TESTING HYPOTHESIS ABOUT MEANS

PROBLEM 1: TESTING HYPOTHESIS ABOUT A POPULATION MEAN

Norepinephrine (also referred to as adrenaline) is a is a hormone and neurotransmitter used as a drug to treat cardiac arrest and other cardiac dysrhythmias resulting in diminished or absent cardiac output. It is found as ampule of 1 ml and contains 1 mg of active substance.

On a given producer, there is some variation from on ampule to another because the filling machinery is not work properly. The distribution of the content is normal with standard deviation of σ = 0.05. A research was done in order to measure the contents of 10 ampules and the results are as follows:

0.94	0.97	1.10	0.99	1.05	0.89	1.02	0.91	0.82	1.04

Is it convincing evidence that the mean contents of Norepinephrine of ampules is less than the advertised 1 ml?

1. What is the H_0?

   H_0: The mean content of Norepinephrine of studied ampules is equal to 1.

2. What is the $H_{1/a}$?

   Two-sided hypothesis: $H_{1/a}$: The mean content of Norepinephrine of studied ampules is equal to 1.
   One-sided hypothesis: $H_{1/a}$: The mean content of Norepinephrine of studied ampules is less than 1.
   One-sided hypothesis: $H_{1/a}$: The mean content of Norepinephrine of studied ampules is higher than 1.

Think about whether your hypotheses are one-sided or two-sided!!!
This will affect the way you calculate probabilities!!

3. What is the sample size?
   - n = 10

1	2	3	4	5	6	7	8	9	10
0.94	0.97	1.10	0.99	1.05	0.89	1.02	0.91	0.82	1.04

4. Find the mean of the sample and its associated confidence interval:
   - m = (0.94+0.97+1.10+0.99+1.05+0.89+1.02+0.91+0.82+1.04)/10 = 0.9730
   - 95% CI = [0.9730 – 1.96*(0.05/sqrt(10)); 0.9730 – 1.96*(0.05/sqrt(10))] = [0.9420; 1.0040]

Hint:
- The formula for arithmetic mean is $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- The formula for 95% confidence interval for mean is: $\overline{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

5. Calculate the z test statistic for this sample (the value of z for the sample allows us to perform a probability calculation to test the significance of this finding against the H_0)

   z = (0.9730-1)/(0.05/sqrt(10)) = -1.7076

Hint: The formula for the z test statistic is: $z = \frac{m-\mu}{\frac{\sigma}{\sqrt{n}}}$

6. Interpret from statistical point of view the obtained result.

Statistical interpretation:

- Critical region $(-\infty; -2.69] \cup [2.96; \infty)$
- Since the calculated z score did not belong to critical region (two-tails test), respectively its value is greater than -2.69 and smaller than +2.69, the H_0 could NOT be rejected.

7. So, is this convincing evidence that the mean contents of ampules are less than the imposed 1 ml?

There is no evidence that the mean content of ampules is less than the imposed 1 ml.

PROBLEM 2: ONE-SAMPLE T TEST

Occupational medicine is a relatively new field in medicine, whereby specific health hazards are identified for particular occupations. One topic of recent interest is the effect of fire fighting on pulmonary function. Suppose a group of 26, 25-35 year-old male fire fighters are identified and change in their pulmonary function over a 5-year period is measured. Over 5 years it is found that the fire fighters have a mean decline in forced expiratory volume (FED, the volume of air expelled in 1 second) of 0.27 liters with a sample standard deviation of 0.32 liters. Can any conclusions be drawn about the occupational exposure if the expected change over 5 years is 0.10 liters in normal male in this age group?

Assume that the decline in FEV is normally distributed with mean μ and unknown variance σ^2.

1. What is the H_0?

    H_0: The mean decline in forced expiratory volume of fire fighters is not different by the mean decline in forced expiratory volume of normal population.

2. What is the $H_{1/a}$?

    $H_{1/a}$ (two-sided): The mean decline in forced expiratory volume of fire fighters is different by the mean decline in forced expiratory volume of normal population.

Think about whether your hypotheses are one-sided or two-sided!!!
This will affect the way you calculate probabilities!!

3. What is the sample size?
    - n = 26

4. Find the mean value of
    - The sample: m = 0.27 liters; s = 0.32 liters
    - The population: μ = 0.10 liters

5. Find a 95% confidence interval for the appropriate parameter. Interpret your result.
    - 95% CI = [0.27-1.96*(0.32/sqrt(26)); 0.27+1.96*(0.32/sqrt(26))] = [0.1470; 0.3930]
    - The formula for 95% confidence interval for mean is: $\overline{X} \pm 1.96 \cdot \frac{s}{\sqrt{n}}$

5. Calculate the t test statistic for this sample:

    t = (0.27-0.10)/(0.32/sqrt(26)) = 2.7089

Hint: The formula for the z test statistic is: $t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$

6. Interpret from statistical point of view the obtained result.

    Critical region for two-sided test: $(-\infty; -2.38\,] \cup [2.38; \infty)$
    Statistical interpretation: H0 is rejected since the t-value belongs to critical region!

7. So, 5 years occupational exposure leads to decrease in the forced expiratory volume compared with reduction in normal male?

    5 years occupational exposure leads to statistically significant decrease in the forced expiratory volume compared with reduction in normal male.

Problem 3: Independent Samples and Known Population Variances – Normal Test

The alkanlinity (mg/L) of water in the upper reaches of Somesului Cald river is known to be normally distributed with a standard deviation of 10 mg/L. Alkalinity readings in the lower reaches of the Somesului Cald river is also known to be normally distributed with a standard deviation of 15 mg/L. Ten alkanility readings are made in the upper reaches of Somesului Cald and twelve in the lower reaches with the following results

Upper (m_1)	90	76	92	87	90	67	80	70	69	78		
Lower (m_2)	120	115	135	87	95	64	68	105	113	79	145	121

Investigate at a 5% significance level the claim that the true alkalinity of water in the lower reaches of the river is greater than that in the upper reaches.

1. What is the H_0?

   H_0: The water alkalinity in the lower reaches of the river is not different by the water alkalinity in the upper reaches.

2. What is the $H_{1/a}$?

   $H_{1/a}$ (one-sided test): The water alkalinity in the lower reaches of the river is greater than the water alkalinity in the upper reaches.

3. Find the mean of alkalinity in upper and lower reaches of the river:
   - m_1 = (90+76+92+87+90+67+80+70+69+78)/10 = 79.90
   - m_2 = (120+115+135+87+95+64+68+105+113+79+145+121)/12 = 103.92

Hint:

- The formula for arithmetic mean is $m = \frac{1}{n}\sum_{i=1}^{n} x_i$

4. Significance level: $\alpha = 0.05$
5. Critical region is: $z \geq 2.09$. If calculated $z \geq 2.09 \rightarrow$ reject H_0.
6. Calculate the z test statistic for this sample (the value of z for the sample allows us to perform a probability calculation to test the significance of this finding against the H_0)

   s_1 = 10 mg/L
   s_2 = 15 mg/L
   m_1 = 79.90 mg/L
   m_2 = 103.92 mg/L

   z = (103.92-79.90)/sqrt(15^2/12+10^2/10) = 4.48

Hint: The formula for the z test statistic is: $z = \frac{(m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

7. Interpret from statistical point of view the obtained result.

   Since the calculated z-value is greater than 2.09 the null hypothesis is rejected.

8. So, alkalinity of water in the lower reaches of the river is greater than that in the upper reaches?

   Alkalinity of water in the lower reaches of the river is greater than that in the upper reaches.

PROBLEM 4: INDEPENDENT SAMPLES AND UNKNOWN BUT EQUAL POPULATION VARIANCES – T-TEST

Suppose a group of 8, 25-35 year-old male smoking fire fighters are identified; this sample have a mean systolic blood pressure of 135 mmHg and a sample deviation of 10 mmHg. A sample of 20 25-35 year-old male non-smoking fire fighters were similarly identified who have mean systolic blood pressure of 125 mmHg and a standard deviation of 10 mmHg. Investigate the difference between means of systolic blood pressure of these two samples.

1. What is the H_0?

   H_0: The difference between means of systolic blood pressure of male smoking fire fighters and non-smoking fire fighters is zero.

2. What is the $H_{1/a}$?

   $H_{1/a}$ (two-sided test): The difference between means of systolic blood pressure of male smoking fire fighters and non-smoking fire fighters is different by zero.

3. What are the sample size?
   - $n_1 = 8$
   - $n_2 = 20$

4. Significance level: $\alpha = 0.05$
   - $m_1 = 135$ mmHg
   - $s_1 = 10$ mmHg
   - $m_2 = 125$ mmHg
   - $s_2 = 10$ mmHg

5. t critical: 2.06
6. Calculate the t test statistic for this sample:

   t = (135-125)/sqrt(10*(1/8+1/20)) = 7.56

Hint: The formula for the z test statistic is: $t = \dfrac{(m_1 - m_2)}{\sqrt{s\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

7. Interpret from statistical point of view the obtained result.

   Since the calculated t-value is greater than 2.06 the null hypothesis is rejected.

8. So, the mean of systolic blood pressure of smoking fire fighters is significantly different compared to the non-smoking fire fighters?

   The mean of systolic blood pressure of smoking fire fighters is significantly different compared to the non-smoking fire fighters.

PROBLEM 5: PAIRED SAMPLES – T-TEST

A study was conduct in order to assess daily therapeutic schemas for treatment of ferriprive anemia (Fe deficiency) in newborn child. There were included into the study 10 breast feed newborn from urban environments. The biochemical expression of the ferriprive anemia is hemoglobin blood level expressed in mg/dl. The data were collected at baseline and after 3 months of treatment and are presented in the table below:

Hemoglobin (mg/dl) baseline	12.1	13.2	10.1	9.2	10.6	12.3	12.3	11.4	10.5	13.1
Hemoglobin (mg/dl) 3 months of treatment	13.2	13.2	13.3	12.1	12.2	13	12.9	12.9	11.2	12.7

Was the treatment efficient?

1. What is the H_0?

   H_0: The mean of hemoglobin at baseline is not statistically different by the mean of hemoglobin after 3 months of treatment.

2. What is the $H_{1/a}$?

   $H_{1/a}$: The mean of hemoglobin at baseline is statistically different by the mean of hemoglobin after 3 months of treatment.

3. Compute the differences between baseline and 3 months determination of hemoglobin levels:

Hemoglobin (mg/dl) baseline	12.1	13.2	10.1	9.2	10.6	12.3	12.3	11.4	10.5	13.1
Hemoglobin (mg/dl) 3 months of treatment	13.2	13.2	13.3	12.1	12.2	13	12.9	12.9	11.2	12.7
d	-1.1	0	-3.2	-2.9	-1.6	-0.7	-0.6	-1.5	-0.7	0.4

4. Significance level: α = 0.05
5. Critical region for two-tailed test: degrees of freedom = 9; t > 2.69
6. Summarize the ***differences***:
   - ***n*** = 10
   - $\bar{d}$ = (-1.1+0+-3.2+-2.9+-1.6+-0.7+-0.6+-1.5+-0.7+0.4)/10 = -1.19
   - s_d = $(((-1.1)^2+0^2+(-3.2)^2+(-2.9)^2+(-1.6)^2+(-0.7)^2+(-0.6)^2+(-1.5)^2+(-0.7)^2+0.4^2)-(-1.19)^2)/9$ = 2.75
   - t = -1.19/(2.75/sqrt(10)) = -1.37

Formulas:

- $t = \dfrac{\bar{d}}{\dfrac{s_d}{\sqrt{n}}}$ - $\bar{d} = \dfrac{(d_1 + d_2 + \ldots + d_n)}{n}$ - $s_d = \sqrt{\left[\sum_{i=1}^{n} d_i^2 - \left(\sum_{i=1}^{n} d_i\right)^2 / n\right] / (n-1)}$

7. Interpret from statistical point of view the obtained result.

   Critical region for two-sided test: (-∞; -2.69] ∪ [2.69; ∞)
   Since the calculated t-value is of -1.37 and did not belong to the critical value, the null hypothesis could not be rejected.

8. So, was the treatment efficient?

   The treatment was not efficient since the mean hemoglobin after 3 months of treatment proved not to be statistically different by the mean hemoglobin at baseline.