Multiple Choice Questions [20 pts]

Please provide your answers to this section below:

1.  
2.  
3.  
4.  
5.  
6.  
7.  
8.  
9.  
10.  

1) An estimator is
   a. an estimate.
   b. a formula that gives an efficient guess of the true population value.
   c. a random variable.
   d. a nonrandom number.

2) An estimate is
   a. efficient if it has the smallest variance possible.
   b. a nonrandom number.
   c. unbiased if its expected value equals the population value.
   d. another word for estimator.

3) An estimator \( \hat{\mu}_y \) of the population value \( \mu_y \) is more efficient when compared to another estimator \( \tilde{\mu}_y \), if
   a. \( E(\hat{\mu}_y) > E(\tilde{\mu}_y) \).
   b. it has a smaller variance.
   c. its c.d.f. is flatter than that of the other estimator.
   d. both estimators are unbiased, and \( \text{var}(\hat{\mu}_y) < \text{var}(\tilde{\mu}_y) \).
4) Among all unbiased estimators that are weighted averages of $Y_1, \ldots, Y_n$, $\bar{Y}$ is

a. the only consistent estimator of $\mu_Y$.
b. the most efficient estimator of $\mu_Y$.
c. a number which, by definition, cannot have a variance.
d. the most unbiased estimator of $\mu_Y$.

5) The $p$-value is defined as follows:

a. $p = 0.05$.
b. $\Pr_{H_0}(|\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}|)$.
c. $\Pr(z > 1.96)$.
d. $\Pr_{H_a}(|\bar{Y} - \mu_{Y,0}| < |\bar{Y}^{act} - \mu_{Y,0}|)$.

6) A large $p$-value implies

a. rejection of the null hypothesis.
b. a large $t$-statistic.
c. a large $\bar{Y}^{act}$.
d. that the observed value $\bar{Y}^{act}$ is consistent with the null hypothesis.

7) The power of the test

a. is the probability that the test actually incorrectly rejects the null hypothesis when the null is true.
b. depends on whether you use $\bar{Y}$ or $\bar{Y}^2$ for the $t$-statistic.
c. is one minus the size of the test.
d. is the probability that the test correctly rejects the null when the alternative is true.
8) The sample covariance can be calculated in any of the following ways, with the exception of:

a. \[
\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}).
\]

b. \[
\frac{1}{n-1} \sum_{i=1}^{n} X_i Y_i - \frac{n}{n-1} \bar{X} \bar{Y}.
\]

c. \[
\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_X)(Y_i - \mu_Y).
\]

d. \(r_{XY}s_X s_Y\), where \(r_{XY}\) is the correlation coefficient.

9) When the sample size \(n\) is large, the 90% confidence interval for \(\mu_Y\) is

a. \(\bar{Y} \pm 1.96SE(\bar{Y})\).

b. \(\bar{Y} \pm 1.64SE(\bar{Y})\).

c. \(\bar{Y} \pm 1.64\sigma_Y\).

d. \(\bar{Y} \pm 1.96\).

10) The following statement about the sample correlation coefficient is true.

a. \(-1 \leq r_{XY} \leq 1\).

b. \(r_{XY}^2 \xrightarrow{p} corr(X_i, Y_i)\).

c. \(|r_{XY}| < 1\).

d. \(r_{XY} = \frac{s_{XY}^2}{s_X^2 s_Y^2}\).
Problems [40 pts]

Instructions: The goal of the problem set is to understand what you are doing rather than just getting the correct result. Please show your work clearly and neatly. Please write your answers in the space provided.

1) Adult males are taller, on average, than adult females. Visiting two recent American Youth Soccer Organization (AYSO) under-12-year-old (U12) soccer matches on a Saturday, you do not observe an obvious difference in the height of boys and girls of that age. You suggest to your little sister that she collect data on height and gender of children in 4th to 6th grade as part of her science project. The accompanying table shows her findings.

### Height of Young Boys and Girls, Grades 4-6, in inches

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th></th>
<th></th>
<th>Girls</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{Y}_{Boys}$</td>
<td>$s_{Boys}$</td>
<td>$n_{Boys}$</td>
<td>$\bar{Y}_{Girls}$</td>
<td>$s_{Girls}$</td>
<td>$n_{Girls}$</td>
</tr>
<tr>
<td></td>
<td>57.8</td>
<td>3.9</td>
<td>55</td>
<td>58.4</td>
<td>4.2</td>
<td>57</td>
</tr>
</tbody>
</table>

(a) Let your null hypothesis be that there is no difference in the height of females and males at this age level. Specify the alternative hypothesis.

(b) Find the difference in height and the standard error of the difference.

(c) Generate a 95% confidence interval for the difference in height.

(d) Calculate the t-statistic for comparing the two means. Is the difference statistically significant at the 1% level? Which critical value did you use? Why would this number be smaller if you had assumed a one-sided alternative hypothesis? What is the intuition behind this?
2) The development office and the registrar have provided you with anonymous matches of starting salaries and GPAs for 108 graduating economics majors. Your sample contains a variety of jobs, from church pastor to stockbroker.

(a) The average starting salary for the 108 students was $38,644.86 with a standard deviation of $7,541.40. Construct a 95% confidence interval for the starting salary of all economics majors at your university/college.

(b) A similar sample for psychology majors indicates a significantly lower starting salary. Given that these students had the same number of years of education, does this indicate discrimination in the job market against psychology majors?

(c) You wonder if it pays (no pun intended) to get good grades by calculating the average salary for economics majors who graduated with a cumulative GPA of B+ or better, and those who had a B or worse. The data is as shown in the accompanying table:

<table>
<thead>
<tr>
<th>Cumulative GPA</th>
<th>Average Earnings $Y$</th>
<th>Standard Deviation $s_y$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B+ or better</td>
<td>$39,915.25$</td>
<td>$8,330.21$</td>
<td>59</td>
</tr>
<tr>
<td>B or worse</td>
<td>$37,083.33$</td>
<td>$6,174.86$</td>
<td>49</td>
</tr>
</tbody>
</table>

Conduct a $t$-test for the hypothesis that the two starting salaries are the same in the population. Given that this data was collected in 1999, do you think that your results will hold for other years, such as 2002?
3) Assume that under the null hypothesis, $\bar{Y}$ has an expected value of 500 and a standard deviation of 20. Under the alternative hypothesis, the expected value is 550. Sketch the probability density function for the null and the alternative hypothesis in the same figure. Pick a critical value such that the p-value is approximately 5%. Mark the areas, which show the size and the power of the test. What happens to the power of the test if the alternative hypothesis moves closer to the null hypothesis, i.e., $\mu = 540, 530, 520$, etc.?

4) Consider the following alternative estimator for the population mean:

$$\tilde{Y} = \frac{1}{n} (\frac{1}{4} Y_1 + \frac{7}{4} Y_2 + \frac{1}{4} Y_3 + \frac{7}{4} Y_4 + \ldots + \frac{1}{4} Y_{n-1} + \frac{7}{4} Y_n)$$

Prove that $\tilde{Y}$ is unbiased and consistent, but not efficient when compared to $\bar{Y}$. 
**EViews Exercise [40 pts]**

The data for this exercise is contained in the excel file “salary.XLS.” It contains 93 observations on employees for a Chicago Bank during the period 1969-1971. In particular, the variable “salary” refers to the starting salary, “education” refers to years of education, “experience” refers to the number of months of previous work experience, “seniority” refers to the number of months the person has been working at the current job with the bank, and “gender” takes the value of one for males and 0 for females.

You will need to import the data into EViews. This can be easily accomplished by opening a new workfile. Select the option corresponding to undated or irregular observations and a range that goes from 1 to 93. Then from the “procs” button, select the import option, and the “excel” option thereafter. The rest is pretty straightforward.

Answer the following questions:

1. Give summary statistics for all variables in one table. Then, subsample according to gender and report the summary statistics with an additional two tables. Next, select all the variables (make sure to undo the subsampling condition) and compute their correlations and include this table with the previous tables in the same page. Comment on all of these results, making particular emphasis on whether you see any evidence of discrimination (nothing formal yet, just hunches). Beware of the experience and seniority levels of each group.

2. Do a formal test of the null hypothesis that females are discriminated against at a 5% confidence level. Do your results hold at a 1% confidence level? Consider now a test of the null hypothesis that males earn $250 more than females on average at a 5% confidence level. Make sure to report these tests in one page and in one paragraph report on the economic significance of your tests (Is there gender bias in pay scales?). *Hints: make sure you consider whether a one-tailed test or a two-tailed test is appropriate.*

3. Load data on the exchange rate between the U.S. dollar and the Japanese yen into EViews. Select monthly frequency and select a data source that begins at least in 1980. You will need to plot these data and compute summary statistics. Regarding the plot, make sure that you label important economic dates (such as oil crises, or other political events) that may help explain some of the long swings in these data. Make sure to label the graph appropriately, including title, units of