PROBLEMS

C.1 Let Y_1, Y_2, Y_3, and Y_4 be independent, identically distributed random variables from a population with mean μ and variance σ^2. Let $\bar{Y} = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$ denote the average of these four random variables.

(i) What are the expected value and variance of $\bar{Y}$ in terms of μ and σ^2?

(ii) Now, consider a different estimator of μ:

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4.$$

This is an example of a *weighted* average of the Y_i. Show that W is also an unbiased estimator of μ. Find the variance of W.

(iii) Based on your answers to parts (i) and (ii), which estimator of μ do you prefer, $\bar{Y}$ or W?

(iv) Now, consider a more general estimator of μ, defined by

$$W_a = a_1Y_1 + a_2Y_2 + a_3Y_3 + a_4Y_4,$$

where the a_i are constants. What condition is needed on the a_i for W_a to be an unbiased estimator of μ?

(v) Compute the variance of the estimator W_a from part (iv).

C.2 This is a more general version of Problem C.1. Let $Y_1, Y_2, \ldots, Y_n$ be n pairwise uncorrelated random variables with common mean μ and common variance σ^2. Let $\bar{Y}$ denote the sample average.

(i) Define the class of *linear estimators* of μ by

$$W_a = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n,$$

where the a_i are constants. What restriction on the a_i is needed for W_a to be an unbiased estimator of μ?

(ii) Find $\text{Var}(W_a)$.

(iii) For any numbers $a_1, a_2, \ldots, a_n$, the following inequality holds: $(a_1 + a_2 + \ldots + a_n)^2/n \leq a_1^2 + a_2^2 + \ldots + a_n^2$. Use this, along with parts (i) and (ii), to show that $\text{Var}(W_a) \geq \text{Var}(\bar{Y})$ whenever W_a is unbiased, so that $\bar{Y}$ is the *best linear unbiased estimator*. [*Hint*: What does the inequality become when the a_i satisfy the restriction from part (i)?]

C.3 Let Y denote the sample average from a random sample with mean μ and variance σ^2. Consider two alternative estimators of μ: $W_1 = [(n - 1)/n]\bar{Y}$ and $W_2 = \bar{Y}/2$.

(i) Show that W_1 and W_2 are both biased estimators of μ and find the biases. What happens to the biases as $n \rightarrow \infty$? Comment on any important differences in bias for the two estimators as the sample size gets large.

(ii) Find the probability limits of W_1 and W_2. {*Hint*: Use properties PLIM.1 and PLIM.2; for W_1, note that plim $[(n - 1)/n] = 1$.} Which estimator is consistent?

(iii) Find $\text{Var}(W_1)$ and $\text{Var}(W_2)$.

(iv) Argue that W_1 is a better estimator than $\bar{Y}$ if μ is "close" to zero. (Consider both bias and variance.)

C.4 For positive random variables X and Y, suppose the expected value of Y given X is $E(Y|X) = \theta X$. The unknown parameter θ shows how the expected value of Y changes with X.

(i) Define the random variable $Z = Y/X$. Show that $E(Z) = \theta$. [*Hint*: Use Property CE.2 along with the law of iterated expectations, Property CE.4. In particular, first show that $E(Z|X) = \theta$ and then use CE.4.]

(ii) Use part (i) to prove that the estimator $W = n^{-1}\sum_{i=1}^{n}(Y_i/X_i)$ is unbiased for W, where $\{(X_i,Y_i): i = 1,2,\ldots,n\}$ is a random sample.

(iii) The following table contains data on corn yields for several counties in Iowa. The USDA predicts the number of hectares of corn in each county based on satellite photos. Researchers count the number of "pixels" of corn in the satellite picture (as opposed to, for example, the number of pixels of soybeans or of uncultivated land) and use these to predict the actual number of hectares. To develop a prediction equation to be used for counties in general, the USDA surveyed farmers in selected counties to obtain corn yields in hectares. Let Y_i = corn yield in county i and let X_i = number of corn pixels in the satellite picture for county i. There are $n = 17$ observations for eight counties. Use this sample to compute the estimate of θ devised in part (ii).

Plot	Corn Yield	Corn Pixels
1	165.76	374
2	96.32	209
3	76.08	253
4	185.35	432
5	116.43	367
6	162.08	361
7	152.04	288
8	161.75	369
9	92.88	206

continued

Plot	Corn Yield	Corn Pixels
10	149.94	316
11	64.75	145
12	127.07	355
13	133.55	295
14	77.70	223
15	206.39	459
16	108.33	290
17	118.17	307

C.5 Let Y denote a Bernoulli(θ) random variable with $0 < \theta < 1$. Suppose we are interested in estimating the *odds ratio*, $\gamma = \theta/(1 - \theta)$, which is the probability of success over the probability of failure. Given a random sample $\{Y_1, \ldots, Y_n\}$, we know that an unbiased and consistent estimator of θ is $\bar{Y}$, the proportion of successes in n trials. A natural estimator of γ is $G = \{\bar{Y}/(1 - \bar{Y})\}$, the proportion of successes over the proportion of failures in the sample.

(i) Why is G not an unbiased estimator of γ?

(ii) Use PLIM.2(iii) to show that G is a consistent estimator of γ.

C.6 You are hired by the governor to study whether a tax on liquor has decreased average liquor consumption in your state. You are able to obtain, for a sample of individuals selected at random, the difference in liquor consumption (in ounces) for the years before and after the tax. For person i who is sampled randomly from the population, Y_i denotes the change in liquor consumption. Treat these as a random sample from a Normal(μ,σ^2) distribution.

(i) The null hypothesis is that there was no change in average liquor consumption. State this formally in terms of μ.

(ii) The alternative is that there was a decline in liquor consumption; state the alternative in terms of μ.

(iii) Now, suppose your sample size is $n = 900$ and you obtain the estimates $\bar{y} = -32.8$ and $s = 466.4$. Calculate the t statistic for testing H_0 against H_1; obtain the p-value for the test. (Because of the large sample size, just use the standard normal distribution tabulated in Table G.1.) Do you reject H_0 at the 5% level? at the 1% level?

(iv) Would you say that the estimated fall in consumption is large in magnitude? Comment on the practical versus statistical significance of this estimate.