

상관계수. 즉 상관계수
~~Covariance~~ Correlation
 ↳ 분산과 공분산의 비

→ 이분산성 변수에서 의미를 가짐

기분

$$\text{분산} = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

$$\text{표준편차} : \frac{X - \bar{X}}{S}$$

$$\text{공분산} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \Rightarrow \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y}) / (n-1)}{\sqrt{\frac{\sum (X_i - \bar{X})^2}{(n-1)}} \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{(n-1)}}} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}} = \text{상관계수} (r)$$

$$-1 \leq r \leq 1$$

$r = 0.084^{**}$ → Age와 balance의 상관계수는 0.084. 유의수준은 0.000 이므로 유의수준 0.05보다 상관계수가 작아
 대분. 상관계수가 작을수록 더욱 작은 상관계수가 존재할 가능성이 있다.

유의수준
 0.000

유의수준
 0.05 : *
 0.01 : **
 0.001 : ***

귀무: $r = 0$; 유의수준이 유의수준보다 작을 경우 상관계수의 값이 존재함.
 (귀무가 아닌 의미를 가짐)

연속형 vs 범주형
 $X_1, X_2 \dots X_n$
 age, duration

Y

y

duration

y...

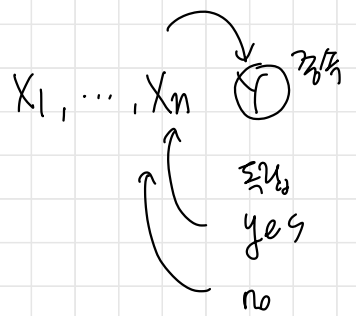
yes
521

no
4000

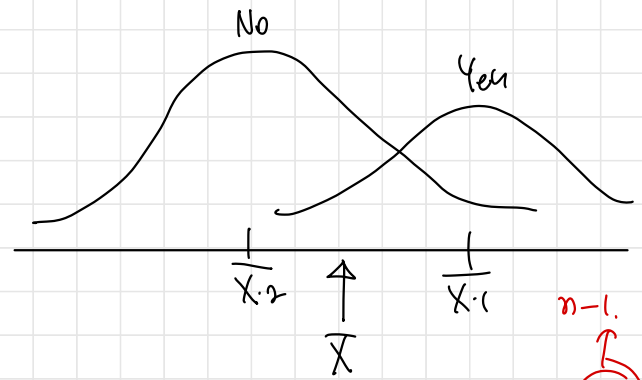
ANOVA

yes	no
X_{11}	X_{12}
\vdots	\vdots
X_{n1}	X_{n2}
$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$
\bar{X}	

$n_1 = 521$
 $n_2 = 4000$



H_0 : 두 집단간의 평균치에 차이가 없다 (연속형 X)



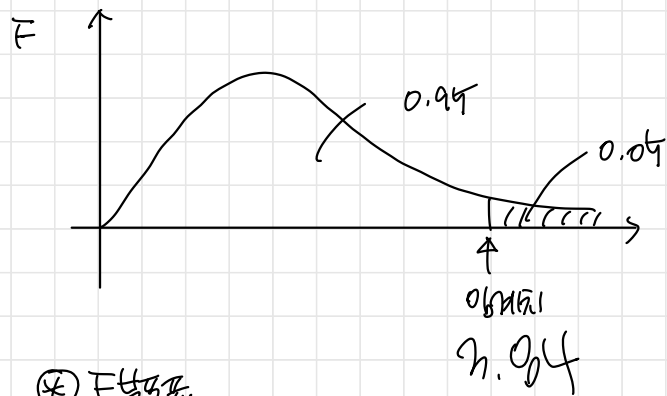
모집단분산(F) = $\frac{\text{집단간 제곱합} / \text{집단수} - 1}{\text{집단내 제곱합} / \text{집단수} - 1}$

★ $\left\{ \begin{array}{l} \sum_j \sum_i (X_{ij} - \bar{X})^2 : \text{총변동 (총평균과의 차이)} \\ \sum_j \sum_i (X_{ij} - \bar{X}_j)^2 : \text{집단내변동 = 집단내 제곱합} \\ \sum_j n_j (\bar{X}_j - \bar{X})^2 : \text{집단간변동 = 집단간 제곱합} \end{array} \right\}$

• $n_0 \text{변동} = \text{집단내 변동} + \text{집단간 변동}$

$n_1 - 1 = 520$
 $n_2 - 1 = 3999$
 $n_1 - 1 + n_2 - 1 = 4519$
 $2 - 1 = 1$

평균이 고정되어
 있다고 가정.



⊛ F 曲线

0.99

중간고사 문제 ppt 1075 01월. 6/26(토) 방 안까지 지킴 *

→ 4.6. 7가지에 대해 이해하는 것.

매우 쉽게 이해할 수 있다.

7가지에 대해 이해: 7가지에 대해 → 7가지에 대해

6가지 " : 6가지에 대해 → 6가지에 대해 (7가지에 대해) 8가지 Job. ~~marital education~~.

7가지 " : 6가지에 대해 ANOVA.
↓

moderating job

contact.

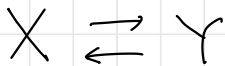
4가지: 7가지에 대해, box-plot.

~~default~~ housing loan.

Contact. part come.

or age balance duration of
campaign plays previous

part come <
contact <
housing <
job <
loan <
marital



반응물

반응물

결과. K^2

H_0

H_2

결과 차이 검정. 분산분석 (ANOVA). F
ANOVA 표 1

H_2

H_2

결과 분석

H_2

H_0

결정분석 결과 분석. 의사결정 나무 분석 (DT)

이제.

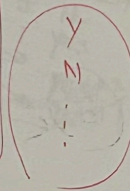
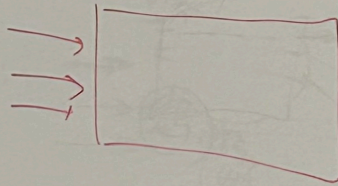
+ 결과. Box-plot 상관관계. 공분산
($H_0 - H_2$) ($H_2 - H_2$)

$$X \iff Y$$

4.521

$x_1 \dots x_n$	Y (yes. No)
	yes
	No
	No
	yes
	⋮

Rules



if \sim then \sim

DT.