Akshay
Lec 7

Date:
Page No.

Inference regarding two populations.

A $\quad$ $B \leftarrow n_B$ $\qquad H_0 = \underset{\sim}{\mu}_A = \underset{\sim}{\mu}_B$

$n_A \rightarrow$ $\quad \underset{\sim}{\bar{X}}_A \qquad \underset{\sim}{\bar{X}}_B \qquad H_1 = \underset{\sim}{\mu}_A \neq \underset{\sim}{\mu}_B$

$\Sigma_A \qquad \Sigma_B$

CR for $\underset{\sim}{\mu}_A - \underset{\sim}{\mu}_B$

$$E(\underset{\sim}{\bar{X}}_A - \underset{\sim}{\bar{X}}_B) = \underset{\sim}{\mu}_A - \underset{\sim}{\mu}_B$$

Multivariate Normal populations, small samples

$\Sigma_A = \Sigma_B$ but unknown

univariate

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ (unknown), small sample

$$\hat{\sigma}^2 = s^2_{pooled} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Multivariate

$\Sigma = \Sigma_A = \Sigma_B =$ unknown

$$\hat{\Sigma} = S_{pooled} = \frac{(n_A-1)S_A + (n_B-1)S_B}{n_A+n_B-2}$$

Date:
Page No.

Small sample.

$$X_A \sim N_p(\underset{\sim}{\mu}_A, \Sigma_A) \quad , \quad X_B \sim N_p(\underset{\sim}{\mu}_B, \Sigma_B)$$

$$\Sigma_A = \Sigma_B = \text{unknown}$$

Statistic $$T^2 = \left[(\underset{\sim}{\bar{X}}_A - \underset{\sim}{\bar{X}}_B) - (\underset{\sim}{\mu}_A - \underset{\sim}{\mu}_B)\right]^T \left[\left(\frac{1}{n_A} + \frac{1}{n_B}\right) S_{pooled}\right]^{-1} \left[(\underset{\sim}{\bar{X}}_A - \underset{\sim}{\bar{X}}_B) - (\underset{\sim}{\mu}_A - \underset{\sim}{\mu}_B)\right]$$

Sampling Dist^n $$\frac{(n_A + n_B - 2)p}{n_A + n_B - p - 1} \cdot F_{p, n_A + n_B - p - 1}^{(\alpha)}$$

$$CR: T^2 \leq$$

Case 2

large sample $$X_A \sim N_p(\underset{\sim}{\mu}_A, \Sigma_A), X_B \sim N_p(\underset{\sim}{\mu}_B, \Sigma_B), \Sigma_A \neq \Sigma_B \text{ unknown}$$

Statistic $$T^2 = \left[(\underset{\sim}{\bar{X}}_A - \underset{\sim}{\bar{X}}_B) - (\underset{\sim}{\mu}_A - \underset{\sim}{\mu}_B)\right]^T \left[\left(\frac{S_A}{n_A} + \frac{S_B}{n_B}\right)\right]^{-1} \left[(\underset{\sim}{\bar{X}}_A - \underset{\sim}{\bar{X}}_B) - (\underset{\sim}{\mu}_A - \underset{\sim}{\mu}_B)\right]$$

χ^2_p $$CR: T^2 \leq \chi^2_p(\alpha)$$

Date:
Page No.

Regression Analysis

Regression analysis is a statistical tool for investigating the relationship between a dependent variable and one or more independent variables.

Example You are marketing analyst at amazon

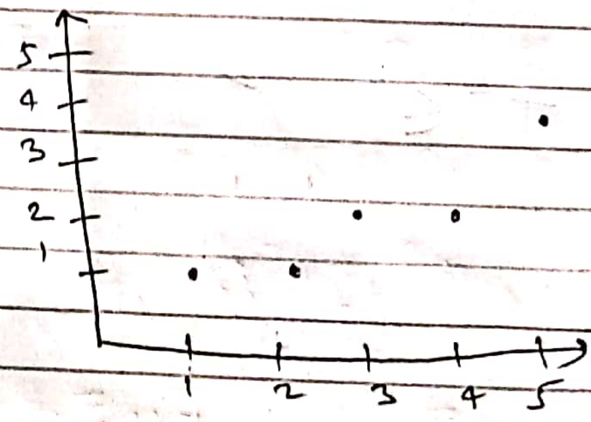
Independent variable / Control var / Regressor var — X

✓Ad	Sales. → dependent variable / response variable
1	1
2	1
3	2
4	2
5	4

What is the relationship between Sales and advertising?

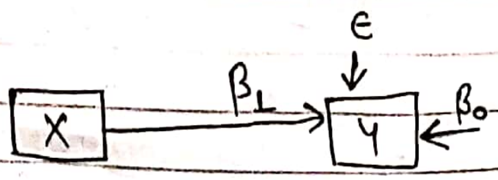
→ Y (response Variable)

X_i	Y_i
1	1
2	1
3	2
4	2
5	4



Simple linear regression model is a model with single regressor X

model $\boxed{Y = \beta_0 + \beta_1 X + \epsilon}$



$Y \rightarrow$ Response variable
$X \rightarrow$ Regressor variable
$\epsilon \rightarrow$ Error

There are other regressors influencing Y which we don't know.
ϵ Random error component.

Change in 1 unit of Cause change of β_1 unit in Y

$\beta_1 \rightarrow$ Slope
$\beta_0 \rightarrow$ Intercept

for given X, the corresponding observation Y consists of $\beta_0 + \beta_1 X$ plus random amount ϵ

(X_i, Y_i) $\quad Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, \ldots n$

1→ ϵ_i is a r.v with zero mean and variance σ^2 (unknown)

$E(\epsilon_i) = 0 \quad , \quad var(\epsilon_i) = \sigma^2$

2→ ϵ_i & ϵ_j are uncorrelated. $Cov(\epsilon_i, \epsilon_j) = 0$

3→ $\epsilon_i \overset{\text{independent}}{\sim} N(0, \sigma^2)$

$Y_i \overset{\text{independent}}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2)$

Homoscedasticity