Statistics for Business **Fall Semester 2023–2024**

Assignment 2

(Release Date: 13/09/2023)

Answer the following questions and hand in your answers by **Friday, September 22, 2023**. These should either be typed and sent electronically or handwritten, scanned and sent electronically (via E-mail at pkonstantinou.aueb@gmail.com). Please make sure that your files are in PDF format and that they do not exceed 5MB in size, otherwise the server might block them!

Sampling Distributions

1. A population has a mean of 160 and a standard deviation of 50. Suppose a simple random sample of size 100 is selected and $\bar{x}$ is used to estimate μ.

   (a) What is the probability that the sample mean will be within ± 4 of lhe population mean?

   (b) What is the probability that the sample mean will be within ± 10 of lhe population mean?

2. The mean annual cost of automobile insurance is €570. Assume that the standard deviation is $\sigma =$€150.

   (a) What is the probability that a simple random sample of automobile insurance policies will have a sample mean within €50 of the population mean for each of the following sample sizes: 30, 50, 100. and 400?

   (b) What is the advantage of a larger sample size when we try to estimate the population mean?

Confidence Intervals

3. The sample mean of monthly food spending was found to be $\bar{x} =$€500 based on a sample of 25 households in Greece. We also know that the sample was randomly drawn from a nornal population with a standard deviation of 15.

   (a) Develop a 99% confidence interval estimate of the mean monthly food expenditure.

(b) How would your answer in (a) change if you change the population standard deviation to 30?

(c) How would your answer in (a) change if you change the population standard deviation to 60? Explain your finding.

(d) How would your answer in (a) change if you change confidence level to 90%? Explain your finding.

4. A simple random sample with $n = 400$ provided a sample mean of 700 and a sample standard deviation of 100.

   (a) Develop a 90% confidence interval for the population mean.

   (b) Develop a 95% confidence interval for the population mean.

   (c) Develop a 99% confidence interval for the population mean.

   (d) What happens to the margin of error and the confidence interval as the confidence level is increased?

5. The mean number of teaching hours for AUEB professors is 32 hours per month. Assume that this mean was based on actual teaching hours for a sample of 49 AUEB professors and that the sample standard deviation was 4.42 hours.

   (a) At 95% confidence, what is the margin of error?

   (b) What is the 95% confidence interval estimate of the population mean **teaching time for the professors**?

   (c) The mean number of teaching hours for professors at UNIPI is 36 hours per month. Use your results from part (b) to discuss differences between the teaching times for the professors at the two universities.

6. Vodafone Inc collected data from the two independent simple random samples of its customers (males and females) about the duration of their calls. In a sample of 36 females it found that the average duration was 40 minutes and in a sample of 49 males it found that the average duration was 35 minutes. Based on data from previous customer demographic studies. the two population standard deviations are assumed known with $\sigma_1 = 9$ minutes and $\sigma_2 = 10$ minutes.

   (a) What is the point estimate of the difference between the mean duration of calls?

   (b) What is the 95% confidence interval for the difference between the two population means?

(c) Suppose that the population variances are equal, that the variances are unknown and that they have been estimated to be $s_1 = 9$ minutes and $s_2 = 10$ minutes. What is the 95% confidence interval for the difference between the two population means?

7. Fifty students who intended to major in engineering were compared with 45 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of a test for the two groups of students

	Verbal		Math	
Engineering	$\bar{x}_1 = 63$	$s_1 = 18$	$\bar{x}_1 = 54$	$s_1 = 12$
Language/literature	$\bar{x}_2 = 60$	$s_2 = 7$	$\bar{x}_2 = 52$	$s_1 = 9$

(a) Construct a 95% confidence interval for the difference in average verbal scores of students majoring in engineering and of those majoring in language/literature.

(b) Construct a 95% confidence interval for the difference in average math scores of students majoring in engineering and of those majoring in language/literature.

(c) Interpret the results obtained in parts (a) and (b).

(d) What assumptions are necessary for the methods used previously to be valid?

Hypothesis Tests

8. A sample of 25 high-school students attending mathematics provided a sample mean of their grade equal to $\bar{x} = 14$ and a sample standard deviation $s = 4.32$. The school director claims that in his experience the mean grade in mathematics is at most 12.

(a) Formulate hypotheses that can be used to challenge the validity of the claim made by the school director.

(b) Compute the test statistic and calculate the approximate p-value for your hypothesis test (use any table for the t distribution).

(c) At $\alpha = 0.05$, what is your conclusion?

(d) What is the rejection rule using the (appropriate) critical value? Whal is your conclusion?

9. The following measures have been obtained on length of time required to complete an assembly procedure using each of two different training methods.

$n_1 = 35$	$n_2 = 40$
$\bar{x}_1 = 13.6$ seconds	$\bar{x}_2 = 10.1$ seconds
$\sum_{i=1}^{35}(x_{1i} - \bar{x}_1)^2 = 919.36$	$\sum_{i=1}^{40}(x_{2i} - \bar{x}_2)^2 = 2817.75$

Is there sufficient evidence to indicate a difference in true mean assembly times for those trained using the two methods? Explain the distribution of the test statistic, and report the p-value for the statistic. At $\alpha = 0.05$ what is your conclusion?

10. A paired-difference experiment was conducted to compare the mean dinner expenditure of individuals when they dine out in Athens and in Thessaloniki (data available in Tab 'Dinner Expenditure' in 2023Assignment2Data.xlsx)

    (a) Do the data provide sufficient evidence to indicate that μ_1 differs from μ_2? Test using $\alpha = .05$.

    (b) Find the approximate p-value for the test and interpret its value.

    (c) Find a 95% confidence interval for $(\mu_1 - \mu_2)$.

    (d) Compare your interpretation of the confidence interval with your test results in part (a).

11. Apple produces iPhone 11 and iPhone 11 Pro. I have collected data from seven retailers, which are summarized below:

Population	1	2	3	4	5	6	7
iPhone 11 :	604.85	615.00	641.50	657.00	667.00	675.40	678.90
iPhone 11 Pro:	719.30	719.87	721.80	795.00	811.80	830.80	835.45

    (a) Do the data provide sufficient evidence to indicate that the mean selling price of iPhone 7 Plus is higher? Test using $\alpha = 0.01$.

    (b) Find the approximate p-value for the test and interpret its value.

    (c) Find a 90% confidence interval for the mean price differential.

12. Two independent random samples of the delay of Aegean and Ryanair flights were drawn (data available in Tab 'Delays (in mins)' in 2023Assignment2Data.xlsx). Assume that both populations are normal.

    (a) Calculate s_p^2, the pooled estimator of σ^2.

    (b) Find a 90% confidence interval for $(\mu_1 - \mu_2)$, the difference between the two population means.

    (c) Test $H_0 : (\mu_1 - \mu_2) \leq 0$ against $H_1 : (\mu_1 - \mu_2) > 0$ for $\alpha = .05$. State your conclusions.

13. Independent random samples of $n_1 = 15$ and $n_2 = 23$ observations were selected from two normal populations with unknown but equal variances:

	X	Y
Sample Size:	15	23
Sample mean:	17.6	24.9
Sample variance:	9	4

(a) Suppose you wish to detect a difference between the population means. State the null and alternative hypotheses for the test.

(b) Find the rejection region for the test in part (a) for $\alpha = .01$.

(c) Find the value of the test statistic.

(d) Calculate the approximate p-value for the test.

(e) Conduct the test and state your conclusions.