Student: ____________________	**Instructor:** Richeng Piao	**Assignment:** Practice Problem Set 3
Date: ____________________	**Course:** ECON 2560 - Applied Econometrics	

1. Consider a random variable *Y*. What is the difference between the sample average $\bar{Y}$ and the population mean?

- **A.** Both the population mean and the sample average $\bar{Y}$ are estimators of the central tendency of the distribution of *Y*.
- **B.** Both the population mean and the sample average $\bar{Y}$ are true measures of the central tendency of the distribution of *Y*.
- **C.** The sample average $\bar{Y}$ is a true measure of the central tendency of the distribution of *Y* whereas the population mean is an estimator of the sample average.
- **D.** The population mean is a true measure of the central tendency of the distribution of *Y* whereas the sample average $\bar{Y}$ is an estimator of the population mean.

Answer: D.
The population mean is a true measure of the central tendency of the distribution of *Y* whereas the sample average $\bar{Y}$ is an estimator of the population mean.

ID: Review Concept 3.1

2. What is the difference between an estimator and an estimate?

- **A.** An estimator is a function of a sample of data to be drawn randomly from a population whereas an estimate is the numerical value of the estimator when it is actually computed using data from a specific sample.
- **B.** Both an estimator and an estimate are functions of a sample of data to be drawn randomly from a population.
- **C.** Both an estimator and an estimate are numerical values computed using data from a specific sample.
- **D.** An estimate is a function of a sample of data to be drawn randomly from a population whereas an estimator is the numerical value of the estimator when it is actually computed using data from a specific sample.

Determine whether the following are examples of estimators, estimates or neither.

A. The sample variance is 7.5. (1) ____________

B. The mean value of random variable Y in a population. (2) ____________

C. The variance of random variable Z in a population. (3) ____________

D. The sample average is $\bar{X}$. (4) ____________

E. The sample average is 89.6. (5) ____________

(1)	(2)	(3)	(4)	(5)
estimate	neither	neither	estimator	estimate
estimator	estimate	estimate	neither	neither
neither	estimator	estimator	estimate	estimator

Answers A.

An estimator is a function of a sample of data to be drawn randomly from a population whereas an estimate is the numerical value of the estimator when it is actually computed using data from a specific sample.

(1) estimate

(2) neither

(3) neither

(4) estimator

(5) estimate

ID: Review Concept 3.2

3. Suppose random variable X follows a normal distribution with mean 10 and variance 16. Consider random samples of different sizes from that population to answer the following questions.

Consider a sample of size n = 10. Download the data attached in the data table and use it to construct your response .

For n = 10, the sample mean = [ ]

(*Round your response to three decimal places*)

Consider a sample of size n = 100. Download the data attached in the data table and use it to construct your response .

For n = 100, the sample mean = [ ]

(*Round your response to three decimal places*)

Consider a sample of size n = 999. Download the data attached in the data table and use it to construct your response .

For n = 999, the sample mean = [ ]

(*Round your response to three decimal places*)

How can you relate your answers above to the law of large numbers?

- **A.** The sample mean is equal to the population mean regardless what the sample size is.
- **B.** As the sample size increases, the positive distance between the sample mean and the population mean increases.
- **C.** The sample size has no effect on the sample mean.
- **D.** As the sample size increases, the sample mean approaches the population mean.

Answers 8.774

9.755

10.108

D. As the sample size increases, the sample mean approaches the population mean.

ID: Review Concept 3.3

4. Let Y be a random variable. In a population, $\mu_Y = 98$ and $\sigma_Y^2 = 40$. Use the central limit theorem to answer the following questions. (***Note****: any intermediate results should be rounded to four decimal places)*

In a random sample of size n = 161, find $\Pr(\bar{Y} < 101)$.

$$\Pr(\bar{Y} < 101) = \square$$

(*Round your response to four decimal places*)

In a random sample of size n = 153, find $\Pr(99 < \bar{Y} < 104)$.

$$\Pr(99 < \bar{Y} < 104) = \square$$

(*Round your response to four decimal places*)

In a random sample of size n = 65, find $\Pr(\bar{Y} > 99)$.

$$\Pr(\bar{Y} > 99) = \square$$

(*Round your response to four decimal places*)

Answers 1.0000

0.0252

0.1012

ID: Exercise 3.1

5. Observe that for a random variable Y that takes on values 0 and 1, the expected value of Y is defined as follows:

$$E(Y) = 0 \times \Pr(Y=0) + 1 \times \Pr(Y=1)$$

and the variance of Y is

$$var(Y) = E\left[\left(Y - \mu_y\right)^2\right] = (0-p)^2 \times \Pr(Y=0) + (1-p)^2 \times \Pr(Y=1) = p^2(1-p) + (1-p)^2 p = p(1-p)$$

Now, suppose that X is a Bernoulli random variable with success probability Pr $(X = 1) = p$ and let $Y_1,\ldots,Y_n$ be i.i.d. draws from this distribution. Also, let $\hat{p}$ be the fraction of successes (1s) in this sample.

Use the information above to answer the following questions.

Show that $\hat{p} = \overline{Y}$.

$$\hat{p} = \frac{\#(\underset{(1)}{\_\_\_\_\_\_})}{n} = \frac{\#(Y_i = \underset{(2)}{\_\_\_\_\_\_})}{n} = \frac{\Sigma_{i=1}^{n} Y_i}{n} = \overline{Y}$$

Show that $\hat{p}$ is an unbiassed estimator of p.

$$E\left(\hat{p}\right) = E\left(\frac{\Sigma_{i=1}^{n} Y_i}{n}\right) = \frac{1}{\underset{(5)}{\_\_\_\_\_\_}} \Sigma_{i=1}^{n} E(\underset{(3)}{\_\_\_\_\_\_}) = \frac{1}{\underset{(6)}{\_\_\_\_\_\_}} \Sigma_{i=1}^{n} \underset{(4)}{\_\_\_\_\_\_} = p$$

Show that $\text{var}(\hat{p}) = \frac{p(1-p)}{n}$.

$$\text{var}\left(\hat{p}\right) = \text{var}\left(\frac{\Sigma_{i=1}^{n} Y_i}{n}\right) = \frac{1}{\underset{(9)}{\_\_\_\_\_\_}} \Sigma_{i=1}^{n} \text{var}(\underset{(7)}{\_\_\_\_\_\_}) = \frac{1}{\underset{(10)}{\_\_\_\_\_\_}} \Sigma_{i=1}^{n} \underset{(8)}{\_\_\_\_\_\_} = \frac{p(1-p)}{n}$$

(1) ○ successes ○ observations ○ variables

(2) ○ 0 ○ 1 ○ 2

(3) ○ Y_i ○ $\overline{Y}$ ○ p

(4) ○ $\overline{Y}$ ○ p ○ $1-p$

(5) ○ n ○ n^2 ○ n^3

(6) ○ n ○ n^2 ○ n^3

(7) ○ Y_i ○ $\overline{Y}$ ○ p

(8) ○ $p(1-p)$ ○ p ○ $1-p$

(9) ○ n ○ n^2 ○ n^3

(10) ○ n ○ n^2 ○ n^3

Answers (1) successes

(2) 1

(3) Y_i

(4) p

(5) n

(6) n

(7) Y_i

(8) $p(1-p)$

(9) n^2

(10) n^2

ID: Exercise 3.2

6. Follow the steps below to show that the sample variance is an unbiased estimator of the population variance when $Y_1,\dots,Y_n$ are i.i.d. with mean μ_Y and variance σ_Y^2.

Step 1. Use the following equation to show that $E\left[\left(Y_i - \bar{Y}\right)^2\right] = \text{var}\left(Y_i\right) - 2\text{cov}\left(Y_i, \bar{Y}\right) + \text{var}\left(\bar{Y}\right)$. For random variables X and Y,

$$\text{var}\,(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2$$

First observe that $E\left[\left(Y_i - \bar{Y}\right)^2\right] = E\left\{\left[\left(Y_i - \mu_Y\right) - \left(\bar{Y} - \mu_Y\right)\right]^2\right\}$ by adding and subtracting μ_Y inside the expression.

Expand this to get,

$= E[(Y_i -$ (1) ______ $)^2 -$ ______ $(Y_i -$ (2) ______ $)($ (3) ______ $- \mu_Y) +$
$(\bar{Y} - \mu_Y)^2]$

Distributing the expectation,

$$= E[(Y_i - \mu_Y)^2] - 2E[(Y_i - \mu_Y)(\bar{Y} - \mu_Y)] + E[(\bar{Y} - \mu_Y)^2]$$

Consider the last expression above. This is equivalent term by term to

= (4) ______ $-\,2$ (5) ______ + (6) ______

Therefore, $E\left[\left(Y_i - \bar{Y}\right)^2\right] = \text{var}\left(Y_i\right) - 2\text{cov}\left(Y_i, \bar{Y}\right) + \text{var}\left(\bar{Y}\right)$. Call this result **(A)**.

Now, use the following equation to show that $\text{cov}\left(\bar{Y}, Y_i\right) = \frac{\sigma_Y^2}{n}$. For random variables X, V and Y,

$$\text{cov}(a + bX + cV,\, Y) = b\sigma_{XY} + c\sigma_{VY}$$

First, observe that by definition of covariance,

$\text{cov}\left(\bar{Y}, Y\right) =$ (7) ______

By definition of $\bar{Y}$, this is equivalent to

$$= E\left[\left(\frac{\Sigma_{j=1}^{n} Y_j}{n} - \mu_Y\right)\left(Y_i - \mu_Y\right)\right] = E\left[\left(\frac{\Sigma_{j=1}^{n}\left(Y_j - \mu_Y\right)}{n}\right)\left(Y_i - \mu_Y\right)\right]$$

Distributing the expectation,

= (8) ______ $E\,[$(9) ______ $]$ + (10) ______ $\sum_{j \neq i} E\left[\left(Y_j - \mu_Y\right)\left(Y_i - \mu_Y\right)\right]$

Consider the last expression above. This is equivalent term by term to

$= \frac{1}{n}$ (11) ______ $+ \frac{1}{n}\sum_{j \neq i}$ (12) ______ $= \frac{\sigma_Y^2}{n}$

Therefore, $\text{cov}\left(\bar{Y}, Y_i\right) = \frac{\sigma_Y^2}{n}$. Call this result **(B)**.

Let s_Y^2 denote the sample variance. Use results **A** and **B** to show that $E\left(s_Y^2\right) = \sigma_Y^2$.

First, observe that by definition of sample variance,

$E\left(s_Y^2\right) = E\,[$(13) ______ $\sum_{i=1}^{n}$ (14) ______ $]$

Using result **A,** this becomes

$$= \frac{1}{n-1} \sum_{i=1}^{n} E\left[\left(Y_i - \bar{Y}\right)^2\right] = \frac{1}{n-1} \sum_{i=1}^{n} [(15) \underline{\hspace{3cm}}]$$

Using result **B,** this simplifies to

$$= \frac{1}{n-1} \sum_{i=1}^{n} [\sigma_Y^2 - 2 \times (16) \underline{\hspace{3cm}} + \frac{\sigma_Y^2}{n}]$$

Finally,

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{n-1}{n} \sigma_Y^2\right) = \sigma_Y^2$$

Since $E\left(s_Y^2\right) = \sigma_Y^2$, the sample variance is an unbiased estimator of the population variance.

(1)
- $\bar{Y}$
- μ_Y

(2)
- $\bar{Y}$
- μ_Y

(3)
- $\bar{Y}$
- Y_i

(4)
- $\text{var}\left(Y_i\right)$
- $\text{var}\left(\bar{Y}\right)$
- $\text{cov}\left(Y_i, \bar{Y}\right)$

(5)
- $\text{var}\left(Y_i\right)$
- $\text{var}\left(\bar{Y}\right)$
- $\text{cov}\left(Y_i, \bar{Y}\right)$

(6)
- $\text{var}\left(Y_i\right)$
- $\text{var}\left(\bar{Y}\right)$
- $\text{cov}\left(Y_i, \bar{Y}\right)$

(7)
- $E\left[\left(\bar{Y} - \mu_Y\right)\left(Y_i - \mu_Y\right)\right]$
- $\text{var}\left[\left(\bar{Y} - \mu_Y\right)\left(Y_i - \mu_Y\right)\right]$
- $\text{corr}\left[\left(\bar{Y} - \mu_Y\right)\left(Y_i - \mu_Y\right)\right]$

(8)
- $\frac{1}{n}$
- $\frac{1}{n^2}$
- $\frac{1}{n^3}$

(9)
- $\left(Y_i - \mu_Y\right)$
- $\left(Y_i - \mu_Y\right)^2$
- $\left(Y_i - \mu_Y\right)^3$

(10)
- $\frac{1}{n}$
- $\frac{1}{n^2}$
- $\frac{1}{n^3}$

(11)
- σ_Y^2
- μ_Y
- $\text{cov}\left(Y_j, Y_i\right)$

(12)
- σ_Y^2
- μ_Y
- $\text{cov}\left(Y_j, Y_i\right)$

(13)
- $\frac{1}{n^2}$
- $\frac{1}{n-1}$
- $\frac{1}{n-2}$

(14)
- $\left(Y_i - \bar{Y}\right)$
- $\left(Y_i - \bar{Y}\right)^2$
- $\left(Y_i - \bar{Y}\right)^3$

(15)
- $\text{var}\left(Y_i\right) - 2\text{cov}\left(Y_i, \bar{Y}\right) + \text{var}\left(\bar{Y}\right)$
- $E\left(Y_i\right) - 2\text{cov}\left(Y_i, \bar{Y}\right) + E\left(\bar{Y}\right)$
- $\text{var}\left(Y_i\right) - 2\text{corr}\left(Y_i, \bar{Y}\right) + \text{var}\left(\bar{Y}\right)$

(16)
- $\frac{\sigma_Y^2}{n}$
- σ_Y^2
- $\frac{\sigma_Y^2}{n^2}$

Answers (1) μ_Y

2

(2) μ_Y

(3) $\overline{Y}$

(4) $\text{var}\left(Y_i\right)$

(5) $\text{cov}\left(Y_i, \overline{Y}\right)$

(6) $\text{var}\left(\overline{Y}\right)$

(7) $E\left[\left(\overline{Y}-\mu_Y\right)\left(Y_i-\mu_Y\right)\right]$

(8) $\frac{1}{n}$

(9) $\left(Y_i-\mu_Y\right)^2$

(10) $\frac{1}{n}$

(11) σ_Y^2

(12) $\text{cov}\left(Y_j, Y_i\right)$

(13) $\frac{1}{n-1}$

(14) $\left(Y_i-\overline{Y}\right)^2$

(15) $\text{var}\left(Y_i\right) - 2\text{cov}\left(Y_i, \overline{Y}\right) + \text{var}\left(\overline{Y}\right)$

(16) $\frac{\sigma_Y^2}{n}$

ID: Exercise 3.18

7. Let $Y_1, \ldots, Y_n$ be i.i.d with mean μ_Y and variance σ_Y^2.

$\bar{Y}$ is an unbiased estimator of μ_Y. Is $\bar{Y}^2$ an unbiased estimator of μ_Y^2?

First, note that for a random variable X

$$\text{var}(X) = E\left(X^2\right) - E(X)^2 = E\left(X^2\right) - \mu_X^2 \text{ and } E\left(X_i X_j\right) = \mu_X^2 \text{ for } i \neq j.$$

So,

$$E\left(\bar{Y}^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E\left(Y_i^2\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j \neq i} E\left(Y_i Y_j\right) = (1)\_\_\_\_\_\_ + (2)\_\_\_\_\_\_ (3)\_\_\_\_\_\_$$

Hence,

- **A.** Yes, $\bar{Y}^2$ is an unbiased estimator of μ_Y^2.
- **B.** No, $\bar{Y}^2$ is not an unbiased estimator of μ_Y^2.

$\bar{Y}$ is a consistent estimator of μ_Y. Is $\bar{Y}^2$ a consistent estimator of μ_Y^2?

$$E\left(\bar{Y}^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E\left(Y_i^2\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j \neq i} E\left(Y_i Y_j\right) = \mu_Y^2 + \frac{1}{n}\sigma_Y^2$$

Hence,

- **A.** Yes, $\bar{Y}^2$ is a consistent estimator of μ_Y^2.
- **B.** No, $\bar{Y}^2$ is not a consistent estimator of μ_Y^2.

(1)
- μ_Y
- μ_Y^2
- σ_Y^2

(2)
- $\frac{1}{n}$
- $\frac{1}{n^2}$
- $\frac{1}{n^3}$

(3)
- μ_Y
- μ_Y^2
- σ_Y^2

Answers (1) μ_Y^2

(2) $\frac{1}{n}$

(3) σ_Y^2

B. No, $\bar{Y}^2$ is not an unbiased estimator of μ_Y^2.

A. Yes, $\bar{Y}^2$ is a consistent estimator of μ_Y^2.

ID: Exercise 3.19

8. An estimator $\hat{\mu}_Y$ of the population value μ_Y is consistent if:

- **A.** $\bar{Y} \xrightarrow{p} 0$.
- **B.** $\hat{\mu}_Y \xrightarrow{p} \mu_Y$.
- **C.** Y is normally distributed.
- **D.** its mean square error is the smallest possible.

Answer: B. $\hat{\mu}_Y \xrightarrow{p} \mu_Y$.

ID: Test A Ex 3.1.1

---

9. Which of the following statements best describes an unbiased estimator?

- **A.** Its value is always equal to the true parameter value.
- **B.** Its value is always the same in repeated sampling for the same sample size.
- **C.** Its value is a function of the sample size.
- **D.** Its average value, over repeated sampling for the same sample size, is equal to the population value.

Answer: D. Its average value, over repeated sampling for the same sample size, is equal to the population value.

ID: Concept Exercise 3.1.1

---

10. Suppose K_1, K_2, and K_3 are three independent and unbiased estimators of μ. Their variances are in the ratio of 5:4:9.

Z_1, Z_2 and Z_3 are three other estimators of μ such that:

$$Z_1 = \frac{(5K_1 + K_2 + K_3)}{7},$$

$$Z_2 = \frac{(K_1 + 5K_2 + K_3)}{7},$$

$$Z_3 = \frac{(K_1 + K_2 + 5K_3)}{7}.$$

Which among Z_1, Z_2, and Z_3 are unbiased estimators of μ?

- **A.** Z_1, Z_2, and Z_3 are all biased.
- **B.** Z_1 and Z_3 are unbiased but Z_2 is biased.
- **C.** Z_2 and Z_3 are unbiased but Z_1 is biased.
- **D.** Z_1, Z_2, and Z_3 are all unbiased.

Which among Z_1, Z_2, and Z_3 would be the best estimator for μ?

- **A.** Z_2 should be chosen since it is both unbiased and has the least variance.
- **B.** Sufficient information is not provided to determine which estimator would be the best.
- **C.** Both Z_1 and Z_2 could be chosen since they are unbiased and have the least variances.
- **D.** Z_1 should be chosen since it is both unbiased and consistent.

Answers D. Z_1, Z_2, and Z_3 are all unbiased.

A. Z_2 should be chosen since it is both unbiased and has the least variance.

ID: Concept Exercise 3.1.2

11. In a survey of 400 likely voters, 214 responded that they would vote for the incumbent and 186 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

Using the survey results, the estimated value of $\hat{p}$ is [ ]. (*Round your response to four decimal places.*)

Using $\hat{p}(1 - \hat{p})/n$ as the estimator of the variance of $\hat{p}$, the standard error of the estimator is [ ]. (*Round your response to four decimal places.*)

The p-value for the test $H_0: p = 0.5$ versus $H_1: p \neq 0.5$ is [ ]. (*Round your response to three decimal places.*)

The p-value for the test $H_0: p = 0.5$ versus $H_1: p > 0.5$ is [ ]. (*Round your response to three decimal places.*)

Why do the p-values for $H_0: p = 0.5$ versus $H_1: p \neq 0.5$ and $H_0: p = 0.5$ versus $H_1: p > 0.5$ differ?

- **A.** $H_0: p = 0.5$ versus $H_1: p \neq 0.5$ is a two-sided test and the p-value is the area in the tails of the standard normal distribution outside ± the calculated t-statistic.
- **B.** $H_0: p = 0.5$ versus $H_1: p > 0.5$ is a two-sided test and the p-value is the area in the tails of the standard normal distribution outside ± the calculated t-statistic.
- **C.** $H_0: p = 0.5$ versus $H_1: p \neq 0.5$ is a one-sided test and the p-value is the area under the standard normal distribution to the right of the calculated t-statistic.
- **D.** $H_0: p = 0.5$ versus $H_1: p > 0.5$ is a one-sided test and the p-value is the area under the standard normal distribution to the left of the calculated t-statistic.

Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey?

- **A.** For the test $H_0: p = 0.5$ versus $H_1: p > 0.5$, we cannot reject the null hypothesis at the 5% significance level. The p-value is larger than 0.05. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.
- **B.** For the test $H_0: p = 0.5$ versus $H_1: p \neq 0.5$, we cannot reject the null hypothesis at the 5% significance level. The p-value is larger than 0.05. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.
- **C.** For the test $H_0: p = 0.5$ versus $H_1: p \neq 0.5$, we can reject the null hypothesis at the 5% significance level. The p-value is less than 0.05. The test suggests that the survey contained statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.
- **D.** For the test $H_0: p = 0.5$ versus $H_1: p > 0.5$, we can reject the null hypothesis at the 5% significance level. The p-value is less than 0.05. The test suggests that the survey contained statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

Answers 0.535

0.0249

0.160

0.080

A.

H_o: $p = 0.5$ versus H_1: $p \neq 0.5$ is a two-sided test and the p-value is the area in the tails of the standard normal distribution outsid ± the calculated t-statistic.

A.

For the test H_o: $p = 0.5$ versus H_1: $p > 0.5$, we cannot reject the null hypothesis at the 5% significance level. The p-value is large than 0.05. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

ID: Exercise 3.3

---

12. A large p-value implies:

- **A.** a large $\bar{Y}^{act}$.
- **B.** rejection of the null hypothesis.
- **C.** a large t-statistic.
- **D.** that the observed value $\bar{Y}^{act}$ is consistent with the null hypothesis.

Answer: D. that the observed value $\bar{Y}^{act}$ is consistent with the null hypothesis.

ID: Test A Ex 3.2.2

---

13. A type II error is:

- **A.** typically smaller than the type I error.
- **B.** the error you make when choosing type II or type I.
- **C.** the error you make when not rejecting the null hypothesis when it is false.
- **D.** cannot be calculated when the alternative hypothesis contains an "=".

Answer: C. the error you make when not rejecting the null hypothesis when it is false.

ID: Test B Ex 3.2.1

---

14. The power of the test:

- A. is the probability that the test correctly rejects the null when the alternative is true.
- B. depends on whether you use $\bar{Y}$ or $\bar{Y}^2$ for the *t*-statistic.
- C. is the probability that the test actually incorrectly rejects the null hypothesis when the null is true.
- D. is one minus the size of the test.

Answer: A. is the probability that the test correctly rejects the null when the alternative is true.

ID: Test B Ex 3.2.2

---

15. A researcher observes the weight of a class of 36 students in high school. The given table shows the data she collected (in kg).

71	70	68	75	70	71
70	72	72	69	67	67
74	74	72	68	71	74
65	71	70	71	71	72
75	77	66	70	73	73
69	67	71	71	72	71

The researcher wants to test whether the mean weight of all high school students in the state (μ) would be 70 kg or not.

So, the null and the alternative hypotheses for the test the researcher wants to conduct are:

- A. The researcher's null hypothesis is H_0: $\mu = 70$ and the alternative hypothesis is H_A: $\mu \neq 70$.
- B. The researcher's null hypothesis is H_0: $\mu \geq 70$ and the alternative hypothesis is H_A: $\mu \leq 70$.
- C. The researcher's null hypothesis is H_0: $\mu = 70$ and the alternative hypothesis is H_A: $\mu \geq 70$.
- D. The researcher's null hypothesis is H_0: $\mu \geq 70$ and the alternative hypothesis is H_A: $\mu \neq 70$.

For a prespecified 5% significance level, the *p*-value of the test the researcher wants to conduct is ____.

(*Round your answer to four decimal places*.)

The calculated *p*-value suggests that we will (1) ____ the null hypothesis.

(1) reject
fail to reject

Answers A. The researcher's null hypothesis is H_0: $\mu = 70$ and the alternative hypothesis is H_A: $\mu \neq 70$.

0.0598

(1) fail to reject

ID: Concept Exercise 3.2.1

---

16. A survey collected data from a random sample of 144 people living in Jade city. The sample average of the distance people travel to reach their workplaces ($\bar{Y}$) is 23.28 km and the standard deviation (s_Y) is 7.48 km.

The standard error of the sample average of the distance people travel to reach their workplaces is [ ] km.

(*Round your answer to two decimal places.*)

Let μ_Y denote the mean of the distance all the people in Jade city travel to reach their workplaces.

The *p*-value of the test $H_0: \mu_Y = 22$ km vs. $H_1: \mu_Y \neq 22$ km is [ ].

(*Round your answer to two decimal places.*)

The *p*-value suggests that at the 5% significance level, we (1) ______ the hypothesis that the mean distance people travel to reach their workplaces in Jade city is 22 km.

(1) ○ reject
○ fail to reject

Answers 0.62

0.04

(1) reject

ID: Concept Exercise 3.2.2

---

17. Suppose a survey of a random sample of 144 smokers, conducted by the Department of Health, suggests that the mean number of cigarettes a person smokes in a day in Smokelandia ($\bar{Y}$) is 2.58 and the standard deviation (s_Y) is 0.58. The Department of Health is concerned about the results of the survey and wants to test whether the mean number of cigarettes a person smokes in a day is 2.52 or not.

The test statistic associated with the above test is [ ].

(*Round your answer to two decimal places.*)

If the Department of Health uses a 5% significance level, the test statistic suggests that we (1) ______ the hypothesis that the mean number of cigarettes a person smokes in a day is 2.52.

Suppose the Department of Health decides to reject the null hypothesis (H_0) if $|\bar{Y} - 2.52| > 0.08$.

The size of the above test is [ ].

(*Round your answer to two decimal places.*)

If $\mu = 2.69$, the power of the above test is [ ].

(*Round your answer to two decimal places.*)

(1) ○ reject
○ fail to reject

Answers 1.24

(1) fail to reject

0.10

0.97

ID: Concept Exercise 3.2.3

---

18. Suppose a random sample of 150 universities is used to test the null hypothesis that the average number of spam emails economics graduate students receive in a month is 46.92. The value of the test statistic is found to be 2.06.

If the null is tested against the alternative hypothesis that the number of spam emails is not 46.92, the smallest significance level at which you can reject the null hypothesis is [ ].

(*Round your answer to four decimal places.*)

If the null is tested against the alternative hypothesis that the number of spam emails is more than 46.92, the smallest significance level at which you can reject the null hypothesis is [ ].

(*Round your answer to four decimal places.*)

If while conducting a hypothesis test you reject the null hypothesis when in fact it was true, you would have made a (1) ________.

(1) ○ Type I error
○ Type II error

Answers 0.0394

0.0197

(1) Type I error

ID: Concept Exercise 3.2.4

---

19. Suppose we want to test whether true population parameter θ equals a certain number. Consider the following hypothesis test:

$$H_0: \theta = c$$
$$H_1: \theta \neq c$$

Where H_0 is the null hypothesis, H_1 is the alternative hypothesis, and c a constant. Let $\bar{\theta}$ be an estimator of θ.

What role does the central limit theorem play in statistical hypothesis testing?

○ **A.** The central limit theorem plays no role in statistical hypothesis testing.

○ **B.** To construct a rejection rule for H_0, it is necessary to know the sampling distribution of $\bar{\theta}$ under the null hypothesis. If the sampling distribution is unknown, the central limit theorem says that it can be approximated by a normal distribution when the sample size n is sufficiently large.

○ **C.** The central limit theorem allows us to directly compute the true value of θ.

○ **D.** To construct a rejection rule for H_0, it is necessary to know the sampling distribution of $\bar{\theta}$ under the null hypothesis. If the sampling distribution is unknown, the central limit theorem says that it can be approximated by a normal distribution when the sample size n is sufficiently small.

Answer: B.

To construct a rejection rule for H_0, it is necessary to know the sampling distribution of $\bar{\theta}$ under the null hypothesis. If the sampling distribution is unknown, the central limit theorem says that it can be approximated by a normal distribution when the sample size n is sufficiently large.

ID: Review Concept 3.4

---

20. Let Y be a random variable. Consider the following hypothesis test:

$$H_0: E(Y) = \mu_{Y,0}$$
$$H_1: E(Y) \neq \mu_{Y,0}$$

Where H_0 is the null hypothesis and H_1 is the alternative hypothesis.

Which of the following statements are true about the null and alternative hypothesis? *(Check all that apply.)*

- ☐ **A.** If the null hypothesis is accepted the statistician declares it to be true and the alternative false. If the null hypothesis is rejected the statistician declares it to be false and the alternative true.
- ☐ **B.** Hypothesis testing entails using data to provide sufficient evidence in favor of both the null and alternative hypothesis.
- ☐ **C.** Hypothesis testing entails using data to provide sufficient evidence against both the null and alternative hypothesis.
- ☐ **D.** Hypothesis testing entails using data to provide sufficient evidence against the null hypothesis and in favor of the alternative hypothesis.
- ☐ **E.** The null hypothesis is that the population mean *E(Y)* takes on a specific value whereas the alternative hypothesis specifies what is true if the null hypothesis is not.

Determine whether the following illustrate a one–sided alternative hypothesis or a two–sided alternative hypothesis.

1. $H_0: E(Y) = \mu_{Y,0}$ vs. $H_1: E(Y) > \mu_{Y,0}$. This example illustrates a (1) __________ alternative hypothesis.

2. $H_0: E(Y) = \mu_{Y,0}$ vs. $H_1: E(Y) \neq \mu_{Y,0}$. This example illustrates a (2) __________ alternative hypothesis.

3. $H_0: E(Y) = \mu_{Y,0}$ vs. $H_1: E(Y) \leq \mu_{Y,0}$. This example illustrates a (3) __________ alternative hypothesis.

A (4) __________ alternative hypothesis states the value is *either* greater than or less than $\mu_{Y,0}$ but not both, whereas a

(5) __________ alternative hypothesis simply states the value is not equal to $\mu_{Y,0}$.

Match the hypothesis testing terminology to its definition.

1. The prescribed rejection probability of a statistical hypothesis test when the null hypothesis is true. (6) __________

2. The probability that the test incorrectly rejects the null hypothesis when it is true. (7) __________

3. The value of the statistic for which the test just rejects the null hypothesis at the given significance level. (8) __________

4. The probability that the test correctly rejects the null hypothesis when the alternative is true. (9) __________

(1) ○ one-sided ○ two-sided

(2) ○ two-sided ○ one-sided

(3) ○ one-sided ○ two-sided

(4) ○ one-sided ○ two-sided

(5) ○ one-sided ○ two-sided

(6) ○ Significance level ○ Size of the test ○ Critical value ○ Power of the test

(7) ○ Significance level ○ Size of the test ○ Critical value ○ Power of the test

(8) ○ Significance level ○ Size of the test ○ Critical value ○ Power of the test

(9) ○ Significance level ○ Size of the test ○ Critical value ○ Power of the test

Answers D.

Hypothesis testing entails using data to provide sufficient evidence against the null hypothesis and in favor of the alternative hypothesis.

, E.

The null hypothesis is that the population mean *E(Y)* takes on a specific value whereas the alternative hypothesis specifies what true if the null hypothesis is not.

(1) one-sided

(2) two-sided

(3) one-sided

(4) one-sided

(5) two-sided

(6) Significance level

(7) Size of the test

(8) Critical value

(9) Power of the test

ID: Review Concept 3.5

---

21. Let *Y* be a random variable. Suppose you are interested in estimating the population mean, *E(Y)*.

Which of the following statements about confidence intervals and hypotheses tests are true? *(Check all that apply.)*

- ☐ **A.** A confidence interval defines a rejection rule, at a specified significance level, for all possible values of the population mean, *E(Y)*, whereas a single hypothesis test is a rejection rule for only one possible value of the population mean.
- ☐ **B.** A confidence interval summarizes the set of hypotheses about the population mean, *E(Y)*, you can and cannot reject at a given significance level.
- ☐ **C.** A single hypothesis test about the population mean, *E(Y)*, contains more information than a confidence interval.
- ☐ **D.** A confidence interval defines a rejection rule, at a specified significance level, for only one possible value of the population mean, *E(Y)*, whereas a single hypothesis test is a rejection rule for all possible values of the population mean.

Answer: A.

A confidence interval defines a rejection rule, at a specified significance level, for all possible values of the population mean, *E(Y)*, whereas a single hypothesis test is a rejection rule for only one possible value of the population mean.

, B.

A confidence interval summarizes the set of hypotheses about the population mean, *E(Y)*, you can and cannot reject at a given significance level.

ID: Review Concept 3.6

---

22. Suppose a researcher believes that the average height of female students in a large local high school is 140 cm. The researcher wants to construct an interval that contains the true average height of all female students in the local high school with a certain prespecified probability. The researcher selects 36 female students at random from the high school. The distribution of heights is known to follow a normal distribution.

The given table shows the data she collected (in cm).

140	145	142	139	135	139
142	147	129	138	147	136
144	140	140	150	134	139
138	139	142	141	143	142
139	140	145	137	146	135
143	140	135	142	143	137

The average height of the female students in the given sample is [ ] cm.

The sample standard deviation is [ ] cm.

(*Round your answers to two decimal places.*)

The standard error of the sample mean height is [ ] cm.

The margin of error in calculating a 95% confidence interval for the average height of all female students in the local high school is [ ] cm.

(*Round your answers to two decimal places.*)

The 95% confidence interval for the average height of all female students in the local high school, μ_X, is: ([ ], [ ]).

(*Round your answers to two decimal places.*)

Answers 140.36

4.22

0.70

1.37

138.99

141.73

ID: Concept Exercise 3.3.1

23. Suppose a new standardized test is given to 97 randomly selected third-grade students in New Jersey. The sample average score $\overline{Y}$ on the test is 61 points, and the sample standard deviation, s_Y, is 10 points. The authors plan to administer the test to all third-grade students in New Jersey.

The 95% confidence interval for the mean score of all New Jersey third graders is ([ ] , [ ]). (*Round your responses to two decimal places.*)

Suppose the same test is given to 194 randomly selected third graders from Iowa, producing a sample average of 65 points and sample standard deviation of 13 points.

The 90% confidence interval for the difference in mean scores between Iowa and New Jersey is ([ ] , [ ]). (*Round your responses to two decimal places.*)

The *p*-value of the test of no difference in means versus some difference is [ ]. (*Round your response to four decimal places.*)

Can you conclude with a high degree of confidence that the population means for Iowa and New Jersey students are different?

- **A.** Because of the extremely low *p*-value, we can reject the null hypothesis with a very high degree of confidence. Hence, the population means for Iowa and New Jersey students are different.
- **B.** Because of the extremely high *p*-value, we cannot reject the null hypothesis with a very high degree of confidence. Hence, the population means for Iowa and New Jersey students are not different.
- **C.** Because of the extremely low *p*-value, we cannot reject the null hypothesis with a very high degree of confidence. Hence, the population means for Iowa and New Jersey students are not different.
- **D.** Because of the extremely high *p*-value, we can reject the null hypothesis with a very high degree of confidence. Hence, the population means for Iowa and New Jersey students are different.

Answers 59.01

62.99

– 6.26

– 1.74

0.0038

A.

Because of the extremely low *p*-value, we can reject the null hypothesis with a very high degree of confidence. Hence, the population means for Iowa and New Jersey students are different.

ID: Exercise 3.10

24. Data on fifth-grade test scores (reading and mathematics) for 423 school districts in California yield $\overline{Y}$ = 659.1 and standard deviation s_Y = 19.9.

The 95% confidence interval for the mean test score in the population is (______ , ______). (*Round your responses to two decimal places.*)

When the districts were divided into districts with small classes (< 20 students per teacher) and large classes (≥ 20 students per teacher), the following results were found:

Class Size	Average Score ($\overline{Y}$)	Standard Deviation (s_Y)	*n*
Small	670.5	19.8	230
Large	663.0	18.3	187

Is there statistically significant evidence that the districts with smaller classes have higher average test scores?

The *t*-statistic for testing the null hypothesis is ______ . (*Round your response to two decimal places.*)

The *p*-value for the test is ______ . (*Round your response to six decimal places.*) **Hint:** Use the Excel function Norm.S.Dist to help answer this question.

Is there statistically significant evidence that the districts with smaller classes have higher average test scores?

The (1) ______ suggests that the null hypothesis (2) ______ with a high degree of confidence. Hence, (3) ______ statistically significant evidence that the districts with smaller classes have higher average test scores.

(1) ○ large p-value
○ small p-value

(2) ○ cannot be rejected
○ can be rejected

(3) ○ there is no
○ there is

Answers 657.2

661

4.01

0.00003

(1) small p-value

(2) can be rejected

(3) there is

ID: Exercise 3.13

25. An efficiency manager is evaluating the quality of sportswear manufactured by two brands - *A* and *B*. She wishes to test whether the difference between the mean number of faulty pieces supplied by the brands in each of their consignments is zero or not. For this, the manager collects data from the two brands. Suppose that the mean number of faulty pieces supplied by the brands are independent of each other.

Let *n* denote the number of consignments for which the data was collected, $\overline{X}$ denote the sample mean of the number of faulty pieces supplied per consignment by each brand, and *S* denote the sample standard deviation in the number of faulty pieces supplied per consignment by each brand. The given table shows the data collected.

Brand	n	$\overline{X}$	S
A	110	15.21	11.2
B	85	10.11	14.1

The manager's null hypothesis would be that the difference between the mean number of faulty pieces supplied per consignment by brands *A* and *B* is (1) __________.

The manager's alternative hypothesis would be that the difference between the mean number of faulty pieces supplied per consignment by brands *A* and *B* is (2) __________.

The standard error of the difference between the sample means is: [ ].

(*Round your answer to two decimal places.*)

The *t*-statistic for testing the null hypothesis is [ ].

(*Round your answer to two decimal places.*)

Given the value of the *t*-statistic, for a prespecified 1% level of significance, we (3) __________ the null hypothesis.

(1) ○ equal to zero
○ not equal to zero
○ less than zero
○ greater than zero

(2) ○ greater than zero
○ less than zero
○ equal to zero
○ not equal to zero

(3) ○ reject
○ fail to reject

Answers (1) equal to zero

(2) not equal to zero

1.87

2.73

(3) reject

ID: Concept Exercise 3.4.1

26. Two shooters, Rodney and Philip, practice at a shooting range. They fire 100 rounds each at separate targets. The targets are marked with circles and each bullet hitting a particular circle gets them a particular number of points. 50 rounds are selected at random. The sample mean scores of Rodney and Philip are 8 and 7, respectively. And, their variances are 0.49 and 0.36, respectively.

The standard error of the difference between their mean scores is [ ].

(*Round your answer to two decimal places*.)

The 95% confidence interval for the difference between the mean scores of Rodney and Philip is: [[ ], [ ]].

(*Enter a minus sign if your answer is negative*. *Round your answer to two decimal places*.)

Answers 0.13

0.75

1.25

ID: Concept Exercise 3.4.2

---

27. Suppose you have a binary randomized controlled experiment designed to measure the causal effect of X on Y. Let the control group be identified by $X = 0$ and the treatment group by $X = 1$.

The differences–of–means estimator, denoted by $E(Y|X = 1) - E(Y|X = 0)$, is an estimator of the causal effect of X on Y because the differences–of–means estimator estimates:

- **A.** the expectation of Y for observations within the treatment group.
- **B.** the change in the expectation of Y as result of an observation being in the treatment group as opposed to the control group.
- **C.** the expectation of Y for observations within the control group.
- **D.** All of the above.

Answer: B.
the change in the expectation of Y as result of an observation being in the treatment group as opposed to the control group.

ID: Review Concept 3.7

---

28. Grades on a standardized test are known to have a mean of 1050 for students in the United States. The test is administered to 433 randomly selected students in Florida; in this sample, the mean is 1063.65 and the standard deviation (s) is 113.40.

The 95% confidence interval for the average test score for Florida students is (____ , ____). (*Round your responses to two decimal places.*)

Is there statistically significant evidence that Florida students perform differently than other students in the United States?

The 95% confidence interval for the average test score for Florida students (1) ____ μ = 1050, so the null hypothesis that μ = 1050 (that Florida students have the same average performance as other students in the United States) (2) ____ at the 5% level.

Another 483 students are selected at random from Florida. They are given a 3-hour preparation course before the test is administered. Their average test score is 1069.95 with a standard deviation of 99.75.

The 95% confidence interval for the change in average test score associated with the prep course is (____ , ____). (*Round your responses to two decimal places.*)

Is there statistically significant evidence that the prep course helped?

- **A.** No, because the 95% confidence interval for the change in average test score associated with the prep course includes $\mu_{prep} - \mu_{non-prep}$ = 0.
- **B.** Yes, because the 95% confidence interval for the change in average test score associated with the prep course does not include $\mu_{prep} - \mu_{non-prep}$ = 0.
- **C.** Yes, because the 95% confidence interval for the change in average test score associated with the prep course includes $\mu_{prep} - \mu_{non-prep}$ = 20.20.
- **D.** No, because the 95% confidence interval for the change in average test score associated with the prep course does not include $\mu_{prep} - \mu_{non-prep}$ = 20.20.

The original 433 students are given the prep course and then are asked to take the test a second time. The average change in their test scores is 9.45 points, and the standard deviation of the change is 63.0 points.

Let X denote the change in the test score. The 95% confidence interval for the change in average test scores (μ_X) is (____ , ____). (*Round your responses to two decimal places.*)

Is there statistically significant evidence that students will perform better on their second attempt after taking the prep course?

- **A.** Yes, because the 95% confidence interval for the change in average test scores does not include μ_X = 9.45.
- **B.** No, because the 95% confidence interval for the change in average test scores includes μ_X = 9.45.
- **C.** Yes, because the 95% confidence interval for the change in average test scores does not include μ_X = 0.
- **D.** No, because the 95% confidence interval for the change in average test scores includes μ_X = 0.

(1) ○ includes
○ does not include

(2) ○ can be rejected
○ cannot be rejected

Answers 1052.97

1074.33

(1) does not include

(2) can be rejected

−7.60

20.20

A.
No, because the 95% confidence interval for the change in average test score associated with the prep course includes $\mu_{prep} - \mu_{non-prep} = 0$.

3.52

15.38

C. Yes, because the 95% confidence interval for the change in average test scores does not include $\mu_x = 0$.

ID: Exercise 3.16

29. The following table contains data for full-time, full-years workers, age 25-34, with a high school diploma or B.A./B.S. as their highest degree for the years 1992 and 2008. Download the data from the tables by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here . Use a statistical package of your choice to answer the following questions.

Compute the sample mean for average hourly earnings (AHE) in 1992.

The sample mean for average hourly earnings (AHE) in 1992 is ____.

(*Round your response to three decimal places*)

Compute the sample mean for average hourly earnings (AHE) in 2008.

The sample mean for average hourly earnings (AHE) in 2008 is ____.

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the population mean of AHE in 1992.

The 95% confidence interval is [____, ____]

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the population mean of AHE in 2008.

The 95% confidence interval is [____, ____]

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the population mean of the change in AHE between 1992 and 2008. (*2008 AHE – 1992 AHE)*

The 95% confidence interval is [____, ____]

(*Round your response to three decimal places*)

In 2008, the value of the Consumer Price Index (CPI) was 215.2. In 1992, the value of the CPI was 140.3. Answer the following questions using AHE measured in real 2008 dollars. That is, adjust the 1992 data for the price inflation that occurred between 1992 and 2008.

Compute the sample mean for the inflation-adjusted average hourly earnings (AHE) in 1992.

The sample mean for the inflation-adjusted average hourly earnings (AHE) in 1992 is ____.

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the population mean of inflation adjusted AHE in 1992.

The 95% confidence interval is [____, ____]

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the population mean of the change of the inflation-adjusted AHE between 1992 and 2008. (*2008 AHE – 1992 AHE)*

The 95% confidence interval is [____, ____]

(*Round your response to three decimal places*)

If you were interested in the change in workers' purchasing power from 1992 to 2008, which results would you use?

- **A.** The results using the inflation-adjusted data.
- **B.** The results using the original data.

Use the 2008 data to construct a 95% confidence interval for the population mean of AHE for high school graduates.

The 95% confidence interval is [____, ____]

(*Round your response to three decimal places*)

Use the 2008 data to construct a 95% confidence interval for the population mean of AHE for workers with a college degree.

The 95% confidence interval is [,]

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the mean of the difference in AHE for workers that are high school graduates and those with a colle
degree in 2008 (*college AHE – high school AHE).*

The 95% confidence interval is [,]

Round your response to three decimal places)

Use the 1992 data expressed in $2008 to construct a 95% confidence interval for the population mean of AHE for high school graduates.

The 95% confidence interval is [,]

Round your response to three decimal places)

Use the 1992 data expressed in $2008 to construct a 95% confidence interval for the population mean of AHE for workers with a college degree.

The 95% confidence interval is [,]

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the mean of the difference in AHE for workers that are high school graduates and those with a colle
degree in 1992 using the data expressed in $2008 (*college AHE – high school AHE).*

The 95% confidence interval is [,]

(*Round your response to three decimal places*)

Consider the change in the mean of real (inflation-adjusted) wages of high school graduates and college graduates between 1992 and 200
answer the following questions. (*Check all that apply.)*

- [ ] **A.** Real wages of college graduates appears to have increased from 1992 to 2008.
- [ ] **B.** Real wages of high school graduates appears to have increased from 1992 to 2008.
- [ ] **C.** The gap between earnings of college and high school graduates appears to have increased from 1992 to 2008.
- [ ] **D.** None of the above are correct.

Answers 11.726

18.536

10.492

12.960

16.447

20.625

4.383

9.237

17.986

16.093

19.879

−2.269

3.369

A. The results using the inflation-adjusted data.

12.394

17.200

21.070

27.216

5.445

13.247

12.933

15.691

19.650

26.894

5.085

12.835

A. Real wages of college graduates appears to have increased from 1992 to 2008. , B.
Real wages of high school graduates appears to have increased from 1992 to 2008. , C.
The gap between earnings of college and high school graduates appears to have increased from 1992 to 2008.

ID: Empirical Exercise 3.1

30. Open the Excel data set, CPS92_08[1]. A detailed description of the data is given in the file, CPS92_08_Description[2]. The Excel file is very large and we do not need to work with such a large data set at this point. Thus, take the first 300 observations for 1992 (cells A2:E301) and the first 300 observations for 2008 (cells A7607:E7906). Suggestion: copy the larger data set to another worksheet and then delete the cells that are not needed. Using this smaller version of CPS92_08 answer the following questions:

The mean for the average hourly earnings (*AHE*) in 1992 and 2008 are, respectively:

- **A.** \$11.63 and \$18.98
- **B.** \$12.09 and \$19.12
- **C.** \$10.85 and \$17.88
- **D.** \$18.98 and \$11.63

The 95% confidence interval for the population mean of *AHE* in 1992 is

- **A.** 9.22 - 11.88
- **B.** 11.45 - 12.74
- **C.** 11.01 - 13.02
- **D.** 10.85 - 12.04

The 95% confidence interval for the population mean of *AHE* in 2008 is

- **A.** 17.50 - 22.10
- **B.** 18.15 - 21.05
- **C.** 17.55 - 20.25
- **D.** 18.02 - 20.22

The Consumer Price Index (CPI) was 215.2 in 2008. The CPI was 140.3 in 1992. Convert the *AHE* in 1992 to 2008 dollars. That is, increase *AHE* in 1992 to what they would be in 2008 based solely on inflation. The mean of *AHE* in 1992 expressed in 2008 dollars is.

- **A.** \$19.06
- **B.** \$18.95
- **C.** \$18.55
- **D.** \$17.55

Get the difference in the means of *AHE* in 2008 and 1992 adjusted for inflation. In other words, get the difference in *AHE*s when both are expressed in 2008 dollars.
Use a two-tailed test to determine whether the population mean of *AHE* adjusted for inflation is different in 1992 and 2008 at the 5% level of significance. Use equation 3.19 in your text to compute the standard error of the difference in means. The null hypothesis is : H_0: $\text{MeanAHE}_{2008} - \text{MeanAHE}_{aj,1992} = 0$

- **A.** The difference, \$0.57; it's NOT statistically different from 0 at the 5% level
- **B.** The difference, \$6.96; it's NOT statistically different from 0 at the 5% level
- **C.** The difference, \$0.57; it's statistically different from 0 at the 5% level
- **D.** The difference, \$6.96; it's statistically different from 0 at the 5% level

The 95% confidence interval for the population difference in the means of *AHE* in 2008 and 1992 used in part (5) is:

- **A.** -0.05 - 1.05
- **B.** - 0.91 - 2.05
- **C.** -0.18 - 1.32
- **D.** 0.20 - 0.95

The difference in the population mean of *AHE* for those with and without a bachelor's degree in 2008 is

- A. $4.87
- B. $6.97
- C. $5.87
- D. $0.52

Obtain the 95% confidence interval for the difference in the population means of *AHE* for those with and without a bachelor's degree compu
in part (7). With the confidence interval, test the following,
H_0: MeanAHE_{bach} - MeanAHE_{nobach} = $2.0

- A. The 95% CI is $3.78- $8.01; do not reject the null
- B. The 95% CI is $1.78- $8.01; do not reject the null
- C. The 95% CI is $4.78- $7.01; reject the null
- D. The 95% CI is $3.78- $8.01; reject the null

Compute the following four sums: $\sum_i \left(X_i - \bar{X}\right)$, $\sum_i \left(X_i - \bar{X}\right)^2$ $\sum_i \left(Y_i - \bar{Y}\right)$ $\sum_i \left(Y_i - \bar{Y}\right)^2$ where X is age and Y is average hourly of earnin
in 2008; The four sums are,

- A. -2.84, 2431, 2.9, 27795
- B. None of the above
- C. 0, 2541, 0, 3221
- D. 0, 2431, 0, 27795

Using Excel compute the following quantity $\sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$ Use this sum along with the sums in part (9) to compute sample covarian
and correlation. They are

- A. 20.22 and 0.45
- B. 6.42 and 0.23
- C. -20.22 and -0.45
- D. -6.42 and -0.23

1: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/cps92_08.xlsx

2: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/CPS92_08_Description.pdf

Answers B. $12.09 and $19.12

B. 11.45 - 12.74

D. 18.02 - 20.22

C. $18.55

A. The difference, $0.57; it's NOT statistically different from 0 at the 5% level

B. - 0.91 - 2.05

C. $5.87

D. The 95% CI is $3.78- $8.01; reject the null

D. 0, 2431, 0, 27795

B. 6.42 and 0.23

ID: Empirical Exercise 3.1 (static)

31. A consumer is given the chance to buy a baseball card for $1, but he declines the trade. If the consumer is now given the baseball card, will he be willing to sell it for $1? Standard consumer theory suggests yes, but behavioral economists have found that "ownership" tends to increase the value of goods to consumers. That is, the consumer may hold out for some amount more than $1 (for example, $1.20) when selling the card, even though he was willing to pay only some amount less than $1 (for example, $0.88) when buying it. Behavioral economists call this phenomenon the "endowment effect." John List investigated the endowment effect in a randomized experiment involving sports memorabilia traders at a sports-card show.

The following table contains data from 148 randomly selected traders who attended a trading card show in Orlando, Florida in 1998. Traders were randomly given one of two sports collectables, say good A or good B, that had approximately equal market value; those receiving good A were then given the option of trading good A for good B with the experimenter; those receiving good B were given the option of trading good B for good A with the experimenter. (Good A was a ticket stub from the game that Cal Ripken Jr. set the record for consecutive games played, and Good B was a souvenir from the game that Nolan Ryan won his 300th game.)

Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here [3]. Use a statistical package of your choice to answer the following questions.

Suppose that, absent any endowment effect, all the subjects prefer good A to good B. What fraction of the experiment's subjects would you expect to trade the good that they were given for the other good? (*Hint:* Because of random assignment of the two treatments, approximately 50% of the subjects received good A and 50% received good B.)

[ ]% of the subjects would trade the good that they were given for the other good.

(*Express your response as a percentage*)

Suppose that, absent any endowment effect, 50% of the subjects prefer good A to good B, and the other 50% prefer good B to good A. What fraction of the subjects would you expect to trade the good that they were given for the other good?

[ ]% of the subjects would trade the good that they were given for the other good.

(*Express your response as a percentage*)

Suppose that, absent any endowment effect, *X*% of the subjects prefer good A to good B, and the other (100-*X*)% prefer good B to good A. Show that you would expect 50% of the subjects to trade the good that they were given for the other good.

A person will trade if he received good A but prefers good B or he received good B and prefers good A. 50% receive good A, of these (100-*X*)% prefer good B; 50% receive good B, of these *X*% prefer good A. Let x = *X*/100. The expected fraction is:

[[ ] $\times (1 - x)$]+ $(0.5x)$ = [ ]

(*Round your response to two decimal places*)

Using the sports-card data , what fraction of the subjects traded the good they were given?

The fraction of trades is [ ]

(*Round your response to two decimal places*)

Is the fraction statistically significant different from 0.50? Compute the *t*-statistic for this test.

The *t*-statistic for this test is [ ]

(*Round your response to three decimal places*)

Is the fraction statistically significant different from 0.50 at the 10% significance level?

- **A.** Yes.
- **B.** No.

Is there evidence of an endowment effect?

- **A.** Yes.
- **B.** No.

Some have argued that the endowment effect may be present, but that it is likely to disappear as traders gain more trading experience. Half of the experimental subjects were dealers, and the other half were nondealers. Dealers have more experience than nondealers.

Repeat the exercise above for dealers and nondealers.

Using the sports-card data , what fraction of *dealers* traded the good they were given?

The fraction of trades is ____

(*Round your response to four decimal places*)

Is the fraction statistically significant different from 0.50? Compute the *t*-statistic for this test.

The *t*-statistic for this test is ____

(*Round your response to three decimal places*)

Using the sports-card data , what fraction of *nondealers* traded the good they were given?

The fraction of trades is ____

(*Round your response to four decimal places*)

Is the fraction statistically significant different from 0.50? Compute the *t*-statistic for this test.

The *t*-statistic for this test is ____

(*Round your response to three decimal places*)

Is there evidence consistent with the hypothesis that the endowment effect disappears as traders gain more experience at the 1% significance level?

- **A.** Yes.
- **B.** No.

3: More Info

Variable Definitions

Variable	Definition
Trade	= 1 if subject traded the good he was given for other good.
Dealer	= 1 if subject was a dealer.

Answers 50

50

0.50

0.50

0.37

−2.679

A. Yes.

A. Yes.

0.4259

−1.091

0.3043

−2.853

A. Yes.

ID: Empirical Exercise 3.2

32. A randomized control experiment to test the utility of a health drink was conducted between two groups of individuals of the same age, height, and weight. The number of individuals in the control group and the treatment group were 105 and 85 respectively. One group (the treatment group) was asked to consume the health drink for a certain period of time and the other group (the control group) was not permitted to consume the drink for the same interval of time. The objective was to check whether the BMI (Body Mass Index) of those who consume the drink remains the same or increases significantly relative to the control group. Let $\overline{Y}_C$ denote the mean BMI of the control group and let $\overline{Y}_T$ denote the mean BMI of the treatment group. $\overline{Y}_C$ was calculated to be 20 whereas $\overline{Y}_T$ was calculated to be 30. Let the population mean for those who consume the drink be μ_T and the population mean for those who don't consume the drink be μ_C.

Which of the following hypotheses should be used to compare the causal effect of the health drink on the BMI in order to fulfill the experiment's objective?

- **A.** $H_0: \mu_T - \mu_C = 10$ vs. $H_1: \mu_T - \mu_C = 0$.
- **B.** $H_0: \mu_T - \mu_C = 0$ vs. $H_1: \mu_T - \mu_C = 10$.
- **C.** $H_0: \mu_T - \mu_C = 0$ vs. $H_1: \mu_T - \mu_C > 0$.
- **D.** $H_0: \mu_T - \mu_C = 0$ vs. $H_1: \mu_T - \mu_C \neq 0$.

Let n_C and s_C be the number of individuals and the standard deviation of the control group, and let n_T and s_T be the number of individuals and the standard deviation of the treatment group, respectively.

Complete the given table.

Group A (Control group)			**Group B (Treatment group)**			**Difference, Treatment vs. Control**				
$\overline{Y}_C$	n_C	s_C	$\overline{Y}_T$	n_T	s_T	$(\overline{Y}_T - \overline{Y}_C)$	$SE(\overline{Y}_T - \overline{Y}_C)$	95% confidence interval	t^{act}	p–value
20	105	26.42	30	85	35.86	10		(,)		

(*Round your answers to four decimal places.*)

Considering a 5% significance level, what can be inferred from the p–value calculated above?

- **A.** There is a significant causal effect of the health drink on the BMI since p–value < 0.05.
- **B.** There is no significant causal effect of the health drink on the BMI since p–value < 0.05.
- **C.** There is a significant causal effect of the health drink on the BMI since p–value > 0.05.
- **D.** There is no significant causal effect of the health drink on the BMI since p–value > 0.05.

Answers C. $H_0: \mu_T - \mu_C = 0$ vs. $H_1: \mu_T - \mu_C > 0$.

4.6665

0.8537

19.1463

2.1429

0.0161

A. There is a significant causal effect of the health drink on the BMI since p–value < 0.05.

ID: Concept Exercise 3.5.1

33. When testing for differences of means, you can base statistical inference on the:

- **A.** Chi-squared distribution with $(n_m + n_w - 2)$ degrees of freedom.
- **B.** Student *t* distribution in general.
- **C.** Student *t* distribution if the underlying population distribution of *Y* is normal, the two groups have the same variances, and you use the pooled standard error formula.
- **D.** normal distribution regardless of sample size.

Answer: C.

Student *t* distribution if the underlying population distribution of *Y* is normal, the two groups have the same variances, and you use the pooled standard error formula.

ID: Test A Ex 3.6.3

34.

When testing for differences of means, the *t*-statistic $t = \frac{\bar{Y}_m - \bar{Y}_w}{SE\left(\bar{Y}_m - \bar{Y}_w\right)}$, where $SE\left(\bar{Y}_m - \bar{Y}_w\right) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}$ has:

- **A.** cannot be computed unless $n_m = n_w$.
- **B.** a student *t* distribution if the population distribution of *Y* is not normal.
- **C.** a normal distribution even in small samples.
- **D.** a student *t* distribution if the population distribution of *Y* is normal.

Answer: D. a student *t* distribution if the population distribution of *Y* is normal.

ID: Test B Ex 3.6.3

---

35. The local government of country X is committed to curb the consumption of tobacco among youngsters owing to its severe health effects. Previous attempts to regulate tobacco consumption have failed to produce any fruitful results. The only possible way out is to tax the consumption of tobacco. Optimal taxation would require a proper knowledge of how price sensitive the consumers are to tobacco consumption. Data is collected by selecting a small sample of individuals at random. It is known that the price elasticity of tobacco consumption follows a normal distribution. Similar studies carried out earlier report that the price elasticity of tobacco consumption has a mean 0.30. The current survey aims to verify whether the value of the mean population price elasticity is actually 0.30 or not.

The given table summarises the findings from the survey:

Number of individuals surveyed (n)	**The standard deviation of the sample** (s_Y)	**The sample mean of price elasticity** $(\bar{Y})$
23	0.32	0.49

Given that the sample size is relatively small, is it reliable to assume a normal approximation to the t-statistic for hypothesis testing?

- **A.** No, since the t-statistic follows an F distribution in small samples.
- **B.** Yes, since a normal approximation to the t-statistic is independent of the sample size.
- **C.** Yes, a normal approximation to the t-statistic can be used for hypothesis testing, but only with $n - 2$ degrees of freedom.
- **D.** Yes, a normal approximation to the t-statistic is reliable if the population distribution is itself normal and the sample size is greater than 15.

Let μ_Y denote the average price elasticity of tobacco consumption among youngsters in the entire country.

Based on the above information, the t-statistic for the test $H_0: \mu_Y = 0.30$ vs. $H_1: \mu_Y \neq 0.30$ is: ______.

(*Round your answer to two decimal places*).

Given the value of the t-statistic, at the 5% level of significance, we (1) ______ the null hypothesis.

The 95% confidence interval for the mean price elasticity of tobacco consumption among youngsters in the entire country is:
(______, ______).

(*Round your answer to two decimal places*).

(1) fail to reject
reject

Answers D.

Yes, a normal approximation to the t-statistic is reliable if the population distribution is itself normal and the sample size is greater than 15.

2.71

(1) reject

0.35

0.63

ID: Concept Exercise 3.6.1

36. Two Junior Reserve Officers' Training Corps (JROTC) cadet groups, *A* and *B*, are training for an interschool rifle shooting competition. Their coach measures their performance every day and awards them certain points according to how they may have fared on that day (A_i and B_i for groups *A* and *B*, respectively). The coach knows the average performance of both the cadet groups. The variance of the performances of both the groups is the same. He wants to forecast their pooled variance to gauge their chances at the competition. Suppose, we select 75 cadets from group *A* and 90 cadets from group *B* at random. So, the number of cadets in each group, denoted by n_a and n_b, would be 75 and 90, respectively. Let the average number of points the random sample of cadets from the groups *A* and *B* win be $\overline{A} = 18$ and $\overline{B} = 11$, respectively.

Which of the following formulas would the coach use to estimate the pooled variance, s^2_{pooled}?

- **A.** $s^2_{pooled} = \dfrac{1}{75 + 90 - 2}\left[\sum_{i=1}^{75} \left(A_i - 18\right)^3 + \sum_{i=1}^{90} \left(B_i - 11\right)^3\right].$
- **B.** $s^2_{pooled} = \dfrac{1}{75 + 90 - 2}\left[\sum_{i=1}^{75} \left(18 - A_i\right)^3 + \sum_{i=1}^{90} \left(11 - B_i\right)^3\right].$
- **C.** $s^2_{pooled} = \dfrac{1}{75 + 90}\left[\sum_{i=1}^{75} \left(A_i - 18\right)^2 + \sum_{i=1}^{90} \left(B_i - 11\right)^2\right].$
- **D.** $s^2_{pooled} = \dfrac{1}{75 + 90 - 2}\left[\sum_{i=1}^{75} \left(A_i - 18\right)^2 + \sum_{i=1}^{90} \left(B_i - 11\right)^2\right].$

Suppose the variances of the performances of both the cadet groups are not same.

Which of the following statements would be correct in this case? (*Check all that apply.*)

- **A.** The null distribution of the pooled *t*-statistic is not a Student *t* distribution, even if the data are normally distributed.
- **B.** The pooled variance estimator would give biased and inconsistent results.
- **C.** The pooled variance estimator would give unbiased but inconsistent results.
- **D.** The null distribution of the pooled *t*-statistic is a standard normal distribution for large samples.

Answers

D. $s^2_{pooled} = \dfrac{1}{75 + 90 - 2}\left[\sum_{i=1}^{75} \left(A_i - 18\right)^2 + \sum_{i=1}^{90} \left(B_i - 11\right)^2\right].$

A. The null distribution of the pooled *t*-statistic is not a Student *t* distribution, even if the data are normally distributed., B. The pooled variance estimator would give biased and inconsistent results.

ID: Concept Exercise 3.6.2

37. The following statement about the sample correlation coefficient is true:

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

B. $-1 \le r_{XY} \le 1$.

C. $r_{XY} = \frac{s_{XY}^2}{s_X^2 s_Y^2}$.

D. $|r_{XY}| < 1$.

Answer: B. $-1 \le r_{XY} \le 1$.

ID: Test A Ex 3.7.4

---

38. You have collected data on the average weekly amount of studying time (T) and grades (G) from the peers at your college.

Changing the measurement from minutes into hours has the following effect on the correlation coefficient:

A. decreases the r_{TG} by dividing the original correlation coefficient by 60.

B. does not change the r_{TG}.

C. results in a higher r_{TG}.

D. cannot be computed since some students study less than an hour per week.

Answer: B. does not change the r_{TG}.

ID: Test A Ex 3.7.5

---

39. Assume that you have 125 observations on the height (H) and weight (W) of your peers in college. Let $S_{HW} = 68$, $S_H = 3.5$, $S_W = 68$.

The sample correlation coefficient is:

A. 0.286.

B. 0.572.

C. 1.22.

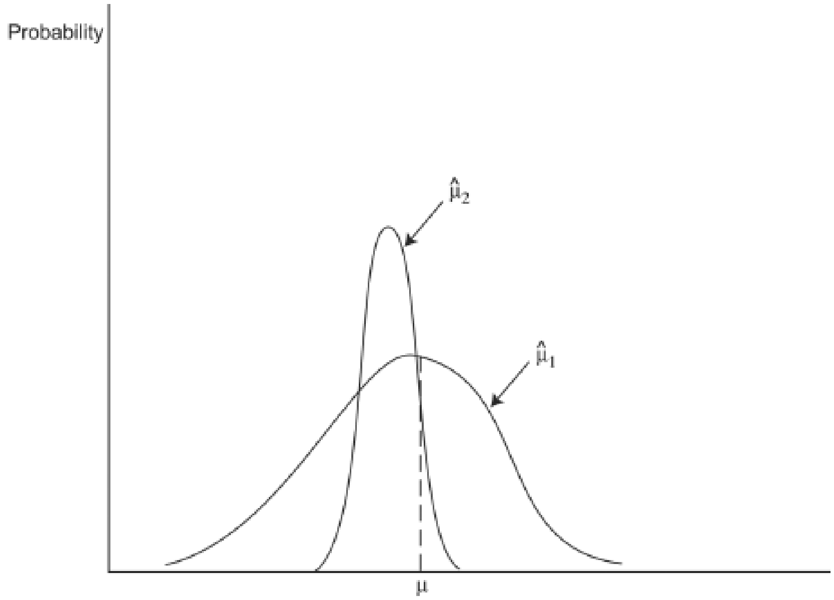
D. Cannot be computed since males and females have not been separated out.

Answer: A. 0.286.

ID: Test B Ex 3.7.4

---

40. A low correlation coefficient implies that:



- A. you should use a tighter scale of the vertical and horizontal axis to bring the observations closer to the line.
- B. the two variables are unrelated.
- C. the line always has a flat slope.
- D. in the scatterplot, the points fall quite far away from the line.

Answer: D. in the scatterplot, the points fall quite far away from the line.

ID: Test B Ex 3.7.5

41. A researcher wants to estimate the relationship between years of educational attainment (X) and the average yearly earnings of individuals (Y). For this, he collects data from a random sample of 109 individuals. From this data, he calculates the sample variances (S_X^2, S_Y^2) and the sample covariance (S_{XY}).

If $S_X^2 = 85.23$, $S_Y^2 = 175.76$, and $S_{XY} = 117.56$, then the sample correlation coefficient will be ______.

(*Round your answer to two decimal places*).

Which of the following statements can be inferred from the value of the sample correlation coefficient calculated above? (*Check all that apply.*)

- ☐ **A.** The points in the scatterplot lie very close to a straight line.
- ☐ **B.** There is a strong linear association between X and Y.
- ☐ **C.** There is a weak linear association between X and Y.
- ☐ **D.** The points in the scatterplot have a steep slope.

Could the researcher make general inferences about the correlation between X and Y from the sample correlation value?

- ○ **A.** Yes, because the sample correlation is a consistent estimator for the population correlation.
- ○ **B.** Yes, because the sample correlation is an unbiased estimator for the population correlation.
- ○ **C.** No, because the sample correlation is not an efficient estimator for the population correlation.
- ○ **D.** No, because the sample correlation value changes from sample to sample.

Answers 0.96

A. The points in the scatterplot lie very close to a straight line., B. There is a strong linear association between X and Y.

A. Yes, because the sample correlation is a consistent estimator for the population correlation.

ID: Concept Exercise 3.7.1

---

42. If the sample correlation between two random variables is zero, then which of the following statements could be the possible reasons behind observing this value? (*Check all that apply.*)

- ☐ **A.** There exists a negative linear relationship between the two variables.
- ☐ **B.** There is no evident relationship between the two variables.
- ☐ **C.** There may exist a non-linear relationship between the two variables.
- ☐ **D.** There exists a positive linear relationship between the two variables.

Considering a large sample, could the researcher make inferences about the population covariance from the sample covariance value?

- ○ **A.** No, because the sample covariance is not an efficient estimator for the population covariance.
- ○ **B.** Yes, because the sample covariance is an unbiased estimator for the population covariance.
- ○ **C.** No, because the value of sample covariance changes from sample to sample.
- ○ **D.** Yes, because the sample covariance is a consistent estimator for the population covariance.

Answers B. There is no evident relationship between the two variables., C. There may exist a non-linear relationship between the two variables.

D. Yes, because the sample covariance is a consistent estimator for the population covariance.

ID: Concept Exercise 3.7.2