

P. Dalgaard, Introductory Statistics with R

CHAPTER 5

One and two Sample tests

발표자: 조현선

순서

- 5.1 One-sample t test
- 5.2 Wilcoxon-signed-rank test
- 5.3 Two-sample t test
- 5.4 Comparison of variance
- 5.5 Two-sample Wilcoxon test
- 5.6 The paired t-test
- 5.7 The matched-pairs Wilcoxon test

순서

- 5.1 One-sample t test 단일표본 검정
- 5.2 Wilcoxon-signed-rank test
- 5.3 Two-sample t test 이표본 검정
- 5.4 Comparison of variance
- 5.5 Two-sample Wilcoxon test
- 5.6 The paired t-test
- 5.7 The matched-pairs Wilcoxon test

순서

- 5.1 One-sample t test
- 5.2 Wilcoxon-signed-rank test
- 5.3 Two-sample t test
- 5.4 Comparison of variance
- 5.5 Two-sample Wilcoxon test
- 5.6 The paired t-test
- 5.7 The matched-pairs Wilcoxon test

단일표본 검정

Comparing continuous data between a group against an a priori stipulated value.

모집단에서 추출된 표본의 평균과 연구자가 이론적 배경이나 경험적 배경에 의하여 설정한 특정한 수를 비교하는 방법

순서

- 5.1 One-sample t test
- 5.2 Wilcoxon-signed-rank test

단일표본 검정

- 5.3 Two-sample t test
- 5.4 Comparison of variance
- 5.5 Two-sample Wilcoxon test
- 5.6 The paired t-test
- 5.7 The matched-pairs Wilcoxon

이표본 검정

Comparing continuous data between
Two groups.

두 모집단으로부터 표본들을 추출하여 표본들의
평균을 비교함으로써 모집단의 평균을 비교하는
통계적 방법

t-test

t 분포에 의존하여 의사결정을 하는 검정

기본 가정

- 1. 종속변수가 양적변수이어야 한다.
- 2. 모집단의 분산, 표준편차를 알지 못할 때 사용한다.
- 3. 모집단 분포가 정규분포이어야 한다.
- 4. 등분산 가정이 충족되어야 한다.

t-test

t 분포에 의존하여 의사결정을 하는 검정

기본 가정

- 1. 종속변수가 양적변수이어야 한다.
- 2. 모집단의 분산, 표준편차를 알지 못할 때 사용한다.
- 3. 모집단 분포가 정규분포이어야 한다.
- 4. 등분산 가정이 충족되어야 한다.

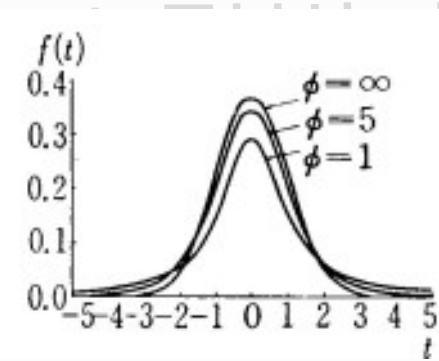
t-test

t 분포에 의존하여 의사결정을 하는 검정

기본 가정

- 1. 종속변수가 양적변수이어야 한다.
- 2. 모집단의 분산, 표준편차를 알지 못할 때 사용한다.

스튜던트 t 분포 : 모집단을 대표하는 표본으로부터 추정된 분산이나 표준편차를 이용



평균이 μ , 표준오차가 S_y / \sqrt{n}

→ 표본의 크기에 따라 변화

→ 표본의 크기가 작을수록

(=자유도 낮을수록) 봉이 낮아지고

꼬리가 높아지는 형태의 분포

t-test

t 분포에 의존하여 의사결정을 하는 검정

기본 가정

- 1. 종속변수가 양적변수이어야 한다.
- 2. 모집단의 분산, 표준편차를 알지 못할 때 사용한다.
- 3. 모집단 분포가 정규분포이어야 한다.



- 4. 등분산 가정이 충족되어야 한다.
만약, 모집단의 분포가 정규분포라는 가정을 충족시키지 못한다면,
비모수 통계 (non-parametric statistics)를 사용해야 한다.

→ Wilcoxon-test

t-test

t 분포에 의존하여 의사결정을 하는 검정

기본 가정

- 1. 종속변수가 양적변수이어야 한다.
- 2. 모집단의 분산, 표준편차를 알지 못할 때 사용한다.
- 3. 모집단 분포가 정규분포이어야 한다.
- 4. 등분산 가정이 충족되어야 한다.



두 집단의 분산이 같지 않아 **등분산 가정**을 충족시키지 **못할 경우에는**
두 독립표본 t검정 대신에
Welch-Aspin 검정을 사용해야 한다.

순서

- 5.1 One-sample t test 단일표본 검정
- 5.2 **Wilcoxon**-signed-rank test
- 5.3 Two-sample t test 이표본 검정
- 5.4 **Comparison of variance**
- 5.5 Two-sample **Wilcoxon** test
- 5.6 The paired t-test
- 5.7 The matched-pairs **Wilcoxon** test

• Wilcoxon signed-rank test

The t tests are fairly robust against departures from the normal distribution especially in large samples, but sometimes you wish to avoid making that assumption.

DISTRIBUTION FREE METHODS
ARE CONVENIENT.

윌콕슨 부호순위 검정

i	X_i	$X_i - 3.7$	$ X_i - 3.7 $	RANK R_i	SIGNED RANK	Z_i
1	5.0	1.3	1.3	5	5	1
2	3.9	0.2	0.2	1	1	1
3	5.2	1.5	1.5	6	6	1
4	5.5	1.8	1.8	7	7	1
5	2.8	-0.9	0.9	3	-3	0
6	6.1	2.4	2.4	9	9	1
7	6.4	2.7	2.7	10	10	1
8	2.6	-1.1	1.1	4	-4	0
9	1.7	-2.0	2.0	8	-8	0
10	4.3	0.6	0.6	2	2	1

$$\begin{aligned}
 W &= (1)(5) + (1)(1) + \dots + (0)(-8) + (1)(2) \\
 &= 5 + 1 + 6 + 7 + 9 + 10 + 2 \\
 &= 40
 \end{aligned}$$

	w_1^*	w_2^*	$P(W \leq w_1^*) = P(W \geq w_2^*)$
$n = 10$	3	52	0.005
	4	51	0.007
	5	50	0.010
	6	49	0.014
	7	48	0.019
	8	47	0.024
	9	46	0.032
	10	45	0.042
	11	44	0.053
	12	43	0.065
	13	42	0.080
$n = 11$	14	41	0.097
	15	40	0.116
	16	39	0.138

The relative merits of distribution-free (or nonparametric) versus parametric methods such as the t test are a contentious issue.

- **Robust**

#비모수 기법의 매력

#순위검정

#검증력이 떨어지는 단점

#But,극단치에 버티는 힘이 강함

P. Dalgaard, Introductory Statistics with R

CHAPTER 6

Regression and Correlation

발표자: 조현선

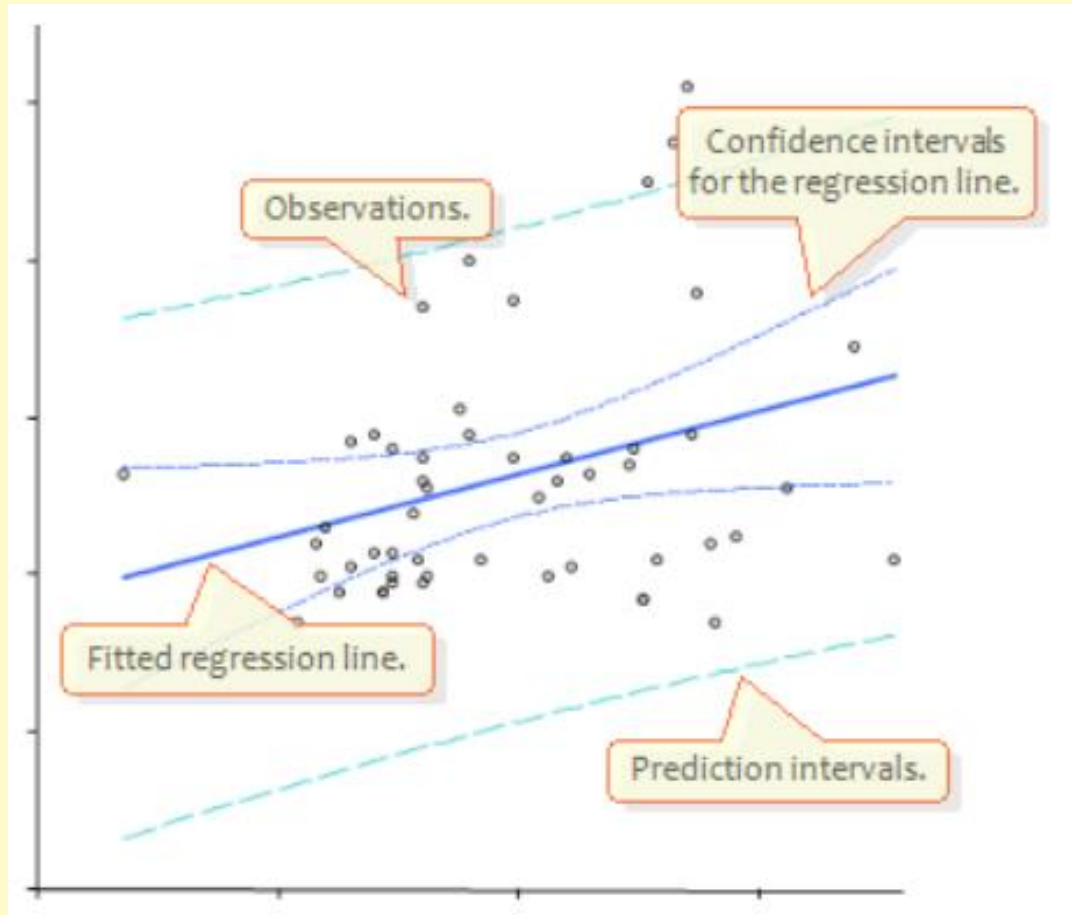
순서

- **6.1** Simple linear regression
- **6.2** Residual and fitted values
- **6.3** Prediction and confidence bands
- **6.4** Correlation
 - 6.4.1 Pearson correlation
 - 6.4.2 Spearman's ρ
 - 6.4.3 Kendall's τ

DATA

- thuesen {ISwR}
- Ventricular shortening velocity
- The thuesen data frame has 24 rows and 2 columns. It contains ventricular shortening velocity and blood glucose for type I diabetic patients.

6.3 Prediction and confidence bands



신뢰구간 :

- 주어진 X값에서 Y의 평균에 대한 추정
- $x = \bar{x}$ 일때 가장 좁고, \bar{x} 에서 멀어질 수록 넓어짐
- X의 각 점에서 신뢰구간 상한과 하한을 구하여 연결하면 회귀직선의 신뢰구간이 곡선 형태로 형성됨

예측구간 :

- 주어진 X값에서 나타날 수 있는 새로운 Y값(확률변수)에 대한 예측

• 6.4.1 Pearson correlation

두 변수가 이변량 정규분포

(bivariate normal distribution)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

두 변수의 선형적 연관성에 관한 측도이고, 정규모집단인 경우로 한정된다.

But,

두 변수 X,Y가 이변량 정규분포를 가정할 수 없을 때는 T-검정을 실시할 수 없다. 피어슨의 표본상관계수는 연속형 변수의 자료가 아닌 경우 즉 순위자료 등에서는 사용할 수 없다.

→ 비모수적 측도 필요

→ 순위상관분석

Spearman's ρ

Kendall's τ

• 6.4.2 Spearman's ρ

- 관측치의 순위에 근거하여 만들어진다.
- r 의 정의에서 X_i, Y_i 대신에 각각의 순위인 R_i, S_i 를 대입한 통계량으로 상관관계를 나타내는 척도로 사용할 수 있다.

$$r_s = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}}$$

• 6.4.3 Kendall's τ [켄달의 순위상관]

- 관측치의 순위에 근거하여 만들어진다.
 - Concordant pair : $(X_i - X_j)$ 와 $(Y_i - Y_j)$ 부호가 같음
 - Disconcordant pair : $(X_i - X_j)$ 와 $(Y_i - Y_j)$ 부호가 다름
- 일치쌍의 수 P 와 불일치쌍의 수를 Q 라고 할 때,
 불일치쌍의 수보다 많은 일치쌍의 수를
 K 라는 켄달 통계량이라고 한다.
 이를 가지고 독립성 검정을 실시한다.

*검정의 대립가설 : 상호 독립이 아니다.

월	1	2	3	4	5	6	7	8	9	10
광고비	1.2	0.8	1	1.3	0.7	0.8	1	0.6	0.9	1.1
매출	101	92	110	120	90	82	93	75	91	105



월	1	2	3	4	5	6	7	8	9	10
광고비 rank	2	7	4	1	9	7	4	10	6	3
매출 rank	4	6	2	1	8	9	5	10	7	3