# Lecture 6: Linear Regression Approach to ANOVA

## STA 106: Analysis of Variance

Xiner Zhou
Department of Statistics, University of California, Davis

# Linear Regression Approach to ANOVA

Regression Formulation to Single-Factor Studies

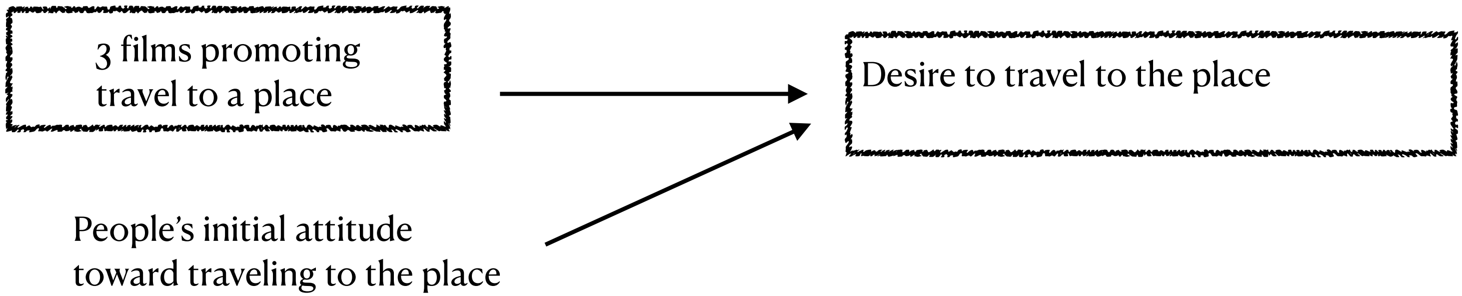
Linear Regression  Formulation to Two-Factor Studies

Analysis of Covariance Model ( ANCOVA)

# Example

(The Travel Promotion Study)

The objective of the study:




The study setup:

15 participants were randomly and equally split into 3 groups, with each group viewing one film.

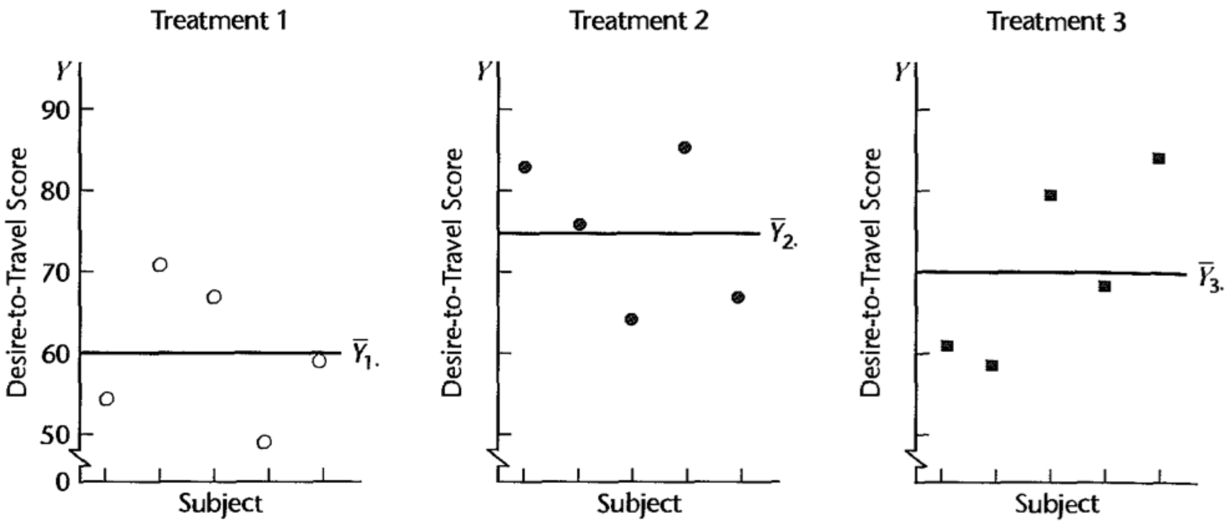
Their initial attitude toward traveling to the place and the desire after viewing the assigned promotion film were recorded.

What type of study?

# Example

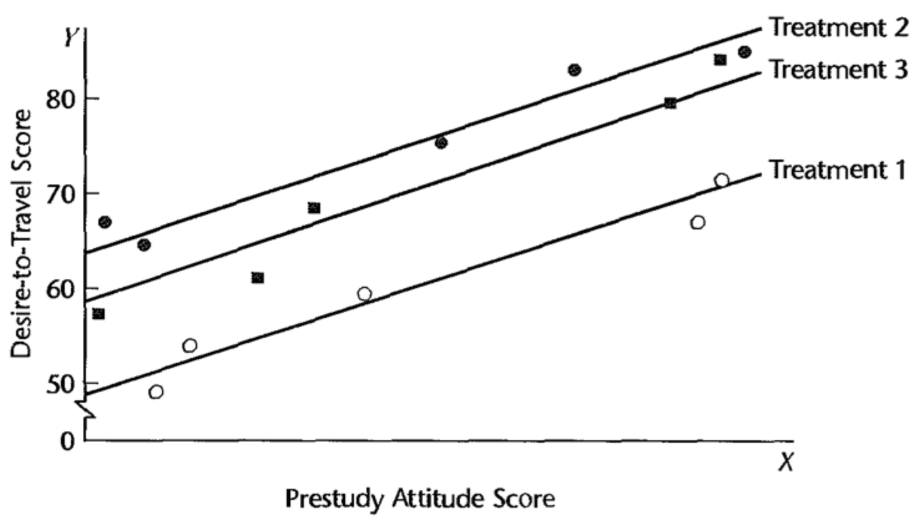
(The Travel Promotion Study)

(a) Error Variability with Single-factor Analysis of Variance Model




Error term: scatter around the treatment means are fairly large

(b) Error Variability with Covariance Analysis Model




Prestudy attitude is highly suggestive of post-study attitude, in fact, they are highly linearly associated.

If we incorporate pre study attitude, the error variability is much smaller, which will lead to more precise inference.

# Analysis of Covariance (ANCOVA)

## ANCOVA

Utilizes the relationship between the response variables and one or more variables, to
reduce the error term variability,
make more precision inference,
and make the study a more powerful one for comparing treatment effects

### Concomitant variables

We call variables added to ANOVA model in order to reduce error term variability:
Concomitant variable or Covariate (natural accompanying or occurring)

Must have relation to the response, otherwise, no good
Human subjects: pre-study score, age, socioeconomic status, aptitude....

### Difference between Blocking versus ANCOVA?

Blocking is used in the design stage, when we have one or more qualitative variables, to reduce the error term variability
ANCOVA is used in the analysis stage

# Single-Factor ANCOVA Model

ANCOVA model starts with ANOVA model, adds one more term reflecting the relation between Y and concomitant variable X:

One-Way ANOVA Model $Y_{ij} = \mu_{.} + \tau_i + \epsilon_{ij}$

$$Y_{ij} = \mu_{.} + \tau_i + \beta X_{ij} + \varepsilon_{ij}$$

Regression coefficient describing relation between Y ~ X

Centered version: $Y_{ij} = \mu_{.} + \tau_i + \beta(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij}$

# Single-Factor ANCOVA Model

ANCOVA Model: $$Y_{ij} = \mu. + \tau_1 I_{ij,1} + \ldots + \tau_{r-1} I_{ij,r-1} + \beta X_{ij} + \varepsilon_{ij}$$

Where

$$I_1 = \begin{cases} 1 \text{ if store received treatment } 1 \\ -1 \text{ if store received treatment } 3 \\ 0 \text{ otherwise} \end{cases}$$

$$I_{r-1} = \begin{cases} 1 \text{ if store received treatment } r-1 \\ -1 \text{ if store received treatment } 3 \\ 0 \text{ otherwise} \end{cases}$$

$$X_{ij} = X_{ij} - \bar{X}_{..}$$

There is no interest in whether the regression coefficient beta is zero or not, we are not interested in finding out its relation with Y.

We just want to use this relation to reduce error variability.

If beta is indeed not zero, i.e. there is a regression relation between X and Y, good
If not, no bias is introduced

# Single-Factor ANCOVA Model

Comparisons of Treatment Effects

ANOVA: $E\left(\eta_{ij}\right) = \mu_i = \mu. + \tau_i$

All observations for ith treatment has the same mean response, therefore comparing treatment effect is just comparing treatment means

ANCOVA: $E\left(\eta_{ij}\right) = \mu. + \tau_i + \beta\left(X_{ij} - \bar{X}_{..}\right)$

Mean response not only depend on the treatment, but also on concomitant variable X

A meaningful measure the effect of treatment 1 versus treatment 2 have to let the value of concomitant X to be the same:

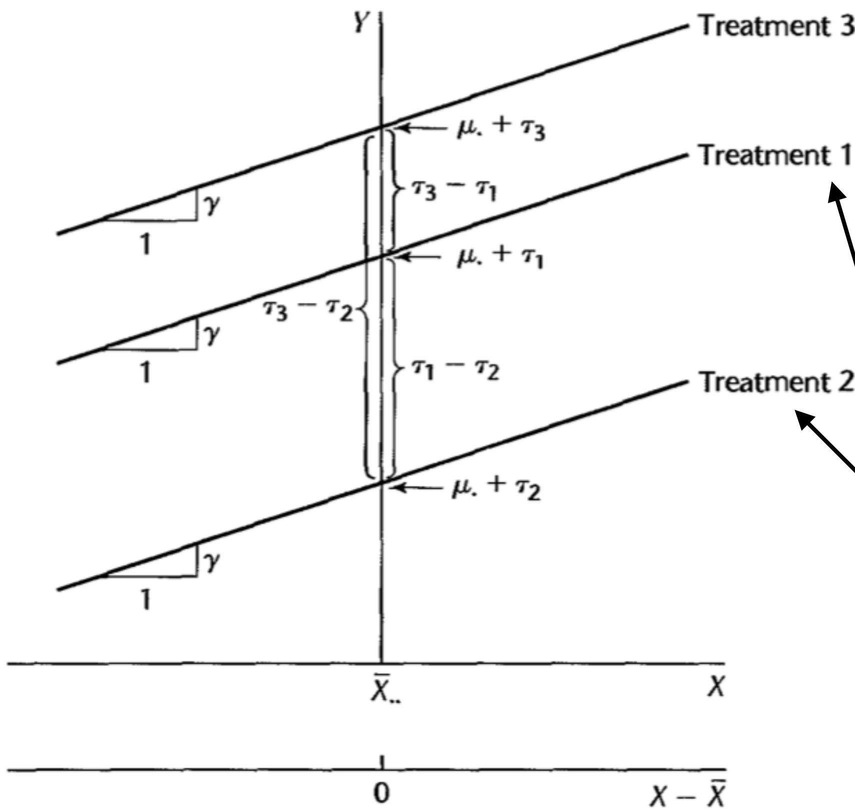
Treatment effect comparison: $\tau_1 - \tau_2$

Measures how much higher the mean response is with treatment 1 versus 2 for any fixed value of X

To test for the presence of treatment or factor effect:

$H_0 : \tau_1 = \ldots = \tau_r = 0$ vs $H_a :$ not all $\tau_i s$ equal zero

# Single-Factor ANCOVA Model




We can no longer speak of the mean response for ith treatment, as it varies with X

**Compare "treatment effect":**

Treatment 1 leads to a higher mean response than treatment 2 by a fixed amount, regardless of what the value X is.

The constant differential effect is reflected by the parallel lines.

If there is no differential treatments, then all lines must be identical. Otherwise, the treatment effects are reflected by the vertical distances.

# Example

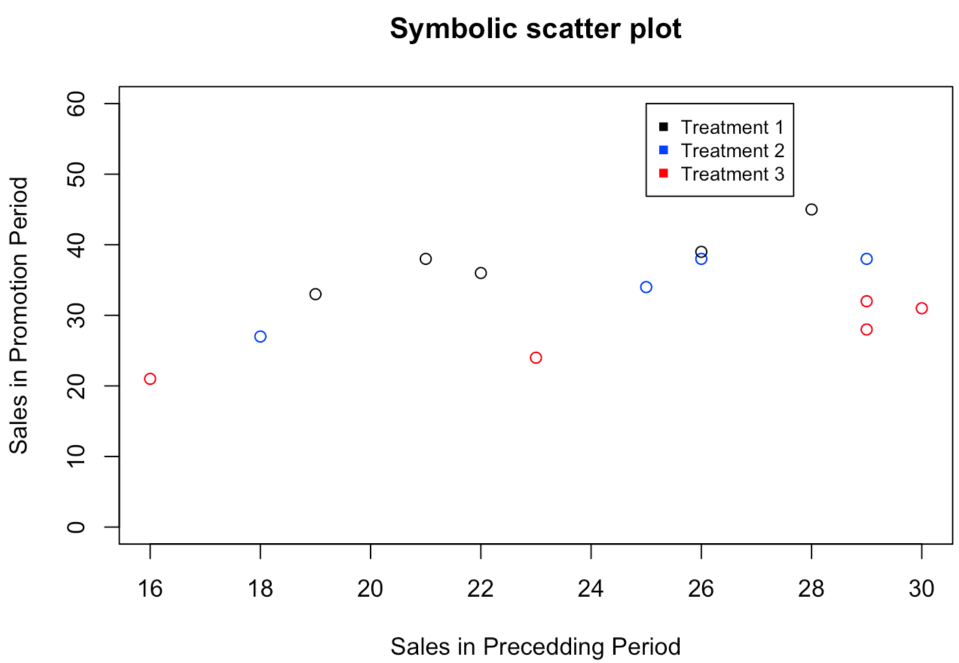
A company studied the effects of three different types of promotions on sales of its crackers:

- Treatment I-Simpling of product by customers in store and regular shelf space
- Treatment 2-Additional shelf space in regular location
- Treatment 3-Special display shelves at ends of aisle in addition to regular shelf space

Fifteen stores were selected for the study, and a completely randomized experimental design was utilized. Each store was randomly assigned one of the promotion types, with five stores assigned to each type of promotion. Other relevant conditions under the control of the company, such as price and advertising, were kept the same for all stores in the study. Data on the number of cases of the product sold during the promotional period, denoted by Y, as are also data on the sales of the product in the preceding period, denoted by X. Sales in the preceding period are to be used as the concomitant variable.Assume that covariance model is applicable.

Does it appear that sales in preceding period is a good concommitant variable?

Does it appear that there are treatment effects?




Sales in predecing period seems to have strong linear relationship with sales in promotion period, therefore it is a good concommitant variable.

Treatment 1 seems to have highest average sales in promotion period, followed by treatment 2, then treatment 1, there appears to be some treatment effects.

# Example

State the regression model equivalent to covariance model for this case; use 1,-1,0 indicator variables. Also state the reduced regression model for testing for treatment effects.

ANCOVA Model: $$Y_{ij} = \mu. + \tau_1 I_{ij1} + \tau_2 I_{ij2} + \gamma X_{ij} + \varepsilon_{ij}$$

Where
$$I_1 = \begin{cases} 1 \text{ if store received treatment } 1 \\ -1 \text{ if store received treatment } 3 \\ 0 \text{ otherwise} \end{cases}$$

$$I_2 = \begin{cases} 1 \text{ if store received treatment } 2 \\ -1 \text{ if store received treatment } 3 \\ 0 \text{ otherwise} \end{cases}$$

$$X_{ij} = X_{ij} - \bar{X}_{..}$$

# Example

Fit the full and reduced regression models and test for treatment effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

$H_0 : \tau_1 = \tau_2 = 0$ vs $H_a :$ not both $\tau_1$ and $\tau_2$ equal zero

$Y_{ij} = \mu. + \tau_1 I_{ij1} + \tau_2 I_{ij2} + \gamma X_{ij} + \varepsilon_{ij}$ Full model

$Y_{ij} = \mu. + \gamma X_{ij} + \varepsilon_{ij}$ Reduced model

$$F^* = \frac{\text{SSE}(R) - \text{SSE}(F)}{df_R - df_F} \div \frac{\text{SSE}(F)}{df_F}$$

Code

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## I1         1 302.500 302.500  86.269 1.538e-06 ***
## I2         1  36.300  36.300  10.352  0.008196 **
## x          1 269.029 269.029  76.723 2.731e-06 ***
## Residuals 11  38.571   3.506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Code

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value  Pr(>F)
## x          1 190.68 190.678  5.4393 0.03641 *
## Residuals 13 455.72  35.056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Code

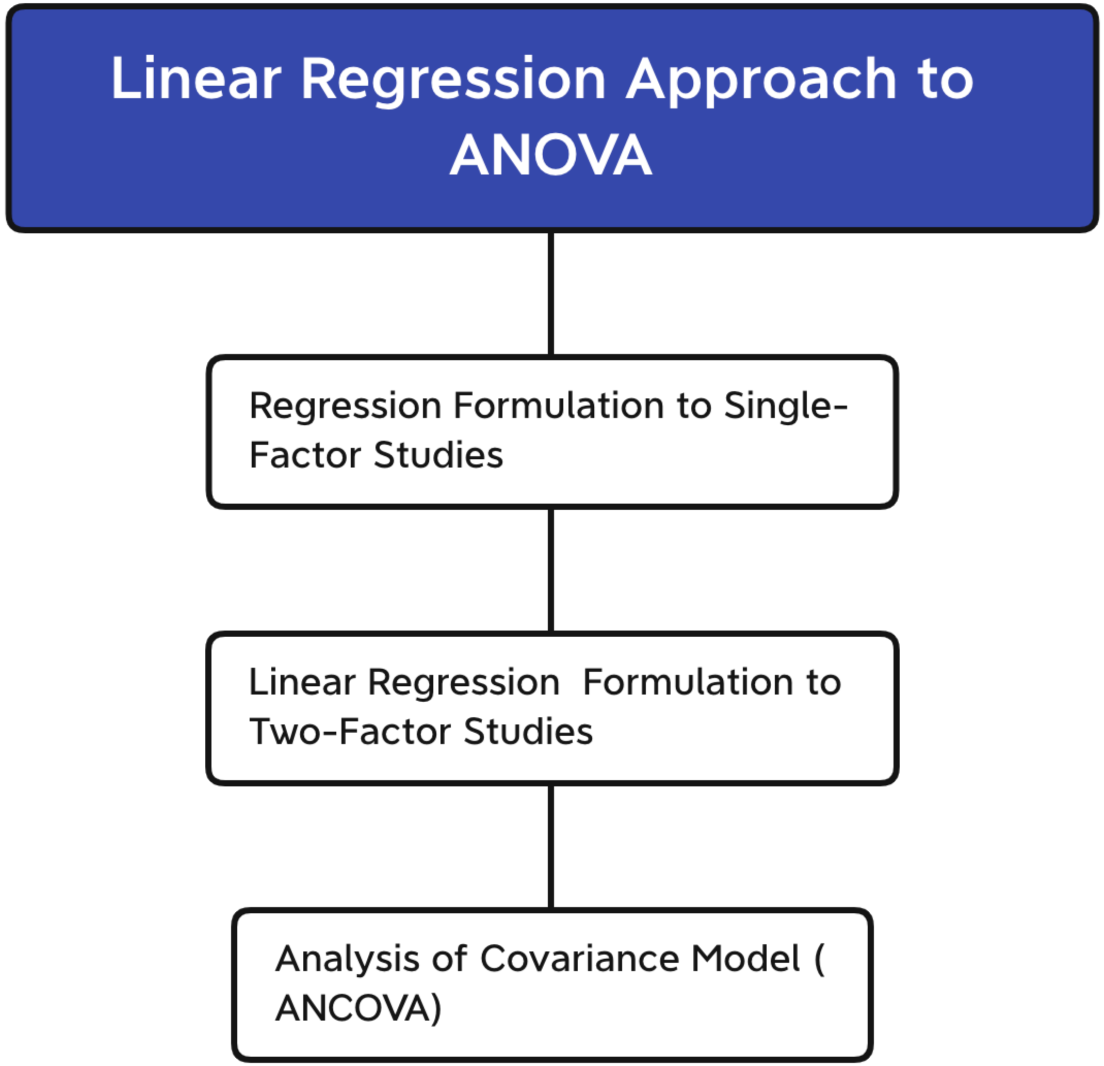
```
## [1] 59.48282
```

Code

```
## [1] 3.982298
```

The level of significance is to be controlled at $\alpha = .05$; hence, we need to obtain $F(.95; 2, 11) = 3.98$. The decision rule therefore is: If $F^* \leq 3.98$, conclude $H_0$ If $F^* > 3.98$, conclude $H_a$ Since $F^* = 59.5 > 3.98$, we conclude $H_a$, that the three cracker promotions differ in sales effectiveness.

# Summary


Linear Regression Approach to ANOVA
Regression Formulation to Single-Factor Studies
Linear Regression Formulation to Two-Factor Studies
Analysis of Covariance Model ( ANCOVA)