PART 1

(1) What is the approximate population standard deviation? **(D)**

(2) For the specific case studied with the simulation, which of these statements is true? **(E)**

(3) How should you expect the simulation results to differ if instead of 100,000 you use only 1,000 simulation draws? **(B)**

(4) If a Binomial distribution has $0.6 \le p \le 0.9$ and $n = 10$, which is definitely true? **(A)**

(5) Suppose a random variable X is Normally distributed. You wish to find P(X > 20). After standardizing you seek P(Z > 3). What does 3 mean? **(E)**

(6) If the degrees of freedom are 12 then what is P(-1.782 < t < 0)? **(C)**

(7) What is P(2 < X < 8)? **(A)**

(8) For a random sample of size 20, what is $P(4 < \overline{X} < 6)$? **(D)**

(9) A population is negatively skewed. You estimate the population median using the sample mean and a sample size much greater than 30. Which properties would this estimator have? **(E) (also accepted (D))**

(10) If the researcher listens in on 20 calls what is the chance that none of them involve an inappropriate remark? **(E)**

(11) If the researcher listens in on 30 calls what is the chance that two of them involve an inappropriate remark? **(A)**

(12) If the researcher listens in on 400 calls what is the chance that more than 24 of them involve an inappropriate remark? **(C)**

(13) Which is a correct statement about what the Central Limit Theorem says? **(A)**

(14) The standard deviation of income is $30,750 and the population is highly skewed. How big of a sample should you collect if you wish to estimate with 96% confidence the average income with a margin of error of plus or minus $1,000? **(D)**

(15) You should expect the width of the confidence interval estimator of the population mean to stay the same (not change) if _______________. **(C)**

(16) You wish to make an inference about the fraction of students in favor of a proposed change in the tuition structure at a very large university. In which of the following cases would the sample proportion be subject to the LEAST amount of sampling error? **(C)**

PART 2
Solutions

(1) (a) First find the covariance between CGPA and bonus.

$$r = \frac{\text{cov}(CGPA, bonus)}{sd(CGPA)sd(bonus)}$$
$$0.4806 * 0.5549993 * 329.9047 = \text{cov}(CGPA, bonus)$$
$$\text{cov}(CGPA, bonus) = 87.996$$

Find the slope of the least squares line.

$$b = \frac{\text{cov}(CGPA, bonus)}{V(CGPA)}$$
$$b = \frac{87.996}{0.3080242} = 285.68$$

Find the intercept of the least squares line.

$$a = \overline{Y} - b\overline{X}$$
$$a = 1751.25 - 285.68 * 3.23125 = 828.15$$

Hence the equation of the least squares line is: $bonus - hat = 828.15 + 285.68 * CGPA$

(b) The intercept has no particular meaning because we can see from the STATA summary that none of the employees had a CGPA of zero. The slope means that employees who have a cumulative GPA that is one point higher are observed to have bonuses that on average are $285.68 higher. We must be careful not to interpret the slope in a way that implies causality but rather the interpretation must simply describe the trends we see in the data.

(c) This is false. There is no way that we can know that the person with the lowest CGPA actually had the lowest bonus. There is a positive correlation but it is not perfect: it is a weak positive correlation. Hence there is a very good chance that the person with the lowest CGPA is not the person with the lowest bonus.

(2) (a) Section 1 uses the iClicker and Section 2 does not. We need to find: $P\left((\overline{X}_1 - \overline{X}_2) < 0\right) = ?$

If claims about the iClicker are true then:

$$\mu_1 = 71$$
$$\sigma_1 = 12$$
$$\sigma_{\overline{X}_1} = \frac{12}{\sqrt{48}} = 1.732$$

According to the given information we know that:

$$\mu_2 = 68$$
$$\sigma_2 = 13$$
$$\sigma_{\overline{X}_2} = \frac{13}{\sqrt{53}} = 1.786$$

Because both sample sizes are bigger than 30 (100 > 30) we can apply the Central Limit Theorem (CLT) to each sample and conclude that each sample mean is Bell shaped and therefore the difference between the sample means is also Bell shaped.

$$P\left((\overline{X}_1 - \overline{X}_2) < 0\right) = P\left(\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < \frac{0 - (71 - 68)}{\sqrt{1.732^2 + 1.786^2}}\right)$$

$$P(Z < -1.206) \approx 0.5 - 0.3869 = 0.1131$$

(b) The probability that the section using the iClicker gets a lower mark than the section not using it is about 11.3% even if the claims about the positive effects of the iClicker are true. This means that there is an 11.3% chance that sampling error explains the discrepancy, which means that sampling error is a plausible explanation.

(3) First calculate the sample mean.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{2*10 + 9*11 + 2*12 + 3*13 + 4*14 + 8*15 + 4*16 + 7*17 + 4*18 + 6*19 + 22}{50}$$

$$\overline{X} = \frac{749}{50} = 14.98$$

Alternatively,

$$\overline{X} = 0.04*10 + 0.18*11 + 0.04*12 + 0.06*13 + 0.08*14 + 0.16*15 + 0.08*16 + 0.14*17 +$$
$$0.08*18 + 0.12*19 + 0.02*22 = 14.98$$

Next,

$$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$14.98 \pm z_{0.10/2} \frac{3}{\sqrt{50}}$$

$$14.98 \pm 1.645 * 0.4243$$

$$14.98 \pm 0.6979$$

Lower Confidence Limit (LCL) = 14.98 – 0.6979 = 14.28
Upper Confidence Limit (UCL) = 14.98 + 0.6979 = 15.68

We are 90% confident that the average time it takes to complete an oil change in the population is within the interval from 14.28 and 15.68.

(4)

$$X \sim N(20, 20^2)$$
$$\mu_{\overline{X}} = \mu_X = 20$$
$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{20}{\sqrt{5}} = 8.9443$$

$$P(\overline{X} < 0) = P\left(Z < \frac{0-20}{8.9443}\right)$$
$$= P(Z < -2.236)$$
$$= 0.5 - 0.4875$$
$$= 0.0125$$

(Students do not have to interpolate using the standard normal table. As the graph below shows the exact area is 0.0127 but 0.0125 is close enough.)

(5) (a) Does playing video games that focus on altruistic behavior (helping others) cause people to behave more altruistically (helpfully) in the real world? [Note: This is the primary research question, which is what the question asked for.]

(b) Studies A and D use experimental data. Study A randomly assigned participants to play either a helpful, violent or neutral video game: X is randomly assigned and is exogenous. In Study D the participants were randomly assigned either a helpful or neutral video game: so again X is exogenous. Studies B and C use observational data. In Studies B and C the participants chose what type of video games they wanted to play: X is endogenous. Specifically X (type of video game) is influenced by underlying personality traits that will also affect a person's behavior.

(c) We can infer a causal relationship from the two experimental studies (Studies A and D) but not from the two observational studies (Studies B and C). Because the type of video game is randomly assigned in the experimental studies any differences in behavior *can* be attributed to the influence of the game. However because users can choose the types of video games to play in the observational studies we cannot know whether the video games changed their behavior. This is because part of the explanation would likely be that people who are the type to be altruistic (helpful) chose to play altruistic (helpful) games and people who are the type to be aggressive choose to play aggressive games.