기초통계학내용 : 스느보드 영상강의 제공 (출석영향X)
2월중으로 꼭보기

수치형이 아닌 경우 → 분산분석, 범주형자료분석

통계적 추론

■ 통계적추론

□ 모집단: 우리의 관심을 끌어 관찰의 대상이 되는 전체 모임

- 관심의 대상인 특성을 알아보고자 모집단 전체를 조사 (?)
- 표본을 추출하여 관찰하고, 이를 통해 모집단의 특성을 파악

+) 기술통계
산점도그리기
계산하기
⋮

□ 통계적추론

- 표본을 통해 모집단의 특성(모수)에 대해 알아보는 과정

■ 추정

▫ 모집단의 특성의 값을 표본을 통해 찾는 과정

■ 검정

▫ 모집단 특성에 관한 기존의 가설을 표본을 통해 검증

3

통계적 추론

□ 확률표본

- $\{X_1, X_2, \cdots, X_n\}$: 가상의 이론 표본 (대문자)
- 모집단이 정규분포 $N(\mu,\ \sigma^2)$ 가정
  - $X_i,\ i=1,\ldots n$는 서로 독립이고 $N(\mu,\ \sigma^2)$의 확률변수
  - $\{x_1, x_2, \cdots, x_n\}$이 확률표본 $\{X_1, X_2, \cdots, X_n\}$의 실제 값 (소문자)
- 모평균 μ의
  1) 점추정 : 평균
  2) 구간추정 : 신뢰구간
  3) 가설검정 : 가설이 적절한지

- 모집단이 정규분포 $N(\mu, \sigma^2)$을 따르고, σ^2이 알려진 경우
- 모집단이 정규분포 $N(\mu, \sigma^2)$을 따르고, σ^2이 알려지지 않은 경우

4

통계적 추론

□ 모집단이 정규분포 $N(\mu, \sigma^2)$을 따르고, σ^2이 알려진 경우

■ 모평균 μ의 추정 (점추정)

- 모평균 μ의 추정량: 확률표본 $\{X_1, X_2, \cdots, X_n\}$의 표본평균 $\overline{X}$

$$\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$$

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

분산이 작을수록 좋은추정량
⇒ 표본이 많을수록 정확해짐
무조건 크다고 좋은것 X 적절한 표본의 크기 필요

불편(비편향)추정량 ⇒ 좋은추정량! 모수와같음

- 모평균 μ의 추정값: 확률표본의 실제 값 $\{x_1, x_2, \cdots, x_n\}$의 평균 $\overline{x}$

$$\overline{x} = (x_1 + x_2 + \cdots + x_n)/n$$

  - 실제 μ와 얼마나 가까운지는 알 수 없음

5

통계적 추론

- 구간추정 (신뢰구간)
  - 확률표본 $X_i \sim N(\mu,\ \sigma^2)$이고 서로 독립

$$\overline{X} \sim N(\mu, \sigma^2/n), \rightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

표준정규분포

$$P\left(-1.96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

$$\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

  - μ의 95% 신뢰구간 ($\overline{x}$는 표본평균의 값)

99%
95%
90%

$$\left(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

  - 실제 μ를 포함하는 구간이 될 가능성은 95%

6

통계적 추론

■ 모평균 μ의 가설 검정 : 가설에 대한 진위여부

- 귀무가설: 기존의 가설 또는 주장인 귀무가설 (H_0)
- 대립가설: 새로이 내세우는 주장을 대립가설이라 한다. (H_1)

$H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$ $\quad$ $H_1 : \mu < \mu_0$ $\quad$ $H_1 : \mu \neq \mu_0$

(단측가설) $\quad$ (양측가설)

1) H_0가 참임을 가정
→ 표준정규분포를 따르는 Z

2) Z에 대한 신뢰구간 구하기

3-1) $\bar{x}$의 위치확인
- 기각역에 속하면 H_0 기각
- 신뢰역에 속하면 H_0 기각 x

3-2) 유의수준 α 검정

$H_0 : \mu = \mu_0$가 사실이라면, $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

$P(Z \geq 1.645) = P(Z \leq -1.645) = 0.05,$

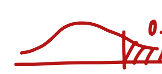
$P(Z \geq 1.96) = P(Z \leq -1.96) = 0.025 \quad \Rightarrow \quad P(|Z| \geq 1.96) = 0.05$

7

통계적 추론

<단측검정>

(i) $H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$인 경우



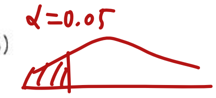
0.05

- $\bar{x}$가 μ_0보다 매우 큰 값이면 H_0를 기각
  - H_0가 사실이면, $P\left(\dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \geq 1.645\right) = 0.05$
- $z^* = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq 1.645$ 이면 H_0를 기각
  - 기각역(critical region): $z^* \geq 1.645$ 또는 $\bar{x} \geq \mu_0 + 1.645\, \sigma / \sqrt{n}$
- 유의수준 $\alpha = 0.05$의 검정: H_0이 사실일 때 $P(z^* \geq 1.645) = 0.05$

(ii) $H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0$인 경우

- $z^* = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq -1.645$ 이면 H_0를 기각 $(\alpha = 0.05)$

$\alpha = 0.05$

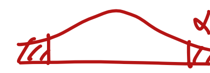


8

통계적 추론

〈양측검정〉

(iii) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$인 경우



$\alpha = 0.025$

- $\bar{x}$가 μ_0보다 매우 크거나 작은 값이면 H_0를 기각
  - H_0가 사실이면 $P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \geq 1.96\right) = P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \leq -1.96\right) = 0.025$
- $|z^*| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq 1.96$ 이면 H_0를 기각 ($\alpha = 0.05$)

  (z^*가 1.96 이상이거나 -1.96 이하이면 H_0를 기각)

◦ $\alpha = 0.10$: 귀무가설을 기각하는 오류가 10번 중에 1번

◦ $\alpha = 0.05$: 귀무가설을 기각하는 오류가 20번 중에 1번

$\Rightarrow$ $\alpha = 0.05$ 일 때 더 신뢰할 수 있는 결론

9

통계적 추론

$\alpha > p \Rightarrow H_0$ 기각
$\alpha < p \Rightarrow H_0$ 기각X (충분히 있을수있는일!)

- 유의확률(p-값, $p-value$) → 오류를 α보다 크지 않게 해주겠다
  - H_0을 기각할 수 있게 되는 가장 작은 유의수준 α의 값
  - H_0이 사실일 때 관찰된 것보다 더 대립가설 쪽으로 치우칠 확률

알고있음 ⇒ 정규분포

예) $H_0 : \mu = 100$, $H_1 : \mu > 100$, $\bar{x} = 101.5$, $\sigma^2 = 25$, $n = 25$

$$z^* = \frac{101.50 - 100}{5/\sqrt{25}} = 1.5, \quad p\text{-값} = P(Z \geq 1.50) = \cancel{0.0668} \ 0.067 \ (p\text{-value})$$

검정통계량

⇒ μ=100일때 (H_0이 참일때) 101.5보다 더 크게 나올확률은 0.067
유의수준 α=0.05보다 크므로 H_0 기각X
즉 "충분히 있을수있는 일이다"

- 유의확률 (p-값)

  (i) $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$: p-값$= P(Z \geq z^*)$

  (ii) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$: p-값$= P(Z \leq z^*)$

  (iii) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$: p-값$= 2P(Z \geq |z^*|)$

10

통계적 추론

t분포

- ◻ 모집단이 정규분포 $N(\mu, \sigma^2)$을 따르고, σ^2이 알려지지 않은 경우

↳ 알아야되는수가됨
표본분산으로추정

- ■ 모평균 μ의 추정
  - ■ 모평균 μ의 추정량: 확률표본 $\{X_1, X_2, \cdots, X_n\}$의 표본평균 $\overline{X}$

$$\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$$

  - – μ의 추정값: $\overline{x} = (x_1 + x_2 + \cdots + x_n)/n$
  - ■ 모분산 σ^2의 추정량

$$S^2 = \Sigma(X_i - \overline{X})^2/(n-1)$$

  - – σ^2의 추정값: $s^2 = \Sigma(x_i - \overline{x})^2/(n-1)$

11

통계적 추론

- ■ 구간추정
  - • 확률표본 $X_i \sim N(\mu,\ \sigma^2)$이고 서로 독립
  - • 모 분산 σ^2을 표본분산 S^2으로 대체하여 $\overline{X}$를 표준화

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

$$P\left(-t(\alpha/2, n-1) < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t(\alpha/2, n-1)\right) = 1 - \alpha$$

$$\overline{X} - t(\alpha/2, n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t(\alpha/2, n-1)\frac{S}{\sqrt{n}}$$

표본평균: $\overline{x}$　　표본표준편차: s

$$\left(\overline{x} - t(\alpha/2, n-1)\frac{s}{\sqrt{n}},\ \overline{x} + t(\alpha/2, n-1)\frac{s}{\sqrt{n}}\right)$$

12

통계적 추론

■ 모평균 μ의 가설검정

- 귀무가설 $H_0 : \mu = \mu_0$
  - 확률표본 $X_i \sim N(\mu,\ \sigma^2)$이고 서로 독립
  - $H_0 : \mu = \mu_0$이 사실이면

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

  - 모분산 σ^2이 알려진 경우와 같은 논리

모분산	σ^2	$\Rightarrow$	표본분산	S^2
분포	$N(0,1)$	$\Rightarrow$	분포	$t(n-1)$

13

통계적 추론

(i) $H_0 : \mu = \mu_0,\quad H_1 : \mu > \mu_0$인 경우

- $t^* = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} \geq t(\alpha, n-1)$ 이면 H_0를 기각
- $p-$값$= P(t(n-1) \geq t^*)$

(ii) $H_0 : \mu = \mu_0,\quad H_1 : \mu < \mu_0$인 경우

- $t^* = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} \leq -t(\alpha, n-1)$ 이면 H_0를 기각
- $p-$값$= P(t(n-1) \leq t^*)$

(iii) $H_0 : \mu = \mu_0,\quad H_1 : \mu \neq \mu_0$인 경우

- $|t^*| = \left| \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} \right| \geq t(\alpha/2, n-1)$ 이면 H_0를 기각
- $p-$값$= 2P(t(n-1) \geq |t^*|)$

14

분포이론

■ 정규분포

$$\boxed{X \sim N(\mu, \sigma^2)}$$

$$E(X) = \mu, \qquad Var(X) = \sigma^2$$

■ 확률밀도함수

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \qquad -\infty < x < \infty$$

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

$$\boxed{aX+b \sim N(a\mu+b,\, a^2\sigma^2)}$$

15

15

분포이론

■ χ^2-분포

$z_1, z_2, \ldots$가 각각 표준정규분포를 따르고 서로 독립

제곱

$z_1^2 \sim \chi^2(1)$: 자유도 1

$z_1^2 + z_2^2 \sim \chi^2(2)$: 자유도 2

$\sum_{i=1}^{k} z_i^2 \sim \chi^2(k)$: 자유도 k

■ t-분포 자유도가 커질수록 정규분포와 유사해짐

$z \sim N(0,1)$, $u \sim \chi^2(k)$, z와 u는 서로 독립

$$\frac{z}{\sqrt{u/k}} \sim t(k)$$

16

16

분포이론

■ F-분포

$u \sim \chi^2(k_1)$, $v \sim \chi^2(k_2)$, u와 v는 서로 독립

$$\frac{u/k_1}{v/k_2} \sim F(k_1, k_2)$$: 자유도 2

– $u \sim t(k)$

↓ 제곱

$u^2 \sim F(1, k)$

예) $t(0.025, 6) = 2.447$

$(t(0.025, 6))^2 = 5.99 \qquad F(0.05, 1, 6) = 5.99$

17

17

? 분포이론

[표 1] 표준정규분포표

$P(Z < 1.5) = .9332$

$P(Z > 1.5) = .0668$

$P(Z > 1.96) = 1 - .975 = .025$

$P(Z < -1.96) = .025$

$P(-1.96 < Z < 1.96) = .95$

0.9

0.05 0.05

$P(-1.645 < Z < 1.645) = .90$

$P(Z > 2.33) = P(Z < -2.33)$

$= 1 - .9901$

N_α	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9308	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

→ 0.995

⇒ 99%

18

분포이론

[표 2] t–분포표

자유도가 6인 t분포에서 2.447보다 클 확률이 0.025

- $t(.025, 6) = 2.447$

$P(t(6) > 2.447)) = .025$

$P(|t(6)| > 2.447)) = .05$

$P(-2.447 < t(6) < 2.447) =$ 0.95

$P(\ t(15)\ > 2.0) =\ ?$

- $t(.025, 15) = 2.131$
- $t(.05,\ 15) = 1.753$

$.025 < P(\ t(15)\ > 2.0) < .05$

↓

$.05\ \ < P(|t(15)| > 2.0) < .10$

예상가능

d.f.	α 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	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	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

19

분포이론

[표 3] χ^2–분포표

- $\chi^2(0.025, 6) = 14.449$

$P(\chi^2(6) > 14.449) = 0.025$

- $\chi^2(0.975, 6) = 1.237$

$P(\chi^2(6) > 1.237) = 0.975$

$P(\chi^2(6) < 1.237) = 0.025$

$P(1.237 < \chi^2(6) < 14.449) = ?$

d.f.	α 0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	3.841	5.024	6.635	7.879
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597
3	0.0717	0.115	0.261	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672

20

분포이론

[표 4 계속] F-분포표 $(\alpha = 0.05)$

- $F(0.05 : 5,15) = 2.90$ $\quad P(F(5,15) > 2.90) = 0.05$
- $F(0.01 : 5,15) = 4.56$ $\quad P(F(5,15) > 4.56) = 0.01$

$P(F(5,15) > 3.5) = ?$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84

21

확률변수

■ 기대값 (확률변수 Y)

(i) 이산형 확률변수 Y : $\quad E(Y) = \sum y\, f(y) = \mu$

(ii) 연속형 확률변수 Y : $\quad E(Y) = \int y\, f(y)\, dy = \mu$

■ 분산 (확률변수 Y)

$$\sigma^2 = Var(Y) = E(Y - E(Y))^2 = E(Y-\mu)^2 = E(Y^2) - \mu^2$$

■ 공분산 (확률변수 Y_1, Y_2)

$$\sigma_{12} = Cov(Y_1, Y_2) = E(Y_1 - \mu_1)(Y_2 - \mu_2) = E(Y_1 Y_2) - \mu_1 \mu_2$$

■ 상관계수 (확률변수 Y_1, Y_2)

$$\rho_{12} = Corr(Y_1, Y_2) = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)}\,\sqrt{Var(Y_2)}}$$

22

확률변수

□ 확률변수 Y (a, b: 상수)

- $E(a) = a$
- $E(aY+b) = aE(Y)+b$
- $Var(aY+b) = a^2 Var(Y)$

□ 확률변수 Y_1, Y_2 (a_1, a_2, b_1, b_2: 상수)

- $E(a_1 Y_1 + a_2 Y_2) = a_1 E(Y_1) + a_2 E(Y_2)$
- $Var(a_1 Y_1 + a_2 Y_2) = a_1^2 Var(Y_1) + a_2^2 Var(Y_2) + 2a_1 a_2 Cov(Y_1, Y_2)$
- $Cov(a_1 Y_1 + b_1,\ a_2 Y_2 + b_2) = a_1 a_2 Cov(Y_1, Y_2)$
- $Corr(a_1 Y_1 + b_1,\ a_2 Y_2 + b_2) = sign\{(a_1 a_2)\}\ Corr(Y_1, Y_2)$

23

확률변수

확률변수가 여러개 있을때

□ 확률변수 $Y_1, \cdots, Y_n$ ($a_1, \cdots, a_n$: 상수)

a가 모두 $\frac{1}{n}$인경우 : 평균

$$a_1 Y_1 + \cdots + a_n Y_n = \sum_{i=1}^{n} a_i Y_i$$: 선형결합 (linear combination)

- $E\left(\sum_{i=1}^{n} a_i Y_i\right) = E(a_1 Y_1 + \cdots + a_n Y_n) = a_1 E(Y_1) + \cdots + a_n E(Y_n) = \sum_{i=1}^{n} a_i E(Y_i)$
- $$Var\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + \sum_{i \neq j} a_i a_j Cov(Y_i, Y_j)$$
$$= a_1^2 Var(Y_i) + a_2^2 Var(Y_2) + \ldots + a_n^2 Var(Y_n)$$
$$+ a_1 a_2 Cov(Y_1, Y_2) + a_1 a_3 Cov(Y_1, Y_3) + \cdots + a_{n-1} a_n Cov(Y_{n-1}, Y_n)$$

$$Var\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(Y_i, Y_j)$$
$$= a_1 a_1 Cov(Y_1, Y_1) + a_1 a_2 Cov(Y_1, Y_2) + \cdots + a_n a_n Cov(Y_n, Y_n)$$

24

확률변수

- $Cov(Y_i, Y_j) = 0$이면

$$Var\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + \sum_{i \neq j} a_i a_j Cov(Y_i, Y_j)$$

$$= \sum_{i=1}^{n} a_i^2 Var(Y_i)$$

$$= a_1^2 Var(Y_i) + a_2^2 Var(Y_2) + \ldots + a_n^2 Var(Y_n)$$

- $Cov(\sum_{i=1}^{n} a_i Y_i, \sum_{j=1}^{m} b_j X_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j)$

- $Cov(Y_1 + Y_2,\ Y_3) = Cov(Y_1,\ Y_3) + Cov(Y_2,\ Y_3)$

25

확률변수

생각해보기

(example)

$n = 3,\ m = 2,\quad \sum_{i=1}^{3} a_i Y_i = a_1 Y_1 + a_2 Y_2 + a_3 Y_3,\quad \sum_{j=1}^{2} b_j X_j = b_1 X_1 + b_2 X_2$

공분산

- $Var\left(\sum_{i=1}^{3} a_i Y_i\right) = \sum_{i=1}^{3} a_i^2 Var(Y_i) + \sum_{i \neq j} a_i a_j Cov(Y_i, Y_j)$

$$= a_1^2 Var(Y_1) + a_2^2 Var(Y_2) + a_3^2 Var(Y_3)$$

$$+ a_1 a_2 Cov(Y_1, Y_2) + a_1 a_3 Cov(Y_1, Y_3) + a_2 a_1 Cov(Y_2, Y_1)$$

$$+ a_2 a_3 Cov(Y_2, Y_3) + a_3 a_1 Cov(Y_3, Y_1) + a_3 a_2 Cov(Y_3, Y_2)$$

같음!

✓ $Cov(\sum_{i=1}^{3} a_i Y_i, \sum_{j=1}^{2} b_j X_j) = \sum_{i=1}^{3}\sum_{j=1}^{2} a_i b_j Cov(Y_i, X_j)$

$$= a_1 b_1 Cov(Y_1, X_1) + a_1 b_2 Cov(Y_1, X_2)$$

$$+ a_2 b_1 Cov(Y_2, X_1) + a_2 b_2 Cov(Y_2, X_2)$$

$$+ a_3 b_1 Cov(Y_3, X_1) + a_3 b_2 Cov(Y_3, X_2)$$

꼭기억할필요X
나중에다시나옴
이해만하기!

26