Block 4. Statistical inference

PROPOSED PROBLEMS

Estimation

Problem 4.1. Calculate the 0.85confidence interval of a random variable with an unknown mean and a standard deviation of 25 if, in a random sample of 50, elements you obtained a mean of 112. How many elements should the sample have so that the error is equal to or less than 2.5?

Problem 4.2. In a random sample with a size of 10 from a normally distributed population, you obtain a mean of 124 and a variance of 21. Calculate the 90% confidence interval for the population mean.

Problem 4.3. In a sample of 200 second-year students at a University in Madrid, 35% stated that they wanted to work between 16 and 20 hours per week to earn some extra money. Calculate the 95% confidence interval for the proportion of all the students in that year who wanted to work to earn some extra money.

Problem 4.4. A random sample of interest rates for personal loans charged by a bank is $12.8\%, 12.2\%, 13.4\%, 11.9\%, and\ 13\%$. Considering that interest rates are normally distributed with a standard deviation of 0.9%, calculate the confidence interval for the mean of the interest rates.

Problem 4.5. In a random sample of 100 computers, 92 meet the manufacturer's specifications. Calculate the 99.5% confidence interval for the proportion of computers that really meet the manufacturer specifications.

Problem 4.6. The paracetamol content in a sample of 20 painkiller tablets is checked, giving a mean of $22mg$ and a standard deviation of $4mg$. Find a 95% confidence interval for the mean of the variable, assuming it is normally distributed in the population.

Problem 4.7. The 95% confidence interval for a population mean of a normally distributed random variable and a known variance is $[126.4, 132.8]$. Calculate the 98% confidence interval for the mean based on the same sample.

Problem 4.8. A random sample with 28 values of a normally distributed random variable has a standard deviation of 6. Calculate the confidence interval for the population standard deviation at a level of 0.98.

Problem 4.9. The distribution of the scores in a test is known to obey a normal law with a mean of 48 and a standard deviation of 10. If you take 100 samples of 25 people each, between what two values are 95% of the samples obtained ?

Problem 4.10. It is suspected that the number of units contained in each dose of a vaccine is less than the 10000 units indicated on their packaging. The laboratory that produces the vaccine claims that this is the mean content. To check this, you take 100 doses at random and determine the number of units each one contains, obtaining a mean of 9940 units and a standard deviation of 120 units. Assuming that the number of units in each dose is normally distributed, what can you say about the information provided by the laboratory for a confidence level of 99%?

Problem 4.11. You want to analyse the standard deviation of the level of benzocaine per pill of a particular medicine. You take a sample of 16 pills, obtaining a mean benzocaine content of 2.8 units per gram and a deviation of 0.4 units per gram. Obtain the confidence intervals for the standard deviation at the levels of 95% and 99%, assuming that the distribution is normal.

Problem 4.12. What must the sample size be to obtain a 95% confidence interval for a population proportion with a maximum margin of error of 0.04?

Problem 4.13. A given survey is used to try to determine a 95% confidence interval for the proportion of citizens in favour of a points-based driving licence, with a margin of error of less than 2%. What must the size of the sample be?

Problems with two populations

Problem 4.14. During the manufacturing of a plane wing, traction resistance tests are performed on two different grades of aluminium bars. From their experience of this process, the company assumes that the standard deviations of the traction resistances are known. The data obtained are $n_1 = 10$, $\bar{x}_1 = 87.6$, $\sigma_1 = 1$, $n_2 = 12$, $\bar{x}_2 = 74.5$, $\sigma_2 = 1.5$. If μ_1 and μ_2 indicate the true mean traction resistance for the two grades of bars, find the 90% confidence interval for the difference in the mean resistance $\mu_1 - \mu_2$.

A/ $[12.22, 13.98]$

Problem 4.15. Two machines are used to fill plastic bottles with a net volume of 16.0 $ounces$. The filling volume can be assumed as normal, with a standard deviation of $\sigma_1 = 0.020$ and $\sigma_2 = 0.025$ $ounces$. An expert suspects that both machines fill the same mean net volume, whether this is 16.0 $ounces$ or not. A random sample of 10 bottles is taken at the output of each machine. Find

Machine 1	Machine 2
16.03	16.02
16.04	15.97
16.05	15.96
16.05	16.01
16.02	15.99
16.01	16.03
15.96	16.04
15.98	16.02
16.02	16.01
15.99	16.00

a 95% confidence interval above the difference of the means. Provide a practical interpretation of this interval.

A/ $[-0.0098, 0.0298]$. There is no significant difference between the means.

Problem 4.16. Two machines are used to fill plastic bottles with washing-up liquid. The standard deviations of the filling volume are $\sigma_1 = 0.10\ fluid\ ounces$ and $\sigma_2 = 0.15\ fluid\ ounces$ for the two machines, respectively. Two random samples of $n_1 = 12$ bottles from machine 1 and $n_2 = 10$ bottles from machine 2 are selected, and the mean filling volumes of the sample are $\bar{x}_1 = 30.87\ fluid\ ounces$ and $\bar{x}_2 = 30.68\ fluid\ ounces$. Assuming normality:

(a) Construct a 90% two-sided confidence interval for the difference between the means in the filling volume. Interpret this interval.

(b) Construct a 95% two-sided confidence interval for the difference between the means in the filling volume. Compare and comment on the width of this interval and the width of the interval in section (a).

(c) Construct a confidence interval of greater than 95% for the mean difference in filling volume. Interpret this interval.

A/ (a) $[0.0987, 0.2813]$, (b) $[0.0812, 0{,}299]$, (c) $[-\infty, 0.2813]$.

Problem 4.17. An article (published in the journal *Hazardous Waste and Hazardous Materials*, Vol. 6, 1989) reported the results of an analysis of calcium weight in standard cement and lead-doped cement. Reduced calcium levels indicate that the hydration mechanism in the cement is blocked, allowing water to attack various locations in the cement structure. Ten samples of standard cement had an average weight percentage of calcium of $\bar{x}_1 = 90.0$, with a sample standard deviation of $s_1 = 5.0$, while 15 samples of the lead-doped cement had an average weight percentage of calcium of $\bar{x}_2 = 87.0$ with a sample standard deviation of $s_2 = 4.0$.

Assuming that the weight percentage of calcium is normally distributed and that both normal populations have the same standard deviation, find a 95% confidence interval for the difference in the averages, $\mu_1 - \mu_2$, for both types of cement.

A/ $[-0.72, 6.72]$. At this confidence level, you cannot conclude that there is a difference in the means. What does this mean?

Problem 4.18. In a random sample of 85 automotive engine crankshaft bearings, 10 have a surface finish that is coarser than that allowed by the specifications. The surface finish process is modified and then a second random sample of 85 bearings is taken. There are 8defective bearings in the second sample. Therefore, given that $n_1 = 85$, $p_1 = 10/85 = 0.12$, $n_2 = 85$, and $p_2 = 8/85 = 0.09$, obtain a confidence interval of approximately 95% for the difference in the proportion of defective bearings produced in the two processes.

A/ $[-0.06, 0.12]$.

Problem 4.19. In a survey, 22 out of 40 randomly chosen men said that a woman could be president of the country and, independently, 33 of 48 randomly chosen women also said the same. Let p_x and p_y be the proportions of all men and women who said that a woman could be elected president. Find the 95% confidence interval for $p_x - p_y$.

A/ $[-0{,}340, 0{,}065]$.

Problem 4.20. In a study of two types of beans, a random sample with a size of 31, obtained from a normal population X, has a sample variance of $S_X^2 = 42.25$. Independently, from a population Y, another random sample with a size of 41 is obtained, which has a sample variance of $S_Y^2 = 23.04$. Assuming that X and Y are independent, find the 95% confidence interval for σ_X^2/σ_Y^2 in the analysis of the variability of the variances.

A/ $[0.95, 3.69]$.

Problem 4.21. The sample values of a normal random variable X are: 10, 12, 18, 27, and 13. The values of another normal random variable Y are: 23, 24, 31, 26, 28, and 30. Find the 90% confidence interval for σ_X/σ_Y.

A/ $[0.93, 5.29]$.

Contrasting a population

Problem 4.22. An operator cuts a batch of PVC tubes at random. Due to their experience at this task, they claim that the lengths of the tubes are normally distributed and that the mean length is $174\ cm$. A quality inspector wants to check this claim and selects a sample of 25 tubes. The data analysis gives the mean and the sample standard deviation as $\bar{x} = 170\ cm$ and $s = 10\ cm$, respectively. If the inspector performs a contrast analysis with a significance level of $\alpha = 0.05$, does the mean length differ from $174\ cm$?

Problem 4.23. The useful life of a 1.5 V battery is a normally distributed random variable, with a mean of 40 hours and a standard deviation of 4 hours. An engineer decides to introduce a chemical compound to improve the efficiency of the batteries. The manufacturing company wants to know if this change affects the useful life of the batteries. For this purpose, they take a sample of 100 batteries, assume that the standard deviation is 4 hours, and obtain a mean useful life of 39.1 hours.

a. If the check is performed at 95% and 99% confidence, can you say that the mean useful life of the batteries changes?
b. What confidence level is used if the null hypothesis is rejected when the value of the statistic under the null hypothesis does not belong in the interval $[39.5, 40.5]$?

Problem 4.24. To reduce the risk damage, in collisions at speeds of above 10 km/h, a car company designed a new bumper system. In a contrast analysis of 12 cars, the mean speed for this reduction was 8 km/h, with a sample standard deviation of 1.5 km/h.

a. Make the appropriate contrast analysis of the statement at a significance level of 0.05.
b. Is there any significant evidence to reject the statement that at speeds of up to 10 km/h the risk of collision is reduced?

Problem 4.25. In a random sample of 125 beer consumers, 68 said that they can easily tell the difference between non-alcoholic beer and normal beer. You want to find if 50% of beer consumers can distinguish between the two types of beer, against the alternative that this percentage is low, with a significance level of 0.05.

a. Is there any statistical evidence for rejecting the null hypothesis?
b. And if you consider the contrary alternative hypothesis?
c. Without performing the two-tailed test, can you predict the result?

Problem 4.26. A report states that in no university faculty is the number of students with a scholarship greater than or equal to 50%. However, in the fine arts faculty they say that it is. To confirm what the arts faculty says, a random sample of 25 students is selected, finding that 17 of them have a scholarship.

a. Make the contrast and indicate whether the statement by the fine arts faculty can be rejected with a significance level of 0.05.
b. Is the statistical value significant at a level of 0.01 ? Justify your answer.

Problem 4.27. The grades for a subject in a university are normally distributed, with a mean of 73 and a standard deviation of 9. An innovative teaching procedure is implemented to introduce changes in teaching that reduce the variability. For this reason, you want to study the effectiveness of the procedure; you choose a random sample of 51 students that are following the new programme and determine a standard deviation of 7.4.

a. If the analysis is performed with a significance level of 0.05, make the most appropriate contrast analysis to demonstrate that the new programme works. In other words, that the standard deviation is lower.
b. If you use a significance level of 0.01, is the result still the same?

Problem 4.28. The regulations in the mineral water market require that any bottle has, on average, $333\ ml$, with a standard deviation of less than $3\ ml$. A control technician takes a sample of 50 bottles of a given water brand to verify the capacity of the bottles. The results obtained in the sample indicate that the mean capacity is $333{,}682\ ml$ and the standard deviation is $3{,}069\ ml$. To verify that the market regulations are met, make the contrasts on the sample that you consider appropriate.

Problem 4.29. In a new filament manufacturing process, you want to compare if it can be reasonably assumed that the variability in thickness is $4\ mm$. For this purpose, a sample of 28 filaments is taken, which gives a variability of $2\ mm$. Make the requested contrast analysis for a level of $\alpha = 0.05$, assuming normality in the thickness of the filaments.

Problem 4.30. A manufacturer of fire protection sprinkler systems claims that the real average activation temperature of their system is 130°C. A sample $n = 9$ system is tested, giving a sample average activation temperature of 131.08 °F. If the activation times are normally distributed with a standard deviation of 1.5 °F, does the data contradict the manufacturer's claims at a significance level of $\alpha = 0.01$?

Contrasting two populations

Problem 4.31. See the statement for problem 4.15:

a. Is the expert right?
b. What is the P-value of the test?
c. What is the power of the test in part (a) for a true difference in the means of 0.04?
d. Assuming that the samples are the same size, what sample size should you use to guarantee that $\beta = 0.05$ if the true difference in the means is 0.04? Assume that $\alpha = 0.05$.

A/ a. It is not rejected. H_0 b. 0.3222 c. 0.9967 d. Assume that the samples are the same size.

Problem 4.32. 9 mice are selected to check whether a new serum can stop leukaemia at an advanced stage. Out of 9 mice, 5 receive the treatment and the others do not. Once the experiment starts, the survival times, expressed in years, are:

With treatment	2.1	5.3	1.4	4.6	0.9
Without treatment	1.9	0.5	2.8	3.1	

Assuming that both are normally distributed, with known variances, and that the level of significance is 0.05, can you say that the serum is effective?

A/ It is not effective.

Problem 4.33. To improve fuel economy, a taxi company wants to replace radial tyres with bias-belted tyres. To check this, they take 12 taxis and test them with both types of wheels, under the same conditions, and record the fuel consumption in kilometres per litre, as shown in the following table:

km/l

Taxi	Radial tyres	Belted tyres
1	4.2	4.1
2	4.7	4.9
3	6.6	6.2
4	7.0	6.9
5	6.7	6.8
6	4.5	4.4
7	5.7	5.7
8	6.0	5.8
9	7.4	6.9
10	4.9	4.7
11	6.1	6.0
12	5.2	4.9

Assuming that the populations are normally distributed, can you say that the taxis have better fuel economy with radial tyres than belted tyres?

In the conclusions, use a P-value.

A/ Yes

Problem 4.34. In a city, it is found that 20 out of 200 women have breast cancer. In a nearby rural town, 15 out of 150 have the same illness. If you use a significance level of 0.06, can you be sure that the incidence of breast cancer in women is greater in the city than in the rural town?

A/ Yes

Problem 4.35. It is said that the product assembly times for men and women have an approximately normal distribution, but the variance in the times for women is less than for men. To verify this claim, a random sample of times is taken for 11 men and 14 women, with standard deviations of 6.1 and 5.3 for men and women, respectively. If you use a significance level of 0.01, can you say that the assembly times for men and women are the same?

A/ Yes.

Problem 4.36. Trying to determine air quality, you measure the amount of sulphur monoxide it contains. You want to know if the measurements performed with two instruments have the same variability. Some records are taken:

Sulphur monoxide

Instrument A	Instrument B
0.86	0.87
0.82	0.74
0.75	0.63
0.61	0.55
0.89	0.76
0.64	0.70
0.81	0.69
0.68	0.57
0.65	0.53

Assuming that the populations are normally distributed, can you say that the instruments have the same variability?

Use a P-value.

A/ Yes.