PART 1

S	O	L	U	T	I	O	N	S

(1) [18 points] Suppose, the producer claims that Cheerios is a power breakfast: eating Cheerios for breakfast would boost test scores on average by 4 percentage points and decrease the variance among students by 1 percentage point. The average on the national test score is known to be distributed with the mean of 59 and standard deviation of 21. To find out whether a Cheerios is a power breakfast, a study is done in a large elementary statistics class; 499 students agree to participate; 250 are randomized to the treatment group, and 249 to the control group. The treatment group is fed Cheerios for breakfast 7 days a week. The control group gets Captain Crunch.[1]

(a) [12 points] If the producer's claim is true, what is the chance that the control group outperforms treatment group by at least 1 percentage point? [Analysis&1-2 sentences]

Solution:

We need to find $P((\overline{X}_1 - \overline{X}_2) \le -1) = ?$

$$\mu_1 = 63 \qquad \mu_2 = 59$$

$$\sigma_1 = \sqrt{21^2 - 1} = 21 \qquad \sigma_2 = 21$$

$$\sigma_{\overline{X}_1} = \frac{21}{\sqrt{250}} = 1.331 \qquad \sigma_{\overline{X}_2} = \frac{21}{\sqrt{249}} = 1.331$$

Because both sample sizes are bigger than 30 (249>30), we can apply the CLT to each sample and conclude that each sample mean is Bell-shaped and therefore the difference between two sample means is also Bell-shaped.

$$P((\overline{X}_1 - \overline{X}_2) < -1) = P(Z < \frac{-1-(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_{X_1}}{n_1} + \frac{\sigma^2_{X_2}}{n_2}}}) = P(Z < \frac{-1-4}{\sqrt{\frac{21^2-1}{250} + \frac{21^2}{249}}}) = P(Z < -2.66)$$

$$P(Z < -2.66) = 0.0039$$

[1] Cheerios and Captain Crunch are registered trademarks. The study is hypothetical.

(b) [4 points] Explain whether the sampling error would be a plausible explanation for such a disappointing result of the study? [1 – 2 sentences]

Solution:
There is only 0.4 percent chance that the treatment group would perform worse than the control. Sampling error (or pure chance) is not a plausible explanation for this disappointing result.

(c) [2 points] What type of study is this? What is an exogenous variable in this study? Explain [2 – 4 sentences]

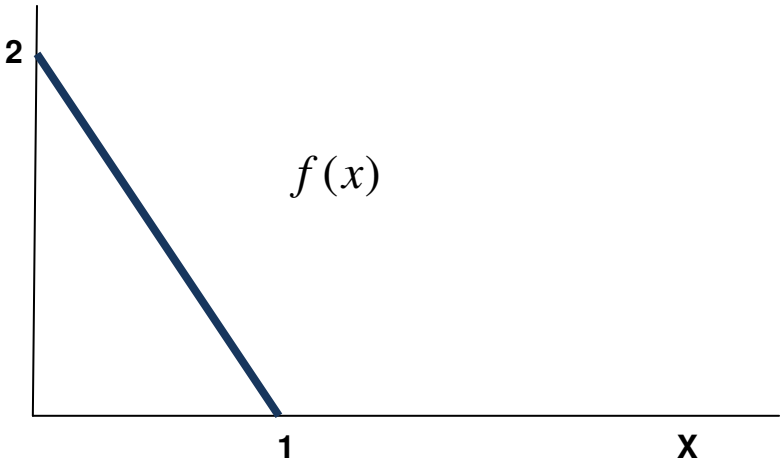
Solution:
This is an experimental study: students were randomly selected into two groups. Exogenous variable is the type of cereal; students *do not choose* whether to eat Cheerios or Captain Crunch for breakfast.

(2) [12 points] A random variable X has the following continuous probability distribution function:

$$f(x) = 2 - 2X \text{ , } \quad 0 \le x \le 1$$
$$f(x) = 0 \quad \text{ , otherwise}$$

(a) [6 points] Graph this probability distribution. Carefully label the graph. How do you know that this is a valid probability density function? [Graph & 1-2 sentences]

Solution:



This is a valid probability density function because two conditions for the probability density function are satisfied: the total area under the curve is equal to 1 and the function f(x) is positive everywhere on its support.

(b) [6 points] Find $P(X \geq 0.5)$ and $P(X \leq 0.25)$. [Analysis only]

Solution:

$$P(X \geq 0.5) = 0.5 * 0.5 * 1 = 0.25$$
$$P(X \leq 0.25) = 1 - 0.5 * 0.75 * 1.5 = 0.4375$$

(3) [12 points] The researcher wants to learn more about whether the number of years of education is a good predictor of future income. A sample of 150 30-old men and women has been randomly drawn and each respondent has been asked how many years of formal education he or she had completed and his or her income (in thousands of $) for previous 12 months. The STATA summary for the sample is presented below. Coefficient of correlation between years of education and current income is 0.66.

(a) [6 points] Find the least squares line (round your answer to the nearest tenth). [Analysis & 1 equation]

Solution:

Need to find: $\hat{Y} = a + bX$

$$b = \frac{\text{cov}(X,Y)}{\text{var}(X)}$$
$$\Rightarrow \text{cov}(X,Y) = corr(X,Y)sd(X)sd(Y)$$
$$\Rightarrow b = \frac{corr(X,Y)sd(X)sd(Y)}{\text{var}(X)} = \frac{corr(X,Y)sd(Y)}{sd(X)}$$
$$b = \frac{0.66 \cdot 20.9}{3.34} \approx 4.13 \qquad a = \overline{Y} - b\overline{X} = 78.1 - 4.13 \cdot 13.2 \approx 23.6$$

OLS line: $\hat{Y} = 23.6 + 4.1X$

(b) [6 points] Interpret coefficients of the OLS line. [3 – 4 sentences]

Solution:
The slope of the OLS line (coefficient b) shows that in the sample of 150 men and women individuals who got one more year of formal education are observed to earn by 4.1 thousand dollars more on average. (Interpretation that implies causal relationship between education and income is incorrect.)

The intercept (coefficient a in the OLS line) has no particular interpretation. It is wrong to say that individuals without formal education earn on average $23,600, because in our sample no one has 0 years of education.

(4) [16 points] In a small town of Cape d'Azur, all households live either in 1- , 2-, or 3-storey houses. Fraction of those who live in 1-storey houses is 45%, those who live in 2-storey houses – 35%, and 20% of the households live in 3-storey houses. Define X to be a random variable which denotes a number of floors in a house of a randomly selected household.

(a) [8 points] Derive probability distribution of X-bar (sample mean) if sample size is 2 (n=2). Show your work. [Analysis only]

Solution:

Students have to list all possible samples of size 2 with their respective probabilities and sample means:

Sample	Mean	Probability
1,2	1.5	0.45*0.35=0.1575
1,3	2	0.45*0.20=0.0900
2,1	1.5	0.45*0.35=0.1575
3,1	2	0.20*0.45=0.0900
3,2	2.5	0.20*0.35=0.0700
2,3	2.5	0.20*0.35=0.0700
1,1	1	0.45*0.45=0.2025
2,2	2	0.35*0.35=0.1225
3,3	3	0.20*0.20=0.0400

Probability distribution of X-bar:

X-bar	P(X-bar)
1	0.2025
1.5	0.3150
2	0.3025
2.5	0.1400
3	0.0400

(b) [8 points] Graph probability distribution of X-bar. Carefully label the graph. Find mean and standard deviation of X-bar. [Analysis &graph]

Solution:

$$\mu_{\bar{X}} = 1 \cdot 0.45 + 2 \cdot 0.35 + 3 \cdot 0.2 = 1.75 \quad ,$$

$$\sigma^2{}_{\bar{X}} = ((1-1.75)^2 \cdot 0.45 + (2-1.75)^2 \cdot 0.35 + (3-1.75)^2 \cdot 0.2)/2 = 0.29375$$

$$\sigma_{\bar{X}} = 0.54198$$

(5) [14 points] The number of coffee cups consumed by the second year students during exam period is normally distributed with population standard deviation of 4 cups per day. A number of cups per day is recorded for a random sample of 60 second year students. Tabulation below summarizes these data.

(a) [10 points] Compute the 99% confidence estimator of the mean number of cups per day. [Analysis only]

Solution:

To find sample mean, students can use one of two methods:

(a) $\overline{X} = (0*4+1*6+2*9+3*20+4*8+5*5+6*4+10*3+12*1)/60 = 3.45$

(b) $\overline{X} = 0*0.667+1*0.10+2*0.15+3*0.3333+4*0.1333+5*0.833$
$+6*0.667+10*0.5+12*0.167 = 3.45$

Confidence interval :

LCL: $\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 3.45 - 2.58 \cdot \frac{4}{\sqrt{60}} = 2.12$

UCL: $\overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 3.45 + 2.58 \cdot \frac{4}{\sqrt{60}} = 4.78$

(b) [4 points] Interpret confidence interval [1 – 2 sentences]

Solution:

We are 99 percent confident that the interval (2.12, 4.78) contains average number of cups per day consumed by the second year students.

(1) Consider two random variables X and Y which are distributed with means 4 and 5 and standard deviations 1 and 4 respectively. If standard deviation of W=X+Y is 3.5, what can we conclude about the relationship between X and Y? **(C)**

(2) In a large company, 25% of employees work for more than 15 years, 30% work for more than 10 and less than 15 years, and 45% work less than 10 years. What is the chance that both of 2 randomly selected employees work more than 10 years? **(D)**

(3) The heights of the 20,000 men at Big State University (BSU) are normally distributed with a mean of 70 inches and a standard deviation of 2.5 inches. How many of these men are at least 6 feet 5 inches tall? (1 foot=12 inches) **(C)**

►**Questions (4)-(5):** Tweedle Dee and Tweedle Dum are running for President of Wonderland. Half of the voters prefer Dee and half prefer Dum.

(4) If Mad Hatter randomly polls 10 voters, what is the probability that at least 60 percent of those polled will prefer Mr. Dee? **(C)**

(5) If Mad Hatter randomly polls 3 voters and X records the proportion of those who prefer Mr. Dee, which best describes the shape of the probability distribution of X? **(E)**

(6) Consider two identically and independently distributed random variables Z and W, and Z ~ U [-2, 4]. What is the mean and standard deviation of Z+W? **(D)**

(7) Find $P(X \leq 12)$ **(C)**

(8) What is the 95^{th} percentile of Y = 2.5+3*X? **(B)**

(9) Which best describes the shape of the distribution of X? **(A)**

(10) Cans of salmon have a nominal net weight of 250 g. However, due to variation in the canning process, the actual net weight has an approximate normal distribution with a mean of 255 g and a standard deviation of 10 g. According to Consumer Affairs, a sample of 16 tins should have less than a 5% chance that the mean weight is less than 250 g. What is the actual probability that a sample of 16 tins will have a mean weight less than 250 g? **(B)**

(11) If Y = 4 – 5*X, what are the mean and standard deviation of X? **(C)**

(12) If 2 students are selected at random, what is the chance that one is from arts and sciences and one is from engineering? **(D)**

(13) If 10 students are selected, what is the chance that half of them are from engineering? **(B)**

(14) If 400 students are randomly selected, what is the chance that more than 80 are from engineering? **(B)**

(15) Which is the expected number of books for each type of student: undergraduate and graduate (round answer to nearest tenth)? **(E)**

(16) If 1% of all checked out books get lost, and it costs $145 to replace one book, how much money does the librarian expect to allocate for monthly replacements if there are 5,000 students who regularly use library services (round answer to nearest dollar)? **(A)**