**Statistical data analysis, Assignment 5**

---

**Problem 1.** Suppose that random samples $X_1, X_2, ..., X_n \overset{i.i.d.}{\sim} F(x|\theta)$, for $\alpha = 0.05$, given a 95% confidnce interval of the population parameter $\theta$, $(T_1(X), T_2(X))$ : random interval, which of the followings are true?

a) $P(T_1 < \theta < T_2) = 0.95$.

b) When $T_1$ and $T_2$ are observed, denoted as $t_1$ and $t_2$, respectively, $P(t_1 < \theta < t_2) = 0.95$.

c) If you repeatedly obtained random samples and constructed 95% C.I.;s, say 100 C.I.'s, then 95% of these C.I.'s will contain $\theta$.

d) All of the above answers are correct.

**Problem 2. R Practice:** Simulate the Confidence interval and plot it.

1. For $X_1, ..., X_{10} \overset{i.i.d.}{\sim} N(\mu, 5)$, simulate 100 95%CI's for population mean and plot 100 CI's on the same graph. This critical value should be *qnorm()* in R.

2. For $X_1, ..., X_{1000} \overset{i.i.d.}{\sim} Pois(\lambda)$, simulate 100 **approximately** 95%CI's for $\lambda$ and plot 100 CI's on the same graph.

The true parameters are specified by yourself.

**Problem 3.** A coin was tossed $n = 1000$ independent times, let M be the number of times it lands heads, the owner claims that the coin is biased at $\alpha = 0.05$ significant level.

(a) If M=482, is this claim correct?

(b) If M=535, is this claim correct?

(c) If M=465, is this claim correct?

**Problem 4.** A survey conducted by a polling company collected the birth months of 12,000 individuals from a certain city. The data were grouped by combining every two months, as shown in the table below:

| Jan & Feb | Mar & Apr | May & Jun | Jul & Aug | Sept & Oct | Nov & Dec |
|-----------|-----------|-----------|-----------|------------|-----------|
| 1968 | 1919 | 2042 | 1932 | 2003 | 2136 |

Determine if the survey data distribution is uniform across all groups by conducting a **chi-square test** at the 0.05 significance level.

**Problem 5.** A survey conducted with 200 visitors at Sun Moon Lake on a Sunday collected data on their arrival and stay durations. The aim is to assess the independence between the time of arrival and the duration of stay. The data are organized into a two-way table as shown below:

| Arrival Time \ Stay Duration | 0-4 hrs | 4-8 hrs | 8+ hrs |
|------------------------------|---------|---------|--------|
| **7:00-10:00** | 14 | 32 | 37 |
| **10:00-13:00** | 19 | 17 | 42 |
| **13:00-16:00** | 12 | 10 | 17 |

1. At a significance level of $\alpha = 0.05$, calculate the chi-square test statistic and the degrees of freedom for this data.

2. Based on the chi-square test results at $\alpha = 0.05$, determine whether arrival and stay durations are independent variables. For analysis in R, construct the table as a matrix, say *M*, and apply *chisq.test(M)*.

**Problem 6. Textbook OpenIntro Statistics, 2019**

- **Chapter 5:** Exercises 27, 28, 32, 36.

- **Chapter 6:** Exercises 15, 21.