

EXPLORATORY DATA ANALYSIS (NUMERICAL DATA)

1. Hypothesis test for:

- Single mean
- Difference between two means for independent samples
- Difference between two means (or paired) samples.

2. Single Mean:

- Step 1: Null and Alternative Hypothesis
 - H_0 : The mean serum amylase level in the population from which the sample was drawn is 120 units/100ml.
 - H_A : The mean serum amylase level in the population from which the sample was drawn is different from 120 units/ 100ml. (two-sided test)
- Step 2: Level of significance
 - $\alpha = 0.05$ ($\alpha/2 = 0.025$)
- Step 3: Check the assumption
 - Population is approximately normally distributed
 - Random sampling
 - Independent variable/sample
- Step 4: Statistical test (one sample t test)
 - $t = (x - \mu_0) / (s/\sqrt{n})$
 - where
 - x = sample mean
 - s = sample standard deviation
 - n = sample size
 - μ_0 = the hypothesis mean
 - t stat has $n-1$ degrees of freedom
- step 5: interpretation

- p-value < 0.025
- reject H_0
- step 6: conclusion
 - The population mean serum amylase is statistically significantly different from 120 units/ 100ml.

3. Difference between two means for independent samples

- Step 1: Null and Alternative Hypothesis
 - H_0 : The mean serum amylase level in hospitalized and healthy subjects are the same ($\mu_1 = \mu_2$)
 - H_A : The mean serum amylase levels in hospitalized and healthy subjects are different ($\mu_1 \neq \mu_2$) → (two-tailed)
- Step 2: Level of significance
 - Alpha = 0.05 (alpha/2 = 0.025)
- Step 3: Check the assumption
 - Two population are normally distributed
 - Two population have equal variance (Levene's test)
 - Both are independent samples/variables
 - Random samples
- Step 4: Statistical test
 - Name of the test = independent t-test
 - t statistic = $(\bar{x}_1 - \bar{x}_2) / \sqrt{[s_p^2 (1/n_1 + 1/n_2)]}$
 - where: $s_p^2 = [(n_1-1)s_1^2 + (n_2-1)s_2^2] / [n_1 + n_2 - 2]$
 - \bar{x}_1, \bar{x}_2 = sample means
 - s_1, s_2 = sample standard deviations
 - n_1, n_2 = sample size
 - t stat has $n_1 + n_2 - 2$ degree of freedom (df)
 - Degree of freedom
 - p-value
 - 95% confident interval (lower border & upper border)

- Step 5: Interpretation
 - P-value < 0.025
 - Reject null hypothesis
- Step 6: Conclusion
 - At the 5% level of significance that the mean serum amylase levels are different in healthy and hospitalized subjects.

4. Difference between two means for dependent (or paired) samples

- Research question: the investigators wanted to determine if treatment with Amynophylline altered the average number of apneic episodes per hour
- Step 1: Null and Alternative Hypothesis
 - H_0 : there is no difference in average number of apneic episodes before and after Amynophylline (no difference = zero)
 - H_A : the average number of apneic episodes before and after was difference (difference not equal to zero) (two-tailed)
- Step 2: Level of significance
 - $\alpha = 0.05$ ($\alpha/2 = 0.025$)
- Step 3: Check the assumption
 - The population are normally distributes
 - The two samples are dependent variables/samples
 - Random sampling
- Step 4: Statistical test (paired t test)
 - $t = \frac{d}{s_d} \sqrt{n}$
 - where
 - d = means of differences
 - s_d = standard deviation of the differences
 - n = number of pairs
 - t stat has $n-1$ degree of freedom
- Step 5: Interpretation

- p-value < 0.025
- reject null hypothesis
- Step 6: Conclusion
 - At the 5% level of significance that the average number of apneic episodes before and after Amyn0phylline were difference.