Concordia University
Department of Economics
ECON 221 – Sections A, AA, B
Instructors: B. Campbell, S. Elliston, F. Rabby
Fall 2013 – ASSIGNMENT 2 - Due Date: Friday, October 25, before 4:00 pm

---

Name: **I.D:** **Section:**

Points Total: 100 points

Question 1
(10 points) A professor collected data on the number of absences in an introductory statistics class of 30 students over the course of a semester. The data are summarized in the table below.

Number of Absences	0 – 2	3 – 5	6 – 8	9 – 11	12– 14	15 - 17
Number of Students	7	11	6	3	2	1

a) **[3 points]** Estimate the sample mean number of absences.

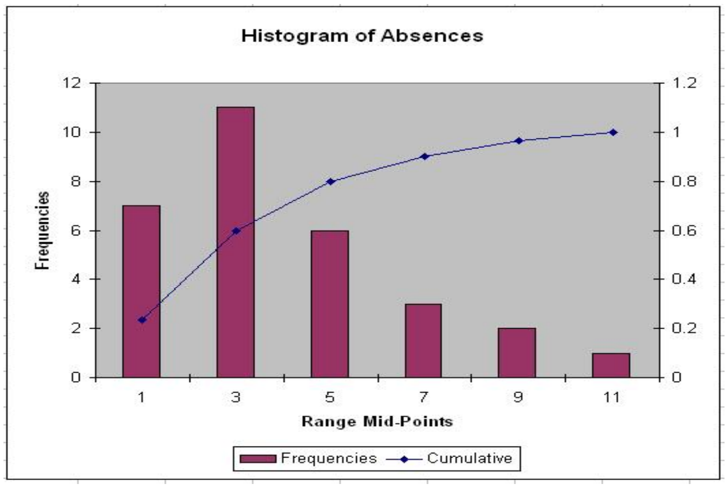
Absences X_i	Students F_i	m_i	$m_i * F_i$	$F_i * (m_i - \overline{X})^2$
0 - 2	7	1	7	63
2 - 4	11	3	33	11
4 - 6	6	5	30	6
6 - 8	3	7	21	27
8 - 10	2	9	18	50
10 - 12	1	11	11	49
Totals	**30**		**120**	**206**

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{120}{30} = 4$$

b) **[4 points]** Estimate the sample standard deviation. (4 points)

$$S^2 = \frac{\sum f_i (m_i - \overline{X})^2}{n-1} = \frac{206}{30-1} = 7.1034 \quad \text{and} \quad S = \sqrt{S^2} = \sqrt{7.1034} = 2.665$$

c) **[3 points]** Make a rough sketch of the histogram.



Question 2

a) **[20 points]** National Car Rental Systems Inc., commissioned the Canadian Automobile Association (CAA) to conduct a survey of the general condition of the cars rented to the public by Hertz, Avis, National and Budget Rent-a Car. CAA officials evaluate each company's cars using a demerit points system. Each car starts with a perfect score of 0 points and incurs demerit points for each discrepancy noted by the inspectors. One measure of the overall condition of a company's cars is the mean of all scores received by the company (i.e., the company's fleet mean score). To estimate the fleet mean score of each rental car company, 10 major airports were randomly selected and 10 cars from each company were randomly rented for inspection from each airport by CAA officials (i.e., a random sample of n = 100 cars from each company's fleet was drawn and inspected).

i. **[3 points]** Describe the sampling distribution of $\bar{x}$, the mean score of a sample of n = 100 rental cars.

Answer: By the Central Limit Theorem, the sampling distribution of $\bar{x}$ is approximately normal with $\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \sigma/\sqrt{n} = \sigma/\sqrt{100}$.

ii. **[3 points]** Interpret the mean of $\bar{x}$ in the context of this problem.

Answer: The mean of the $\bar{x}$ distribution, is the mean of the fleet mean score.

iii. **[3 points]** Assume that $\mu = 30$ and $\sigma = 60$ for one rental car company. For this company, find $P(\bar{x} \geq 45)$.
Answer: $\mu_{\bar{x}} = \mu = 30 \; and \; \sigma_{\bar{x}} = \sigma/\sqrt{n} = 60/\sqrt{100} = 6$

$$P(\bar{x} \geq 45) = P\left(z \geq \frac{45-30}{6}\right) = P(z \geq 2.5) = 1 - 0.9938 = 0.0062$$

iv. **[3 points]** If $P(\bar{x} > 30) = 0.2$, what is μ?

Answer: $P(\bar{x} > 30) = P\left(z > \frac{30-\mu}{6}\right) = 0.2.\; We\ know\ that\ P(z > 0.84) = 0.2.$
$Thus, \frac{30-\mu}{6} = 0.84 \Rightarrow \mu = 24.96$

v. **[4 points]** Refer to part **iii**. The company claims that their true fleet mean score "couldn't possibly be as high as 30." The sample mean score tabulated by the CAA for this company was 45. Does this result tend to support or refute the claim? Explain.
Answer: The sample mean of 45 tends to refute the claim. If the true fleet mean were as high as 30, then, observing a sample mean of 45 or higher would be extremely unlikely (from part **iii** the probability is only 0.0062). Thus, we would infer that the true mean is actually not 30, but something higher.

b) **[4 points]** Will the sampling distribution of $\bar{x}$ always be approximately normally distributed? Explain.
Answer: The sampling distribution is approximately normal only if the sample size is sufficiently large to allow the researcher to invoke the Central Limit Theorem or if the population being sampled from is normal.

Question 3

a) **[15 points]** According to a recent Pew Internet and American Life Project Survey (October 2010), 67% of adults who use the Internet have paid to download music. In a random sample of n = 1,000 adults who use the Internet, let $\hat{p}$ represent the proportion who have paid to download music.

i. **[4 points]** Find the mean and standard deviation of the sampling distribution of $\hat{p}$.

Answer: $E(\hat{p}) = p = 0.67$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.67(1-0.67)}{1000}} = 0.0149$$

ii. **[5 points]** What does the Central Limit Theorem say about the shape of the sampling distribution of $\hat{p}$.

Answer: By the Central Limit Theorem, the sampling distribution of $\hat{p}$ will be approximately normal since the sample size is sufficiently large.

iii. **[3 points]** What is the probability that less than 75% of adults who use the Internet have paid to download music?

Answer: $P(\hat{p} < 0.75) = P\left(z < \frac{0.75-0.67}{0.0149}\right) = P(z < 5.38) \cong 1$

iv. **[3 points]** What is the probability that more than 50% of these adults have paid to download music?

Answer: $P(\hat{p} > 0.5) = P\left(z > \frac{0.5-0.67}{0.0149}\right) = P(z > -11.43) \cong 1$

b) **[5 points]** Due to inaccuracies in drug-testing procedures (e.g., false positives and false negatives) in the medical field, the results of a drug test represent only one factor in a physician's diagnosis. Yet, when Olympic athletes are tested for illegal drug use (i.e., doping), the results of a single positive test are used to ban the athlete from competition. In *Chance* (Spring 2004), University of Texas biostatisticians D. A. Berry and L. Chastain demonstrated the application of Baye's Rule for making inferences about testosterone abuse among Olympic athletes. They used the following example: In a population of 1,000 athletes, suppose 100 are illegally using testosterone. Of the users, suppose 50 would test positive for testosterone. Of the nonusers, suppose 9 would test positive. If an athlete tests positive for testosterone, use Baye's Rule to find the probability that the athlete is really doping.

Answer: Define the following events:

U: {Athlete uses testosterone}

P: {Test is positive}

$$P(U|P) = \frac{P(U\cap P)}{P(P)} = \frac{P(P|U)P(U)}{P(P|U)P(U)+P(P|U')P(U')} = \frac{0.5(0.1)}{0.5(0.1)+0.01(0.9)} = 0.847$$

4. We have $x_1, x_2 \sim N(\mu, \sigma^2)$

Define
$$X = (x_1 + x_2)/2$$
$$Y = (x_1 + 3x_2)/4$$
$$Z = (x_1 + 2x_2)/3$$

(a) $E(X) = E[(x_1+x_2)/2] = \frac{1}{2}(E(x_1) + E(x_2)) = \frac{2\mu}{2} = \mu$

$E(Y) = E[(x_1+3x_2)/4] = \frac{1}{4}[E(x_1) + 3E(x_2)] = \frac{4\mu}{4} = \mu$

$E(Z) = E[(x_1+2x_2)/3] = \frac{1}{3}[E(x_1) + 2E(x_2)] = \frac{3\mu}{3} = \mu$

All unbiased

(b) Compute variances

$Var(X) = Var\left(\frac{x_1+x_2}{2}\right) = \frac{1}{4}[Var(x_1) + Var(x_2)] = \frac{2}{4}\sigma^2 = \frac{\sigma^2}{2}$

$Var(Y) = Var\left(\frac{x_1+3x_2}{4}\right) = \frac{1}{16}[Var(x_1) + 9\,Var(x_2)] = \frac{10}{16}\sigma^2 = \frac{5}{8}\sigma^2$

$Var(Z) = Var\left(\frac{x_1+2x_2}{3}\right) = \frac{1}{9}(Var(x_1) + 4\,Var(x_2)] = \frac{5}{9}\sigma^2$

$Var(X) < Var(Z) < Var(Y)$

(c) Relative Efficiency

Relative efficiency of X:

relative to $Y \equiv \frac{Var(Y)}{Var(X)} = \frac{5/8\,\sigma^2}{1/2\,\sigma^2} = \frac{10}{8}$

relative to $Z \equiv \frac{Var(Z)}{Var(X)} = \frac{5/9\,\sigma^2}{1/2\,\sigma^2} = \frac{10}{9}$

Question 5

[20 points] The amount of time spent (in minutes) for the completion of the 4th assignment in Statistics by a random sample of 10 students gave the following results: 215, 182, 193, 208, 210, 176, 197, 188, 218, 213.

a) **[5 points]** Calculate the sample mean and sample standard deviation.

X_1	$(X_i - \mu)^2$
215	225
182	324
193	49
208	64
210	100
176	576
197	9
188	144
218	324
213	169
2000	**1984**

$$\overline{X} = \frac{\sum X_i}{n} = \frac{2000}{10} = 200$$

$$s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1} = \frac{1984}{10-1} = 220.44 \Rightarrow s = 14.85$$

b) **[5 points]** Specify the appropriate assumptions and find a 95% confidence interval for the population mean time spent by students on the 4th Assignment.

Answer: *We must assume a normally distributed population. Thus, in the place of the unknown population variance, we use the estimated sample variance and follow the t-distribution for the limits of the confidence interval.*

For 10-1=9 degrees of freedom and a 95% confidence interval, the t-cutoff points are $\pm$ *2.262.*

$$\overline{X} - t_{v,\alpha/2} * \frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{v,\alpha/2} * \frac{s}{\sqrt{n}} \qquad or$$

$$200 - 2.262 * \frac{14.85}{\sqrt{10}} < \mu < 200 + 2.262 * \frac{14.85}{\sqrt{10}}$$

or $200 \pm 2.262*(14.85 / 3.1623)$ *or* 200 ± 10.62

or $189.38 < \mu < 210.62$

The mean time spent for the 4^{th} *Assignment by all students should be anywhere between 189.38 and 210.62 minutes with 95% probability.*

c) **[5 points]** Find a 90% confidence interval for the population mean time spent by students on the 4th Assignment.

Answer:

$$200 - 1.833 * \frac{14.85}{\sqrt{10}} < \mu < 200 + 1.833 * \frac{14.85}{\sqrt{10}}$$

or $200 \pm 1.833*(14.85 / 3.1623)$ *or* 200 ± 8.61 *or* $191.39 < \mu < 208.61$

The mean time spent for the 3rd Assignment by all students should be anywhere between 191.32 and 208.61 minutes with 90% probability.

d) **[5 points]** Compare your findings in (b) and (c).

Note that the interval in (c) is narrower than the one in (b).

Question 6

EXCEL problem**.** **(15 points)**

a) Simulate 10,000 draws from a standard normal distribution for each of 5 variables. Square each of these; you should have 5 columns, each of length 10000, with squared N(0,1's) as entries. (Use *Random Number Generator* in Data Analysis.)

b) Create a new column of length 10,000 by adding the five entries across each row and dividing this sum by 5.

c) Sort the last column from low to high. Using this sorted series, fill in the following Table as the answer to this problem, along with an EXCEL output of the first twenty rows of the six variables. Also, compute the average and variance of this last row to be included in the Table.

Simulation of the Chi-square Distribution ($X_{5,\alpha}^2$) with 5 Degrees of Freedom

$X_{5,\alpha}^2$	α = .99	α = .975	α = .95	α = .05	α = .025	α = .01
True Value						

Simulated Value						
Simulated Mean						
Simulated Variance						