

2022-2 확률과통계 기말고사

*** ※ 풀이과정(80%)과 정답(20%)이 모두 맞아야 만점입니다. ***

* 각 문제 끝에 표시된 숫자는 문제에 배정된 점수이다.

1. 50명의 대학생을 랜덤으로 선택하여 키를 측정한 결과 평균 174.5 cm, 표준편차 6.9 cm 이다. 학생의 키는 정규분포를 따른다고 가정한다. 모평균에 대한 95% 신뢰구간을 구하라. (5)

모표준편차를 모르기 때문에 t 분포 이용하여 값을 계산하면

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$174.5 - 2.009 \cdot \frac{6.9}{\sqrt{50}} \leq \mu \leq 174.5 + 2.009 \cdot \frac{6.9}{\sqrt{50}}$$

$$172.5396 \leq \mu_1 - \mu_2 \leq 176.4604$$

$$\therefore (172.5396, 176.4604)$$

2. 10개의 초콜릿 에너지바(energy bar)의 칼로리를 측정하고 평균을 계산하니 230 이다. 이 초콜릿 에너지바 칼로리의 모표준편차는 15 이다. 초콜릿 에너지바 칼로리는 정규분포를 따른다고 가정할 때, 모평균에 대한 99% 신뢰구간을 구하라. (5)

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$230 - 2.58 \cdot \frac{15}{\sqrt{10}} \leq \mu \leq 230 + 2.58 \cdot \frac{15}{\sqrt{10}}$$

$$217.7620 \leq \mu_1 - \mu_2 \leq 242.2380$$

$$\therefore (217.7620, 242.2380)$$

3. K 라는 식물 재배에 필요한 영양소를 파악하기 위하여 소량의 질소비료를 투입한 것 9개와 투입하지 않은 것 9개의 줄기 무게를 측정한 자료를 이용하여 다음과 같은 결과를 구했다. 줄기 무게는 정규분포를 따른다고 가정한다. A 는 질소비료를 투입하지 않은 것, B 는 소량의 질소비료를 투입한 것으로 구분한다.

A - 평균 0.395556 분산 0.005828

B - 평균 0.454444 분산 0.008478

(1) A 와 B 의 모분산이 동일한지 유의수준 0.05에서 가설검정하라. (10)

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$H_1 : \sigma_1^2 \neq \sigma_2^2$, 검정통계량을 계산하면

$$F_0 = \frac{S_2^2}{S_1^2} \sim F_{n_1-1, n_2-1}$$

$$= \frac{(0.008478)}{(0.005828)} = 1.4547, \text{ or } \frac{S_1^2}{S_2^2} = \frac{(0.005828)}{(0.008478)} = 0.6874$$

기각역은 (1.77, 4.43)이므로 H_0 기각 못 함

(2) A 와 B 의 모평균차에 대한 95% 신뢰구간을 구하라. (10)

σ_1^2 과 σ_2^2 이 같다고 할 수 있으므로

$$Sp^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(8 \cdot 0.005828) + (8 \cdot 0.008478)}{9 + 9 - 2} = 0.007153$$

$$Sp = 0.08457$$

$$(\bar{x}_1 - \bar{x}_2) - t_{0.025, 16} Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{0.025, 16} Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$0.058888 - 2.120 \cdot 0.08457 \sqrt{\frac{2}{9}} \leq \mu_1 - \mu_2 \leq 0.058888 + 2.120 \cdot 0.08457 \sqrt{\frac{2}{9}}$$

$$-0.02563 \leq \mu_1 - \mu_2 \leq 0.14340$$

$\therefore (-0.02563, 0.14340)$ (or $(-0.14340, 0.02563)$ 도 성립)

(3) A의 모평균과 B의 모평균이 같은지 양측검정을 통하여 판정하라. 이때 유의수준은 0.01로 하라. (10)

σ_1^2 과 σ_2^2 이 같다고 할 수 있으므로

$$H_0 : \mu_1 = \mu_2$$

$H_1 : \mu_1 \neq \mu_2$, 검정통계량을 계산하면

$$\frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(0.454444 - 0.395556)}{0.08457 \sqrt{\frac{1}{9} + \frac{1}{9}}} = 1.4771 \text{ (or } -1.4771),$$

기각역을 구하면 $t_{0.005, 16} = 2.921$, ± 2.921 이므로 H_0 기각할 수 없음

4. 다음 문제에서 단순회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \sim iidN(0, \sigma^2)$ 이 사용된다. (소계: 35)

아파트 건물에서 사용되는 창호(window)의 외부온도($^{\circ}F$)와 열손실율(%)의 관계에 대한 자료이다.

외부온도, X($^{\circ}F$) 열손실율, Y(%)

X	Y
20	86
20	80
40	77
50	62
50	75
60	33
60	38
70	43
70	30

다음의 계산결과를 활용하라.

X	Y	
평균 48.89	평균 58.22	$SS_E = 856.78$
분산 361.1	분산 495.94	$SS_{xy} = -2998.57$

4-1. 최소자승법을 이용하여 $\hat{\beta}_0$ 과 $\hat{\beta}_1$ 을 구하고, 외부온도가 32°F 일 때 열손실을 예측값을 구하라. (10)

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} = \frac{-2998.57}{361.1 \cdot 8} = -1.037$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 58.22 - (-1.037) \cdot 48.89 = 108.92$$

외부온도가 32°F 일때를 구하기 위해 X에 32 대입 하면 75.736 이 된다.

4-2. 오차 분산 σ^2 에 대한 추정값을 구하라. (5)

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = \frac{856.78}{7} = 122.3971$$

4-3. $H_0 : \beta_1 = 0$ 와 $H_1 : \beta_1 \neq 0$ 에 대하여 유의수준 0.01에서 검정하라. (10)

$$\hat{\sigma} = \sqrt{122.3971} = 11.0633$$

$$SE_{\hat{\beta}_0} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_{xx}}} = 11.0633 \sqrt{\frac{1}{9} + \frac{2390.23}{2894.4}} = 10.71$$

$$SE_{\hat{\beta}_1} = \hat{\sigma} \sqrt{\frac{1}{SS_{xx}}} = 11.0633 \sqrt{\frac{1}{2894.4}} = 0.205$$

$$t_0 = \frac{\hat{\beta}_1}{SE_{\hat{\beta}_1}} = \frac{-1.037}{0.205} = -5.058$$

기각역 : $t_{0.005,7} = 3.499$ 또는 $-t_{0.005,7} = -3.499$

즉, $-5.058 < -3.499$ 이므로 H_0 기각, 유의미한 독립변수임

4-4.. 결정계수 R^2 를 구하라. (5) W

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SS_E}{SST} = 1 - \frac{856.78}{3967.556} = 0.7841$$

4-5. 아래의 잔차그림표를 이용하여 회귀진단을 실시하라. (5)

잔차그림표를 통해 잔차들이 선형성을 위반하는 것과 outlier 가 존재함을 알 수 있다.

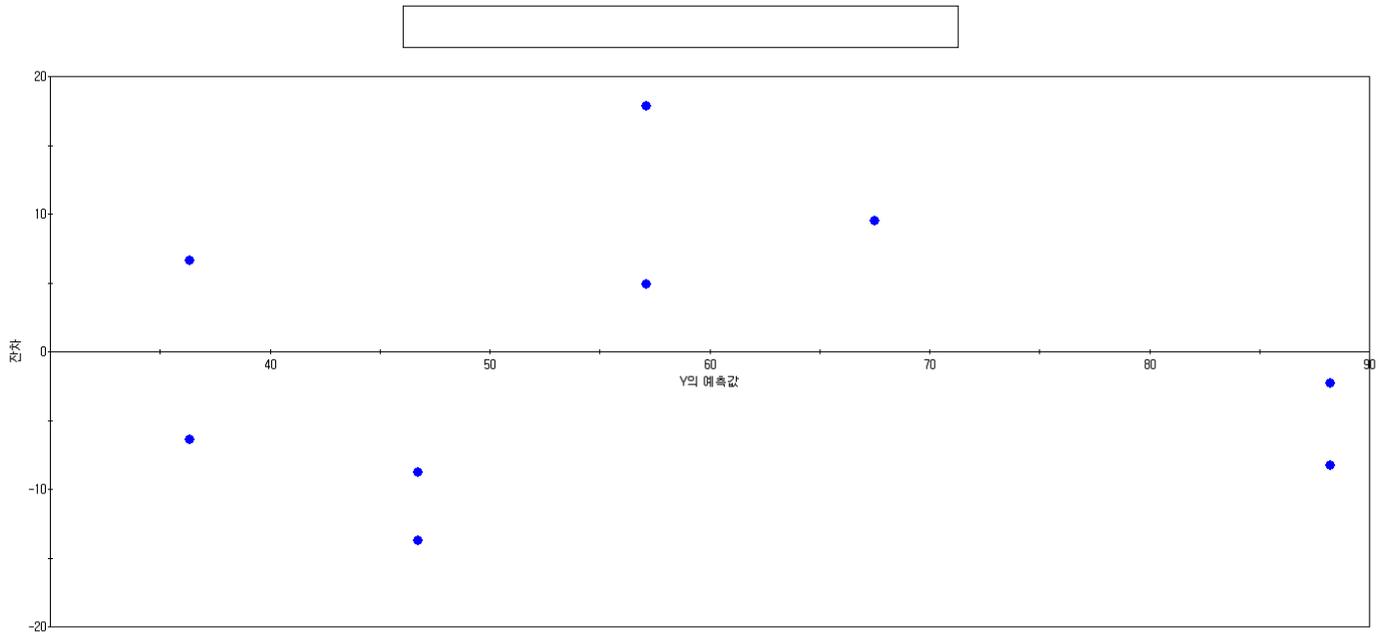
5. 다음은 분산분석과 실험계획법에 관한 문제이다. (소계: 25)

전기자동차의 1회 충전 주행거리는 아주 중요한 특성이다. 한 자동차 연구기관에서 4종류의 전기자동차(A,B,C,D)의 1회 충전 주행거리의 평균이 동일한지를 확인하기 위하여 실험을 하고 다음과 같은 데이터를 확보했다.

데이터는 정규분포를 따른다고 가정한다.

	A	B	C	D
1	380	450	460	324
2	395	482	470	308
3	446	495	490	312
4	470	520	495	320
평균	422.75	486.75	478.75	316.0

5-1. 4종류의 전기자동차 1회 충전 주행거리 모평균의 동일성을 검정하기 위한 귀무가설과 대립가설을 수립하라. (5)



$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1 : 최소한 하나의 평균은 다르다.

5-2. 아래의 ANOVA Table에서 빈칸을 완성하고, 유의수준 0.01에서 분산분석의 결론을 내리라. (10)

[ANOVA Table]

Source	SS	DF	MS	F0	p-value
Factor	(74334.69)	(3)	(24778.23)	33.423	0.000004
Error	(8896.2)	(12)	741.35		
Total	(83230.89)	(15)			

*741.354 는 eZ SPC이용한 결과임_

$$MSA = MSE * F0$$

$$SSA = DFA * MSA$$

$$SSE = DFE * MSE$$

$$SST = SSA + SSE \quad (\text{이렇게 구할 수도 있다})$$

또는 R을 이용하여 결과를 인용한 경우도 가능함

P-value가 0.000004로 작으므로 H_0 기각, 최소한 하나의 평균은 다르다.

5-3. 유의수준 0.01에서 다중비교를 위한 LSD 값을 구하고, 브랜드 A와 D는 모평균이 같은지 다른지 판정하라. (10)
LSD 값을 계산하면

$$t_{12, 0.005} \cdot \sqrt{MSE} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_4}}$$

$$= 3.055 \cdot \sqrt{741.35} \cdot \sqrt{\frac{1}{4} + \frac{1}{4}}$$

$$= 58.81783$$

브랜드 A와 D의 샘플 평균의 차가 106.75 이므로 LSD 값보다 큰 값을 가진다.

따라서 통계적으로 모평균이 다르다고 판정할 수 있다.