## As you answer questions 1-3, begin by stating what case the problem falls under.

1.      An examination was given to a random sample of 10 students from a public school (population 1) and a random sample of 15 students from a Catholic school (population 2).  In the public school sample the mean grade was 70 and in the Catholic school sample the mean grade was 74.  Test scores are known to be normally distributed in both schools.  For each of the following situations, (i) indicate whether the two schools differ from each other at the .05 level of significance (use a two-tailed test and the 5 step hypothesis testing procedure) (ii) construct the 95% confidence interval for the difference between the average scores of the two schools, and tell whether the c.i. indicates that you should accept or reject $H_0$.  (HINT:  In each case, the computed test statistic is very close to the acceptance region - sometimes just within it, sometimes just outside it.)  After doing the calculations by hand, use Stata to verify your answers. (You'll need to have Stataquest installed to answer all the problems.)

      a.      $\sigma_1 = 4$, $\sigma_2 = 6$
      b.      $s_1 = 4$, $s_2 = 6$, $\sigma_1$ and $\sigma_2$ are equal but unknown
      c.      $s_1 = 4$, $s_2 = 6$, $\sigma_1$ and $\sigma_2$ are unequal and unknown [HINT: Degrees of freedom = about 23.  Confirming this is OPTIONAL]

2.      Nine students in each of two statistics classes (groups A and B) were carefully matched according to age, sex, grade point average, and SAT scores.  The two groups were then taught statistics by two different methods (but with the same instructor).  Their scores on the (same) final exam were as follows.  (To save you some drudgery, I have computed the last 5 columns for you – but you'll have to figure out what to do with that information.  There are a couple of ways to solve the problems.)

| Pair | A Score | B Score | A - B = D | $(A - B)^2 = D^2$ | $(A - \overline{A})^2$ | $(B - \overline{B})^2$ | $(A - \overline{A})(B - \overline{B})$ |
|------|---------|---------|-----------|-------------------|------------------------|------------------------|----------------------------------------|
| 1 | 76 | 83 | -7 | 49 | 1.78 | 4 | 2.67 |
| 2 | 85 | 91 | -6 | 36 | 106.78 | 100 | 103.33 |
| 3 | 83 | 72 | 11 | 121 | 69.44 | 81 | -75.00 |
| 4 | 84 | 93 | -9 | 81 | 87.11 | 144 | 112.00 |
| 5 | 65 | 75 | -10 | 100 | 93.44 | 36 | 58.00 |
| 6 | 71 | 87 | -16 | 256 | 13.44 | 36 | -22.00 |
| 7 | 60 | 75 | -15 | 225 | 215.11 | 36 | 88.00 |
| 8 | 76 | 79 | -3 | 9 | 1.78 | 4 | -2.67 |
| 9 | 72 | 74 | -2 | 4 | 7.11 | 49 | 18.67 |
| $\Sigma$ | 672 | 729 | -57 | 881 | 596 | 490 | 283 |

      (a)      Test $H_0$: $\mu_A$ - $\mu_B$ = 0 vs. $H_A$: $\mu_A$ - $\mu_B$ <> 0 at $\alpha$ = .05.
      (b)      Construct the 95% c.i. for E(A) - E(B).  Indicate whether, according to the c.i., you should accept or reject $H_0$.

3.      A sample poll of 300 voters from district A and 200 voters from District B showed that 56% and 48% respectively were in favor of a given candidate.

      (a)     Using $\alpha = .05$, test whether the two districts significantly differ in their support of the candidate.  Use our hypothesis testing procedures.

      (b)     Construct the approximate 95% c.i. for $p_1 - p_2$.  According to the c.i., should $H_0$ be accepted or rejected?

Do the following AFTER you have completed the rest of homework # 5.  You may want to use 2sample.sps as a template and modify as needed.  Or, you can use the SPSS menus and the SPSS Data editor.  Confirming some or all of your answers with Stata is OPTIONAL but recommended.

4.      A.      Confirm your answers to problem #2 by using SPSS.  Recall that you will have 9 cases, each of which has data for 2 variables (ASCORE and BSCORE).  You will use the SPSS paired samples t-test routine.  If using the pull-down menus in SPSS, after having entered the data you would choose ANALYZE/COMPARE MEANS/PAIRED SAMPLES T-TEST.

      B.      Using the same data as in problem 2, see what happens when Groups A and B are erroneously treated as though they are independent.  Rework problem 2A – For both cases II and III, use SPSS to test whether the means of the two groups are significantly different, i.e. Test $H_0$: $\mu_A - \mu_B = 0$ vs. $H_A$: $\mu_A - \mu_B <> 0$ at $\alpha = .05$.  Recall that you will have 18 cases, each of which has data for 2 variables: SCORE and GROUP (1 = Group A, 2 = Group B).  You will use the SPSS independent samples t-test routine.  If using the pull-down menus, in SPSS you would choose ANALYZE/COMPARE MEANS/INDEPENDENT SAMPLES T-TEST.  Discuss how and why this mis-specification affects your results. [NOTE: If you are also solving this using Stata, keep in mind that the `unpaired` option on the `ttest` command can save you the trouble of entering the data twice.]

      C.      Using SPSS, treat problem #3 as though it fell under 2 sample tests, case II, $\sigma$'s unknown but assumed equal.  Compare SPSS's results with your earlier calculations.  If you've done things right, the results should be almost identical.  Note that, although there are 500 cases, you actually only need to enter 4 lines of data for each of 3 variables: DISTRICT (1 = District A, 2 = District B), FAVOR (1 = Favor Candidate, 0 = Does not favor), and FREQ (the number of cases that fall into each of the four possible combinations of DISTRICT and FAVOR, e.g. 168 people in district A favor the candidate.) Look at 2Sample.sps if not sure how to set up the data.  If using the pull-down menus, in SPSS you would choose ANALYZE/COMPARE MEANS/INDEPENDENT SAMPLES T-TEST. [NOTE: If you are also using Stata, remember that the `prtesti` command can be used for Case V.]