# ECON STAT-1 LECTURE NOTES | CHAPTERS 7, 8 & 9

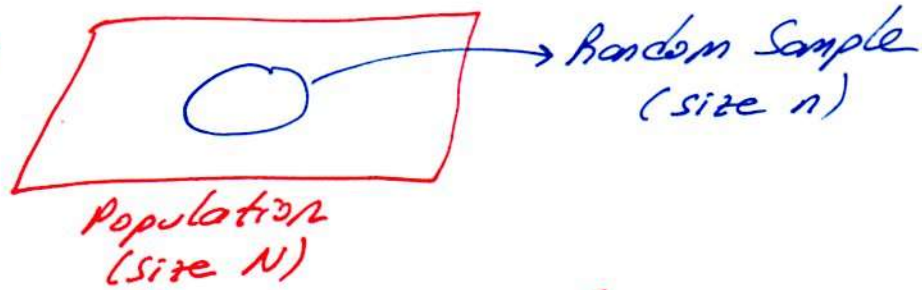
## SAMPLING DISTRIBUTIONS;

* We have 3 sampling distributions to learn.
  (i) MEAN (ii) VARIANCE (iii) PROPORTION

We use sampling distributions to find probabilities about sample statistics.

## SAMPLE vs. POPULATION.

Remember;




| | Population Parameters (Unknown Constants) | Sample statistics (Known Variables) |
|---|---|---|
| Size ↳ Not a parameter or statistics | $N$ | $n$ |
| Mean | $\mu$ | $\bar{X}$ |
| Variance | $\sigma^2$ | $s^2$ |
| Proportion | $p$ | $\hat{p}$ |



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

lecturemania
success maximizer

## SAMPLE MEAN

$$\boxed{\bar{X} = \frac{\sum X_i}{n}}$$

Random Sample: $X_1, X_2, \dots, X_n$

$\mu = E(X_i)$ and $\sigma^2 = Var(X_i)$ for $i = 1, 2, \dots, n$

Then; $E(\bar{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n} E(\sum X_i) = \frac{1}{n} E(X_1 + X_2 + \dots + X_n)$

$= \frac{1}{n}(\mu + \mu + \dots + \mu) = \frac{1}{n} \cdot n \cdot \mu = \mu$

$Var(\bar{X}) = Var\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} \cdot Var(\sum X_i) = \frac{1}{n^2} Var(X_1 + \dots + X_n)$

$= \frac{1}{n^2} \cdot (\sigma^2 + \sigma^2 + \dots + \sigma^2) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$

So; $E(X_i) = \mu$ $\quad Var(X_i) = \sigma^2$

$E(\bar{X}) = \mu$ $\quad Var(\bar{X}) = \frac{\sigma^2}{n}.$

* $X_i \sim Normal(\mu; \sigma^2)$

$\bar{X} \sim Normal\left(\mu; \frac{\sigma^2}{n}\right)$ → Sampling distribution of Normal Variables.

* Single Unit: $Z = \frac{X - \mu}{\sigma}$

Sample Mean; $\boxed{Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}}$

Central Limit Theorem;

Even if $X_i$ does not follow Normal Distribution, for large $n$ (practically $n > 30$) $\bar{X}$ approximately follows normal distribution. So, $Z$-table can be used to find probabilities.



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

7.14 A random sample of 16 junior managers in the offices of corporations in a large city center was taken to estimate average daily commuting time for all such managers. Suppose that the population times have a normal distribution with mean 87 minutes and standard deviation 22 minutes.

a. What is the standard error of the sample mean commuting time?
b. What is the probability that the sample mean is less than 100 minutes?
c. What is the probability that the sample mean is more than 80 minutes?
d. What is the probability that the sample mean is outside the range 85 to 95 minutes?
e. Suppose that a second (independent) random sample of 50 junior managers is taken. Without doing the calculations, state whether the probabilities in parts (b), (c), and (d) would be higher,

7.14) $X \sim Normal(\mu = 87;\ \sigma^2 = 22^2)$

$n = 16$

$\sigma = 22$ → standard Deviation

a) $Var(\bar{X}) = \sigma_{\bar{x}}^2 = \dfrac{22^2}{16}$

$\sigma_{\bar{x}} = \dfrac{22}{\sqrt{16}} = 5,5$ → Standard Error of the sample mean

b) $P(\bar{X} < 100) = P\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \dfrac{100-87}{5,5}\right)$

$= P(Z < 2,36) = 0,5 + 0,4909$

$= 0,9909$

c) $P(\bar{X} > 80) = P\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \dfrac{80-87}{5,5}\right) = P(Z > -1,27)$

$= 0,5 + 0,3980 = 0,8980$

d) $P(85 < \bar{X} < 95) = P\left(\dfrac{85-87}{5,5} < Z < \dfrac{95-87}{5,5}\right)$

$= P(-0,36 < Z < 1,45) = 0,1406 + 0,4265 = 0,5671$

e) (b), (c), (d) will be higher.

7.16 Assume that the standard deviation of monthly rents paid by students in a particular town is \$40. A random sample of 100 students was taken to estimate the mean monthly rent paid by the whole student population.

a. What is the standard error of the sample mean monthly rent?
b. What is the probability that the sample mean exceeds the population mean by more than \$5?
c. What is the probability that the sample mean is more than \$4 below the population mean?
d. What is the probability that the sample mean differs from the population mean by more than \$3?

7.16) $\sigma = 40\ ;\ n = 100$

a) $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{40}{\sqrt{100}} = 4$

b) $P(\bar{X} > \mu+5) = P\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \dfrac{\mu+5-\mu}{4}\right)$

$= P(Z > 1,2) = 0,5 - 0,3849 = 0,1151$

c) $P(\bar{X} < \mu-4) = P\left(Z < \dfrac{\mu-4-\mu}{4}\right)$

$= P(Z < -1) = 0,5 - 0,3413 = 0,1587$

d) $P(\bar{X} > \mu+3) + P(\bar{X} < \mu-3) = 2.P(Z > 0,75) = 2.0,2734 = 0,5468$

7.18 An industrial process produces batches of a chemical whose impurity levels follow a normal distribution with standard deviation 1.6 grams per 100 grams of chemical. A random sample of 100 batches is selected in order to estimate the population mean impurity level.

a. The probability is 0.05 that the sample mean impurity level exceeds the population mean by how much?
b. The probability is 0.10 that the sample mean impurity level is below the population mean by how much?
c. The probability is 0.15 that the sample mean impurity level differs from the population mean by how much?

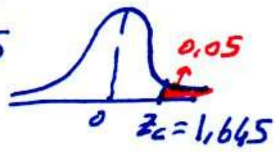
7.18) $X \sim Normal(\mu; \sigma^2 = 1{,}6^2)$

$n = 100$

a) $P(\bar{X} > \mu + \epsilon) = 0{,}05 \Rightarrow \epsilon = ?$

$P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{\mu+\epsilon-\mu}{1{,}6/\sqrt{100}}\right) = 0{,}05$

$P\left(Z > \underbrace{\frac{\epsilon}{0{,}16}}_{=1{,}645}\right) = 0{,}05$




$\epsilon = 0{,}16 \cdot 1{,}645 = 0{,}2632$

b) $P\left(Z < \underbrace{\frac{-\epsilon}{0{,}16}}_{=-1{,}28}\right) = 0{,}10 \Rightarrow \epsilon = 1{,}28 \cdot 0{,}16 = 0{,}2048$

c) $P(\bar{X} > \mu + \epsilon) + P(\bar{X} < \mu - \epsilon) = 0{,}15$

$P(\bar{X} > \mu + \epsilon) = 0{,}075$

$P\left(Z > \underbrace{\frac{\epsilon}{0{,}16}}_{=1{,}44}\right) = 0{,}075 \Rightarrow \epsilon = 1{,}44 \cdot 0{,}16 = 0{,}2304$

7.19 The price-earnings ratios for all companies whose shares are traded on the New York Stock Exchange follow a normal distribution with a standard deviation 3.8. A random sample of these companies is selected in order to estimate the population mean price-earnings ratio.

How large a sample is necessary in order to ensure that the probability that the sample mean differs from the population mean by more than 1.0 is less than 0.10?

7.19) $X \sim Normal(\mu; \sigma^2 = 3{,}8^2)$

$n = ?$

$P(\bar{X} > \mu + 1) + P(\bar{X} < \mu - 1) < 0{,}10$

n should be at least;

$P(\bar{X} > \mu + 1) = 0{,}05$

$P\left(Z > \underbrace{\frac{1}{3{,}8/\sqrt{n}}}_{=1{,}645}\right) = 0{,}05$

$\frac{\sqrt{n}}{3{,}8} = 1{,}645 \quad n = 39{,}07$

$\boxed{n = 40}$ ← Round UP!

# SAMPLE PROPORTION

Let $X$: # of success

$X \sim \text{Binomial}(n, p)$

$E(X) = n.p \; ; \; Var(X) = n.p.(1-p)$

$\hat{p}$: Sample Proportion

$$\hat{p} = \frac{X}{n} \; ; \; E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} \cdot np = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n}$$

Then; $E(\hat{p}) = p \; ; \; Var(\hat{p}) = \dfrac{p(1-p)}{n}$

* Remember; when $np(1-p) > 9$, we can use Normal approximation to binomial distribution.

So; $\hat{p} \sim \text{Normal}\left(\mu_{\hat{p}} = p \; ; \; \sigma_{\hat{p}}^2 = \dfrac{p(1-p)}{n}\right)$

Sampling Distribution of $\hat{p}$.

Using Standard Normal Random Variable $Z$;

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} \; ; \; \boxed{Z = \frac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}}$$



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

7.33 An administrator for a large group of hospitals believes that of all patients 30% will generate bills that become at least 2 months overdue. A random sample of 200 patients is taken.

a. What is the standard error of the sample proportion that will generate bills that become at least 2 months overdue?
b. What is the probability that the sample proportion is less than 0.25?
c. What is the probability that the sample proportion is more than 0.33?
d. What is the probability that the sample proportion is between 0.27 and 0.33?

7.33) $p=0{,}30\ ;\ n=200$

a) $\sigma_{\hat{p}}=\sqrt{\dfrac{p(1-p)}{n}}=\sqrt{\dfrac{0{,}30.0{,}70}{200}}=0{,}0324$

b) $P(\hat{p}<0{,}25)=P\left(\dfrac{\hat{p}-p}{\sigma_{\hat{p}}}<\dfrac{0{,}25-0{,}30}{0{,}0324}\right)$

$=P(Z<-1{,}54)=0{,}5-0{,}4382$

$=0{,}0618$

c) $P(\hat{p}>0{,}33)=P\left(Z>\dfrac{0{,}33-0{,}30}{0{,}0324}\right)=P(Z>0{,}93)$

$=0{,}5-0{,}3238=0{,}1762$

d) $P(0{,}27<\hat{p}<0{,}33)=P(-0{,}93<Z<0{,}93)$

$=2.0{,}3238=0{,}6476$

7.38 A random sample of 100 voters is taken to estimate the proportion of a state's electorate in favor of increasing the gasoline tax to provide additional revenue for highway repairs. What is the largest value that the standard error of the sample proportion in favor of this measure can take?

7.38) $n=100$

$\sigma_{\hat{p}}^{2}=\dfrac{p(1-p)}{100}=\dfrac{\overbrace{0{,}5.0{,}5}^{\text{max. value of } p(1-p)}}{100}$

$\sigma_{\hat{p}}=\dfrac{0{,}5}{10}=0{,}05$

7.39 In Exercise 7.38 above, suppose that it is decided that a sample of 100 voters is too small to provide a sufficiently reliable estimate of the population proportion. It is required instead that the probability that the sample proportion differs from the population proportion (whatever its value) by more than 0.03 should not exceed 0.05. How large a sample is needed to guarantee that this requirement is met?

7.39) $\epsilon=0{,}03$

$P(\hat{p}>p+0{,}03)+P(\hat{p}<p+0{,}03)<0{,}05$

n should be at bost;

$P(\hat{p}>p+0{,}03)=0{,}025$

$P\left(\dfrac{\hat{p}-p}{\sigma_{\hat{p}}}>\dfrac{p+0{,}03-p}{0{,}5/\sqrt{n}}\right)=P\left(Z>\underbrace{\dfrac{0{,}03\sqrt{n}}{0{,}5}}_{=1{,}96}\right)=0{,}025$

round up!

$\sqrt{n}=\dfrac{0{,}5.\ 1{,}96}{0{,}03}=32{,}67\Rightarrow n=1067{,}1\Rightarrow\boxed{n=1068}$

## SAMPLE VARIANCE

$$\boxed{s^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2}$$

* Remember that this was the definition formula.

We have another formula that;

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$ which is calculation formula.

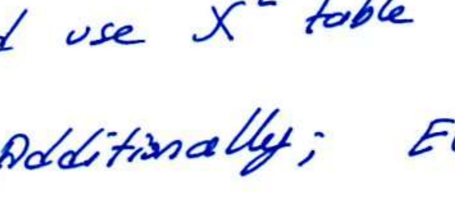
We have;

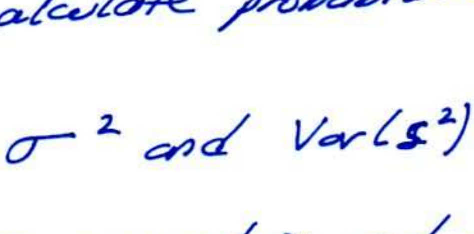
$$\boxed{\frac{(n-1) \cdot s^2}{\sigma^2} \sim \chi^2_{(n-1)}}$$

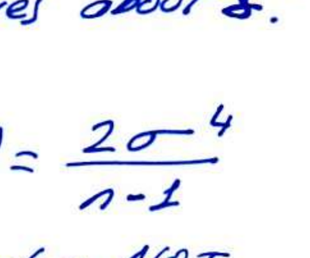
$\chi^2$ (read chi-square) distribution is new for us, with degrees of freedom $n-1$. Here, we make a conversion like $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ ; $\chi^2 = \frac{(n-1) \cdot s^2}{\sigma^2}$ and use $\chi^2$ table to calculate probabilities about $s^2$.

* Additionally; $E(s^2) = \sigma^2$ and $Var(s^2) = \frac{2\sigma^4}{n-1}$

* $\chi^2$ distribution is NOT symmetric, and does NOT have negative values. So, we look two sides seperately






Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

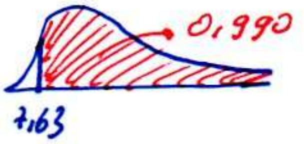
7.53 A mathematics test of 100 multiple-choice questions is to be given to all freshmen entering a large university. Initially, in a pilot study the test was given to a random sample of 20 freshmen. Suppose that, for the population of all entering freshmen, the distribution of number of correct answers would be normal with variance 250.

a. What is the probability that the sample variance would be less than 100?
b. What is the probability that the sample variance would be more than 500?

7.53) $X \sim Normal(\mu; \sigma^2 = 250); n = 20$

a) $P(s^2 > 100) = P\left(\frac{(n-1)s^2}{\sigma^2} > \frac{19.100}{250}\right)$

$= P(\chi^2 > 7,6) \cong 0.990$

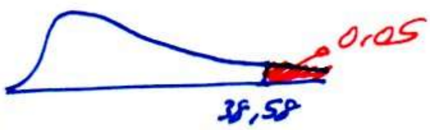



d.of f = $v = n-1 = 19$

Then; $P(s^2 < 100) = 0,010$

b) $P(s^2 > 500)$

$= P\left(\frac{(n-1)s^2}{\sigma^2} > \frac{19.500}{250}\right) = P(\chi^2 > 38) \cong 0,05$



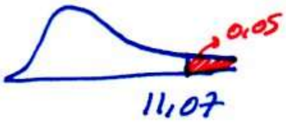

7.57 A production process manufactures electronic components with timing signals whose duration follows a normal distribution. A random sample of six components was taken, and the durations of their timing signals were measured.

a. The probability is 0.05 that the sample variance is bigger than what percentage of the population variance?
b. The probability is 0.10 that the sample variance is less than what percentage of the population variance?

7.57) $X \sim Normal \; ; n = 6 ; d.of f = 5$

a) $P(s^2 > p.\sigma^2) = P\left(\frac{(n-1)s^2}{\sigma^2} > \frac{5p\sigma^2}{\sigma^2}\right)$

$= P(\chi^2 > 5p) = 0,05$



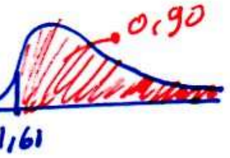

$5p = 11,07$

$p = 2,214$

$= 221,4\%$

b) $P(s^2 < p.\sigma^2) = P\left(\frac{(n-1)s^2}{\sigma^2} < \frac{5p.\sigma^2}{\sigma^2}\right) = P(\chi^2 < 5p) = 0,10$

$P(\chi^2 > 5p) = 0.90$




$5p = 1,61$

$p = 0,322 = 32,2\%$

7.58 A random sample of 10 stock market mutual funds was taken. Suppose that rates of returns on the population of all stock market mutual funds follow a normal distribution.

a. The probability is 0.10 that sample variance is bigger than what percentage of the population variance?

b. Find any pair of numbers, $a$ and $b$, to complete the following sentence: The probability is 0.95 that the sample variance is between $a$% and $b$% of the population variance.

c. Suppose that a sample of 20 mutual funds had been taken. Without doing the calculations, indicate how this would change your answer to part (b).

7.58) $X \sim$ Normal; $n=10$; d.o.f $=9$

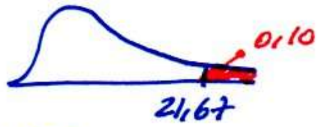
a) $P(s^2 > p.\sigma^2) = P\left(\frac{(n-1).s^2}{\sigma^2} > \frac{9p\sigma^2}{\sigma^2}\right)$

$= P(\chi^2 > 9p) = 0{,}10$

$9p = 14{,}68$

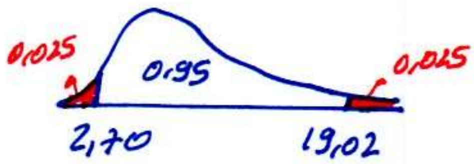
$p = 1{,}6311$

163,11%




b) $P(a.\sigma^2 < s^2 < b.\sigma^2) = P\left(\frac{9a\sigma^2}{\sigma^2} < \frac{(n-1).s^2}{\sigma^2} < \frac{9b\sigma^2}{\sigma^2}\right)$

$= P(9a < \chi^2 < 9b) = 0.95$




$9a = 2{,}70$ $\quad$ $9b = 19{,}02$

$a = 0{,}3$ $\quad$ $b = 2{,}1133$

$a = 30\%$ $\quad$ $b = 211{,}33\%$

c) The interval in part (b) will be smaller because more information would yield more concrete result.

# ESTIMATION : SINGLE POPULATION

* A point estimator $\hat{\theta}$ is an unbiased estimator of $\theta$ if $E(\hat{\theta}) = \theta$

* If $\hat{\theta}$ is a biased estimator of $\theta$, the bias is;

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

* Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of $\theta$. The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is the ratio of their variances.

$$\text{Relative Efficiency} = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

* $\bar{X} = \frac{\sum X_i}{n}$ is unbiased estimator for $\mu$

$s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1}$ is unbiased estimator for $\sigma^2$

$\hat{p} = \frac{\sum X_i}{n}$ is unbiased estimator for $p$

where $X_i = \begin{cases} 1 & \text{if unit is satisfying the condition} \\ 0 & \text{otherwise} \end{cases}$

* Remember;

| | Population | Sample |
|---|---|---|
| Mean | $\mu$ | $\bar{X}$ |
| Variance | $\sigma^2$ | $s^2$ |
| Proportion | $p$ | $\hat{p}$ |

$s_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n}$ is unbiased estimator for $\sigma_{\hat{p}}^2$

$s_{\bar{x}}^2 = \frac{s^2}{n}$ is unbiased estimator for $\sigma_{\bar{x}}^2$



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

8.4 A random sample of 12 blue-collar employees in a large manufacturing plant found the following figures for number of hours of overtime worked in the last month:

22 16 28 12 18 36 23 11 41 29 26 31

Use unbiased estimation procedures to find point estimates for:

a. The population mean
b. The population variance
c. The variance of the sample mean
d. The population proportion of blue-collar employees working more than 30 hours of overtime in this plant in the last month
e. The variance of the sample proportion of blue-collar employees working more than 30 hours of overtime in this plant in the last month

| $X_i$ | $X_i^2$ | $y_i = X_i > 30$ |
|---|---|---|
| 22 | $22^2$ | 0 |
| 16 | $16^2$ | 0 |
| 28 | $28^2$ | 0 |
| 12 | $12^2$ | 0 |
| 18 | $18^2$ | 0 |
| 36 | $36^2$ | 1 |
| 23 | $23^2$ | 0 |
| 11 | $11^2$ | 0 |
| 41 | $41^2$ | 1 |
| 29 | $29^2$ | 0 |
| 26 | $26^2$ | 0 |
| + 31 | $31^2$ | 1 |
| $\Sigma X_i = 293$ | $\Sigma X_i^2 = 8097$ | $\Sigma y_i = 3$ |

$\boxed{n=12}$

a) $\bar{X} = \frac{\Sigma x_i}{n} = \frac{293}{12} = 24,42$

b) $s^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1} = \frac{8097 - \frac{293^2}{12}}{11} = 85,72$

c) $s_{\bar{x}}^2 = \frac{s^2}{n} = \frac{85,72}{12} = 7,14$

d) $\hat{p} = \frac{\Sigma y_i}{n} = \frac{3}{12} = 0,25$

e) $s_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0,25 \cdot 0,75}{12} = 0,015625$

8.6 Suppose that $x_1$ and $x_2$ are random samples of observations from a population with mean $\mu$ and variance $\sigma^2$. Consider the following three point estimators, X, Y, Z, of $\mu$:

$$X = \frac{1}{2}x_1 + \frac{1}{2}x_2 \qquad Y = \frac{1}{4}x_1 + \frac{3}{4}x_2 \qquad Z = \frac{1}{3}x_1 + \frac{2}{3}x$$

a. Show that all three estimators are unbiased.
b. Which of the estimators is the most efficient?
c. Find the relative efficiency of X with respect to each of the other two estimators.

8.6) a) $E(x_i) = \mu$ ; $Var(x_i) = \sigma^2$

$E(X) = E\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$

$E(Y) = E\left(\frac{1}{4}x_1 + \frac{3}{4}x_2\right) = \frac{1}{4}\mu + \frac{3}{4}\mu = \mu$

$E(Z) = E\left(\frac{1}{3}x_1 + \frac{2}{3}x_2\right) = \frac{1}{3}\mu + \frac{2}{3}\mu = \mu$

b) $\sigma_X^2 = \frac{\sigma^2}{4}$ $\sigma_Y^2 = \frac{10\sigma^2}{16}$ $\sigma_Z^2 = \frac{5\sigma^2}{9}$ then $\sigma_X^2$ is min, X is most efficient.

c) $\frac{Var(Y)}{Var(X)} = \frac{10/16}{1/4} = 2,5$ ; $\frac{Var(Z)}{Var(X)} = \frac{5/9}{1/4} = 2,22$

Confidence Interval for Population mean: $\mu$

(i) Population variance: $\sigma^2$ is known;

$100.(1-\alpha)\%$ C.I. for $\mu$ is;

$$\left(\bar{X} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\,;\ \bar{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$$

* Margin of the error: $ME = z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$

Then, confidence interval is $\bar{X} \pm ME$

(ii) Population Variance: $\sigma^2$ is unknown;

$100.(1-\alpha)\%$ C.I. for $\mu$ is;

$$\left(\bar{X} - t_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\,;\ \bar{X} + t_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) \quad d.of\,f = n-1$$

$ME = t_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \Rightarrow \bar{X} \pm ME$ is C.I.

* Note that, as $n$ gets larger, $t$ values come closer to $z$ values. Then, for $n > 30$, namely for large samples, confidence intervals for both cases are very similar.

Confidence interval for Population proportion: $p$

$100.(1-\alpha)\%$ C.I. for $p$ is;

$$\left(\hat{p} - z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,;\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$ME = z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow \hat{p} \pm ME$ is C.I.



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

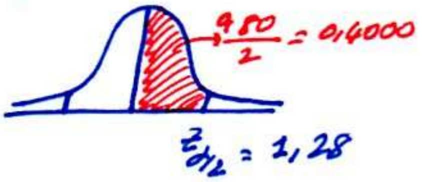
8.13 A personnel manager has found that historically the scores on aptitude tests given to applicants for entry-level positions follow a normal distribution with a standard deviation of 32.4 points. A random sample of nine test scores from the current group of applicants had a mean score of 187.9 points.

a. Find an 80% confidence interval for the population mean score of the current group of applicants.
b. Based on these sample results, a statistician found for the population mean a confidence interval extending from 165.8 to 210.0 points. Find the confidence level of this interval.

8.13) $X \sim Normal(\mu; \sigma^2 = 32{,}4^2)$

a) $n = 9$ ; $\bar{X} = 187{,}9$

$1-\alpha = 80\%$

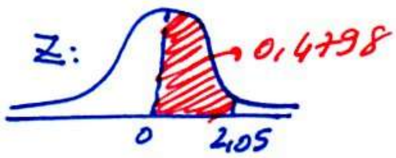



80% C.I. for $\mu$ is; $187{,}9 \pm 1{,}28 \cdot \frac{32{,}4}{\sqrt{9}}$

$(174{,}08 ; 201{,}72)$

b) $187{,}9 - z_{\alpha/2} \cdot \frac{32{,}4}{\sqrt{9}} = 165{,}8$

$z_{\alpha/2} \cdot \frac{32{,}4}{3} = 22{,}1 \Rightarrow z_{\alpha/2} = 2{,}05$




$1-\alpha = 2 \cdot 0{,}4798 = 0{,}9596 \approx 0{,}96 = 96\%$

8.26 There is concern about the speed of automobiles traveling over a particular stretch of highway. For a random sample of seven automobiles radar indicated the following speeds, in miles per hour:

| | | | | | | |
|---|---|---|---|---|---|---|
| 79 | 73 | 68 | 77 | 86 | 71 | 69 |

Assuming a normal population distribution, find the margin of error of a 95% confidence interval for the mean speed of all automobiles traveling over this stretch of highway.

8.26)

| $X_i$ | $X_i^2$ |
|---|---|
| 79 | $79^2$ |
| 73 | $73^2$ |
| ⋮ | ⋮ |
| 69 | $69^2$ |
| $\Sigma X_i = 523$ | $\Sigma X_i^2 = 39321$ |

$\boxed{n = 7}$

$\bar{X} = \frac{523}{7} = 74{,}71$ ; $s^2 = \frac{39321 - 523^2/7}{6} = 40{,}9$ ; d.of f $= n-1 = 6$

$1-\alpha = 0{,}95$

$\alpha/2 = 0{,}025$


95% C.I. for $\mu$ is;

$74{,}71 \pm 2{,}447 \cdot \sqrt{\frac{40{,}9}{7}}$

$(68{,}79 ; 80{,}63)$



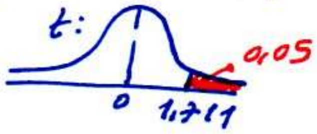
Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

8.28 A business school placement director wants to estimate the mean annual salaries five years after students graduate. A random sample of 25 such graduates found a sample mean of \$42,740 and a sample standard deviation of \$4,780. Find a 90% confidence interval for the population mean, assuming that the population distribution is normal.

8.28) $n = 25;\ \bar{X} = 42{,}74;\ s = 4{,}78$

$1-\alpha = 0{,}90$

$\alpha/2 = 0{,}05;\ d.of\,f = n-1 = 24$




90% C.I. for $\mu$ is;

$42{,}74 \pm 1{,}711 \cdot \frac{4{,}78}{\sqrt{25}}$

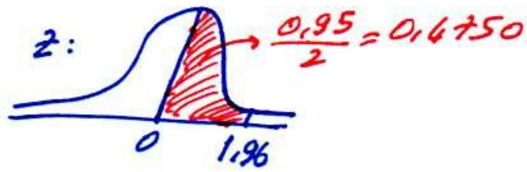
$(41{,}104\ ;\ 44{,}376)$

8.39 In a presidential election year, candidates want to know how voters in various parts of the country will vote. Suppose that 420 registered voters in the Northeast are asked if they would vote for a particular candidate if the election were held today. From this sample 223 indicated that they would vote for this particular candidate. What is the margin of error? Determine the 95% confidence interval estimate of this candidate's support in the Northeast.

8.39) $n = 420\ ;\ X = 223$

$\hat{p} = \frac{223}{420} = 0{,}531$

$s_{\hat{p}} = \sqrt{\frac{0{,}531 \cdot (1-0{,}531)}{420}} = 0{,}0244$




95% C.I. for $p$ is;

$0{,}531 \pm 1{,}96 \cdot 0{,}0244$

$(0{,}483\ ;\ 0{,}579)$

8.38 Of a random sample of 198 marketing students 98 rated a case of résumé inflation as unethical. Based on this information (Reference 2), a statistician computed for the population proportion a confidence interval extending from 0.445 to 0.545. What is the confidence level of this interval?

8.38) $n = 198\ ;\ X = 98\ ;\ \hat{p} = \frac{98}{198} = 0{,}495$

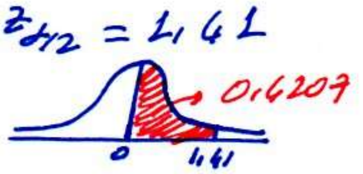
$(1-\alpha) \cdot 100\%$ C.I. for $p$ is;

$(0{,}445\ ;\ 0{,}545)$

$s_{\hat{p}} = \sqrt{\frac{0{,}495 \cdot (1-0{,}495)}{198}} = 0{,}0355$

$0{,}495 + z_{\alpha/2} \cdot 0{,}0355 = 0{,}545$

$z_{\alpha/2} = 1{,}41$




$1-\alpha = 2 \cdot 0{,}4247 = 0{,}8494 \approx 85\%$



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

Confidence Intervals for the difference between two Normal Population Means;

(i) Dependent samples; $d_i = X_i - Y_i$ : Difference

$$\left(\bar{d} - t_{n-1;\alpha/2} \cdot \frac{S_d}{\sqrt{n}} \; ; \; \bar{d} + t_{n-1;\alpha/2} \cdot \frac{S_d}{\sqrt{n}}\right)$$

whee $S_d = \dfrac{\Sigma d_i^2 - (\Sigma d_i)^2/n}{n-1}$

(ii) Independent Samples, Known Population Variances

$$\left((\bar{X}-\bar{Y}) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \; ; \; (\bar{X}-\bar{Y}) + z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$$

(iii) Independent Samples, Unknown Population Variances, assumed Equal

$$\left((\bar{X}-\bar{Y}) - t_{\alpha/2} \cdot \sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}} \; ; \; (\bar{X}-\bar{Y}) + t_{\alpha/2} \cdot \sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}\right)$$

whee $S_p^2 = \dfrac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2}$ ; d.of f $= n_x + n_y - 2$

Confidence Intervals for the Difference between Two Population Proportions;

$$(\hat{p}_x - \hat{p}_y) \pm ME$$

whee $ME = z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \dfrac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

Confidence Intervals for the Variance of a Normal Distribution: $\sigma^2$

$$\left( \frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} ; \frac{(n-1)s^2}{\chi^2_{n-1;(1-\alpha/2)}} \right)$$

SAMPLE SIZE DETERMINATION;

For Mean; with known Variance;

* Remember; $ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Then; $n = \left( \frac{z_{\alpha/2} \cdot \sigma}{ME} \right)^2$

For Population Proportion;

* Remember; $ME = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

max. value of $\hat{p}(1-\hat{p}) = 0,5 \cdot 0,5 = 0,25$

$$ME = z_{\alpha/2} \cdot \sqrt{\frac{0,25}{n}}$$

then; $n = \left( \frac{0,5 \cdot z_{\alpha/2}}{ME} \right)^2$

* If we have some prior knowledge about $p$, such that $0,1 < p < 0,3$; take $p$ with closer bound to $0,5$ ( $p = 0,3$ for this example)

if $0,6 < p < 0,9$, take $p = 0,6 \rightarrow n = 0,6 \cdot 0,4 \cdot \left( \frac{z_{\alpha/2}}{ME} \right)^2$



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

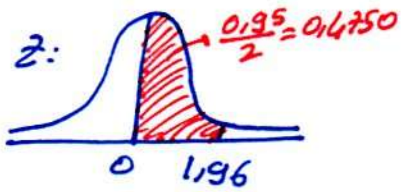
9.6 A manufacturer knows that the numbers of items produced per hour by machine A and by machine B are normally distributed with a standard deviation of 8.4 items for machine A and a standard deviation of 11.3 items for machine B. The mean hourly amount produced by machine A for a random sample of 40 hours was 130 units; the mean hourly amount produced by machine B for a random sample of 36 hours was 120 units. Find the 95% confidence interval for the difference in mean parts produced per hour by these two machines.

9.6)

| Machine A | Machine B |
|---|---|
| $\sigma_A = 8{,}4$ | $\sigma_B = 11{,}3$ |
| $n_A = 40$ | $n_B = 36$ |
| $\bar{X}_A = 130$ | $\bar{X}_B = 120$ |

95% C.I. for $\mu_A - \mu_B$; $\sigma_A^2$ and $\sigma_B^2$ known.

$$(130-120) \pm 1{,}96 \cdot \sqrt{\frac{8{,}4^2}{40} + \frac{11{,}3^2}{36}}$$




$(5{,}483\ ;\ 14{,}517)$

9.5 A random sample of six salespersons that attended a motivational course on sales techniques was monitored in the three months before and the three months after the course. The table shows the values of sales (in thousands of dollars) generated by these six salespersons in the two periods. Assume that the population distributions are normal. Find an 80% confidence interval for the difference between the two population means.

| Salesperson | Before Course | After Course |
|---|---|---|
| 1 | 212 | 237 |
| 2 | 282 | 291 |
| 3 | 203 | 191 |
| 4 | 327 | 341 |
| 5 | 165 | 192 |
| 6 | 198 | 180 |

9.5)

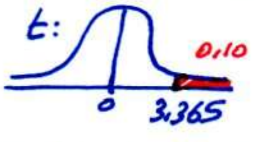
| Salesperson | $d_i$ | $d_i^2$ |
|---|---|---|
| 1 | $212-237 = -25$ | $(-25)^2$ |
| 2 | $282-291 = -9$ | $(-9)^2$ |
| 3 | $203-191 = 12$ | $12^2$ |
| 4 | $327-341 = -14$ | $(-16)^2$ |
| 5 | $165-192 = -27$ | $(-27)^2$ |
| 6 | $198-180 = +18$ | $18^2$ |

$\Sigma d_i = -45$ $\quad \Sigma d_i^2 = 2099$

$$\bar{d} = \frac{-45}{6} = -7{,}5$$

$$s_d^2 = \frac{2099 - \frac{(-45)^2}{6}}{5} = 352{,}3$$

$1-\alpha = 80\% \Rightarrow \alpha/2 = 0{,}10$




d.of f = 6-1 = 5

80% C.I. for $\mu_B - \mu_A$ is;

$$-7{,}5 \pm 3{,}365 \cdot \sqrt{\frac{352{,}3}{6}}$$

$(-33{,}28\ ;\ 18{,}28)$

This is C.I. for decrease in sales. we can also construct C.I. for increase in sales: $\mu_A - \mu_B$ : 7,5 ∓ ME



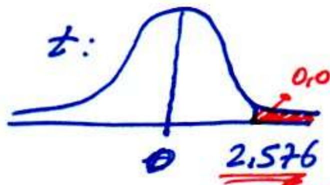
Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

9.14 Suppose, for a random sample of 200 firms that revalued their fixed assets, the mean ratio of debt to tangible assets was 0.517 and the sample standard deviation was 0.148. For an independent random sample of 400 firms that did not revalue their fixed assets, the mean ratio of debt to tangible assets was 0.489 and the sample standard deviation was 0.159. Find a 99% confidence interval for the difference between the two population means.

9.14)

| Revalued Firms | NOT revalued Firms |
|---|---|
| $n_1 = 200$ | $n_2 = 400$ |
| $\bar{X}_1 = 0,517$ | $\bar{X}_2 = 0,489$ |
| $s_1 = 0,148$ | $s_2 = 0,159$ |

$d.\text{of } f = 200 + 400 - 2 = 598$




$1 - \alpha = 0,99$

$\alpha/2 = 0,005$

Note that, this value is the same as for Z table.

$$S_p^2 = \frac{199 \cdot 0,148^2 + 399 \cdot 0,159^2}{598} = 0,0242$$

Assuming $\sigma_1^2 = \sigma_2^2$;

99% C.I. for $\mu_1 - \mu_2$ is;

$$(0,512 - 0,489) \pm 2,576 \cdot \sqrt{\frac{0,0262}{200} + \frac{0,0262}{400}}$$

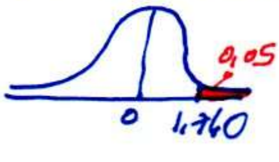
$(-0,0591 \; ; \; 0,0649)$

9.15 A researcher intends to estimate the effect of a drug on the scores of human subjects performing a task of psychomotor coordination. The members of a random sample of 9 subjects were given the drug prior to testing. Their mean score was 9.78, and the sample variance was 17.64. An independent random sample of 10 subjects was used as a control group and given a placebo prior to testing. The mean score in this control group was 15.10, and the sample variance was 27.01. Assuming that the population distributions are normal with equal variances, find a 90% confidence interval for the difference between the population mean scores.

9.15)

| Control Group | Used Drug Group |
|---|---|
| $n_1 = 10$ | $n_2 = 9$ |
| $\bar{X}_1 = 15,10$ | $\bar{X}_2 = 9,78$ |
| $s_1^2 = 27,01$ | $s_2^2 = 17,64$ |

$d.\text{of } f = 10 + 9 - 2 = 17$




$1 - \alpha = 0,90$

$\alpha/2 = 0,05$

Assuming $\sigma_1^2 = \sigma_2^2$;

90% C.I. for $\mu_1 - \mu_2$ is;

$$S_p^2 = \frac{9 \cdot 27,01 + 8 \cdot 17,64}{17} = 22,6$$

$$(15,10 - 9,78) \pm 1,760 \cdot \left(\sqrt{\frac{22,6}{10} + \frac{22,6}{9}}\right)$$

$(1,52 \; ; \; 9,12)$

Note that, the 90% C.I does NOT contain 0. This means that drug significantly decreases (because $\mu_1 - \mu_2$ is increase) psychomotor ~~condition~~ coordination.



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

9.20 "Would you use the library more if the hours were extended?" From a random sample of 138 freshmen, 80 indicated that they would use the school's library more if the hours were extended. In an independent random sample of 96 sophomores, 73 responded that they would use the library more if the hours were extended. Estimate the difference in proportion of first-year and second-year students responding affirmatively to this question. Use a 95% confidence level.

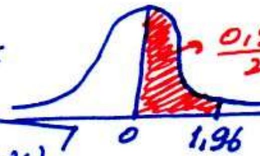
9.20) Freshman Sophomore

$X_1 = 80 \qquad X_2 = 73$

$n_1 = 138 \qquad n_2 = 96$

$\hat{p}_1 = \frac{80}{138} = 0{,}580 \qquad \hat{p}_2 = \frac{73}{96} = 0{,}760$

95% C.I. for $p_1 - p_2$ is; Z:




$$(0{,}580 - 0{,}760) \pm 1{,}96 \cdot \sqrt{\frac{0{,}58 \cdot (1-0{,}58)}{138} + \frac{0{,}76 \cdot (1-0{,}76)}{96}}$$

$(-0{,}300 \; ; \; -0{,}629)$

9.21 A random sample of 100 men contained 61 in favor of a state constitutional amendment to retard the rate of growth of property taxes. An independent random sample of 100 women contained 54 in favor of this amendment. The confidence interval

$$0.04 < P_x - P_y < 0.10$$

was calculated for the difference between the population proportions. What is the confidence level of this interval?

9.21) Men Woman

$X_1 = 61 \qquad X_2 = 54$

$n_1 = 100 \qquad n_2 = 100$

$\hat{p}_1 = \frac{61}{100} = 0{,}61 \qquad \hat{p}_2 = \frac{54}{100} = 0{,}54$

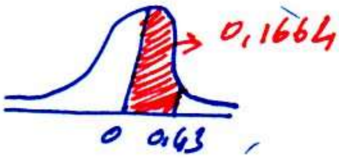
$(1-\alpha) \cdot 100\%$ C.I. is

$(0{,}04 \; ; \; 0{,}10)$

$$(0{,}61 - 0{,}54) - z_{\alpha/2} \cdot \sqrt{\frac{0{,}61 \cdot (1-0{,}61)}{100} + \frac{0{,}54 \cdot (1-0{,}54)}{100}} = 0{,}10$$

$$0{,}0697 \, z_{\alpha/2} = 0{,}03$$

$$z_{\alpha/2} = 0{,}43$$




$1 - \alpha = 2 \cdot 0{,}1664 = 33{,}3\%$

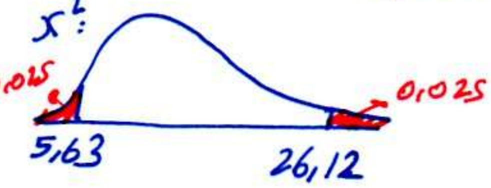
9.30 A manufacturer is concerned about the variability of the levels of impurity contained in consignments of raw material from a supplier. A random sample of 15 consignments showed a standard deviation of 2.36 in the concentration of impurity levels. Assume normality.

a. Find a 95% confidence interval for the population variance.
b. Would a 99% confidence interval for this variance be wider or narrower than that found in part (a)?

9.30) a) $n = 15$ ; $s^2 = 2{,}36^2$ ; $1-\alpha = 0{,}95$

$\alpha/2 = 0{,}025$; $1-\alpha/2 = 0{,}975$

d.of f = 14




95% C.I. for $\sigma^2$ is;

$$\left(\frac{14 \cdot 2{,}36^2}{26{,}12} ; \frac{14 \cdot 2{,}36^2}{5{,}63}\right)$$

$$(2{,}99 ; 13{,}85)$$

9.31 A manufacturer bonds a plastic coating to a metal surface. A random sample of nine observations on the thickness of this coating is taken from a week's output, and the thicknesses (in millimeters) of these observations are as follows:

19.8 21.2 18.6 20.4 21.6 19.8 19.9 20.3 20.8

Assuming normality, find a 90% confidence interval for the population variance.

| $X_i$ | $X_i^2$ |
|---|---|
| 19,8 | $19{,}8^2$ |
| 21,2 | $21{,}2^2$ |
| ⋮ | ⋮ |
| + 20,8 | $20{,}8^2$ |
| $\Sigma X_i = 182{,}4$ | $\Sigma X_i^2 = 3702{,}9$ |

$\boxed{n=9}$

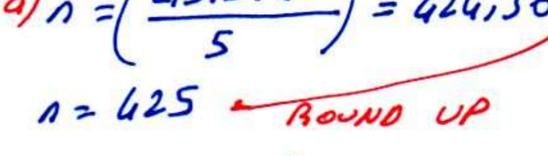
$$s^2 = \frac{3702{,}9 - \frac{182{,}4^2}{9}}{8}$$

$s^2 = 0{,}788$

$1-\alpha = 90\%$

$\alpha/2 = 0{,}05$

d.of f = 8




90% C.I. for $\sigma^2$ is;

$$\left(\frac{8 \cdot 0{,}788^2}{15{,}51} ; \frac{8 \cdot 0{,}788^2}{2{,}73}\right)$$

$$(0{,}320 ; 1{,}820)$$

9.32 How large a sample is needed to estimate the mean of a normally distributed population for each of the following?

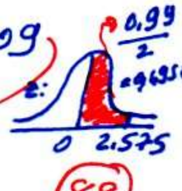
a. $ME = 5$; $\sigma = 40$; $\alpha = 0.01$
b. $ME = 10$; $\sigma = 40$; $\alpha = 0.01$
c. Compare and comment on your answers to parts (a) and (b).

9.32) a) $n = \left(\frac{2{,}575 \cdot 40}{5}\right)^2 = 424{,}36$

$n = 425$ ← ROUND UP

b) $n = \left(\frac{2{,}575 \cdot 40}{10}\right)^2 = 106{,}09$

$n = 107$ ← ROUND UP



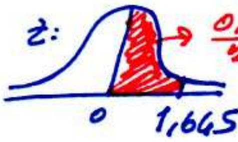

c) ME smaller ⇒ More concrete confidence interval ⇒ Larger n.



Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

9.37 The student government association at a university wants to estimate the percentage of the student body that supports a change being considered in the academic calendar of the university for the next academic year. How many students should be surveyed if a 90% confidence interval is desired and the margin of error is to be only 3%?

9.37)




$$n = \left(\frac{0,5 \cdot 1,645}{0,03}\right)^2 = 751,67$$

$n = 752$ ← Round up

9.46 The supervisor of an orange juice bottling company is considering the purchase of a new machine to bottle 16 fl. oz. (473 mL) bottles of 100% pure orange juice and wants an estimate of the difference in the mean filling weights between the new machine and the old machine. Random samples of bottles of orange juice that had been filled by both machines were obtained. Do the following data indicate that there is a difference in the mean filling weights between the new and the old machines? Discuss assumptions.

| | New Machine | Old Machine |
|---|---|---|
| Mean | 470 mL | 460 mL |
| Standard deviation | 5 mL | 7 mL |
| Sample size | 15 | 12 |

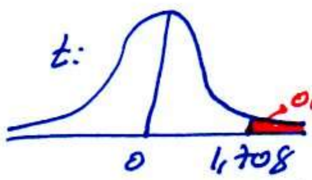
9.46)

| New Mach. | Old Mach. |
|---|---|
| $\bar{X}_1 = 470$ | $\bar{X}_2 = 460$ |
| $s_1 = 5$ | $s_2 = 7$ |
| $n_1 = 15$ | $n_2 = 12$ |

Let $\alpha = 0,10$ ; d.of f $= 15+12-2 = 25$

$1-\alpha = 0,90$

$\alpha/2 = 0,05$




Assuming $\sigma_1^2 = \sigma_2^2$

90% C.I. for $\mu_1 - \mu_2$ is;

$$s_p^2 = \frac{14 \cdot 5^2 + 11 \cdot 7^2}{25} = 35,56$$

$$(470 - 460) \pm 1,708 \cdot \sqrt{\frac{35,56}{15} + \frac{35,56}{12}}$$

$(6,06 \,;\, 13,95)$

The interval for increase using New Machine does NOT contain 0. Therefore, we can say that new machine fills more than old machine at $\alpha = 0,10$.