

[2015-01] 통계학 (강좌) 중간고사 1 (16:00~18:00)

※ 답안지에 소속, 학번, 이름을 빠짐없이 기록하였는지 확인 후, 다음 물음에 대한 정답을 반드시 풀이 과정과 함께 잘 정리하여 제출하세요. 부정행위 (계산기 부정사용 포함) 적발 시 즉시 퇴실 조치할 것입니다.

1. (총 10점) 다음은 한 프로야구팀에서 랜덤하게 선택된 20명의 선수들에 대한 선수별 연봉 (단위: 백만원)과 선수별 평균 홈런 수에 대한 자료이다. 자료를 이용하여 물음에 답하시오.

연봉	34	56	89	150	120	350	250	170	200	270
평균홈런수	13	17	18	23	24	48	30	25	29	35
연봉	160	80	60	180	230	330	380	450	110	210
평균홈런수	25	20	18	26	28	40	47	50	19	30

- (1) (3점) 평균 홈런수에 대한 줄기-잎 그림을 그려라. 단, 줄기의 단위를 10으로 하여라.

(풀이) 주어진 자료에 대한 줄기-잎 그림은 다음과 같다.

5 | 0

4 | 078

3 | 005

2 | 03455689

1 | 37889

- (2) (3점) 평균 홈런수의 표본 평균, 표본 표준편차, 표본 중앙값을 구하여라.

(풀이) ① 표본평균 : $\frac{1}{20} \sum_{i=1}^{20} x_i = \frac{565}{20} = 28.25$

② 표본표준편차:

$$s^2 = \frac{1}{19} \{(13^2 + \dots + 30^2) - 20(\bar{x})^2\} = \frac{1}{19} \{(18161) - 20(28.25)^2\} = 115.7763$$

따라서 표본표준편차는 $s = \sqrt{115.7763} = 10.75994$

③ 표본 중앙값 :

$$\hat{Q}_2 = (20+1) \times \frac{1}{2} = 10.5 \text{ 번째 값이므로 } \frac{x_{(10)} + x_{(11)}}{2} = \frac{25+26}{2} = 25.5 \text{ 이다.}$$

- (3) (4점) 연봉과 홈런의 표본상관계수를 구하여라. (단, 연봉의 표본평균은 193.5이고 표본표준편차는 116.1852이다.)

$$\begin{aligned} \text{(풀이) 표본공분산 } s_{12} &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} (\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}) \\ &= \frac{1}{19} (132876 - 20 \cdot 193.5 \cdot 28.25) = 1226.013 \end{aligned}$$

$$\Rightarrow \text{표본상관계수 } r = s_{12} / s_1 s_2 = 1226.013 / (116.1852 \cdot 10.75994) \approx 0.981$$

2. (총 10점) 별카페에서 판매하고 있는 음료는 총 10개이고 음료의 가격은 1000원짜리 음료가 3개, 500원짜리 음료가 4개, 그리고 100원짜리 음료가 3개로 구성되어 있다. 한 쌍의 커플이 이 카페에 들어서 서로 다른 음료를 주문한다고 하자. 다만, 선택은 늘 여성이 먼저 한 다음 남성이 한다고 가정하고, 남녀 모두 랜덤하게 음료를 선택한다고 하자. 이 때, 남녀가 선택하는 각각의 음료의 가격을 확률변수 X, Y 로 정의하도록 하자. (X : 남, Y : 여)

(1) (3점) 확률변수 X, Y 의 결합확률분포표를 작성하시오.

(풀이) 결합확률분포표는 다음과 같다.

$X \backslash Y$	100	500	1000	합
100	$1/15$	$2/15$	$1/10$	$3/10$
500	$2/15$	$2/15$	$2/15$	$2/5$
1000	$1/10$	$2/15$	$1/15$	$3/10$
합	$3/10$	$2/5$	$3/10$	1

(2) (3점) 두 확률변수는 서로 독립인가? 그렇다면 혹은 그렇지 않다면 그 이유를 설명하라.

(풀이) $P(X=100, Y=100)=1/15$ 이고,

$P(X=100) \cdot P(Y=100) = (3/10) \cdot (3/10) = 9/100$ 이므로

$P(X=100, Y=100) \neq P(X=100) \cdot P(Y=100)$ 이다.

즉, 모든 순서쌍에 대해 $P(X, Y) = P(X)P(Y)$ 가 성립하지 않으므로 두 확률변수는 독립이 아니다.

(3) (4점) 한 쌍의 커플 손님이 왔을 때 위와 같은 방법으로 음료를 주문한다면, 이 커플이 카페에서 사용하는 금액의 기댓값은 얼마인가?

(풀이) 사용금액을 $g(X, Y) = X + Y$ 두면 그 기댓값은 $E[g(X, Y)] = E[X + Y]$ 이다.

$(X + Y)$ 의 확률분포를 구해보면 다음과 같다.

$g(X, Y) = X + Y$	200	600	1000	1100	1500	2000	합계
$P(\cdot)$	$1/15$	$4/15$	$2/15$	$1/5$	$4/15$	$1/15$	1

따라서 위의 확률분포표를 이용하면 사용금액의 기댓값은 다음과 같다.

$$E[X + Y] = 200 \cdot 1/15 + 600 \cdot 4/15 + 1000 \cdot 2/15 + 1100 \cdot 1/5 + 1500 \cdot 4/15 + 2000 \cdot 1/15 = 1060$$

3. (10점) 주머니 안에 3개의 동전이 들어있는데, 이 3개의 동전은 동전을 던졌을 때 앞면이 나올 확률이 모두 다르게 만들어져있다. 즉, 이들 동전을 던졌을 때 앞면이 나올 확률은 각각 0.5, 0.6, 0.2 의 값을 갖는다. 이 때, 주머니에서 1개의 동전을 랜덤하게 선택한 후, 선택된 동전을 세 번 던지는 경우를 생각해보자.

(1) (5점) 랜덤하게 선택한 1개의 동전을 세 번 던진 결과가 순서대로 (앞, 뒤, 뒤)일 확률을 구하시오.

(풀이) C_1, C_2, C_3 를 각각 앞면이 나올 확률이 0.5, 0.6, 0.2 인 동전이 선택되는 사건이라고 하자. 각 동전은 랜덤하게 선택되므로 $P(C_1) = P(C_2) = P(C_3) = 1/3$ 이다.

따라서 구하는 확률은 다음과 같다.

$$P(HTT) = \sum_{i=1}^3 P(HTT \cap C_i) = \sum_{i=1}^3 P(HTT|C_i)P(C_i) \\ = (0.5^3 + 0.6 \times 0.4^2 + 0.2 \times 0.8^2) \times \frac{1}{3} = 0.1163$$

(2) (5점) 동전을 세 번 던진 결과가 순서대로 (앞, 뒤, 뒤)가 나왔을 때, 이 동전이 앞면이 나올 확률이 0.5의 값을 갖는 동전일 확률을 구하시오.

sol) 앞면이 나올 확률이 0.5인 동전이 선택되는 사건을 C_1 이라고 하자. 그렇다면 구하는 확률은 다음과 같다.

$$P(C_1|HTT) = \frac{P(C_1 \cap HTT)}{P(HTT)} = \frac{P(HTT|C_1)P(C_1)}{P(HTT)} = \frac{(0.5^3)(1/3)}{0.1163} = 0.35827$$

4. (총 15점) 확률변수 X, Y 가 다음과 같은 결합분포를 갖는다. 물음에 답하시오.

$P(X=x, Y=y)$		x	
		2	4
y	1	0.1	0.15
	3	0.2	0.3
	5	0.1	0.15

(1) (4점) X 와 Y 의 주변확률분포를 각각 구하여라.

(풀이) 주변확률분포는 다음과 같다.

X	2	4	합계
$P(\cdot)$	0.4	0.6	1

Y	1	3	5	합계
$P(\cdot)$	0.25	0.5	0.25	1

(2) (8점) 공분산 $Cov(X+Y, X-Y)$ 를 구하여라.

$$\begin{aligned} \text{(풀이)} \quad Cov(X+Y, X-Y) &= E((X+Y)(X-Y)) - E(X+Y)E(X-Y) \\ &= E(X^2 - Y^2) - (E(X) + E(Y))(E(X) - E(Y)) \\ &= E(X^2) - E(Y^2) - (E(X))^2 + (E(Y))^2 \\ &= E(X^2) - (E(X))^2 - (E(Y^2) - (E(Y))^2) \\ &= Var(X) - Var(Y) \end{aligned}$$

$$Var(X) = E(X^2) - E(X)^2 = (2^2 \cdot 0.4 + 4^2 \cdot 0.6) - (2 \cdot 0.4 + 4 \cdot 0.6)^2 = 0.96,$$

$$\begin{aligned} Var(Y) &= E(Y^2) - E(Y)^2 \\ &= (1^2 \cdot 0.25 + 3^2 \cdot 0.5 + 5^2 \cdot 0.25) - (1 \cdot 0.25 + 3 \cdot 0.5 + 5 \cdot 0.25)^2 = 2 \end{aligned}$$

$$\therefore Cov(X+Y, X-Y) = 0.96 - 2 = -1.04$$

(3) (3점) $E[(X-Y)^2]$ 을 구하여라.

(풀이) 결합확률 분포가 다음과 같다.

$g(X, Y) = (X - Y)^2$	1	9	합계
$P(\cdot)$	0.75	0.25	1

$$\text{따라서 } E[(X-Y)^2] = 1 \cdot 0.75 + 9 \cdot 0.25 = 3$$

5. (10점) 확률변수 X 에 대해 $X \sim B(100, 0.5)$ 를 가정하자. $\hat{p} = \frac{X}{100}$ 이라고 할 때, 이항분포의 정규근사를 이용하여

$$P(\hat{p} - 0.3 \geq k) < 0.102$$

을 만족하는 k 의 범위를 구하여라.

(풀이) $X \sim B(100, 0.5)$ 일 때, $np > 5$, $n(1-p) > 5$ 의 조건을 만족하므로 중심극한정리에 의해, 근사적으로

$$\hat{p} \sim N(p, p(1-p)/n) = N(0.5, 0.5^2/100)$$

을 만족한다. 따라서

$$P(\hat{p} - 0.3 \geq k) = P\left(\frac{\hat{p} - 0.5}{\sqrt{0.25/100}} \geq \frac{k - 0.2}{\sqrt{0.25/100}}\right) = P\left(Z \geq \frac{k - 0.2}{\sqrt{0.25/100}}\right) < 0.102$$

를 만족하는 k 의 범위를 구하면 된다.

표준정규분포표를 이용하면 $z_{0.102} = 1.27$ 이므로 $\frac{k - 0.2}{\sqrt{0.25/100}} \geq z_{0.102} = 1.27$ 이다.

따라서 이를 만족하는 k 는 다음과 같다.

$$k > 1.27 \sqrt{0.25/100} + 0.2 = 0.2635 \quad \therefore k > 0.2635$$

6. (총 10점) 철수가 학교까지 걸어가는 시간은 평균이 30분이고 표준편차가 5분인 정규분포를 따르고, 영수가 학교까지 걸어가는 시간은 평균이 25분이고 표준편차가 4분인 정규분포를 따른다. 철수와 영수가 학교까지 걸어가는 시간은 서로 독립이라고 가정한다.

(1) (5점) 철수와 영수가 동시에 출발할 때, 철수가 영수보다 먼저 도착할 확률은 얼마인가?

(풀이) $X \sim N(30, 5^2)$, $Y \sim N(25, 4^2)$ 이고 X 와 Y 는 서로 독립이므로

$X - Y \sim N(30 - 25, 5^2 + 4^2)$ 이다. 따라서 구하는 확률은 다음과 같다.

$$P(X - Y < 0) = P\left(\frac{(X - Y) - 5}{\sqrt{5^2 + 4^2}} < \frac{-5}{\sqrt{5^2 + 4^2}}\right) = P(Z < -0.780) = 0.217$$

(2) (5점) 철수가 영수보다 먼저 도착할 확률이 90%가 되려면 철수는 영수보다 몇 분 먼저 출발해야 하는가?

(풀이) 철수가 영수보다 m 분 먼저 출발한다면 $X - Y \sim N(5 - m, 5^2 + 4^2)$ 이 된다.

$$P(X-Y < 0) = P\left(\frac{X-Y-5+m}{\sqrt{5^2+4^2}} < \frac{-5+m}{\sqrt{5^2+4^2}}\right) = P\left(Z < \frac{-5+m}{\sqrt{5^2+4^2}}\right) = 0.9$$

표준정규분포표를 이용하면 $z_{0.1} \simeq 1.28$ 이므로 $\frac{-5+m}{\sqrt{5^2+4^2}} = 1.28$ 을 만족하는 m 의 값을 구하면

된다. 따라서 구하는 값은 $m = 1.28 * \sqrt{5^2+4^2} + 5 = 13.2$ 이다.

7. (총 15점) 확률변수 X_1, \dots, X_{20} 가 서로 독립이고 평균이 μ , 분산이 σ^2 인 정규분포에서 생성되었다고 할 때, 다음 물음에 답하라.

(1) (10점) $P\left(\frac{\bar{X}_1 - \mu}{\sigma/\sqrt{18}} > y_1\right) = P\left(\frac{V_1/10}{V_2/10} > y_2\right) = 0.95$ 를 만족하는 y_1, y_2 를 각각 구하여라. 단,

$$\bar{X}_1 = \frac{1}{9} \sum_{i=1}^9 X_i, \quad \bar{X}_2 = \frac{1}{11} \sum_{i=10}^{20} X_i, \quad V_1 = \frac{1}{2} \sum_{i=1}^9 (X_i - \bar{X}_1)^2, \quad V_2 = \frac{1}{2} \sum_{i=10}^{20} (X_i - \bar{X}_2)^2 \text{ 이다.}$$

(풀이) $\bar{X}_1 \sim N(\mu, \sigma^2/9)$ 이므로

$$P\left(\frac{\bar{X}_1 - \mu}{\sigma/\sqrt{18}} > y_1\right) = P\left(\frac{\bar{X}_1 - \mu}{\sigma/\sqrt{9}} > \frac{y_1}{\sqrt{2}}\right) = P\left(Z > \frac{y_1}{\sqrt{2}}\right) = 0.95 \text{ where } Z \sim N(0,1)$$

$$\therefore y_1 = \sqrt{2} \times (-z_{0.05}) = \sqrt{2} \times (-1.645) = -2.3264$$

$V_1 \sim \chi^2(8), V_2 \sim \chi^2(10)$ 이므로

$$P\left(\frac{V_1/10}{V_2/10} > y_2\right) = P\left(\frac{(V_1/8) \times (8/10)}{V_2/10} > y_2\right) = P\left(F > \frac{5}{4} y_2\right) = 0.95 \text{ where } F \sim F(8,10).$$

$$\therefore y_2 = \frac{4}{5} F_{0.95}(8,10) = 0.8 \times \frac{1}{F_{0.05}(10,8)} = 0.8 \times 0.2988 = 0.2390$$

(2) (5점) $P\left(\frac{\bar{X} - \mu}{\sqrt{V/10}} > y_3\right) = 0.95$ 를 만족하는 y_3 를 구하라.

$$\text{단, } \bar{X} = \frac{1}{20} \sum_{i=1}^{20} X_i, \quad V = \sum_{i=1}^{20} (X_i - \bar{X})^2 \text{ 이다.}$$

(풀이) $\bar{X} \sim N(\mu, \sigma^2/20), V/\sigma^2 \sim \chi^2(19)$ 이므로

$$\begin{aligned} P\left(\frac{\bar{X} - \mu}{\sqrt{V/10}} > y_3\right) &= P\left(\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{20}} \sqrt{\frac{\sigma^2}{20}}}{\sqrt{\frac{V}{19}} \sqrt{\frac{19}{10}}} > y_3\right) = P\left(\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{20}} \sqrt{\frac{1}{20}}}{\sqrt{\frac{V}{19\sigma^2}} \sqrt{\frac{19}{10}}} > y_3\right) \\ &= P\left(T > y_3 \frac{\sqrt{20 \times 19}}{\sqrt{10}}\right) = 0.95 \text{ where } T \sim t(19) \end{aligned}$$

$$\therefore y_3 = \sqrt{\frac{10}{20 \times 19}} \times (-t_{0.05}(19)) = \frac{1}{\sqrt{38}} \times (-1.729) = -0.2805$$

표준 정규 분포표 $P(Z \leq z)$ [illegible]

t 분포표

$$t_{\alpha} : P(T \geq t_{\alpha}) = \alpha, T \sim t(df)$$

df \ α	0.10	0.05	0.025	0.01
1	3.078	6.31	12.71	31.82
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583
17	1.333	1.740	2.110	2.567
18	1.330	1.734	2.101	2.552
19	1.328	1.729	2.093	2.539
20	1.325	1.725	2.086	2.528

χ^2 분포표

$$\chi^2_{\alpha} : P(\chi^2 \geq \chi^2_{\alpha}) = \alpha, \chi^2 \sim \chi^2(df)$$

df \ α	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	4.61	5.99	7.82	9.21
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
19	7.63	8.91	10.12	11.65	27.02	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57

F 분포표

$$F_{0.05} : P(F \geq F_{0.05}) = 0.05, F \sim F(df_1, df_2)$$

	df_1												
df_2	1	2	3	4	5	6	7	8	9	10	11	12	13
1	161.4	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69
2	18.51	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.405	19.413	19.419
3	10.12	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.763	8.745	8.729
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280