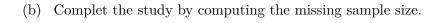
Term Test 1: Practice Problems

- **Problem 1.** We consider a population of N units with bernoulli random variables $\{0,1\}$ data values. Suppose we choose a simple random sample without replacement of n units from this population. Let $p = \frac{1}{N} \sum_{i=1}^{N} y_i$ be the population proportion, and $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i p)^2$ denotes the population variance. Let $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i \hat{p})^2$ be the sample proportion and the sample variance, respectively.
 - (a) Is \hat{p} unbiased estimator fo p? Justify.
 - (b) Show that σ^2 can be writtn as $\sigma^2 = p(1-p)$.
 - (c) Show that s^2 can be writen as $s^2 = \frac{n}{n-1}\widehat{p}(1-\widehat{p})$.
 - (d) Is the sample variance s^2 unbiased estimator for σ^2 ? Find $E(s^2)$
 - (e) Compute $V(\widehat{p})$. Given the result in (d), find an unbiased estimator for $V(\widehat{p})$ in term of \widehat{p} .

Problem 2.	To estimate the proportion of voters in favor of a controversial proposition, a simple random sample of XXXXX eligible voters was contacted and questioned. Of these, 552 reported that they favored the proposition. The study also reports a margin of error \pm 3%, 19 out of 20. The number of eligible voters in the population is approximately 1,800,000.			
	(a) What "a margin of error \pm 3%, 19 out of 20" means ?.			



(c) Using (b), estimate the population proportion in favor.

(d) Give a 95% confidence interval for population proportion.

Problem 3.	Consider the following data from a simple random sample of size $n=4$ from a population of size $N=250$, in which y is the variable of interest and x is an auxiliary variable. The population mean of the x's is 3.9.			
	(a)	_	ggest two types of estimators for estimating the mean of y. Summarize some perties of each estimator.	
	(b)	resu	e data obtained were recorded in variables y and x, analyzed in R and produced alts presented below (see next page). Based on the R output, answer the following estions.	
		(i)	Estimate mean of y using the simple estimator, and estimate the variance of the estimator.	
		(ii)	Estimate mean of y using the ratio estimator, and estimate the variance of the estimator.	
		(iii)	Estimate mean of y using the regression estimator, and estimate the variance of the estimator.	
		(iv)	Estimate mean of y using the difference estimator, and estimate the variance of the estimator.	
		(v)	Based on the data, which estimator appears preferable in this situation?	

R output:

```
> # Population
> N<-250
> mu_x<-3.9
> # SRS of size n=4
> n<-4
> y<-c(150, 100, 200, 140)
> x<-c(4,2,4,3)
> ysum<-sum(y); ysum</pre>
[1] 590
> xsum<-sum(x); xsum
[1] 13
> s2_y<-var(y); s2_y
[1] 1691.667
> r<-(ysum/xsum); r</pre>
[1] 45.38462
> (ysum/xsum)*mu_x
[1] 177
> s2_r<-var(y-r*x); s2_r
[1] 478.501
> (1 - n/N)*(s2_y/n)
[1] 416.15
> (1 - n/N)*(s2_r/n)
[1] 117.7112
> (1 - n/N)*(1/mu_x)^2*(s2_r/n)
[1] 7.739069
> (1 - n/N)*(1/mu_x)*(s2_r/n)
[1] 30.18237
> cor(x,y)
[1] 0.8676399
> # regression of y on x
> fitreg<-lm(y~x)</pre>
> coef(fitreg)
(Intercept)
   26.36364
               37.27273
> yhat<-fitted(fitreg)</pre>
> ehat<-y-yhat
> mean(y) + coef(fitreg)[2]*(mu_x-mean(x))
171.7273
> MSE<-sum( ehat^2 )/(n-2)
> (1 - n/N)*(MSE/n)
[1] 154.3091
> (1 - n/N)*(MSE/(n-1))
[1] 205.7455
> (1 - n/N)*(MSE/(n-2))
[1] 308.6182
```