

## Lecture 11 - Hypothesis Testing 3

Sim, Min Kyu, Ph.D., [mksim@seoultech.ac.kr](mailto:mksim@seoultech.ac.kr)



서울과학기술대학교 IT정책전문대학원

## 1 I. 공분산과 상관계수 (Covariance and Correlation)

## 2 II. 모상관계수에 대한 검정 (단일모집단)

## 3 III. 카이제곱 (Chi-square, $\chi^2$ ) 분포

## 4 IV. 모분산에 대한 검정 (단일모집단)

## I. 공분산과 상관계수 (Covariance and Correlation)

## Recap - The taxonomy

	단일모집단	두개의 모집단
모평균에 대한 검정	L09, Sec. I	L09, Sec. III
모비율에 대한 검정	L09, Sec. II	L10, Sec. II
모상관계수에 대한 검정	L11, Sec. II	
모분산에 대한 검정	L11, Sec. IV	

## Definition

- 공분산(covariance)

- 두 확률변수가 얼마나 같이 움직이는지에 대한 경향을 측정한다.
- 각 변수의 편차의 곱의 기대값으로 정의된다.

- 상관계수(correlation coefficient)

- 두 확률변수의 상관의 정도를 -1과 1사이의 수치로 나타낸 값이다.
- 공분산을 각각의 두 변수의 표준편차로 나눈 값이다.

- Recap

- $Var(X + Y) = Var(X) + Var(Y) + 2 \underbrace{Corr(X, Y)sd(X)sd(Y)}_{Cov(X, Y)}$
- $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2 \underbrace{\rho_{XY}\sigma_X\sigma_Y}_{\sigma_{XY}}$

- Covariance

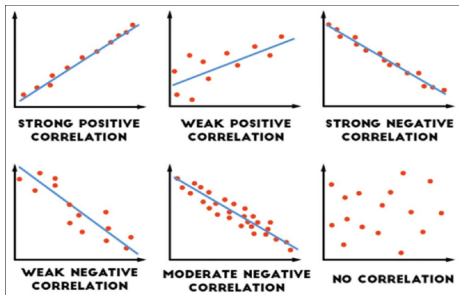
- $Cov(X, Y) = Corr(X, Y)sd(X)sd(Y)$
- $\sigma_{XY} = \rho_{XY}\sigma_X\sigma_Y$

- Correlation

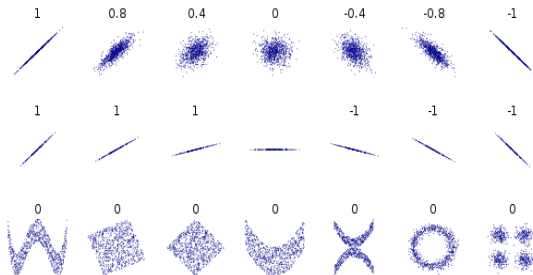
- $Corr(X, Y) = \frac{Cov(X, Y)}{sd(X)sd(Y)}$
- $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$

# Correlation

## 상관관계의 다른 샘플의 형태



## 상관관계의 다른 샘플의 형태



# Parameters and their sample estimators

## Parameters (true, population)

- (true) Mean
  - $EX = \mu_X$
- (true) Variance
  - $Var(X) = \sigma_X^2 = E[(X - EX)^2]$  ( $= E[(X - EX)(X - EX)] = Cov(X, X)$ )
- (true) Covariance
  - $Cov(X, Y) = \sigma_{XY} = E[(X - EX)(Y - EY)]$
- (true) Correlation
  - $Corr(X, Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$



## Sample Estimators

- (sample) Mean

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

- (sample) Variance

- $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X})(X_i - \bar{X}))$

- (sample) Covariance

- $s_{XY} = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X})(Y_i - \bar{Y}))$

- (sample) Correlation

- $r_{XY} = \frac{s_{XY}}{s_X s_Y}$

## Summary

True Parameter	Sample Estimator
Mean	
Variance	
Covariance	
Correlation	

# Some math

## On sample variance

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X})(X_i - \bar{X})) \quad (1)$$

$$= \frac{1}{n-1} \left( \sum_{i=1}^n (X_i X_i - X_i \bar{X} - \bar{X} X_i + \bar{X} \bar{X}) \right) \quad (2)$$

$$= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - \sum_{i=1}^n X_i \bar{X} - \sum_{i=1}^n \bar{X} X_i + \sum_{i=1}^n \bar{X}^2 \right) \quad (3)$$

$$= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - \bar{X} \sum_{i=1}^n X_i - \bar{X} \sum_{i=1}^n X_i + n \bar{X}^2 \right) \quad (4)$$

$$= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n \bar{X}^2 - n \bar{X}^2 + n \bar{X}^2 \right) \quad (5)$$

$$= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n \bar{X}^2 \right) \quad (6)$$

## On sample covariance

$$\begin{aligned}
 s_{XY} &= \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X})(Y_i - \bar{Y})) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ( \quad ) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ( \quad ) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ( \quad ) \\
 &= \frac{1}{n-1} \sum_{i=1}^n ( \quad ) \\
 &= \frac{1}{n-1} \left( \sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y} \right)
 \end{aligned}$$

- 같은 이유로 아래의 성질이 성립됨

- Variance

- $Var(X) = \sigma_X^2 = E[(X - EX)(X - EX)] = E(X^2) - (EX)^2$

- Covariance

- $Cov(X, Y) = \sigma_{XY} = E[(X - EY)(X - EY)] = E(XY) - EX \cdot EY$

# Sample statistics를 구할때에 편리한 “big S” 표기

## Notations

$$\begin{aligned}
 SXX &= \sum (X_i - \bar{X})^2 \\
 &= \sum X_i^2 - n\bar{X}^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n} \\
 SY Y &= \sum (Y_i - \bar{Y})^2 \\
 &= \sum Y_i^2 - n\bar{Y}^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \\
 SXY &= \sum ((X_i - \bar{X})(Y_i - \bar{Y})) \\
 &= \sum X_i Y_i - n\bar{X}\bar{Y} = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}
 \end{aligned}$$

## Sample Estimators들의 “big S” 표기

- (sample) Mean

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

- (sample) Variance

- $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X})(X_i - \bar{X})) = \frac{S_{XX}}{n-1}$

- (sample) Covariance

- $s_{XY} = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X})(Y_i - \bar{Y})) = \frac{S_{XY}}{n-1}$

- (sample) Correlation

- $r_{XY} = \frac{s_{XY}}{s_X s_Y}$

I. 공분산과 상관계수 (Covariance and Correlation)

oooooooooooo●

II. 모상관계수에 대한 검정 (단일모집단)

ooooo

III. 카이제곱 (Chi-square,  $\chi^2$ ) 분포

oooooo

IV. 모분산에 대한 검정 (단일모집단)

ooooo



## II. 모상관계수에 대한 검정 (단일모집단)

## $\rho_{XY}$ 에 대한 검정

- 분포: 두 확률 변수  $X$ 와  $Y$ 의 분포가 정규분포와 가까울 때에 성립

$$\sqrt{n-2} \frac{r_{XY}}{\sqrt{1-r_{XY}^2}} \sim t_{n-2}$$

- 검정통계량

$$T = \sqrt{n-2} \frac{r_{XY}}{\sqrt{1-r_{XY}^2}}$$

- $100(1-\alpha)\%$  신뢰구간

$$-t_{\alpha/2, n-2} \leq \sqrt{n-2} \frac{\rho_{XY}}{\sqrt{1-\rho_{XY}^2}} \leq t_{\alpha/2, n-2}$$

- 주의

- $\rho_{XY}$ 에 관하여 풀어내기가 쉽지 않다.
- (Fisher's z-transformation이라는 기법을 사용해야 함.)
- 그렇기에 가설 검정은  $\rho$ 와 0에 대한 비교만이 가능하다.

- Case 1. 양측 검정

- 가설:  $H_0 : \rho_{XY} = 0$  vs  $H_A : \rho_{XY} \neq 0$
- 기각역:  $|T| \geq t_{\alpha/2, n-2}$

- Case 2. 단측 검정

- 가설:  $H_0 : \rho_{XY} \leq 0$  vs  $H_A : \rho_{XY} > 0$
- 기각역:  $T \geq t_{\alpha, n-2}$

- Case 3. 단측 검정

- 가설:  $H_0 : \rho_{XY} \geq 0$  vs  $H_A : \rho_{XY} < 0$
- 기각역:  $T \leq -t_{\alpha, n-2}$

## Exercise 1

- 10명을 대상으로 왼손과 오른손의 악력을 측정한 결과가 아래와 같다.
- 양손의 악력에 상관관계가 있는지 90% 수준에서 검정하라.

```
X <- c(25, 24, 35, 34, 33, 26, 40, 38, 23, 41) # Left
Y <- c(22, 27, 40, 41, 35, 35, 47, 42, 24, 38) # right
```

```
SXY <- sum((X-mean(X))*(Y-mean(Y)))
```

```
SXY
```

```
## [1] 446.1
```

```
sd(X)
```

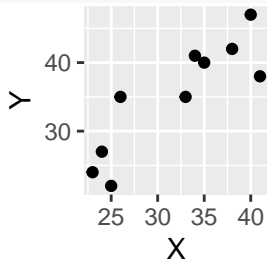
```
## [1] 6.871034
```

```
sd(Y)
```

```
## [1] 8.279157
```

```
library(ggplot2)
```

```
qplot(X,Y)
```



I. 공분산과 상관계수 (Covariance and Correlation)

oooooooooooooooo

II. 모상관계수에 대한 검정 (단일모집단)

oooo●

III. 카이제곱 (Chi-square,  $\chi^2$ ) 분포

oooooo

IV. 모분산에 대한 검정 (단일모집단)

ooooo

### III. 카이제곱 (Chi-square, $\chi^2$ ) 분포

# Development

## 정규분포 review

- $Z \sim N(0, 1)$  이므로  $EZ = 0$  와  $Var(Z) = 1$  이 성립한다.
- $Var(Z) = E(Z^2) - (EZ)^2$  이므로  $E(Z^2) = 1$  도 성립한다.

## 정규분포의 제곱합

- 표준정규분포 따르는 변수  $k$  개가 있다고 하자. 이를  $Z_1, Z_2, \dots, Z_k$  라 쓴다.
- 이들의 제곱합을  $Q$  라고 정의하자. 즉,

$$Q := Z_1^2 + Z_2^2 + \dots + Z_k^2$$

- $EQ$  는 얼마인가?

## 카이제곱 (Chi-square, $\chi^2$ ) 분포

- 표준정규분포의 제곱합  $Q (= Z_1^2 + Z_2^2 + \dots + Z_k^2)$ 는 자유도가  $k$ 인 카이제곱 분포를 따른다. 즉,

$$Q \sim \chi_k^2$$

- $EQ = k$ 이고  $Var(Q) = 2k$ 이다.

## 카이제곱의 pdf (몰라도 됨)

### Probability density function [\[ edit \]](#)

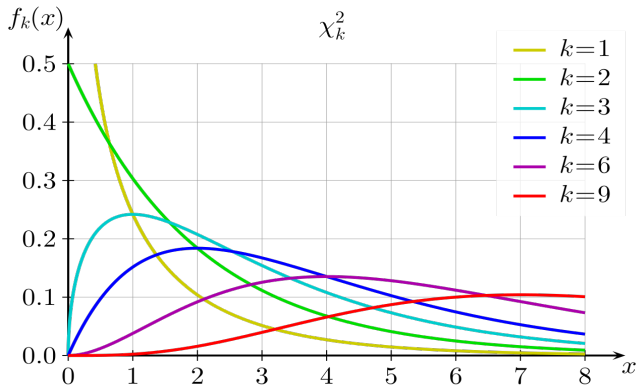
The [probability density function](#) (pdf) of the chi-square distribution is

$$f(x; k) = \begin{cases} \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

where  $\Gamma(k/2)$  denotes the [gamma function](#), which has closed-form values for integer  $k$ .



## 자유도에 따른 카이제곱의 pdf (몰라도 됨)



## 카이제곱 분포표

	P										
DF	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620

I. 공분산과 상관계수 (Covariance and Correlation)

○○○○○○○○○○○○○○○○

II. 모상관계수에 대한 검정 (단일모집단)

○○○○○

III. 카이제곱 (Chi-square,  $\chi^2$ ) 분포

○○○○○●

IV. 모분산에 대한 검정 (단일모집단)

○○○○○

## IV. 모분산에 대한 검정 (단일모집단)

## 단일모집단의 모분산에 대한 검정

- 분포: 모집단  $X \sim N(\mu, \sigma^2)$ 에서 추출된  $n$ 개의 표본  $\{X_1, X_2, \dots, X_n\}$ 에 대해서 아래의 성질이 성립한다.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

- 검정통계량

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

- $100(1-\alpha)$  신뢰구간

$$\chi_{1-\alpha/2, n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{\alpha/2, n-1}^2$$

- ( $\sigma$ 에 대하여 풀면)

$$\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2}$$

- Case 1. 양측검정

- 가설:  $H_0 : \sigma^2 = \sigma_0^2$  vs  $H_A : \sigma^2 \neq \sigma_0^2$
- 기각역:  $\chi_0^2 \leq \chi_{1-\alpha/2, n-1}^2$  or  $\chi_0^2 \geq \chi_{\alpha/2, n-1}^2$

- Case 2. 단측검정

- 가설:  $H_0 : \sigma^2 \leq \sigma_0^2$  vs  $H_A : \sigma^2 > \sigma_0^2$
- 기각역:  $\chi_0^2 \geq \chi_{\alpha, n-1}^2$

- Case 3. 단측검정

- 가설:  $H_0 : \sigma^2 \geq \sigma_0^2$  vs  $H_A : \sigma^2 < \sigma_0^2$
- 기각역:  $\chi_0^2 \leq \chi_{1-\alpha, n-1}^2$

## Exercise 2

- 앞의 문제 Exercise 1에서 주어진 왼손의 악력은 아래와 같다.

```
X
```

```
## [1] 25 24 35 34 33 26 40 38 23 41
```

- 왼손의 악력의 모분산이 25와 다른지 90% 수준에서 검정하라.

```
length(X) # n
```

```
## [1] 10
```

```
var(X) # sample variance
```

```
## [1] 47.21111
```

"This note is made with Rmarkdown"

```
## [1] "This note is made with Rmarkdown"
```