실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

정규분포 및 그 파생분포

$X_1, X_2, \ldots, X_n$ 이 정규분포 $N(\mu, \sigma^2)$에서의 확률표본이라고 하자.

확률표본(random sample): 표본내의 모든 자료가 독립이고 분포가 같음.
iid(independent and identically distributed)라고도 함.

표준화

$E(X) = \mu$, $Var(X) = \sigma^2 < \infty$인 임의의 확률변수 X에 대해서

$$Z = \frac{X - \mu}{\sigma}$$

를 표준화라고 한다. 표준화된 확률변수 Z는 $E(Z) = 0$, $Var(Z) = 1$이다.

$X \sim N(\mu, \sigma^2)$이면 $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$이며 기댓값이 0 분산이 1인 정규분포는 표준정규분포라고 한다.

정규분포의 선형조합은 정규분포를 갖는다.

정규분포 및 그 파생분포

표본평균의 분포

$X_1, X_2, \ldots, X_n$**이** $E(X_i) = \mu$, $Var(X_i) = \sigma^2$인 확률표본이면 표본평균 $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$의 기댓값과 분산은 각각 $E(\overline{X}) = \mu$, $Var(\overline{X}) = \frac{\sigma^2}{n}$ 이다.

- $X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$에서의 확률표본이면 표본평균은 다시 정규분포를 갖는다. 즉, $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
- $\overline{X}$를 표준화하면 $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$이다.

정규분포 및 그 파생분포

카이제곱 분포(chi square distribution)

χ

- 정의: $Z_1, Z_2, \ldots, Z_\nu$ 이 $N(0,1)$에서의 확률표본이면
  $\chi^2_\nu = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2$의 분포를 자유도 ν인 카이제곱분포라고 한다.
- 독립인 카이제곱분포 $\chi^2_{\nu_1}$과 $\chi^2_{\nu_2}$의 합은 다시 카이제곱분포이며 자유도는 각각의 자유도의 합이다. 즉, $\chi_{\nu_1} + \chi_{\nu_2} \sim \chi_{\nu_1+\nu_2}$ (Cochran 정리)
- $X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$에서의 확률표본이면 $\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{\sum (X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$ 이다.
- μ 대신 $\overline{X}$를 사용한 $\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$이다. 단, $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

$\to \frac{\sum x_i}{n}$

$\uparrow x_i -$

정규분포 및 그 파생분포

t-분포

- **정의**: Z와 χ^2_ν이 독립이고 각각 표준정규분포와 자유도 ν인 카이제곱분포를 따르면

  $T = \dfrac{Z}{\sqrt{\chi^2_\nu/\nu}} \sim t_\nu$ 이다. t-분포의 자유도는 카이제곱분포의 자유도에 의해 결정된다.

- $X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$에서의 확률표본이면 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$이고 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

  이므로 $T = \dfrac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}/(n-1)}} = \dfrac{\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma}}{\sqrt{\dfrac{S^2}{\sigma^2}}} = \dfrac{\sqrt{n}(\overline{X}-\mu)}{\sqrt{S^2}}$

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

정규분포 및 그 파생분포

F-분포

- **정의**: $\chi^2_{\nu_1}$과 $\chi^2_{\nu_2}$를 독립이고 각각 자유도가 ν_1, ν_2인 카이제곱분포를 따르는 확률변수라고 하면 $F = \dfrac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2} \sim F_{\nu_1,\nu_2}$ 표시.
- 정의에 의해서 $F_{\nu_1,\nu_2} = \dfrac{1}{F_{\nu_2,\nu_1}}$이다. 따라서 $F_{\nu_1,\nu_2;\alpha} = \dfrac{1}{F_{\nu_2,\nu_1;1-\alpha}}$이다. 예를 들어

  $$F_{8,2;0.95} = 0.0516 = \frac{1}{F_{8,2;0.05}} = \frac{1}{19.3710} = 0.0516$$

  (R-코드로 확인)

```
> qf(0.05, 2, 8)
[1] 0.05162358
> qf(0.95, 8, 2)
[1] 19.37099
```

정규분포 및 그 파생분포

```
> 1/qf(0.95, 8, 2)
[1] 0.05162358
```

- 자유도 ν인 t분포의 제곱은 $F_{1,\nu}$과 같은 분포. $t_\nu^2 = \left(\frac{Z}{\sqrt{\chi_\nu^2/\nu}}\right)^2 = \frac{Z^2}{\chi_\nu^2/\nu} = \frac{\chi_1^2/1}{\chi_\nu^2/\nu} \sim F_{1,\nu}$

$F_{1,10;0.05} = 4.964603 = t_{10,0.025}^2 = 2.228139^2$

(R-언어)

```
> qf(0.95, 1,10)
[1] 4.964603
> qt(0.975,10)
[1] 2.228139
> qt(0.975,10)^2
[1] 4.964603
```