

통계학 (강좌) 중간고사 2 (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 조치**할 것입니다.

※ 적절한 계산과정이 없으면 정답으로 인정하지 않습니다.

※ 답은 소수점 셋째자리까지 쓰시오.

1. (총 10점 : 각 2점, 틀리면 1점 추가 감점) 다음 물음에 대하여 맞으면 O 틀리면 X로 답하시오.

(1) 모평균에 대한 검정에서 랜덤 표본의 크기가 커지면 제 2종 오류도 함께 증가한다. (X)

(2) 유의수준 0.05에서 귀무가설을 기각하지 못했더라도 유의수준 0.01에서는 귀무가설을 기각 할 수 있다. (X)

(3) 제 1종 오류와 제 2종 오류의 합은 1이다. (X)

(4) $\hat{\sigma}^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n}$ 은 모분산 σ^2 의 불편추정량이다. (X)

(5) $F \sim F(k_1, k_2)$ 일 때, $F_{1-\alpha}(k_1, k_2) = \frac{1}{F_{\alpha}(k_2, k_1)}$ 이다. (O)

2. (10점) 철도민영화에 대하여 찬반여론을 조사하려고 한다.

(1) (5점) 찬성비율에 대한 신뢰구간 추정에서 95% 오차한계를 2% 이내로 하기 위해서는 전국에 걸쳐 최소 몇 명을 임의로 조사해야 하는가?

sol) $n \geq \frac{1}{4} \left(\frac{z_{\alpha/2}}{d} \right)^2 = \frac{1}{4} \left(\frac{1.96}{0.02} \right)^2 = 2401$ 이므로 최소 2401명을 조사해야 한다.

(2) (5점) 전국에 걸쳐 임의로 1,000명을 조사해 찬성자들이 300명이라고 할 때, 찬성비율에 대한 추정량과 95%신뢰구간을 구하여라.

sol) 추정량 : $\hat{p} = \frac{300}{1000} = 0.3$

신뢰구간 : $0.3 \pm 1.96 \times \sqrt{\frac{0.3 \times 0.7}{1000}} = (0.272, 0.328)$

3. (10점) 색감의 차이가 감정변화에 미치는 영향을 연구하기 위하여 14명을 랜덤으로 선택하여 이들을 60초 간격으로 보라색과 초록색에 반복적으로 노출시키는 실험을 6분간 지속하였다. 각 색이 변할 때마다 최초 12초간 피부에 나타나는 전기반응을 측정하여, 각 색별로 평균을 취한 후, 이것을 최종 자료로 선택하였다. 다음 자료를 이용하여 보라색과 초록색 사이에 감정변화에 미치는 영향이 존재하는지를 적절한 방법을 통하여 유의수준 5%에서 검정하시오.

단, 자료는 모두 정규분포 가정을 만족한다고 하자.

사람	1	2	3	4	5	6	7	8	9	10	11	12	13	14
보라	3.1	3.7	4.0	3.2	3.6	3.5	4.2	3.8	3.7	3.4	3.6	3.8	3.4	3.4
초록	2.2	2.7	3.1	2.9	3.3	2.6	2.9	2.8	3.2	2.5	3.5	3.1	2.3	3.5

sol) 주어진 자료는 대응 표본 자료이므로 적절한 분석 방법은 대응 비교 방법이다.

보라색에 대한 실험 결과를 X_i , 초록색에 대한 실험 결과를 Y_i 라고 하자. 동일한 짝(pair)내에서 두 실험 결과의 차이를 나타내는 새로운 변수 $D_i = X_i - Y_i$ 를 정의하도록 하자. 이 때, 귀무가설과 대립가설은 다음과 같다.

$$H_0: \mu_D = 0 \quad vs \quad H_1: \mu_D \neq 0$$

차이의 모집단이 정규분포를 따른다는 가정 하에 검정통계량은 다음과 같다.

$$T = \frac{\bar{D} - \mu_D}{s_D / \sqrt{n_D}} = \frac{0.7 - 0}{0.413 / \sqrt{14}} = 6.342$$

유의수준 0.05에서 양측 검정의 기각역은 $|T| \geq t_{0.025}(13) = 2.16$ 이고 검정통계량은 기각역에 속하므로 귀무가설을 기각할 수 있다. 따라서 색감의 차이는 감정 변화에 영향을 미친다고 말할 수 있는 충분한 통계적 근거가 있다.

4. (30점) 왼손잡이는 우뇌발달과 밀접한 관련이 있다는 가설이 있다. 이 가설이 옳은지 알아보기 위해 왼손잡이와 오른손잡이 그룹에서 각각 표본을 추출하여 조사한 결과가 다음과 같다.

주사용 손	우뇌발달		계
	평균	평균이상	
왼손잡이	36	59	95
오른손잡이	74	31	105
계	120	80	200

(1) (5점) 왼손잡이 그룹에서 평균 이상의 우뇌 발달을 보이는 비율이 50%를 넘는다고 말할 수 있는가? 유의수준 5%에서 이를 검정하시오.

sol) 왼손잡이 그룹에서 평균 이상의 우뇌발달 비율을 p 라고 하면,

표본비율은 $\hat{p} = 59/95 = 0.621$ 이고 귀무가설과 대립가설은 다음과 같다.

$$H_0: p = 0.5 \quad vs \quad H_1: p > 0.5$$

$np_0 = 200 \times 0.5 = 100 > 5$, $n(1-p_0) = 200 \times (1-0.5) = 100 > 5$ 이므로 표본비율 \hat{p} 은 근사적으로 정규분포를 따르고 검정통계량은 다음과 같다.

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.621 - 0.5}{\sqrt{0.5(1-0.5)/200}} = 3.422$$

유의수준 5%에서 기각역은 $Z \geq z_{0.05} = 1.645$ 이고 검정통계량은 기각역에 속하므로 귀무가설을 기각할 수 있다. 따라서 왼손잡이 그룹에서 평균 이상의 우뇌발달 비율은 50%를 넘는다고 말할 수 있다.

(2) (5점) 왼손잡이 그룹과 오른손잡이 그룹에서 평균 이상의 우뇌발달을 보이는 비율을 각각 p_1, p_2 라고 하자. 왼손잡이가 우뇌발달을 더욱 촉진한다고 말할 수 있는 지를 유의수준 5%에서 검정하시오.

sol) 주어진 문제에 대해 가설을 세우면 다음과 같다.

$$H_0 : p_1 - p_2 = 0 \quad \text{vs} \quad H_1 : p_1 - p_2 > 0$$

모비율의 추정치는 각각 $\hat{p}_1 = 59/95 = 0.621$, $\hat{p}_2 = 31/105 = 0.295$ 이고, 공통표본비율은 $\hat{p} = (59+31)/(95+105) = 0.45$ 이다. 표본비율의 정규근사 조건인

$$n_1 \hat{p} = 95 \times 0.45 > 5, \quad n_1 (1 - \hat{p}) = 95 \times (1 - 0.45) > 5,$$

$$n_2 \hat{p} = 105 \times 0.45 > 5, \quad n_2 (1 - \hat{p}) = 105 \times (1 - 0.45) > 5$$

을 모두 만족하므로 표본비율의 차이 역시 근사적으로 정규분포를 따른다고 할 수 있으며 검정통계량은 다음과 같다.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.621 - 0.295) - 0}{\sqrt{0.45(1-0.45)\left(\frac{1}{95} + \frac{1}{105}\right)}} = 4.628$$

유의수준 5%에서 기각역은 $Z \geq z_{0.05} = 1.645$ 이고 검정통계량은 기각역에 속하므로 귀무가설을 기각할 수 있다. 따라서 왼손잡이가 우뇌 발달을 더욱 촉진한다고 말할 수 있다.

(3) (5점) 오른손잡이 그룹과 왼손잡이 그룹에서 평균 이상의 우뇌 발달을 보이는 비율의 차이에 대한 95% 신뢰구간을 구하시오.

sol) 주어진 자료에서 $\hat{p}_1 = 59/95 = 0.621$, $\hat{p}_2 = 31/105 = 0.295$ 이다. 두 모비율의 차이에 대한 95% 신뢰구간은 다음과 같다.

$$\begin{aligned} & (\hat{p}_1 - \hat{p}_2) \pm z_{0.025} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= (0.621 - 0.295) \pm 1.96 \sqrt{\frac{0.621(1-0.621)}{95} + \frac{0.295(1-0.295)}{105}} \\ &= 0.326 \pm 1.96 \times 0.067 \\ &= 0.326 \pm 0.131 \\ &= (0.195, 0.457) \end{aligned}$$

(4) (10점) (1)번에서 실제로 왼손잡이 중에서 평균이상의 우뇌발달을 보이는 비율이 60%라고 할 때, 제 2종 오류가 발생할 확률은 얼마인가?

sol) 제 2종 오류 =

$$\begin{aligned}
& P\left(\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} \leq z_{0.05} \mid p=0.6\right) \\
&= P\left(\frac{\hat{p}-0.5}{\sqrt{0.5(1-0.5)/95}} \leq 1.645 \mid p=0.6\right) \\
&= P\left(\frac{\hat{p}-0.6}{\sqrt{0.6(1-0.6)/95}} \times \frac{\sqrt{0.6(1-0.6)/95}}{\sqrt{0.5(1-0.5)/95}} + \frac{0.6-0.5}{\sqrt{0.5(1-0.5)/95}} \leq 1.645 \mid p=0.6\right) \\
&= P\left(Z \leq \left(1.645 - \frac{0.6-0.5}{\sqrt{0.5(1-0.5)/95}}\right) \times \frac{\sqrt{0.5(1-0.5)/95}}{\sqrt{0.6(1-0.6)/95}}\right) \\
&= P(Z \leq -0.3106) \\
&= 0.3780
\end{aligned}$$

(5) (5점) 주어진 자료를 카이제곱 검정을 이용하여 검정하려고 한다. 이 경우의 검정통계량 값은 얼마인지 (2)번의 결과를 이용하여 구하고 그 근거를 간략하게 제시하시오.

sol) 주어진 자료의 동질성 검정의 검정통계량의 분포는 자유도 1의 카이제곱 분포를 따르고, 카이제곱분포는 표준정규분포의 제곱이다. 따라서 (2)에서 구한 검정 통계량값 4.628의 제곱인 21.418이 될 것이다.

4. (20점) 그룹 A와 그룹 B의 모평균에 차이가 존재하는지를 검정해 보려고 한다. 다음은 두 그룹의 요약통계량이다. 아래의 물음에 답하시오. 단, 모집단은 각각 독립적으로 정규분포를 따른다고 가정하자.

Group	A	B
표본평균	17.04	16.76
표본표준편차	0.247	0.317
표본 수	10	10

(1) (10점) 모평균 검정에 앞서 두 그룹의 등분산 여부를 검정해보려고 한다. 유의수준 10%에서 이를 검정하시오.

sol) $H_0 : \sigma_A^2 = \sigma_B^2$, $H_1 : \sigma_A^2 \neq \sigma_B^2$

$$F = \frac{S_A^2}{S_B^2} = \frac{0.247^2}{0.317^2} = 0.607$$

$$\text{기각역 : } F \geq F_{0.05}(9,9) = 3.1789 \text{ or } F \leq F_{0.95}(9,9) = \frac{1}{F_{0.05}(9,9)} = \frac{1}{3.1789} = 0.3146$$

검정통계량은 기각역에 속하지 않으므로 H_0 를 기각하지 못하고, 등분산이라고 결론을 내릴 수 있다.

(2) (10점) (1)번에서 얻은 결론을 바탕으로 두 그룹의 모평균의 차이($\mu_A - \mu_B$)가 0.2가 아니라고 할 수 있는지를 유의수준 5%에서 검정하시오. 단, 필요시 다음과 같은 t분포의 근사자유도를 사용하라.

$$df = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{1}{n_1-1}(S_1^2/n_1)^2 + \frac{1}{n_2-1}(S_2^2/n_2)^2}$$

sol)

$$H_0: \mu_A - \mu_B = 0.2 \quad , \quad H_1: \mu_A - \mu_B \neq 0.2$$

$$T = \frac{(\overline{X_A} - \overline{X_B}) - (\mu_A - \mu_B)}{S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \quad , \quad S_p^2 = \frac{9 \times 0.247^2 + 9 \times 0.317^2}{10 + 10 - 2} = 0.08075$$

$$T = \frac{(17.04 - 16.76) - 0.2}{0.2842 \sqrt{\frac{1}{10} + \frac{1}{10}}} = 0.6284$$

$$\text{기각역} : |T| \geq t_{0.025}(18) = 2.101$$

검정통계량 0.6284는 기각역 에 속하지 않으므로, H_0 를 기각하지 못하고 두 그룹의 모평균의 차이가 0.2가 아니라고 할 근거는 없다.

5. (20점) 고등학교 내신 성적과 대학 1학년 성적 간에 관계가 있는가를 알기 위하여, 한 대학에서는 학생들의 고등학교 내신 등급을 이용하여 학생들을 세 그룹으로 나눈 후 각 그룹별로 표본 학생들을 선정하여 조사한 결과가 다음과 같다.

내신등급	대학 1학년 평점				합계
	A	B	C	D	
1~2등급	22	22	10	6	60
3~4등급	11	18	36	11	76
5급 이하	8	20	23	13	64
합계	41	60	69	30	200

(1) (5점) 이 자료로부터 고등학교 내신 성적과 대학 1학년 성적이 관계가 있다고 할 수 있는지를 검정하고자 할 때, 적절한 가설을 세우시오.

sol) 주어진 자료는 동질성 검정을 위한 표본 설계 방법을 따르고 있다. 따라서 적절한 가설은 다음과 같다.

H_0 : 고등학교 내신등급에 따른 대학 1학년 평점은 동일하다.

H_1 : 고등학교 내신등급에 따른 대학 1학년 평점은 동일하지 않다.

(2) (15점) 아래의 계산과정은 (1)의 가설을 검정하기 위한 과정이다. 빈 칸을 채우고(10점), (1)의 가설을 유의수준 5%에서 검정하시오(5점).

기대도수	A	B	C	D
1~2등급	① 12.3	18.0	20.70	9.0

3~4등급	② 15.58	22.8	26.22	11.4
5급 이하	③ 13.12	19.2	22.08	9.6

$\frac{(O-E)^2}{E}$	A	B	C	D
1~2등급	④ 7.650	0.889	5.531	1.000
3~4등급	⑤ 1.346	1.011	3.648	0.014
5급 이하	1.998	0.033	0.038	1.204

sol) $\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 24.362$, $\chi_{0.05}^2(6) = 12.59$

기각역 : $\chi^2 > 12.59$ 이므로, 검정통계량이 기각역에 속하여 H_0 를 기각하고, 내신등급에 따라 평점은 동일하지 않다고 결론 내린다.

(3) (5점) 실제로 이 자료는 전체 학생 중 표본으로 선정된 학생 200명에 대해서 고등학교 내신성과 대학 1학년 학점을 조사한 자료였다고 한다. 이 경우, 고등학교 내신 성적과 대학 1학년 학점이 관련이 있다고 할 수 있는가? 적절한 가설을 제시하고 유의수준 5%에서 이를 검정하시오.

sol) 만약 이 자료가 전체 200명 학생에 대한 조사 결과라고 한다면 이는 독립성 검정을 위한 표본설계를 따르고 있다. 따라서 적절한 가설은 다음과 같다.

H_0 : 고등학교 내신등급과 1학년 평점은 독립이다.

H_1 : 고등학교 내신등급과 1학년 평점은 독립이지 않다.

검정 통계량 : $\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 24.362$, $\chi_{0.05}^2(6) = 12.59$

기각역 : $\chi^2 > 12.59$ 이므로, 검정통계량이 기각역에 속하여 H_0 를 기각하고, 내신등급과 평점은 독립적이지 않다고 결론 내린다.

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

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

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	12	15	20
1	161.44	199.50	215.70	224.58	230.16	233.98	236.76	238.88	240.54	241.88	243.90	245.94	248.01
2	18.512	19.000	19.164	19.246	19.296	19.329	19.353	19.371	19.384	19.395	19.412	19.429	19.445
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242