YONSEI UNIVERSITY, COLLEGE OF COMMERCE AND ECONOMICS

연세대학교 18편입 기출 복원문제 상경계열(경제, 응통)

- Statistics -

1. [40 points] Consider a lined sheet of paper that is ruled with equally spaced parallel lines of distance 1. Suppose a needle of length l is randomly dropped on the paper. Let X be the distance from the midpoint of the needle to the nearest line and Y be the shortest incline angle of the needle from the line perpendicular to the parallel lines, as illustrated below.



Because the needle is randomly dropped on the paper, we assume that $X \sim \text{Unif}(0, \frac{1}{2})$ and $Y \sim \text{Unif}(0, \frac{\pi}{2})$, and X and Y are independent. When $X \sim \text{Unif}(a, b)$, the probability density function (pdf) of X is given by $f_X(x) = 1/(b-a)$ for a < x < b and 0 otherwise.

(a) [10 points] When l < 1, show that the event $\left\{X < \frac{l \cos Y}{2}\right\}$ is equivalent to the event that the needle comes to lie in a position where it intersects one of the parallel lines.

(b) [10 points] When l < 1, calculate the probability of the needle crossing one of the parallel lines on the paper.

(c) [10 points] When $l \ge 1$, find a condition of Y such that the needle always crosses one of the parallel lines on the paper, and calculate the corresponding probability.

(d) [10 points] When $l \ge 1$, calculate the probability of the needle crossing one of the parallel lines on the paper, using the fact that $\sin(\arccos x) = \sqrt{1-x^2}$.

2. [**60 points**] Annual salary data of female workers are collected in 7 companies. Table 1 provides a sample size for each of the 7 companies, and Figure 1. shows the side-by-side boxplot of annual salary of female workers for the 7 companies.

Table 1: Sample size of Female workers for 7 companies

Company	А	В	С	D	Е	F	G
Sample size	5	6	9	3	6	9	9



Figure 1: side by side Box plot of annual saraly Female workers for 7 companies

Companies A and B are with female CEOs and the other companies are with male CEOs. Suppose that the Secretary of Labor worried about low annual salary for female workers in company G. In his rebuttal, the male CEO of company G pointed out that among the other six companies, the two companies (A and B) with female CEOs had the highest average salary. The male CEO thus argued that the difference in annual salary of female workers between company G and the rest of companies was really due to excessive salary for female workers in companies with the female CEOs and the difference between company G and the rest of companies with male CEOs can be explained by chance.

This problem asks you to carry out a statistical analysis to investigate whether there is any statistical evidence to support the claims of the male CEO of company G. There are three key questions:

- Q₁: Is there any evidence that there are differences between companies, with female CEOs and those with male CEOs in terms of annual salary of female workers?
- Q₂: Is there any evidence that there are differences between company G and the rest of companies with male CEOs in terms of annual salary of female worker?
- Q₃: Is there any evidence that there are differences among the other four companies with male CEOs in terms of annual salary of female workers?

NOTE: Use the following notation in your answers. Let \overline{y}_0 and n_i denote the sample mean and sample size of group *i* respectively. Without the subscript, \overline{y} and *n* denote the corresponding quantities for the entire sample as on single group. The subscript *i* can take (letter) value A, B, C, D, E, F or G representing the seven companies.. The subscript can also take *f*, *m*, and *o*, representing the combined group of companies with male CEOs except company G. For example

$$n_m = n_C + n_D + n_E + n_F + n_G = \sum_{i=C}^G n_i$$
 and $n_o = \sum_{i=C}^F n_i$

Here we have used the alphabetic order among the seven letters {A, B, C, D, E, F, G}.

(a) [20 points] Based on the information provided in Table 1 and Table 2 below, complete Table 2. Briefly describe how you complete each column.

(For the last column, provide the range of a p-value based on the F-table attached to the exam.)

Source of variance	\mathbf{SS}	df	MS	F-stat	p-value
Between groups	1800	[]	[]	[]	[]
female vs. male CEOs	600	[]	[]	[]	[]
among female CEOs	[]	[]	[]	[]	[]
among male CEOs	1000			[]	
company G vs. others	[]		[]	[]	[]
among other companies	90	[]	[]	[]	[]
Within groups	[]	[]	[]		
Total	3600				

Table 2: Analysis of variance (ANOVA) Table

(b) [10 points] Show that the between-group sum of squares,

$$SS_b = \sum_{i=A}^G n_i \left(\overline{y}_i - \overline{y}\right)^2$$

is decomposed into the sum of squares for female vs. male CEOs $(SS_{f,m})$, the sum of squares for among female CEOs (SS_f) , and the sum of squares for male CEOs (SS_m) , and establish that

$$SS_b = SS_{f,m} + SS_f + SS_m$$

(c) [10 points] Futher decompose SS_m into the sum of squares for company G vs other $(SS_{G,o})$ and the sum of squares for among other companies (SS_o) . That is, find the correct expression for $SS_{G,o}$ and SS_o , and establish that

$$SS_m = SS_{G,o} + SS_o$$

(d) [10 points] Based on Table 2, report your findings to answer the three questions: Q_1 , Q_2 and Q_3 . Be specific about which hypothesis test goes with which question.

(e) [10 points] In answering Q_1 , an issue arises as whether we whould include company G in the male group. Given that goal of your analysis, what is your choice and why? Based Figure 1 and Table 2, summarize your conclusion regarding whether there is any statistical evidence to support the claims of the male CEO of company G.

- Mathematics for $Economics^1$ -

1. [**20 points**] 수열 $< a_n > e a_n = \int_{-\infty}^n \left(\frac{|x|}{x^2 + 1}\right) dx$ 로 정의된다. n이 무한대로 다가갈 때, 이 수열의 극한값을 계산하여라.

2. [**20** points] 닫힌 구간 [*a*,*b*]에서 잘 정의된 연속함수 *y* = *f*(*x*)를 생각해보자 (여기서 *a*,*b* ∈ ℝ). 구간 [*a*,*b*]에서 함수 *y* = *f*(*x*)의 곡선의 길이를 구하는 공식을 유도하여라. 그리고 그 공식을 이용하여, 구간 [$\frac{\sqrt{2}}{4}$,1]에서 함수

$$y = (1 - x^{\frac{2}{3}})^{\frac{3}{2}}$$

의 곡선의 길이를 계산하여라.

3. [**10** points] 다음과 같이 정의된 수열 *a_n* 에 대해

$$a_n = \sum_{k=1}^n b_k$$

n이 무한대로 다가갈 때, 실수 a로 수렴한다고 가정하자.n이 무한대로 다가갈 때, b_n 의 극한 값을 계산하여라.

4. [**20 points**] $m \times n$ 행렬 A와, $n \times n$ 행렬 Q (det(Q) $\neq 0$) 를 생각해보자. 행렬 A의 열벡터들에 의해 생성되는 벡터공간을 V_A 라 하고, 행렬 AQ의 열벡터들에 의해 생성되는 벡터공간을 V_{AQ} 라고 하자.

"두 벡터공간 VA과 VAQ는 동일한 벡터공간이다."

가 참인지 거짓인지 말하고, 설명하시오.

5. [20 points] 3 × 1 벡터 x_i 들로 이루어진 동적시스템(dynamic system)이 시간 i (i = 1,2,...)에 따라 다음의 규칙 x_{i+1} = Ax_i 을 만족한다고 하자.여기서 A는

$$A = \begin{pmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix}$$

이다. Find a stable solution $(= x_i^*)$ statisfying that $x_{i+1}^* = Ax_i^*$ will neither diverge nor collapse to zero as *i* goes to infinity.

6. [10 points] $n \times n$ 행렬 A, B 에 대해, $A = P^{-1}BP$ 를 만족하는 non-singular 행렬 P가

¹실제론 영어로 출제되었지만, 통계 복원하느라 간략하게 적었기 때문에 한글로 복원.

존재한다면 행렬 A와 B는 닮음 (similarity)이라 한다.

"만약 A와 B가 닮음이면, $\det(A) = \det(B)$ "

가 참인지 거짓인지 말하고, 설명하시오.