Advanced Mathematics – WS2021 – Lab 11 solution

Exercise 1 – bathymetric data

The date set *depths.txt* contains the bathymetric depths in meters for an ocean basin part.

Find the mean depth, the standard deviation and the 95% confidence interval on the mean depth.

Precision to the nearest meter since that's the precision of the given data:

$$\bar{d} = -4393.70\, m$$

$$s_d = 909.53\, m$$

$$n = 261$$

For $\alpha = 0.05$ the z table tells us that the z-values for the tails is ±1.96, so the confidence interval on our sample mean becomes:

$$\mu = \bar{d} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = -4393.70 \pm 110.35cm$$

What is the probability that a random depth measurement will be shallower than -4000m?

We convert -4000m to normal scores using the mean and standard deviation and find z = 0.43. The probability that a value is shallower (i.e., higher z) than that equals the area under the normal curve from z to +infinity, which is 33.4 %.

Exercise 2 – plane booking

A small plane doing the Stuttgart-Frankfort link can accommodate 30 people every day. Statistics show 20% of customers who have booked do not come. Let X be the random variable: "number of customers who are present at the counter among 30 people who have reserved".

What is the law of X (only the general form will be given)? What is its expectation and standard deviation?

The law of X is the binomial law n = 30, p = 0.8, E(X) = 24, s = 2.19.

Give a confidence interval at the 95% threshold, making it possible to estimate the number of customers to expect.

A confidence interval at the 95% threshold, making it possible to estimate the number of customers to be expected: this is for the frequency: 0.657; 0.943. Or between 20 and 28 people. It is a large fork due to n small.

Exercise 3 – water contents of soils

The water contents of soils (in volume %) were measured at two sites A and B around Montpellier, France. There are reproduced in *soilwater.txt*. At the 99% level of confidence, do the soils at the two sites have different water content?

Set H_0 “they have the same content”. Compute $1 - \alpha/2$ for a two-tails test with $\alpha = 0.01$. Use a 99% confidence level to fin the critical t value at 2.609. Now, find the absolute value of the statistic test (1.79) which is less than the critical value. Thus, we accept the hypothesis H_0.

Exercise 4 – toxic algae on beaches

We are interested in the problem of toxic algae that reaches certain beaches in France. After study, we note that 10% of beaches are affected by this type of algae and we want to test the influence of new chemical releases on the appearance of these algae. For that, 50 beaches close to the chemical rejection zones are observed. We then count 10 beaches affected by the harmful algae. Can you answer the question: “With the risk $\alpha = 0.05$, have the chemical releases significantly changed the number of beaches affected?”.

Set H_0 "chemical discharges do not modify the number of beaches reached by algae". Note $p_0 = 0.1$ the theoretical proportion of beaches reached by green algae before chemical releases; p the theoretical proportion of beaches reached by green algae after chemical releases and f the observed frequency in the sample.

Consider then the random variable X_i, i<50, which has two modalities: 1 if the beach is reached, 0 otherwise. It is a Bernoulli variable, so the total number of beaches reached in the sample is a random variable which, with H_0, obeys a binomial distribution of parameters n = 50, $p_0 = 0.1$.

With H_0, "$p = p_0 = 0.1$" the variable "sample mean" is $\overline{X} = \frac{\sum_{i=1}^{i=50} X_i}{n}$. The observed frequency is 10/50, obeys a law that can be approached by a law normal of parameters: mean p_0 and standard deviation $\sqrt{\frac{p_0(1-p_0)}{50}}$.

Now, we determine the associated confidence interval: I ≈ [0.017; 0.183]. We acknowledge that the observed frequency is in the rejection zone (non-chemical): 0.2 is not in the confidence interval at the 95% threshold.

We can therefore reject H_0 and conclude, at risk 0.05, that chemical releases modify in a significant way number of beaches affected by algae.

Exercise 5 – sulfur dioxide emissions

An environmental scientist measures the sulfur dioxide emissions from an industrial plant over an 80-day period. The amounts in tons per day are given in the file *sulfur.txt*.

Bin the data using the categories less than 10, 10-15, 15-20, 20-25, 25 and above. Plot the histogram and indicate the counts.

Region	Interval	Counts
A	0-10	6
B	10-15	15
C	15-20	24
D	20-25	24
E	25+	11

The scientist wonders if the emissions are well described by the expected normal distribution. What are the mean and standard deviation for the raw data?

$$\bar{x} = 18.83$$

$$s \sim \sigma = 5.72$$

$$n = 80$$

The scientist decides to use a χ^2test. What are the expected counts in each bin?

The expected counts per bin equals the area (i.e., probability) under a normal distribution with the same mean and standard deviation within the bin interval multiplied by the total number of counts. We get this table where the areas sum to 1 and the expected counts sum to 80.

Region	Interval	z-range	Area	Expected
A	0-10	-3.29 to -1.54	0.061	4.9
B	10-15	-1.54 to -0.67	0.191	15.2
C	15-20	-0.67 to 0.21	0.330	26.4
D	20-25	0.21 to 1.08	0.278	22.3
E	25+	1.08 to ∞	0.140	11.2

Test whether or not the binned data are indistinguishable from a normal distribution at the 95% level of confidence. Should the scientist reject H_0?

H_0: data is from a normal distribution. With the observed counts given in first question and the expected counts before we find.

$$\chi^2 = \sum_{j=1}^{5} \frac{\left(O_j - E_j\right)^2}{E_j} = 1.37$$

$$v = 5 - 1 - 2 = 2$$

$$\chi^2_{\alpha;2} = 5.99$$

This means we cannot reject the null hypothesis; the data appear to be normally distributed.