MA409: Statistical Data Analysis (SAS)

Assignment 2

Note: For problems 3 and 4, please finish by hand calculation and provide necessary details. For problems 5, please work with SAS and provide the SAS code in a separate ***.sas file*** as well as the outputs from SAS (using screenshots) together with problems 1-4 in a ***PDF file***.

1. Suppose that the strength of a type of steel produced by a factory follows a normal distribution $N(\theta, 16)$, the requirement is that the expected strength should be greater than 110 (Pa). One day, the factory conducted inspection of production for quality control. n pieces of steel were sampled, and the strengths were measured and recorded as $X_1, X_2, \dots, X_n$. Prove that a test with level α for testing $H_0: \theta = 110$ vs. $H_1: \theta > 110$ is also a test with level α for testing $H_0^*: \theta \leq 110$ vs. $H_1: \theta > 110$. (10 points)

2. Statistical inference has played a crucial role in scientific research and the most common statistical index in scientific literature is the p-value. However, many surveys show that researchers misunderstand, misinterpret, misuse p-values. In 2016, the *American Statistical Association* released a statement in *The American Statistician* warning against the misuse of p-values (click here for the statement). Read the six principles on page 4-5 of the file (i.e., page 131-132) (no need to read page 1-3 of the file). State your understanding of the six principles in **Chinese**. Try not to simply translate each principle. (15 points)

3. 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:

$$n_1 = 17, \bar{X} = 13.5, S_1 = 5.5$$

$$n_2 = 19, \bar{Y} = 10.5, S_2 = 4.5$$

   (1) Test for equal variance: $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 \neq \sigma_2^2$ at $\alpha = 0.05$. Choose the proper test to apply, compute the test statistic, provide the rejection region, compute the p-value of the test, and state your conclusion. (5 points)

   (2) Assuming $\sigma_1^2 = \sigma_2^2$, construct a 95% confidence interval for $\mu_1 - \mu_2$. (5 points)

   (3) Suppose that the underlying true mean difference between the two population is $\mu_1 - \mu_2 = 2.5$ and it is given that $\sigma_1 = \sigma_2 = \sigma = 5$, compute the Type II error rate if the **two-sample z-test** is applied to test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 > 0$ with samples of sizes $n_1 = 17$ and $n_2 = 19$. Use $\alpha = 0.05$. (10 points)

(4) Suppose that the underlying true mean difference between the two population is $\mu_1 - \mu_2 = 2.5$ and it is given that $\sigma_1 = \sigma_2 = \sigma = 5$, compute the minimum total sample size $n = n_1 + n_2$ such that the power is at least 0.8 when applying the **two-sample z-test** to test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 > 0$. Use $\alpha = 0.05$. (10 points)

4. A beverage company has launched two new beverages with different formulas (配方). 12 consumers have been randomly sampled and asked to taste the two beverages and rate them. The scoring ranges from 1 to 10 (from dislike to like), results are:

Consumer	1	2	3	4	5	6	7	8	9	10	11	12
Beverage A	10	8	6	8	7	5	1	3	9	7	6	9
Beverage B	6	5	2	2	4	6	4	5	9	8	5	7

Use the **Wilcoxon signed-rank test** to test whether the scores of the two beverages differ significantly at $\boldsymbol{\alpha = 0.01}$, i.e., $H_0: m_D = 0$ vs. $H_1: m_D \neq 0$ (m_D is the population median of the difference between the scores of the two beverages). (10 points)

5. The following table gives the racial characteristics of 326 individuals convicted of homicide (杀人罪) in 20 Florida counties during 1976-1977, racial characteristics of their victims (受害者), 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 and Exact confidence intervals of the estimate. (6 points)

(3) Test the hypothesis that the proportion in (2) exceeds 0.09 at $\boldsymbol{\alpha = 0.1}$: state the null and alternative hypotheses, value of the test statistic and the p-values (using both the z-test and the exact version). Present your conclusion clearly. (6 points)

(4) Test the hypothesis that the proportion of White convicts who received death penalty is different from that of Black convicts at $\alpha = 0.05$: 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. (6 points)

(5) Irrespective of the convict's race, does it appear that the proportion of death penalty is higher when the victim's race is White? Carry out an appropriate statistical test at $\alpha = 0.05$: 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. (6 points)

(6) Based on your conclusions in (4) and (5), state your thinking about the existence of racial discrimination. (6 points)