Last Name:

First Name:

Student #:

Instructor: Prof. Murdock

Duration: 110 minutes

Allowed Aids: A non-programmable calculator; Aid sheets are given to you with this test

Format: This test has 6 questions worth 100 points in total. Write your answers on these test papers.

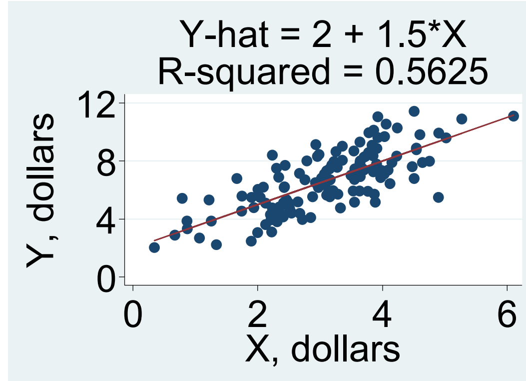
Instructions:

- Write your answers clearly, concisely, and completely below each question. You do not have to fill all of the blank space: a generous amount is provided for your convenience.
- For each question I give a guide for your response in brackets. It indicates what is expected: a quantitative analysis, a graph, and/or a written response. For example, "What would the least squares intercept be and how should it be interpreted? [Analysis & 1 – 2 sentences]"
    - Please take the guides seriously: they are there to help you write the best answer.
    - It is a bad idea to write a long paragraph or a couple of words if the guide asks for 1 – 2 sentences. **Make sure to write actual sentences (not short-hand or bullet lists)**.
    - If the guide says analysis is expected **make sure to show your work and reasoning**.
    - If the guide says a graph is expected **make sure your graphs are fully labeled**.
- To best demonstrate your understanding, focus on directly answering the questions asked.
    - Apply your skills to the specific situation presented with the question.
    - Reproducing examples or discussion from class or our course materials is not an effective strategy for earning points because these will not directly address the specific situation presented with the question.
    - Extraneous analysis does not earn positive marks if it is correct and earns negative marks if incorrect, which is another reason to focus on the question that is asked.
- For questions with multiple parts, attempt each part even if you had trouble with earlier parts.
- Please manage your time. If a question is worth 10 points spend roughly 10 minutes on it and if it is worth 24 points then spend roughly 24 minutes on it.

Question:	**1**	**2**	**3**	**4**	**5**	**6**	**Raw Total**	**Mark**
Point Value:	*12*	*15*	*9*	*16*	*24*	*24*	*100*	
Points Earned:								

(1) [12 pts] The NPR story "Why Seeing (The Unexpected) Is Often Not Believing" has interviews with researchers that estimate that only one-third of subjects notice an obvious fight when focused on something else. Suppose this result is based on 42 subjects. Find an interval estimate and clearly and precisely interpret it. Mention the margin of error. [analysis & 2 – 3 sentences]

(2) [15 pts] For two variables, each measured in dollars, consider this scatter diagram, least squares line, and R^2. For this specific case explain whether the following statements are true or false.



(a) [5 pts] The standard deviation of the residuals (i.e. the standard error of estimate) is bigger than zero and the assumption of homoscedasticity is met. [T/F, 1 – 2 sentences]

(b) [5 pts] A one standard deviation increase in X is associated with three-fourths of a standard deviation increase in Y. [T/F, 1 – 2 sentences]

(c) [5 pts] Because the slope is steeper than one (i.e. steeper than a 45 degree line) there is no regression to the mean. [T/F, 2 – 3 sentences]

(3) [9 pts] Complete this letter about our textbook with <u>quantitative</u> evidence. [analysis & 1 sentence]

Dear Sharpe et al,

When discussing stratified sampling on page 36 in ***Business Statistics***, 1st Canadian ed. you state:

> Suppose we want to survey how shoppers feel about a potential new anchor store at a large suburban mall. The shopper population is 60% women and 40% men, and we suspect that men and women have different views on their choice of anchor stores. If we use simple random sampling to select 100 people for the survey, we could end up with 70 men and 30 women.

In this scenario it is unreasonable to worry about a sample with so few women (30) and so many men (70). [<u>You fill in here</u>]. It may be a typo as I would not have the same concern if it said 70 women and 30 men.

Sincerely, Your Name

(4) [16 pts] The Globe and Mail article “Peak coffee: A cup of trouble” discusses trends in the coffee market. Some excerpts:

> “Coffee prices - which went through a prolonged slump between 2000 and 2006 - have doubled since the middle of last year, with Arabica topping a 34-year high of $3.08 (U.S.) a pound earlier this month.”
>
> “Changing weather patterns have wreaked havoc on coffee supply, particularly the Arabica strain, which is grown in the Americas and Africa and which makes the best coffee. Brazil and Colombia are the top two producers of Arabica, but experts say the crops are not keeping up with skyrocketing demand in emerging markets like China, India and South America, as well as among consumers in Europe and North America.”

(a) [9 pts] Explain why each of the three incorrect choices is an incorrect answer to this multiple-choice question. [3 – 5 sentences]

The coffee industry has recently been in the news. Globally both prices and coffee consumption are increasing. Data from this industry are observational and this confounds easy estimation of the demand relationship. If we imagine experimental data to estimate demand what would it involve?

(A) randomly varying prices
(B) holding demand shifters and supply shifters fixed so there is no variation
(C) setting competition in the industry to either perfect competition or monopoly
(D) making sure that nothing other than price affects each person’s coffee consumption choice

(b) [7 pts] Do the Globe and Mail excerpts suggest that the elasticity of demand for Arabica coffee is positive? Explain. Incorporate course concepts and terms into your explanation and to apply them to this particular case: for example, is the price of Arabica coffee exogenous? [3 – 4 sentences]

(5) [24 pts] Suppose that of all courses in North American universities 20% have no required textbook, 50% have one, 10% have two, and 20% have three required textbooks. You randomly sample 76 courses.

(a) [8 pts] Graph the population distribution. Include the expected value and standard deviation. [analysis & 1 graph]

(b) [10 pts] Graph the sampling distribution of the sample mean. Fully label it. Include the expected value, standard deviation and sample size. Show your work and reasoning. [analysis & 1 graph]

(c) [6 pts] What is the chance of randomly selecting one course with more than one required textbook? What is the chance of randomly selecting 76 courses that on average have more than one required textbook? [analysis]

(6) [24 pts] For all parts, consider a Uniform population from 0 to 1. For parts (a), (b), and (d) also consider these Monte Carlo simulation results related to the sampling distributions of two statistics.

Table 6.1: Simulation results (100,000 simulation draws) with sampling from a Uniform population (X ~ U[0,1])

		Mean	Median	1st Percentile	Std. Dev.
(1)	**Sample Mean, n = 10**	0.4996960	0.4996450	0.2902270	0.0911602
(2)	**Sample Mean, n = 30**	0.4998154	0.4998790	0.3785691	0.0526348
(3)	**Sample Mean, n = 100**	0.4999106	0.5000367	0.4327796	0.0289642
(4)	**Sample Median, n = 10**	0.4992013	0.4988463	0.1943745	0.1376757
(5)	**Sample Median, n = 30**	0.4996024	0.5000049	0.3011423	0.0869820
(6)	**Sample Median, n = 100**	0.4998567	0.4999025	0.3856713	0.0494345

(a) [6 pts] For a sample size of 10 what is the exact expected value and standard deviation for the sample mean? Why aren't the results in row (1) exactly equal to these? [analysis, 1 sentence]

(b) [6 pts] Do these simulation results suggest that the sample median is an unbiased, consistent, and relatively efficient estimator of μ? Incorporate relevant simulation results. [3 – 4 sentences]

► For parts (c) and (d) consider a random sample with sample size 100,000 randomly drawn from the population. Here is a STATA summary of this sample.

```
                                x
      -------------------------------------------------------------
            Percentiles      Smallest
       1%     .0094768       8.08e-06
       5%     .0498024        .000015
      10%     .1002335        .000018       Obs              100000
      25%     .2502917        .000019       Sum of Wgt.      100000

      50%     .5000015                      Mean            .4997975
                              Largest       Std. Dev.       .2884576
      75%     .7494019        .999961
      90%     .8996469       .9999702       Variance        .0832078
      95%     .9489594        .999977       Skewness       -.0001539
      99%     .9903972       .9999818       Kurtosis        1.803577
```

(c) [6 pts] Does this STATA summary illustrate the Law of Large Numbers? Explain. Include specific references to the STATA summary in your explanation. [2 - 3 sentences]

(d) [6 pts] How does this STATA summary relate to the Monte Carlo simulation in Table 6.1? Why is the 1st percentile in this STATA summary smaller than those in Table 6.1? [2 - 3 sentences]

EXTRA SPACE: If you use this, clearly indicate the question number and part and make a note in the original space directing the grader here.

Population Mean: $$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Sample Mean: $$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Population Variance: $$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Sample Variance: $$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1} = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}]$$

Population s.d.: $\sigma = \sqrt{\sigma^2}$

Sample s.d.: $s = \sqrt{s^2}$

Population coefficient of variation: $CV = \frac{\sigma}{\mu}$

Sample coefficient of variation: $cv = \frac{s}{\overline{X}}$

Population covariance: $$\sigma_{XY} = \frac{\sum_{i=1}^{N} (x_i - \mu_X)(y_i - \mu_Y)}{N}$$

Sample covariance: $$s_{XY} = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{n-1}$$

Population coefficient of correlation: $$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Sample covariance (shortcut): $$s_{XY} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

Sample coefficient of correlation: $$r = \frac{s_{XY}}{s_X s_Y} = \frac{\sum_{i=1}^{n} z_{Xi} z_{Yi}}{n-1}$$

Least squares line / linear regression line: $\hat{y} = a + bx$ $\quad b = \frac{\operatorname{cov}(x, y)}{s_x^2} = r\frac{s_Y}{s_X}$ $\quad a = \overline{y} - b\overline{x}$

Standard deviation of the residuals: $$s_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{SSE}{n-2}}$$

Coefficient of Determination: $$R^2 = \frac{[\operatorname{cov}(x, y)]^2}{s_x^2 s_y^2} = (r)^2$$

$$SST = \sum_{i=1}^{n} (y_i - \overline{y})^2 \quad SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 \quad SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \quad SST = SSR + SSE$$

$$SSE = (n-1)\left(s_y^2 - \frac{(s_{xy})^2}{s_x^2}\right)$$

Coefficient of Determination: $$R^2 = \frac{SSR}{SST} \quad R^2 = 1 - \frac{SSE}{SST} \quad R^2 = 1 - \frac{SSE}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

Conditional Probability: $P(A|B) = \dfrac{P(A\ and\ B)}{P(B)}$

Complement Rules: $P(A^C) = 1 - P(A)$

$P(A^C | B) = 1 - P(A | B)$

Multiplication Rule: $P(A \text{ and } B) = P(A|B)P(B)$

Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Expected Value: $E[X] = \mu = \sum_{all\ x} xp(x)$

Variance: $V[X] = E[(X-\mu)^2] = \sigma^2 = \sum_{all\ x} (x-\mu)^2 p(x)$

Covariance: $E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = \sum_{all\ x} \sum_{all\ y} (x-\mu_X)(y-\mu_Y) p(x,y)$

Laws of Expected Value:
E[c] = c
E[X+c] = E[X] + c
E[cX] = cE[X]
E[a+bX+cY] = a + bE[X] + cE[Y]

Laws of Variance:
V[c] = 0
V[X+c] = V[X]
V[cX] = c^2V[X]
V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bcCOV[X,Y]
V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc*s_x*s_y*r

Laws of Covariance:
COV[X, c] = 0
COV[a+bX, c+dY] = bdCOV[X,Y]

Combinatorial formula: $C_x^n = \dfrac{n!}{x!(n-x)!}$

Binomial probability: $p(x) = \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

$for\ x = 0, 1, 2, \ldots, \text{n}$

Mean of Binomial: np **Variance of Binomial:** $np(1-p)$

Uniform density function: $f(x) = \dfrac{1}{b-a} \quad where\ \ a \le x \le b$

Mean of Uniform: $\dfrac{a+b}{2}$

Variance of Uniform: $\dfrac{(b-a)^2}{12}$

Sample Mean:

$$\mu_{\overline{X}} = E[\overline{X}] = \mu_X$$

$$\sigma_{\overline{X}}^2 = V[\overline{X}] = \frac{\sigma_X^2}{n}$$

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

Sample Proportion:

$$E[\hat{P}] = p$$

$$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}}$$

Difference Between Two Sample Means:

$$\mu_{\overline{X}_1 - \overline{X}_2} = E[\overline{X}_1 - \overline{X}_2] = \mu_1 - \mu_2$$

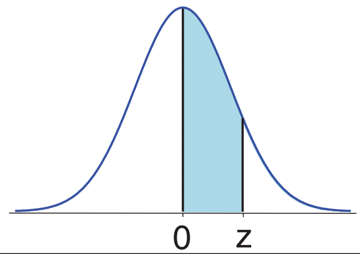
$$\sigma_{\overline{X}_1 - \overline{X}_2}^2 = V[\overline{X}_1 - \overline{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Inference about *p* (population proportion):

CI estimator: $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$

Sample size to estimate *p* ± *τ*: $n = \left(\dfrac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{\tau}\right)^2$



Normal Probabilities:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990