10% more ICS graduates than other social workers are altruistic. Given this fact, how much power did you have as you tested the hypotheses given in part a?

5. There are two competing theories of old-age-despair. Global Neglect Theory argues that old people’s despair will be lessened if others (who could be relatives of the old people) do kind things for them. Relative Neglect Theory argues that old people’s despair will increase if they believe that relatives are neglecting them. Professor I. Hatemom (an outspoken proponent of relative neglect theory) warns, “Nursing homes should be hesitant to make old people’s lives too comfortable. Kindness by non-relatives (such as the nursing staff) will only increase the despair of those whose families already neglect them.” You remain uncommitted to either theory and design an experiment to test whether either theory is true.

You obtain visitation records from 1000 Iowa nursing homes and randomly sample one old resident from each—ensuring only that the resident is in good health and that the resident has had no visitors for at least 5 years. The 1000 residents are randomly divided into two equal groups. The 500 residents in one group receive a bouquet of flowers (“Complements of the State of Iowa”) each month for the two years of your study; the other 500 residents receive no flowers. At the end of two years you find that 30 of the residents, who received flowers, had committed suicide and 15 of the residents, who had not received flowers had committed suicide. (NOTE: You should not use chi-square in answering this problem.)

a. Taking suicide as your measure of despair, give the appropriate null and alternative hypotheses.

b. Using the .05 level of significance, state your rejection rule.

c. Which theory do the data support? Justify your answer in the light of your findings. (Hint: What two pieces of information are needed to justify your answer?)

d. Imagine that among healthy Iowan nursing home residents (who have had no visitors for at least 5 years) the probability that a bouquet-per-month recipient commits suicide is .02 greater than the probability that a non-flower-recipient commits suicide. What is the power of the hypothesis test that you have performed in the previous parts of this problem? (Be sure to use the significance level and variance estimates from previous parts of this problem.)

36
6. You are doing research on worker morale in two companies. About two years ago one of the companies introduced a policy according to which workers who show "creativity" in their jobs receive higher pay increases than noncreative workers. The purpose of this policy was to provide workers with a sense that rewards for their work were allocated fairly. Pay increases for workers in both companies were announced last month. After these announcements you asked the following question to two randomly sampled groups of workers (one from each company):

On a scale from 1 to 10, where 1 is the least fair of companies and 10 is the most fair of companies, how fair do you think your company is in allocating rewards to its workers?

(Note: For the purposes of this analysis, assume that responses to the "fairness question" yield a 10-point interval-level measure.) Means and standard deviations on this measure are as follows:

<table>
<thead>
<tr>
<th>Company</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creativity rewarded</td>
<td>7.1</td>
<td>4.8</td>
<td>21</td>
</tr>
<tr>
<td>Creativity not rewarded</td>
<td>4.6</td>
<td>6.7</td>
<td>25</td>
</tr>
</tbody>
</table>

a. Do you have evidence at the .05 significance level that the policy of rewarding creativity achieved its purpose? (Use the above data to justify your answer.)

b. If the company that rewarded creativity really did have workers with a 2-point higher average score on the fairness scale than did workers in the other company, what is the probability of making a Type II error in an analysis of a different pair of samples (with the same standard deviations and sample sizes) than those analyzed in part a? (Use the .05 significance level and show your work.)

7. You design an experiment to test whether physical contact increases motivation. You randomly divide a group of 6th grade children into two groups of twenty-five children each. Children from each group are individually asked to perform a task. Each child's performance is then assigned a motivation score from 1 to 100 points (by a panel of judges), according to how energetically the child accomplished the task. (High scores correspond to high motivation.) The two groups are treated identically, except that for each of the children in one group, the experimenter touches the child's shoulder just before it begins the task. Children from the other group are not touched. Your data on the children's motivation scores are as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Touched Mean</th>
<th>Touched Standard Deviation</th>
<th>Not touched Mean</th>
<th>Not touched Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>35.89</td>
<td>87.48</td>
<td>32.62</td>
<td>52.60</td>
</tr>
</tbody>
</table>

a. What are your null and alternative hypotheses?

b. What do you conclude? I.e., do your data suggest that physical contact increases motivation at the .05 level of significance? (Justify your answer.)
c. For just this part of the question, imagine that you do not know the motivation scores for the two group means. Imagine also that you know that when 6th graders are touched, they have motivation scores five points higher on the average than 6th graders who are not touched. What is the power of your test of the hypotheses listed in part "a."? (Use the .05 level of significance and the above standard deviation measures.)

8. You wish to determine whether telephone interviewers make fewer errors when they use CATI (Computer-Aided Telephone Interviewing) techniques instead of traditional noncomputer-aided telephone interviewing techniques. You hire 50 interviewers. You randomly select half of them and train these 25 in CATI techniques. The other 25 interviewers receive traditional training as telephone interviewers.

You do not tell your interviewers that they are the subjects of your research. Unknown to the interviewers, they each interview the same 10 people, who are shills in the experiment. (That is, each of these 10 shills is actually someone cooperating with you in your research.) Each shill gives the same 100 answers on his/her interview; each shill gives answers that differ from those given by the other 9 shills. Thus each interviewer should record the same 100 answers as data from each of the 10 interviews. Note that if an interviewer correctly recorded each answer from each subject, the interviewer would have a total of 1,000 correct answers. If no answers were correctly recorded by an interviewer, the interviewer would have a "correctness score" of zero. Interviews are tape recorded to assure that shills give the 100 answers they are supposed to give. Your data are as follows:

Table 1: Mean Number of Items Correctly Recorded with Variance and Sample Size for Interviewers Who Did versus those Who Did Not Use CATI Interviewing Techniques.

<table>
<thead>
<tr>
<th>Interviewers</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATI Interviewers</td>
<td>990</td>
<td>10.2</td>
<td>25</td>
</tr>
<tr>
<td>Other Interviewers</td>
<td>978</td>
<td>16.8</td>
<td>25</td>
</tr>
</tbody>
</table>

a. State the appropriate null and alternative hypotheses.

b. What is your rejection rule? (Use the .05 level of significance.)

c. What do you conclude? I.e., do your data suggest that CATI systems decrease interviewer errors? (Justify your answer.)

9. You wish to determine if women who work outside the home have worse marriages than women who work at home as homemakers. From a list of all Iowa residents you randomly sample 250 people. Each of these people is contacted to determine their gender, marital status, and working hours. After discarding all males, single people, and part-time workers, you arrive at your final sample of 25 married women who have full-time jobs outside their homes and 61 married women who work at home as homemakers. Each of these 86 women completes a questionnaire which includes a series of items (the renowned Roberts Marital Quality Index) that assess the quality of the respondent's marriage. Scores on the Roberts Marital Quality Index range from 0 = an awful marriage to 100 = a marriage made in heaven (that is, a "wonderful"
Means and variances on the Roberts Marital Quality Index are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work outside home</td>
<td>84.3</td>
<td>12.8</td>
</tr>
<tr>
<td>Work at home</td>
<td>79.8</td>
<td>3.9</td>
</tr>
</tbody>
</table>

a. What are your null and alternative hypotheses?

b. What would be your rejection rule? (Use the .05 significance level throughout.)

c. What do you conclude? I.e., what do your data suggest about the hypothesized relation between Iowa women's marital quality and their working status?

d. Assume that women who work outside the home do in fact have worse marriages than women who work at home. In particular, working women have Roberts Marital Quality Index scores that are 3 points lower on average than the scores of women who work at home as homemakers. Assuming the variance estimate derived in part b, what would be the power of your hypothesis test?