STAT3004: Basic Methods in Biomedical Statistics
Assignment 1
Due date: 28 September 2020, 5pm

The following problems are based on questions that can be found in *Fundamentals of Biostatistics, 8th Ed.* by Bernard Rosner, unless stated otherwise.

1. **Problems 7.56-7.59 in Rosner**
   Left ventricular mass (LVM) is an important risk factor for subsequent cardiovascular disease. A study is proposed to assess the relationship between childhood blood pressure levels and LVM in children as determined from echocardiograms. The goal is to stratify children into a normal bp group (¡ 80th percentile for their age, gender, and height) and an elevated bp group (= 90th percentile for their age, gender, and height) and compare change in LVM between the 2 groups. Before this can be done, one needs to demonstrate that LVM actually changes in children over a 4-year period. To help plan the main study, a pilot study is conducted where echocardiograms are obtained from 10 random children from the Bogalusa Heart Study at baseline and after 4 years of follow-up. The data are given in Table 1

ID	Baseline LVM (g)	4-year LVM (g)	Change (g)
1	139	163	24
2	134	126	-8
3	86	142	56
4	98	96	-2
5	78	111	33
6	90	108	18
7	102	167	65
8	73	82	9
9	93	77	-16
10	162	172	10
Mean	105.5	124.4	18.9
sd	29.4	35.2	26.4

Table 1: Pilot data on left ventricular mass

   (a) What test can be used to assess if there is a change in mean LVM over 4 years?
   (b) Implement the test in (a) and provide a two-tailed p-value.
   (c) Provide a 95% CI for the change in LVM over 4 years.
   (d) Since this was a pilot study, the main question of interest is how many subjects would be needed to detect an increase of 10 g in mean LVM over 4 years using a twosided test with $\alpha = 0.05$ and power = 80%? Hint: Assume that the estimated variance of change in LVM in the pilot study is the true variance of change in LVM.

2. **Problems 7.12-7.16 in Rosner**
Suppose the incidence rate of myocardial infarction (MI) was 5 per 1000 among 45- to 54-year-old men in 2000. To look at changes in incidence over time, 5000 men in this age group were followed for 1 year starting in 2010. Fifteen new cases of MI were found.

   (a) Using the critical-value method with $\alpha = 0.05$, test the hypothesis that incidence rates of MI changed from 2000 to 2010.

   (b) Report a p-value corresponding to your answer in (a).

   (c) Suppose that 25% of patients with MI in 2000 died within 24 hours. This proportion is called the 24-hour case-fatality rate. Of the 15 new MI cases in the preceding study, 5 died within 24 hours. Test whether the 24-hour case-fatality rate changed from 2000 to 2010.

   (d) Suppose we eventually plan to accumulate 50 MI cases during the period 2010-2015. Assume that the 24-hour case-fatality rate is truly 20% during this period. How much power would such a study have in distinguishing between case-fatality rates in 2000 and 2010-2015 if a two-sided test with significance level 0.05 is planned?

   (e) How large a sample is needed in (d) to achieve 90% power?

3. **Problems 8.2-8.6 in Rosner**
The sample mean and standard deviation calcium intake among 25 females between 12 and 14 years of age below the poverty level are 6.56 and 0.64 respectively. Similarly, the sample mean and standard deviation calcium intake among 40 females between 12 and 14 years of age above the poverty level are 6.80 and 0.76 respectively. A colleague tells you the standard deviations of the two groups are not significantly different.

   (a) Test whether there is a significant difference in means between the two groups. Provide a p-value.

   (b) Provide a 95% CI for the difference in means between the two groups.

4. **Problems 8.168-8.171 in Rosner**
The EPIC-Norfolk study, a study of diet and cancer in Great Britain, was performed to assess the relationship between dietary intake of vitamin C, plasma levels of vitamin C (in blood), and other predictors. One hypothesis is that smokers might have different vitamin C intake and vitamin C plasma levels than nonsmokers. Dietary intake of vitamin C was obtained using 7-day diet records in which a subject recorded what he or she ate in real time and a computer program was used to estimate nutrient intake based on the diet record data. The data in Table 2 were obtained for current smokers and nonsmokers.

   (a) What test can be used to compare the standard deviation of diet record vitamin C intake between current smokers vs. nonsmokers? Perform this test and identify whether there is a significant difference between the two variances.

Group	Mean vit C intake (mg/day)	sd (mg/day)	n
Nonsmokers	92.5	50.4	306
Smokers	57.0	26.3	17

Table 2: Smoking and Vitamin C study

(b) What test can be performed to compare the mean diet record vitamin C intake between the two groups? Perform the test and report a two-tailed p-value.