1. A study was undertaken to evaluate the effect of percutaneous transluminal coronary angioplasty (PTCA) in patients with one-vessel coronary artery disease. One hundred and seven patients with coronary artery disease were given PTCA. Patients were given exercise tests at baseline and after 6 months of follow-up. Exercise tests were performed up to maximal effort until symptoms (such as angina) were present. The “change” in the duration of exercise was calculated. “Change” is defined as the 6 month test minus the baseline test. The mean change was 2.1 minutes and the standard deviation of the changes was 3.1.

Please answer the following questions regarding this study.
(10 points total, 2 points for each of a – e)

You may use a computer (STATA for example) to assist you with relevant computations.

(a) What statistical test can be performed to see if there has been a statistically significant change in duration of exercise for this group of patients given PTCA?

(b) Perform the test you suggested in part a, report a p-value and interpret this result in a sentence.

(c) Computed a 95% confidence interval for the mean change in exercise duration.

(d) Do the findings of the p-value an confidence interval agree in terms of the population difference?

(e) Can it be concluded from this study that PTCA is effective in increasing exercise duration? Are there any limitations or weaknesses in this study for answering that question?

2. Many times in public health and medical research, the interest is in comparing efficacies of differing treatments, therapies, or public health outreach programs. An optimal design for such studies involves starting with a random sample of subjects from the population of interest and then randomizing subjects to the various treatment/therapy/intervention groups. Randomization can be accomplished in a number of ways (we will discuss this in way more detail in SR2): an example for randomizing subjects to two groups is flipping a coin to determine whether a subject is assigned to one of two groups being compared. The hope is that with randomization the only systematic difference in subjects in the groups being compared is the grouping itself (ie: intervention or control etc.). This makes it easier to attribute any differences in the efficacy measurements to the treatment/program as opposed to other differences between subjects in the groups being compared.

An article from the September 1, 2009 edition of the American Journal of Public Health presents the results of result comparing two approaches to counseling related to sexual health. From the article abstract:
Objective. We evaluated the efficacy of a brief, clinic-based, safer sex program administered by a lay health adviser for young heterosexual African American men newly diagnosed with a sexually transmitted disease (STD).

Methods. Subsequent to STD diagnosis, eligible men (N=266; aged 18–29 years) were randomized to either a personalized, single-session intervention (delivered by a lay health adviser) or standard of care. We conducted behavioral assessments at baseline and 3 months postintervention (retention was 74.1%). We also conducted a 6-month clinic record review.

As part of the article, the authors present a table comparing demographic differences in the subjects randomized to the intervention or “placebo” (current standard of care) groups. Frequently this is done to demonstrate that randomization “worked”: i.e. that there were no statistical or qualitative differences in the characteristics of subjects assigned to the various groups being compared.

| TABLE 1—Demographic and Other Baseline Attributes of Enrollees in a Risk-Reduction Intervention Evaluation of African American Men Aged 18 to 29 Years Newly Diagnosed With an STD, by Group Assignment; Southern United States, September 2004 Through May 2006 |
|---------------------------------|---------------------------------|---------------------------------|
| **Age, y**                      | **Intervention (n=141), Mean ±SD or No. (%)** | **Control (n=135), Mean ±SD or No. (%)** | **P** |
|                                 | 23.1 ± 3.4                      | 23.5 ± 3.1                      | .49 |
| **Net monthly income $1,000**   | 10 (73.9)                      | 23 (33.9)                      | .22 |
| **Current relationship is monogamous** | 68 (48.6)          | 60 (46.9)                      | .92 |
| **Current relationship is not monogamous** | 60 (42.1)          | 54 (43.2)                      | .86 |
| **Previously taught how to use condoms** | 127 (93.3)       | 119 (88.7)                      | .72 |
| **Multiple STDs diagnosed at baseline** | 41 (29.3)           | 27 (21.1)                      | .19 |
| **Baseline diagnosis included chlamydia** | 56 (39.3)           | 59 (43.9)                      | .83 |
| **Baseline diagnosis included gonorrhea** | 87 (61.7)           | 76 (56.3)                      | .44 |
| **Determined condom use skills** | 3.69 ± 2.4                     | 2.60 ± 1.67                    | <.001 |
| **Number of sexual partners, past 3 months** | 2.01 ± 2.73                 | 3.06 ± 2.43                    | .69 |
| **Unprotected acts of sexual intercourse, past 3 months** | 16.0 ± 7.3           | 14.3 ± 11.9                    | .72 |
| **Risk concerns last time sexual intercourse occurred** | 74 (52.5)           | 53 (42.4)                      | .19 |

Note. STD = sexually transmitted disease. All results pertain to men who self-identified as heterosexual.

a) How many subjects were randomized to the Intervention Group?
b) How many subjects were randomized to the Control Group?
c) What is the sample mean and standard deviation of age for those randomized to the Intervention group?
d) What is the sample mean and standard deviation of age for those randomized to the intervention group?
e) What are the competing (null and alternative) hypotheses being tested by the p-value of 0.49 reported in the “age” row of the table?
f) What is the interpretation of the reported p-value?

3. A researcher wishes to determine if vitamin E supplements could increase cognitive ability among elderly women. In 1999 the researcher recruits a sample of elderly women age 75-80. At the time of the enrollment into the study, the women were randomized to either take Vitamin E, or a placebo for six months. Randomization involved flipping a coin and assigning a subject to the “Vitamin E” group if the coin came up “heads” and to the “placebo” group if the coin came up “tails”. At the end of the six month period, the women were given a cognition test. Higher scores on this test indicate better cognition. The mean and standard deviation of the test scores of 81 women who took vitamin E supplements was 27 and 6.9 respectively. The mean and standard deviation of the test scores of the 90 women who took placebo supplements was 24 and 6.2, respectively.

Please answer the following items:
(14 points total, 2 points for each of a –g)

You may use a computer (STATA for example) to assist you with relevant computations.

a) What statistical can be used to compare the mean scores?
b) Perform the test you suggested in part a, report a p-value and interpret this result in a sentence.
c) Compute a 95% confidence interval for the mean difference in cognition test scores between Vitamin E and placebo groups.
d) Do the findings of the p-value an confidence interval agree in terms of the population difference?

e) Are there limitations to this study for drawing conclusions about whether vitamin E can enhance cognitive ability in elderly women?

f) What might you conclude from these study results?
g) Suppose I tell you that the women were not randomized to Vitamin E or placebo groups. Instead, the study was conducted by taking a random sample of 171 elderly women, and then classifying them based on their self-reported Vitamin E consumption (yes or no) at the time of selection. Would this affect your scientific conclusions based on the statistical analysis you performed in parts b and c? (This is an extension question: we haven’t formally covered this, but I just want you to think about it –see question 2, and we will formally cover in term 2: we will grade gently)
4. **Project: Hypothesis Testing in the Scientific Literature**

Find two articles in a scientific journal on a topic of interest to you that makes use of statistical hypothesis or significance testing to support its scientific conclusions. Journals you might look through include *New England Journal of Medicine, Lancet, American Journal of Public Health, Lancet, British Medical Journal*, etc.

For each of the two articles you select, please answer the following items.
*(12 points total: 4 points for each of parts a–d, 2 points for e for each of the two articles)*

(a) What is the citation of your article (author, title, journal, date of volume, pages)?

(b) What type of statistical analysis (analyses) was (were) used to perform the hypothesis test(s)?

(c) How are the results of the significance (hypothesis) testing reported? (e.g., do they say $p = .07$, $p < .05$, or “not significant,” etc.)

(d) Are confidence intervals included, and if so for what parameter (population mean, proportion, difference in means, etc.)?

(e) In a few sentences write down the main scientific findings of the study. This description should make clear how statistical significance testing was used to support the study’s conclusions (e.g., “the mean blood pressure of individuals in high stress occupations was significantly higher than individuals in low stress occupations ($p = .01$”).

**Sample Quiz Questions (Please note: these are a required part of the homework):** Choose the correct answer for the following multiple-choice questions, and justify (explain) your choice:
*(6 points total: 1 point for each correct answer, 1 point for each correct justification)*

4. Measurements on babies of mothers who used marijuana during pregnancy were compared to measurements on babies of mothers who did not. The sample mean head circumference was larger in the group who were not exposed to marijuana and the 95% confidence interval for the difference in mean circumference between the 2 groups was .61 to 1.19 cm. What can be said about the (2-sided) $p$-value for testing the null hypothesis of equal population means?

   (a) The $p$ value is greater than .05
   (b) The $p$ value is equal to .05
   (c) The $p$ value is less than .05
   (d) The $p$ value is 0
   (e) Cannot be determined from the information given.
5. A study of 100 patients is performed to determine if cholesterol levels are lowered after 3 months of taking a new drug. Cholesterol levels are measured on each individual at the beginning of the study and 3 months later. The cholesterol change is calculated which is the value at 3 months minus the value at the beginning of the study. On average the cholesterol levels among these 100 patients decreased by 15.0 (mmol/L) and the standard deviation of the changes in cholesterol was 40 (mmol/L). What can be said about the 2 sided p-value for testing the null hypothesis of no change in cholesterol levels?

(a) The p value is less than .05
(b) The p value is greater than .05
(c) The p value is equal to .05
(d) Cannot be determined from the information given

6. The standard error of a statistic (such as a sample mean or sample standard deviation) is

(a) the mean of the sampling distribution for the statistic
(b) measures how variable the statistic is in value across random samples of the same size (from the same population)
(c) the statistic divided by the square root of the sample size (\(\frac{\text{statistic}}{\sqrt{n}}\))