

# Final Exam, 통계처리입문, 2023F

(Suggested Solution)

Your Name:

Email:

- 읽는 사람이 답안의 유도 과정에 동의할 수 있도록 풀이 과정을 명시할 것
- 여러가지의 답안을 제시한 경우에 더 낮은 점수를 받을 수 있는 답안으로 채점함
- 단면의 A4 용지 cheat sheet이 허용됨
- 계산기 사용이 허용됨

---

Sum

---

1  
2  
3  
4  
5  
6  
7  
8  
9

---

---

### **Problem 1**

모수(parameter)와 추정량(estimator)의 관계를 서술하라. (틀린 진술이 있으면 감점함. 3문장 이하로 간결하게 작성할 것)[5pts]

Suggested Answer: 추정량은 모집단의 부분집합인 표본집단에서 얻어진 수치를 이용해 계산된 값으로서 모수를 추정하는데 사용된다.

### **Problem 2**

검정통계량과 기각역의 관계에 대해서 서술하라. (틀린 진술이 있으면 감점함. 3문장 이하로 간결하게 작성할 것)[5pts]

Suggested Answer: 검정통계량이 기각역에 포함되면 귀무가설을 기각하고 대립가설을 채택한다.

### **Problem 3**

가설검정에서 유의수준  $\alpha$ 의 의미를 서술하라. (틀린 진술이 있으면 감점함. 3문장 이하로 간결하게 작성할 것)[5pts]

Suggested Answer: 가설검정에서 1종 오류의 확률을 의미한다. 즉, 귀무가설을 잘못 기각하는 오류를 범할 확률을 의미한다.

---

#### Problem 4

표본의 크기( $n$ )가 각각 40인 두 독립 표본으로부터  $\bar{X}_1 = 32.45, s_1 = 4.3$  및  $\bar{X}_2 = 41.45, s_2 = 5.25$ 를 얻었다. 두 모평균의 차이  $\mu_1 - \mu_2$ 의 95% 신뢰구간을 구하라.[10pts]

Suggested Answer:

$$\mu_1 - \mu_2 \in [(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{s_1^2/n_1 + s_2^2/n_2}] \text{ - from L9}$$

$$\mu_1 - \mu_2 \in [(32.45 - 41.45) \pm t_{0.025, df} \sqrt{4.3^2/40 + 5.25^2/40}]$$

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} = \frac{(4.3^2/40 + 5.25^2/40)^2}{\frac{(4.3^2/40)^2}{40-1} + \frac{(5.25^2/40)^2}{40-1}} \approx 75$$

$$\mu_1 - \mu_2 \in [(32.45 - 41.45) \pm 1.995 \sqrt{4.3^2/40 + 5.25^2/40}] = [-9 \pm 2.14] = [-11.14, -6.86]$$

(자유도 60에 대응되는 2.00과 자유도 120에 대응되는 1.98 사이의 아무 값이나 써도 정답으로 처리함. 위의 솔루션에서는 linear interpolation을 이용해 자유도 75에 대응되는 1.995를 사용함.)

---

### Problem 5

확률 변수  $X$  와  $Y$ 에 대해서 5개의 쌍을 관찰하여 다음의 관찰값들을 얻었다. 모상관관계가 0과 같은지 90% 수준에서 검정하라. 귀무가설과 대립가설을 각각 명시하고, 가설 검정을 수행하고 결론을 진술하라.[10pts]

- (1, 1)
- (2, 6)
- (3, 3)
- (4, 7)
- (5, 6)

Suggested Answer:

Sample correlation  $r_{XY} = 0.693$

$$\text{검정통계량 } T = \sqrt{n-2} \frac{r_{XY}}{\sqrt{1-r_{XY}^2}} = \sqrt{5-2} \frac{0.693}{\sqrt{1-0.693^2}} = \sqrt{3} \cdot 0.961 = 1.665 - \text{from L11}$$

기각역  $|T| \geq t_{0.05,3} = 2.353$ 이며, 검정통계량이 기각역에 포함되지 않으므로 귀무가설을 기각할 수 없음. 즉, 상관관계가 있다는 충분한 통계적 근거가 있지 않아서 귀무가설을 기각할 수 없다. 여기에서 귀무가설은 상관관계가 없다는 것, 혹은 모상관관계가 0과 같다. 대립가설은 상관관계가 존재한다는 것, 혹은 모상관관계가 0와 다르다.

---

## Problem 6

다이아몬드의 가격( $Y$ )을 캐럿 단위로 측정하는 다이아몬드의 중량( $X$ )을 이용해 선형회귀 분석을 하려고 한다. 50개의 sample을 이용해서 아래의 정보를 얻었다. 선형회귀 방정식  $\hat{Y}_i = \alpha + \beta X_i$ 의 회귀 계수  $\hat{\alpha}$  과  $\hat{\beta}$ 을 구하라.[10pts]

```
sum(X)
## [1] 243869

sum(Y)
## [1] 44.67

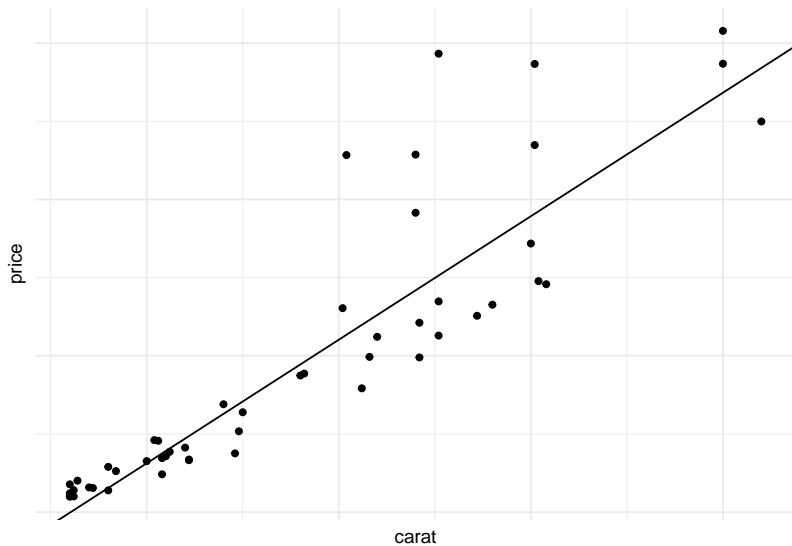
sum(X*X)
## [1] 2141557629

sum(X*Y)
## [1] 314388.8
```

n <- 50	
S_XY <- sum(X*Y)-n*mean(X)*mean(Y)	
S_XX <- sum(X^2)-n*mean(X)^2	
S_XY	## [1] 96516.21
S_XX	## [1] 952115846

---

[참고 그림]



정답은 아래의 아웃풋 블럭에 포함되어 있음:

```
lm(dataset$carat ~ dataset$price)

##
## Call:
## lm(formula = dataset$carat ~ dataset$price)
##
## Coefficients:
## (Intercept) dataset$price
## 0.3989789  0.0001014
```

---

### Problem 7

특정질환과 연관된 생체 수치를 낮춘다고 주장하는 신약의 효과를 알아보기 위하여 6명의 환자를 대상으로 복용전과 후의 수치를 비교하였다. 실험 전의 생체 수치를 A변수로, 실험 후의 생체 수치를 B변수로 정하였다. 신약이 생체 수치를 낮추는지 95% 신뢰 수준에서 검정하라.[10pts]

```
A <- c(250, 220, 204, 222, 206, 259)
B <- c(242, 217, 200, 211, 189, 240)
X <- A-B
mean(X)

## [1] 10.33333

sd(X)

## [1] 6.623192

c(sum(A), sum(B))

## [1] 1361 1299

c(mean(A), mean(B))

## [1] 226.8333 216.5000

c(sd(A), sd(B))

## [1] 22.78962 21.26735

c(sum(A^2), sum(B^2), sum(A*B))

## [1] 311317 283495 296976
```

Suggested answer:

쌍대 집단이므로 L11의 9의 ”단일모집단 모표준편차를 모르는 경우”로 환원된다.

귀무가설  $H_0 : \mu_A \leq \mu_B \Leftrightarrow \mu_X \leq 0$ 과 대립가설  $H_1 : \mu_A > \mu_B \Leftrightarrow \mu_X > 0$ 로 가설을 세운다.

따라서 검정통계량  $T = \frac{\bar{X}-0}{s/\sqrt{n}} = \frac{10.3}{6.62/\sqrt{6}} = 3.81$ 이며, 이는 기각역인  $T \geq t_{\alpha,n-1} = t_{0.05,5} = 2.015$ 에 포함된다.

따라서, 귀무가설을 기각하고 대립가설을 채택한다. 즉, 신약이 생체 수치를 낮춘다는 충분한 통계적인 근거가 있다.

(주. 2023F에는 단측검정이 아닌 양측검정을 수행한 경우에도 감점하지 않았음. 이 경우  $t_{0.025,5} = 2.571$  수치가 사용됨.)

---

## Problem 8

아래의 목적에 부합하는 R code를 작성하라. [5pts]

- Run a linear regression with
  - `price_twd_msq` as the response variable,
  - `n_convenience` as the explanatory variable, and
  - `taiwan_real_estate` as the dataset.

Correct answer:

```
lm(taiwan_real_estate$price_twd_msq ~ taiwan_real_estate$n_convenience)
```

(or)

```
lm(price_twd_msq ~ n_convenience, data=taiwan_real_estate)
```

## Problem 9

아래의 목적에 부합하는 R code를 작성하라. [5pts]

- Find the median life expectancy for each continent, saving it as a column of `medianLifeExp`.
- The dataset `gapminder` has columns of
  - `lifeExp`, which has information on life expectancy of each country
  - `continent`, which indicates which continent each country belongs to

즉, 작성된 R code는 아래의 아웃풋을 생성해야 한다.

```
## # A tibble: 5 x 2
##   continent medLifeExp
##   <fct>        <dbl>
## 1 Africa       47.8
## 2 Americas     67.0
## 3 Asia         61.8
## 4 Europe       72.2
## 5 Oceania      73.7
```

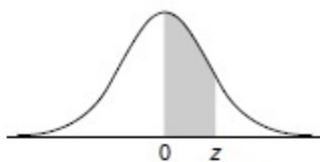
Correct answer:

```
gapminder %>% group_by(continent) %>% summarize(medLifeExp=median(lifeExp))
```

"The following is a table for standard normal distribution"

**Table AIV.2 Standard Norms Table**

Area between 0 and  $z$



$$P(0 < Z < 1.55)$$

<b><math>z</math></b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

"The following is a table for t-distribution"

[표 A-2] t-분포표

$P\{T \geq t_{(q; v)}\} = q$

자유도 <i>v</i>	꼬리 확률 <i>q</i>									
	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.792	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
$\infty$	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

---

"Examinations are formidable even to the best prepared, for the greatest fool may ask more than the wisest man can answer. - Charles Caleb Colton"