Practice Problems for Midterm

- Clearly show all of your work, including formulas used in calculations.
- If you ignore the finite population correction factor, explain why ignoring it is justified.
- For approximate confidence intervals, use $z_{.005}=2.576$, $z_{.025}=1.962$, and $z_{.05}=1.645$.
- Keep full digits during calculation, round your answers, record the units for your estimates.

1. A survey is conducted to find the average weight of cows in a region. A list of all farms is available for the region, and 80 farms are selected at random. Then the weight of each cow at the 80 selected farms is recorded.
   Describe the following terms.

   a) Target population

   b) observation unit/Element

   c) Sampling frame

   d) Sampling unit

   e) Sampled population

   f) Characteristic of interest

   g) possible sources of selection bias or inaccuracy of responses

2. A simple random sample of 8 households was selected from a total of 88 households in a (very) small town. Each household was asked how many people lived in the household. The sample results are:

Sample Household	Number of Household Members
1	2
2	6
3	1
4	3
5	2
6	4
7	3
8	3
Total	24

   (a) Estimate the mean number of people per household in this town.
   (b) Estimate the variance of the estimate in (a).
   (c) Construct an approximate 90% confidence interval for the mean number of persons per household in the town.
   (d) Provide an interpretation of the confidence interval calculated in (c).

3. Consider the same simple random sample of 8 households selected from a total of 88 households in a small town. Each household was also asked how many pets the household has. The objective of this question is to estimate the proportion of households in the town with pets.
   (a) A proportion is a special case of a mean. Define a response variable that has a population mean equal to the proportion of households in the town that have at least one pet.
   (b) The sample data for the question about number of pets are:

Sample Household	Number of pets
1	0
2	4
3	0
4	1
5	0
6	1
7	1
8	0
Total	7

   Estimate the proportion of households with at least one pet.
   (c) Estimate the standard error of the estimate in (b).

4. A manufacturer wants to estimate 2003 sales for its 20 product categories. The company selects a SRS of 6 categories from a list of the 20 product categories. Divisions responsible for each of the selected categories were asked to provide early sales figures for 2003. Last year's 2002 sales figures are available for all 20 product categories, and the total 2002 sales is $674 billion. The data for 2002 and 2003 sales for each of the selected product categories are provided below, in billions of dollars.

	Sales in Billions of Dollars	
Product Category	2002	2003
Paper towels and toilet paper	21	26
Diapers	63	91
Laundry soaps	35	47
Household cleaning products	60	70
Baked foods	16	17
Snack foods	50	76

(a) Define the variable of interest y_i and the auxiliary variable x_i for this problem. Be specific.

(b) Suppose you want to use regression estimation. What is the assumed model for a regression estimator? Define all notations.

(c) What is the population mean for x, for this problem?

(d) Use the above SAS PROC CORR output (below) to obtain a regression estimate of the total 2003 sales for all product categories.

The CORR Procedure

2 Variables: SALE03 SALE02

Covariance Matrix, DF = 5

	SALE03	SALE02
SALE03	861.9000000	568.7000000
SALE02	568.7000000	397.3666667

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
SALE03	6	54.50000	29.35813	327.00000	17.00000	91.00000
SALE02	6	40.83333	19.93406	245.00000	16.00000	63.00000

Pearson Correlation Coefficients, N = 6
Prob > |r| under H0: Rho=0

	SALE03	SALE02
SALE03	1.00000	0.97176 0.0012
SALE02	0.97176 0.0012	1.00000

(e) Estimate the standard error of the estimated 2003 sales in (d). Assume $s_e^2 = 60$.

5. You want to estimate a population mean with a 95% margin of error less than or equal to 10. Previous research suggests that the population variance is about 10,000. Assume that the population size is very large and that you will use a SRS.
   (a) What size sample should you use?
   (b) If the population size is 1000, how large a sample should you use?
6. An Omaha suburb consists of 2800 adult women and 3600 adult men, which are listed individually in a database. Researchers want to estimate the number of volunteer activities adult residents in this suburb are currently involved in. Independent samples of size 100 are selected from each gender, using SRS within the gender. Each adult is asked how many volunteer activities she or he is currently involved in. The summary data are in

the following table.

	Sample mean for the number of volunteer activities per person	Estimated variance of the sample mean
Women	2.0	0.04
Men	1.0	0.02

(a) Estimate the total number of volunteer activities that adults in the community are currently involved in.
(b) Estimate the standard error for your estimate in (a).