

MA409: Statistical Data Analysis (SAS)

Assignment 2 (Mar 19 – Apr 09)

Note: Please work on 2.3-2.5 by hand calculation (critical values and p-values can be obtained with any software) and 2.6 by SAS procedures.

2.1 Explain the meaning of p-value=0.05. (5 points)

2.2 Prove the Central Limit Theorem based on the i.i.d. assumption. Assume X_1, X_2, \dots, X_n are independent and identically distributed random variables with mean μ and finite variance σ^2 (not necessarily follow a normal distribution), show that the sample mean $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$ asymptotically follows a normal distribution $\mathcal{N}(\mu, \sigma^2/n)$. (Hint: you may use the characteristic function and the Levy's continuity theorem.) (10 points)

2.3 Suppose we have an i.i.d. sample of size $n = 25$ from $\mathcal{N}(\mu, \sigma^2)$ with sample mean $\bar{X} = 0.04$ and known population variance $\sigma^2 = 0.04$.

- (1) Test the hypothesis: $H_0: \mu = 0$ vs. $H_1: \mu > 0$ at $\alpha = 0.05$. Choose the proper test to apply, compute the test statistic, provide the rejection region, and compute the p-value of the test. (10 points)
- (2) If the underlying population mean is $\mu = 0.05$, compute the Type II error rate of the test in (1) given $\alpha = 0.05$. (10 points) type1 error?
- (3) If the sample size increases to $n = 100$ and the underlying population mean is $\mu = 0.05$, compute the Type II error rate of the test in (1) given $\alpha = 0.05$. (5 points)

2.4 Let $X_i \sim_{\text{i.i.d.}} \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_i \sim_{\text{i.i.d.}} \mathcal{N}(\mu_2, \sigma_2^2)$ are two independent samples. The corresponding sample size, sample mean, and sample standard deviation are given below:

$$\begin{aligned} n_1 &= 18, \bar{X} = 13.5, S_1 = 5 \\ n_2 &= 12, \bar{Y} = 9.5, S_2 = 6 \end{aligned}$$

- (1) Test for equal variance: $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 \neq \sigma_2^2$ at $\alpha = 0.05$. (5 points)
- (2) Assuming $\sigma_1^2 = \sigma_2^2$, construct a 95% confidence interval for $\mu_1 - \mu_2$. (5 points)

2.5 The following data, in tons, are the amounts of sulfur oxides emitted by a large industrial plant in 40 days:

17 15 20 29 19 18 22 25 27 9
24 20 17 6 24 14 15 23 24 26

Use the sign test to test: $H_0: m = 21.5$ vs. $H_1: m < 21.5$ at $\alpha = 0.01$ (m is the population median). (10 points)

2.6 The following table gives the racial characteristics of 326 individuals convicted of homicide in 20 Florida counties during 1976-1977, racial characteristics of their victims, and whether they received the death penalty or not.

Convict's Race	Victim's Race			
	White		Black	
	Death Penalty		Death Penalty	
	Yes	No	Yes	No
White	19	132	0	9
Black	11	52	6	97

- (1) Create a SAS dataset based on the table above with four variables: 1. convict's race; 2. victim's race; 3. death penalty or not; 4. number of convicts in each group defined by the previous three variables. (5 points)
- (2) Estimate the proportion of homicide convicts who received death penalty, irrespective of the races of the convict and victim. Construct the 95% Wald, Wilson, and Exact confidence intervals of the estimate. (5 points)
- (3) Test the hypothesis that the proportion in (2) exceeds 0.08 at $\alpha = 0.05$: state the null and alternative hypothesis, value of the test statistic, the p-value (using both the z-test and the exact version) and your conclusion clearly. (5 points)
- (4) Test the hypothesis that the proportional of Black convicts who receive death penalty is different from that of White convicts at $\alpha = 0.1$: state the null and alternative hypothesis, the name of the test you are using, the value of the test statistic, the p-value and your conclusion clearly. (10 points)
- (5) Irrespective of the convict's race, does it appear that the death penalty depends on the victim's race? Carry out an appropriate statistical test at $\alpha = 0.01$: state the null and alternative hypotheses, the name of the test you are using, the value of the test statistic, the p-value and your conclusion clearly. (10 points)
- (6) Based on your conclusions in (4) and (5), state your thinking about racial discrimination. (5 points)