Multivariate Data Analysis

(MGT513, BAT531, TIM711)

Lecture 8

UNIST
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY
2009

Ch. 11 Analysis of Variance

Analysis of Variance (ANOVA)

1. Analysis of Variance (**ANOVA**)
2. Analysis of Covariance (**ANACOVA**; **ANCOVA**)
3. Multiple [Multivariate] Analysis of Variance (**MANOVA**)

Analysis of Variance (ANOVA)

1. Analysis of Variance (ANOVA)
   – To explore the dependence relationship between a **continuous** response Y, and **discrete** or **categorical** predictors (experimental factor) $\boldsymbol{X}$
   – To test for an effect of the experimental factor on the dependent variable

Analysis of Variance (ANOVA)

2. Analysis of Covariance (ANACOVA; ANCOVA)
   – To explore the dependence relationship between a **continuous** response Y, and **discrete** and also **continuous** predictors $\boldsymbol{X}$
3. Multiple [Multivariate] Analysis of Variance (MANOVA)
   – To explore the dependence relationship between **multiple(multivariate) continuous** response $\boldsymbol{Y}$, and **discrete** or **categorical** predictors (experimental factor) $\boldsymbol{X}$

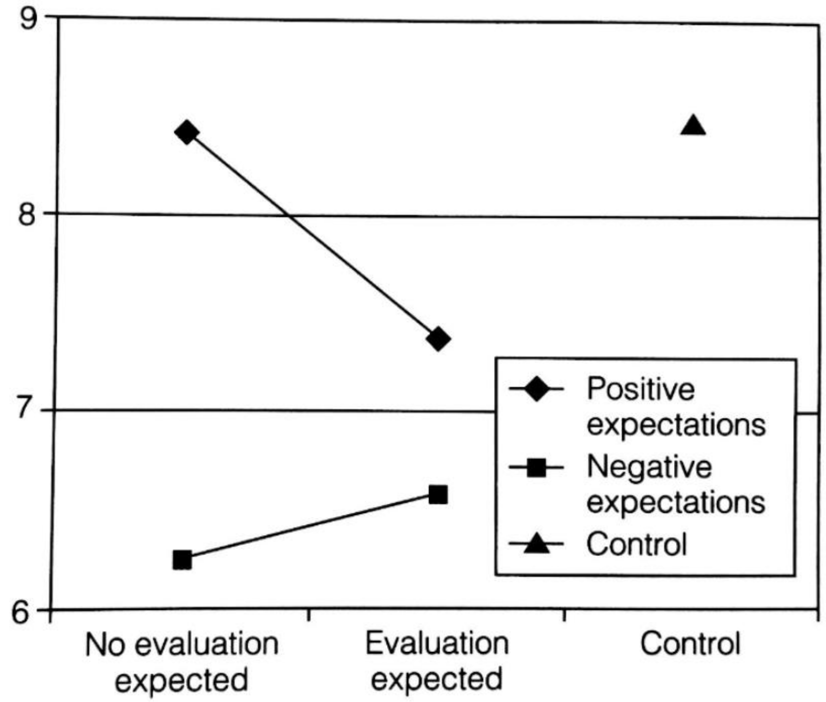
Potential Applications: Customer Satisfaction Ratings

- Example: testing the impact of ”expecting to evaluate” on customer satisfaction ratings (Two-Factor ANOVA)
  - Factor 1: Evaluation expectation (expected / not expected)
  - Factor 2: Shopping experience expectations (positive /negative)
- Factor 1 main effect: expecting to evaluate reinforces consumers’ tendency to focus on and overweight negative aspects when processing information (*negativity enhancement*)
- Factor 2 main effect: positive expectations condition generated more favorable ratings than negative expectations condition

Potential Applications: Customer Satisfaction Ratings

- Interaction effect: according to negativity enhancement, the effect of expecting to evaluate should be pronounced when expectations are positive.

FIGURE 11.1
Effect of expecting to evaluate and quality expectations on overall satisfaction (Source: Ofir and Simonson, 2001)



ANOVA: One-Factor Design

Example: Finnish Liquor Data

TABLE 11.1 Summary statistics for Finnish liquor data: Number of traffic accidents in each municipality across three treatment groups

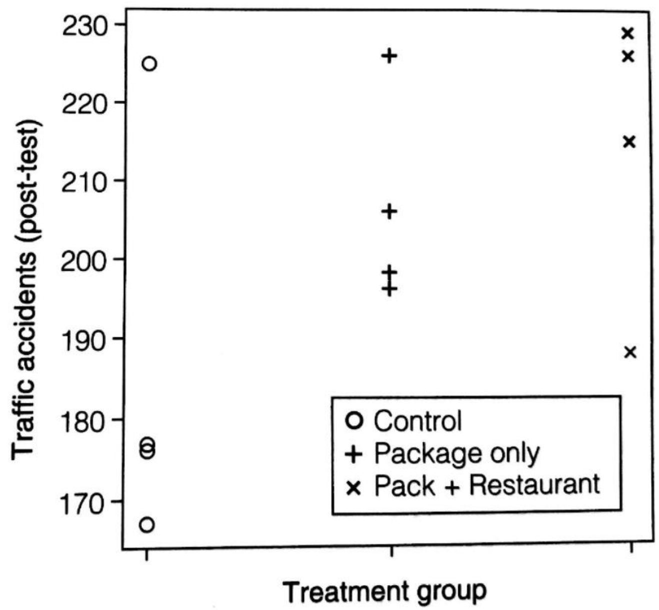
Control	Package Stores Only	Package Stores and Restaurants
177	226	226
225	196	229
167	198	215
176	206	188
$\bar{Y}_{.1} = 186.5$	$\bar{Y}_{.2} = 206.5$	$\bar{Y}_{.3} = 214.5$

ANOVA: One-Factor Design

Example: Finnish Liquor Data

- $n = 12\ (n_1 = n_2 = n_3 = 4)$ – Balanced design
- ANOVA: To test for differences in the mean of the dependent variable across groups **Treatment effects**
- Compare treatment effects

FIGURE 11.2
Plot of Finnish liquor data from Wildt and Ahtola (1978)



$$\bar{Y}_{.1} = 186.5\ (s_1 = 26.2)$$
$$\bar{Y}_{.2} = 206.5\ (s_2 = 13.7)$$
$$\bar{Y}_{.3} = 214.5\ (s_3 = 18.7)$$

ANOVA: One-Factor Design

Notation

- Y_{ij} = the ith observation in the treatement group j
- $\bar{Y}_{.j} = \frac{1}{n_j}\sum_i Y_{ij}$ = the sample mean for treatment group j
- $\bar{Y}_{..} = \frac{1}{n}\sum_i \sum_j Y_{ij}$ =
  the overall sample mean across all observations

ANOVA: One-Factor Design

- ANOVA Intuition
  - ANOVA tests for the effect of a factor by looking for differences across the means of the treatment groups.
  - The test is based on a comparison of two different estimates of the within-group variance σ^2: one estimated within groups (denoted s_W^2) and one estimated across groups (denoted s_A^2).
  - When there are differences across groups, s_A^2 will be much larger than s_W^2, and the test will be significant.

ANOVA: One-Factor Design

- Let $\mu_j =$ the true mean of the dependent variable for group j.

  The null hypothesis for **No effects** is given by

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3 = \cdots \mu_m$$

  The alternative hypothesis is

$$H_1\text{: At least one } \mu_j \text{ is different}$$

ANOVA: One-Factor Design

- Let σ_j^2 = the true variance of the observations in treatment j.

  Homoscedasticity Assumption (of within group variances)

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots \sigma_m^2 = \sigma^2$$

- In general,

$$\underbrace{Y_{ij} - \bar{Y}_{..}}_{\text{overall variation}} = \underbrace{\bar{Y}_{.j} - \bar{Y}_{..}}_{\text{across-group variation}} + \underbrace{Y_{ij} - \bar{Y}_{.j}}_{\text{within-group variation}}$$

ANOVA: One-Factor Design

- **Within-group sample variance** with $df = n - m$

$$s_W^2 = \frac{1}{m}\sum_j \frac{\sum_i (Y_{ij} - \bar{Y}_{.j})^2}{n_j - 1} \quad \text{and} \quad E(s_W^2) = \sigma^2$$

- The within-group estimate of σ^2 is formed by looking at the variation of the individual observations around the group means. The expected value of is equal to σ^2, <u>*whether or not H_0 is true*.</u>

ANOVA: One-Factor Design

- **Across-group sample variance** with $df = m - 1$

$$s_A^2 = \frac{\sum_j n_j(\bar{Y}_{.j} - \bar{Y}_{..})^2}{m-1} \quad \text{and} \quad E(s_A^2) = \sigma^2 \text{ under } H_0$$

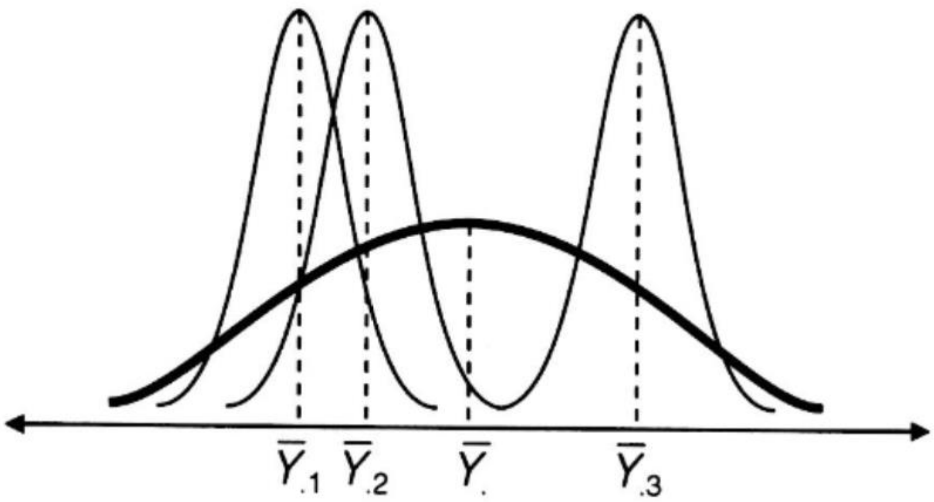
- The across-group estimate of σ^2 is formed by looking at the variation of the group means around the overall mean. The expected value of is equal to σ^2 *only when H_0 is true*.
- When H_0 is false, then s_A^2 will be larger than σ^2 because it is a function of the squared differences between the group means and the overall mean.

ANOVA: One-Factor Design

- When the group means are different, the estimate s_A^2 reflects not only the variability due to sampling error but also any systematic differences due to the treatment effects.
- The larger the differences in group means, the more inflated s_A^2 becomes.
- We use an F-test [with $(m - 1, n - m)$ degrees of freedom] to determine whether the ratio $\frac{s_A^2}{s_W^2}$ is significantly different from 1 — that is, to test whether the difference between s_A^2 and s_W^2 is large enough for us to conclude with sufficient confidence that it could not be attributable to random chance.

ANOVA: One-Factor Design

FIGURE 11.3
Distributions of observations when null hypothesis is false



ANOVA: One-Factor Design

Example: Finnish Liquor Data (revisited)

TABLE 11.2 ANOVA Table for Finnish liquor data

Source	Sum of Squares	*df*	Mean Square	*F*-Value	Pr > *F*
Across	1696.17	2	848.08	2.08	0.1810
Within	3670.75	9	407.86		
Total	5366.92	11			

ANOVA: Two-Factor Design

Two Factor Model: *Two* levels X *Three* levels

FIGURE 11.4
Depiction of 2 × 3 experimental design matrix

$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.13}$	$\bar{Y}_{.1.}$
$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.23}$	$\bar{Y}_{.2.}$
$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{..3}$	$\bar{Y}_{...}$

ANOVA: Two-Factor Design

Notation

- Y_{ijk}= the ith observation in level j of factor 1 and level k of factor 2
- $\bar{Y}_{.j.}$ = the mean value of Y for level j of factor 1
- $\bar{Y}_{..k}$ = the mean value of Y for level k of factor 2
- $\bar{Y}_{.jk}$ = the mean value of Y for cell $(j, k)$of the design
- $\bar{Y}_{...}$ = the overall mean value across all observations

ANOVA: Two-Factor Design

- In general,

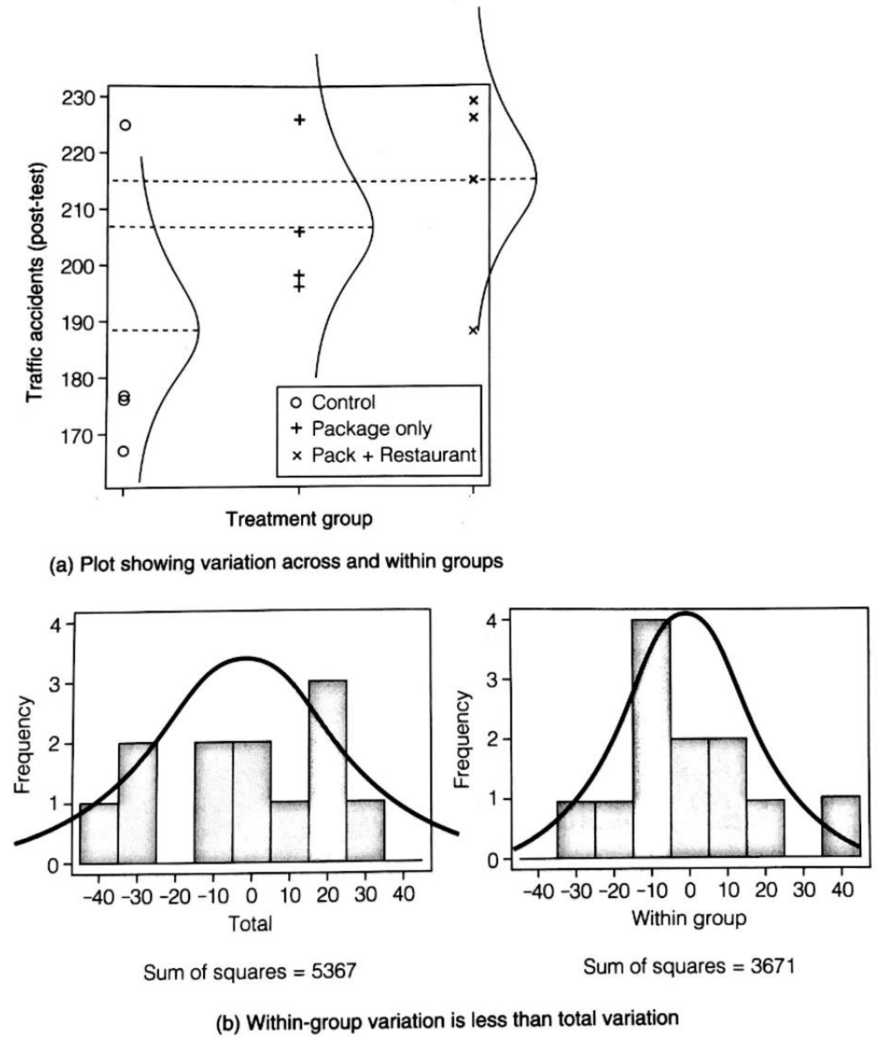
$$\underbrace{Y_{ijk} - \bar{Y}_{...}}_{\text{overall variation}} = \underbrace{\bar{Y}_{.j.} - \bar{Y}_{...}}_{\text{variation for factor 1}} + \underbrace{\bar{Y}_{..k} - \bar{Y}_{...}}_{\text{variation for factor 2}}$$

$$+ \quad \bar{Y}_{.jk} \underbrace{- \bar{Y}_{.j.} - \bar{Y}_{..k} + \bar{Y}_{...}}_{\text{interaction}}$$

$$+ \quad \underbrace{Y_{ijk} - \bar{Y}_{.jk}}_{\text{within-group variation}}$$

ANOVA: Relationship to Regression

FIGURE 11.5
Total variation is made up of across-group and within-group variation



ANOVA: Relationship to Regression

- F-test for the significance of the regression model = F-test from ANOVA
  - $F = \frac{R^2}{1-R^2} \times \frac{n-p-1}{p}$
  - $F = \frac{SS_R/SS_T}{1-SS_R/SS_T} \times \frac{n-p-1}{p}$

    Let $SS_E = SS_T - SS_R$
  - $F = \frac{SS_R/SS_T}{SS_E/SS_T} \times \frac{n-p-1}{p}$
  - $F = \left(\frac{SS_R/p}{SS_E/(n-p-1)}\right)$

ANOVA Assumptions

Assumptions:

1. Independent samples
2. Normality – The responses for each group have a normal population distribution (or the distributions of the residuals are normal).
3. Equality (or "homogeneity") of variances, called homoscedasticity.

Mechanics: ANOVA (One-Factor Model)

- For $i = 1, \ldots, n_j$ and $j = 1, \ldots, m$

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

where μ: overall mean

τ_j: the jth treatment effect

ε_{ij}: the $(i, j)th$ error $\sim$ iid $(0, \sigma^2)$

- Assume balanced design:

$$n_1 = n_2 = \cdots = n_m$$

- **Side condition** to get unique solution:

$$\sum_j \tau_j = 0 \text{ and fix } \tau_1 = 0$$

Mechanics: ANOVA (One-Factor Model)

- The null hypothesis for **No Effects** can be rewritten as

$$H_0: \tau_1 = \tau_2 = \tau_3 = \cdots \tau_m = 0$$

- Test $\frac{s_A^2}{s_W^2}$ is significantly different from 1: the ratio will be greater than 1 of the null is false.
- Within-group sample variance (s_W^2)
  - The expectation of the sample variance for the treatment group j, given by $s_j^2 = \frac{1}{n-1}\sum_i \left(Y_{ij} - \bar{Y}_{.j}\right)^2$, is $E\left(s_j^2\right) = \sigma^2$
  - $s_W^2 = \frac{1}{m}\sum_i s_j^2$
  - $E(s_W^2) = \sigma^2$

Mechanics: ANOVA (One-Factor Model)

- Across-group sample variance (s_A^2)
  - $E(s_A^2) = E\left[\frac{1}{(m-1)}\sum_j n_j\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2\right]$
    - $\bar{Y}_{.j} = \mu + \tau_j + \frac{1}{n_j}\sum_j \varepsilon_{ij}$
    - $\bar{Y}_{..} = \mu + \frac{1}{m}\sum_j \tau_j + \frac{1}{n}\sum_i \sum_j \varepsilon_{ij} = \mu + \frac{1}{n}\sum_i \sum_j \varepsilon_{ij}$
  - $E(s_A^2) = E\left[\frac{1}{(m-1)}\sum_j n_j\left(\tau_j + \frac{1}{n_j}\sum_j \varepsilon_{ij} - \frac{1}{n}\sum_i \sum_j \varepsilon_{ij}\right)^2\right]$
  - $E(s_A^2) = \frac{1}{(m-1)}\sum_j n_j E(\tau_j^2) + \frac{1}{(m-1)} E\left[\sum_j n_j\left(\frac{1}{n_j}\sum_j \varepsilon_{ij} - \frac{1}{n}\sum_i \sum_j \varepsilon_{ij}\right)^2\right]$

Mechanics: ANOVA (One-Factor Model)

- Across-group sample variance (s_A^2) (continued)
  - Since $E\left(\tau_j^2\right) = \tau_j^2$ and $\frac{1}{(m-1)} E\left[\sum_j n_j \left(\frac{1}{n_j} \sum_j \varepsilon_{ij} - \frac{1}{n} \sum_i \sum_j \varepsilon_{ij}\right)^2\right] = \sigma^2$
  - $E(s_A^2) = \sigma^2 + \frac{1}{(m-1)} \sum_j n_j \tau_j^2$
- When the null hypothesis is true, $\tau_j = 0$ for all j and $\frac{1}{(m-1)} \sum_j n_j \tau_j^2$ becomes 0.
- The greater $\frac{1}{(m-1)} \sum_j n_j \tau_j^2$, the greater F-ratio and the greater the likelihood of rejecting the null hypothesis.

Mechanics: ANOVA (One-Factor Model)

One-Factor ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*-ratio
Across	SS_A	$m-1$	$MS_A = SS_A/(m-1)$	MS_A/MS_W
Within	SS_W	$n-m$	$MS_W = SS_W/(n-m)$	
Total	SS_T	$n-1$		

where

- $SS_A = \sum_j n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$
- $SS_W = \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2$
- $SS_T = \sum_j \sum_i (Y_{ij} - \bar{Y}_{..})^2$

- $F = \frac{MS_A}{MS_W} = \frac{s_A^2}{s_W^2} \sim F(m-1, n-m)$

Mechanics: ANOVA (Two-Factor Model)

- For $i = 1, \ldots, n_{jk}$ and $j = 1, \ldots, m_a$ and $k = 1, \ldots, m_b$

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

where μ: overall mean

α_j: treatment effect of level j of factor 1

β_k: treatment effect of level k of factor 2

$(\alpha\beta)_{jk}$: interaction effect between α_j and β_k

$\varepsilon_{ijk} \sim \text{iid}\ (0, \sigma^2)$

m_a: treatment levels for factor 1

m_b: treatment levels for factor 2

Mechanics: ANOVA (Two-Factor Model)

- Assume balanced design:

$$n_{11} = n_{12} = \cdots = n_{m_a,m_b}$$

- **Side condition** to get unique solution:

$$\sum_j \alpha_j = \sum_k \beta_k = \sum_j (\alpha\beta)_j = \sum_k (\alpha\beta)_k = 0$$

- The null hypotheses:
  - $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \cdots \alpha_{m_a} = 0$
  - $H_0: \beta_1 = \beta_2 = \beta_3 = \cdots \beta_{m_b} = 0$
  - $H_0: (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots (\alpha\beta)_{m_a,m_b} = 0$

Mechanics: ANOVA (Two-Factor Model)

Two-Factor ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*-ratio
Across	SS_A	$m_a m_b - 1$	$MS_A = SS_A/(m_a m_b - 1)$	MS_A/MS_W
Factor 1	$SS(\alpha)$	$m_a - 1$	$MS(\alpha) = SS(\alpha)/(m_a - 1)$	$MS(\alpha)/MS_W$
Factor 2	$SS(\beta)$	$m_b - 1$	$MS(\beta) = SS(\beta)/(m_b - 1)$	$MS(\beta)/MS_W$
Interact	$SS(\alpha\beta)$	$(m_a - 1)(m_b - 1)$	$MS(\alpha\beta)$ $= SS(\alpha\beta)/(m_a - 1)(m_b - 1)$	$\frac{MS(\alpha\beta)}{MS_W}$
Within	SS_W	$n - m_a m_b$	$MS_W = SS_W/(n - m_a m_b)$	
Total	SS_T	$n - 1$		

Mechanics: ANOVA (Two-Factor Model)

where

- $SS(\alpha) = \sum_j n_{j.}(\bar{Y}_{.j.} - \bar{Y}_{...})^2$
- $SS(\beta) = \sum_k n_{.k}(\bar{Y}_{..k} - \bar{Y}_{...})^2$
- $SS(\alpha\beta) = \sum_j \sum_k n_{jk}\,(\bar{Y}_{.jk} - \bar{Y}_{.j.} - \bar{Y}_{..k} - \bar{Y}_{...})^2$
- $SS_A = SS(\alpha) + SS(\beta) + SS(\alpha\beta) = \sum_j \sum_k n_{jk}(\bar{Y}_{.jk} - \bar{Y}_{...})^2$
- $SS_W = \sum_i \sum_j \sum_k (\bar{Y}_{ijk} - \bar{Y}_{.jk})^2$
- $SS_T = \sum_i \sum_j \sum_k (\bar{Y}_{ijk} - \bar{Y}_{...})^2$

- Test statistics (*F* test)
  - $F(\textbf{Across}) = \frac{MS_A}{MS_W} \sim F(m_a m_b - 1, n - m_a m_b)$
  - $F(\textbf{Factor 1}) = \frac{MS(\alpha)}{MS_W} \sim F(m_a - 1, n - m_a m_b)$
  - $F(\textbf{Factor 2}) = \frac{MS(\beta)}{MS_W} \sim F(m_b - 1, n - m_a m_b)$
  - $F(\textbf{Interact}) = \frac{MS(\alpha\beta)}{MS_W} \sim F((m_a - 1)(m_b - 1), n - m_a m_b)$

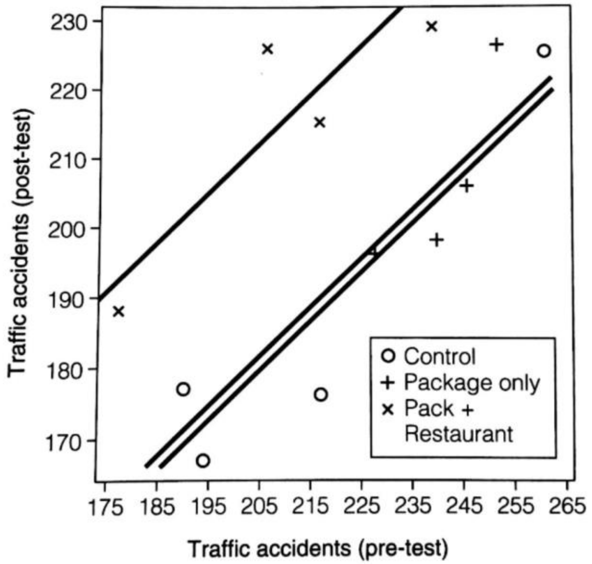
ANCOVA (ANACOVA)

- ANCOVA: To explore the relationship between univariate continuous response Y and discrete factors $\boldsymbol{X}$ and continuous predictors Z
- Example: Finnish Liquor Data

TABLE 11.3 Number of traffic accidents during test year (Y) and in previous year (Z)

Control		Package Stores Only		Package Stores and Restaurants	
Y	Z	Y	Z	Y	Z
177	190	226	252	226	206
225	261	196	228	229	239
167	194	198	240	215	217
176	217	206	246	188	177

FIGURE 11.6 Plot of Y versus Z for Finnish liquor data



ANCOVA (ANACOVA)

Two type of effects of a covariate

(1) **Reducing within-group variance:** First panel of Figure 11.9

- Z explains only within-group variation in Y
- No across-group correlation
- Z reduces the ”noise” within groups, *increasing* the likelihood of finding a significant effect due to the experimental treatment

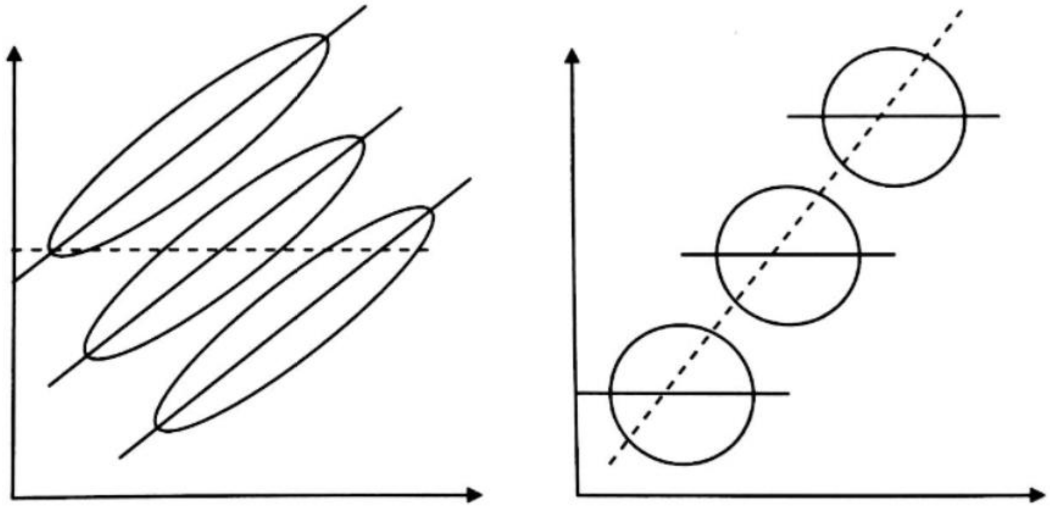
(2) **Explaining differences in means across treatment groups:** Second panel of Figure 11.9

- Z explains only across-group variation in Y
- No within-group correlation
- Adjusting for the effect of Z serves to eliminate some of the apparent differences across groups, thereby *decreasing* the likelihood of finding a significant effect due to the experimental treatment

ANCOVA (ANACOVA)

Two type of effects of a covariate

FIGURE 11.9
Different relationships between *Y* and *Z*



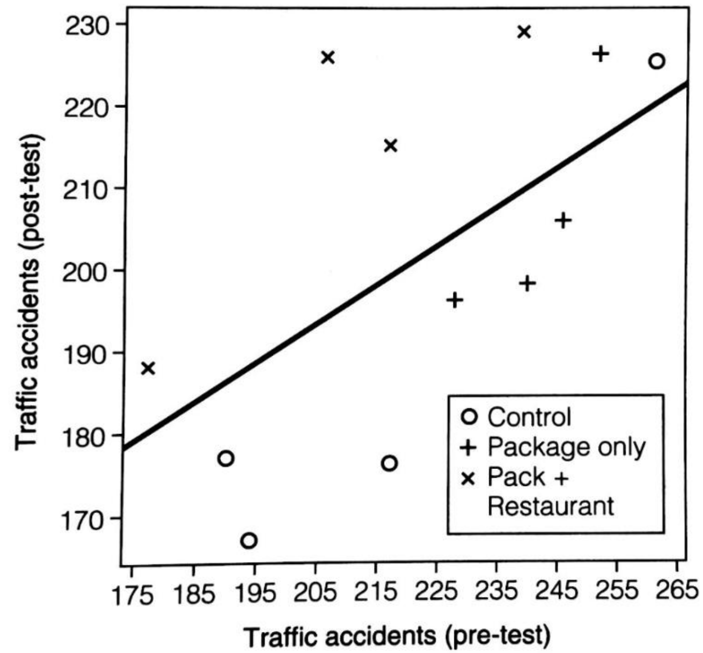
$SS_T(Y_{adj}) = SS_T(Y)$
$SS_W(Y_{adj}) << SS_W(Y)$

$SS_T(Y_{adj}) << SS_T(Y)$
$SS_W(Y_{adj}) = SS_W(Y)$

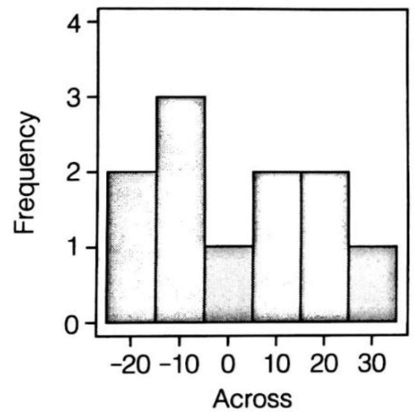
ANCOVA (ANACOVA)

- Example: Finnish Liquor Data

FIGURE 11.7
Adjusting *Y* across groups for variation attributable to *Z*



(a) Regression of *Y* on *Z* across groups

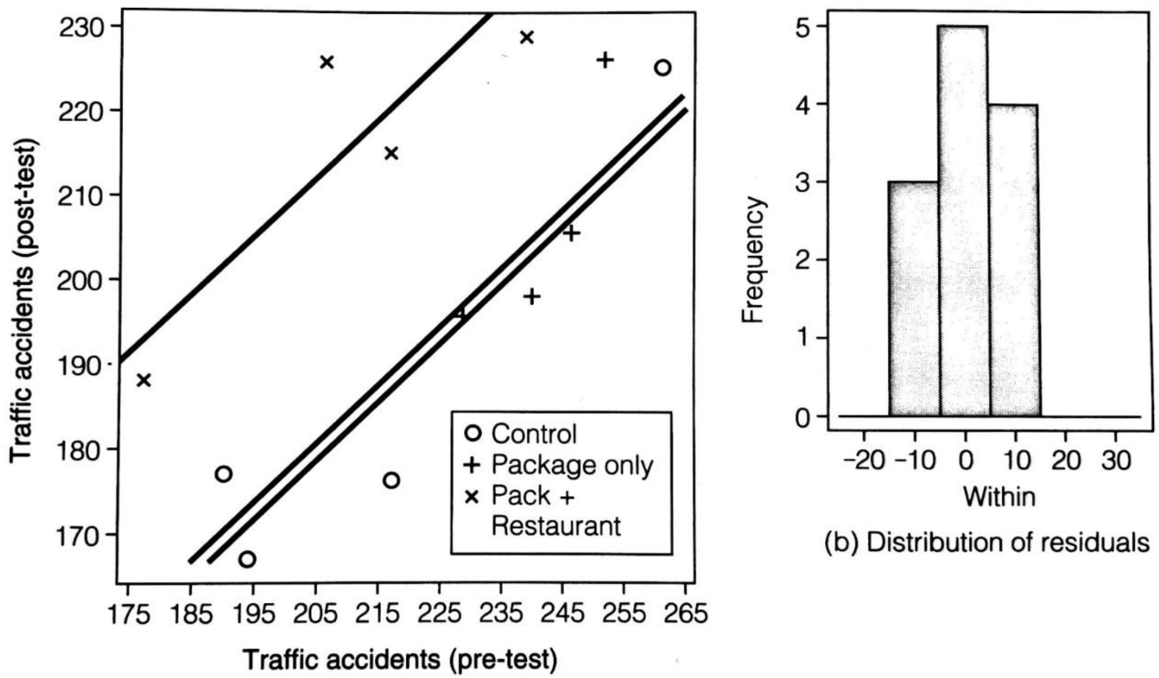


(b) Distribution of residuals

ANCOVA (ANACOVA)

- Example: Finnish Liquor Data

FIGURE 11.8
Adjusting *Y* within groups for variation attributable to *Z*



(a) Regression of *Y* on *Z* across groups

(b) Distribution of residuals

ANCOVA (ANACOVA)

- Example: Finnish Liquor Data

TABLE 11.4 ANCOVA table for Finnish liquor data

Source	Sum of Squares	*df*	Mean Square	*F*-Value	Pr > *F*
Across	2401.76	2	1200.88	12.83	0.0032
Within	748.68	8	93.58		
Adjusted Total	3150.44	10			

Least Squares Means

	$\overline{Y}_{adj}$
Control	191.17
Package only	192.46
Package and restaurant	223.62

ANCOVA Assumptions

Assumptions

1. Independent samples
2. Normality – The responses for each group have a normal population distribution (or the distributions of the residuals are normal).
3. Equality (or "homogeneity") of variances, called homoscedasticity.
4. The lines expressing these linear relationships are all parallel (homogeneity of regression slopes)
5. The covariate is independent of the treatment effects (i.e. the covariant and independent variables are independent)

Mechanics: ANCOVA

- For $i = 1, \ldots, n_j$ and $j = 1, \ldots, m$

$$Y_{ij} = \mu + \tau_j + \beta Z_{ij} + \varepsilon_{ij}$$

where μ: overall mean

τ_j: the jth treatment effect

Z_{ij}: the $(i, j)^{th}$ standardized covariate

ε_{ij}: the $(i, j)^{th}$ error $\sim$ iid $(0, \sigma^2)$

- **Side condition** to get unique solution:

$$\sum_j \tau_j = 0$$

Mechanics: ANCOVA

- $SS_T(Y) = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$
- $SS_T(Z) = \sum_i \sum_j (Z_{ij} - \bar{Z}_{..})^2$
- $SS_T(YZ) = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})(Z_{ij} - \bar{Z}_{..})$
- $SS_W(Y) = \sum_i \sum_j (Y_{ij} - \bar{Y}_{.j})^2$
- $SS_W(Z) = \sum_i \sum_j (Z_{ij} - \bar{Z}_{.j})^2$
- $SS_W(YZ) = \sum_i \sum_j (Y_{ij} - \bar{Y}_{.j})(Z_{ij} - \bar{Z}_{.j})$

Mechanics: ANCOVA

- The total sum of squares for the adjusted dependent variable comes from regressing Y on Z alone

$$SS_T\left(Y_{adj}\right) = SS_T(Y) - b_T SS_T(YZ)$$

where b_T: estimated regression coefficent between Y and Z

$$b_T = \frac{SS_T(YZ)}{SS_T(Z)}$$

and thus

$$SS_T\left(Y_{adj}\right) = SS_T(Y) - \frac{SS_T(YZ)^2}{SS_T(Z)} = \frac{SS_T(Y)SS_T(Z) - SS_T(YZ)^2}{SS_T(Z)}$$

and if $\mathrm{cov}(Y, Z) = 0$, then $SS_T\left(Y_{adj}\right) = SS_T(Y)$

Mechanics: ANCOVA

- The within-group sum of squares for the adjusted dependent variable comes from regressing Y on Z after explaining the different intercept terms for each group

$$SS_W\left(Y_{adj}\right) = SS_W(Y) - b_W SS_W(YZ)$$

where b_W: within-group slope coefficient

$$b_W = \frac{SS_W(YZ)}{SS_W(Z)}$$

and thus

$$SS_W\left(Y_{adj}\right) = SS_W(Y) - \frac{SS_W(YZ)^2}{SS_W(Z)} = \frac{SS_W(Y)SS_W(Z) - SS_W(YZ)^2}{SS_W(Z)}$$

If there is no relationship between Y and Z within group $[SS_W(YZ) = 0]$, then $SS_W\left(Y_{adj}\right) = SS_W(Y)$

Mechanics: ANCOVA

- Therefore, the across-group sum of squares for the adjusted dependent variable is

$$SS_A(Y_{adj}) = SS_T(Y_{adj}) - SS_W(Y_{adj})$$

Mechanics: ANCOVA

One-Factor ANCOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*-ratio
Regression	$\frac{SS_T(YZ)^2}{SS_T(Z)}$	1	$MS(Reg) = \frac{SS_T(YZ)^2}{SS_T(Z)}$	
Factor	SS_A	$m-1$	$MS_A = SS_A/(m-1)$	MS_A/MS_W
Within	SS_W	$m(n-1)-1$	$MS_W = SS_W/(m(n-1)-1)$	
Total	SS_T	$mn-1$		

where

- $SS_W = SS_W(Y_{adj}) = SS_W(Y) - \frac{SS_W(Y)^2}{SS_W(Z)}$
- $SS_A = SS_T(Y_{adj}) - SS_W(Y_{adj})$

- $F = \frac{MS_A}{MS_W} \sim F(m-1, m(n-1)-1)$

Sample Problem: Marketing a New Product

Testing Effects of Price and Advertising (ANOVA)

- Interested in Sales(Y) according to three levels(low, medium, high) of Price (X_1) and two levels(low, high) of Advertising(X_2)

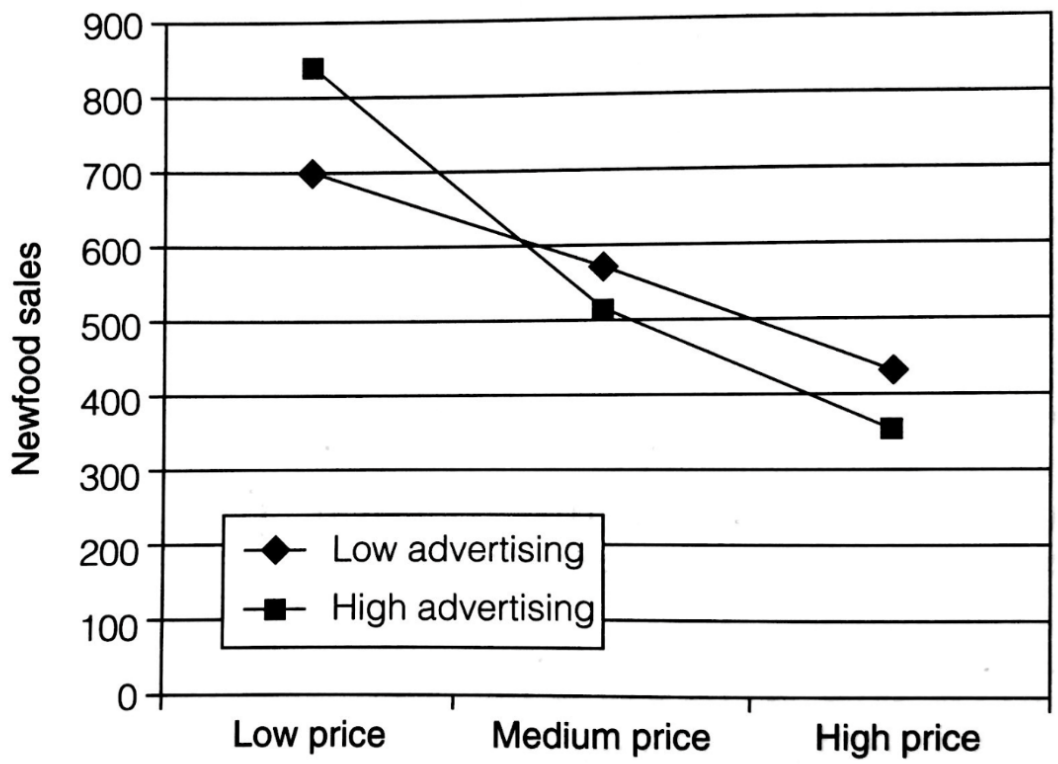
TABLE 11.6 Total sales of Newfood by store (Two-factor design: advertising [low, high] by price [low, medium, high])

	Low Price	Med Price	High Price	
Low adv	620	475	413	$\overline{Y}_{1.} = 570.25$
	774	544	556	
	623	579	395	
	776	706	382	
	$\overline{Y}_{11} = 698.25$	$\overline{Y}_{12} = 576.00$	$\overline{Y}_{13} = 436.50$	
High adv	955	472	294	$\overline{Y}_{2.} = 572.58$
	669	701	378	
	596	482	355	
	1208	388	373	
	$\overline{Y}_{21} = 857.00$	$\overline{Y}_{22} = 510.75$	$\overline{Y}_{23} = 350.00$	
	$\overline{Y}_{.1} = 777.62$	$\overline{Y}_{.2} = 543.38$	$\overline{Y}_{.3} = 393.25$	$\overline{Y}_{..} = 571.42$

Sample Problem:
Marketing a New Product

Testing Effects of Price and Advertising (ANOVA)

FIGURE 11.10 Plot of average Newfood sales for each treatment group



Sample Problem:
Marketing a New Product

Testing Effects of Price and Advertising (ANOVA)

TABLE 11.7 ANOVA table for Newfood

Source	Sum of Squares	*df*	Mean Square	*F*-Value	Pr > *F*
PRICE	600412.58	2	300206.29	14.78	0.0002
ADV	32.67	1	32.67	0.00	0.9685
PRICE × *ADV*	73850.08	2	36925.04	1.82	0.1909
Error	365562.50	18	20309.03		
Total	1039857.83	23			

Sample Problem:
Marketing a New Product

Controlling for Store Volume (ANCOVA)

- Interested in Sales(Y) according to three levels(low, medium, high) of Price (X_1) and two levels(low, high) of Advertising(X_2)
- Differences in Sales (Y) due to store size: Let's consider "store size" (Z) as a covariate

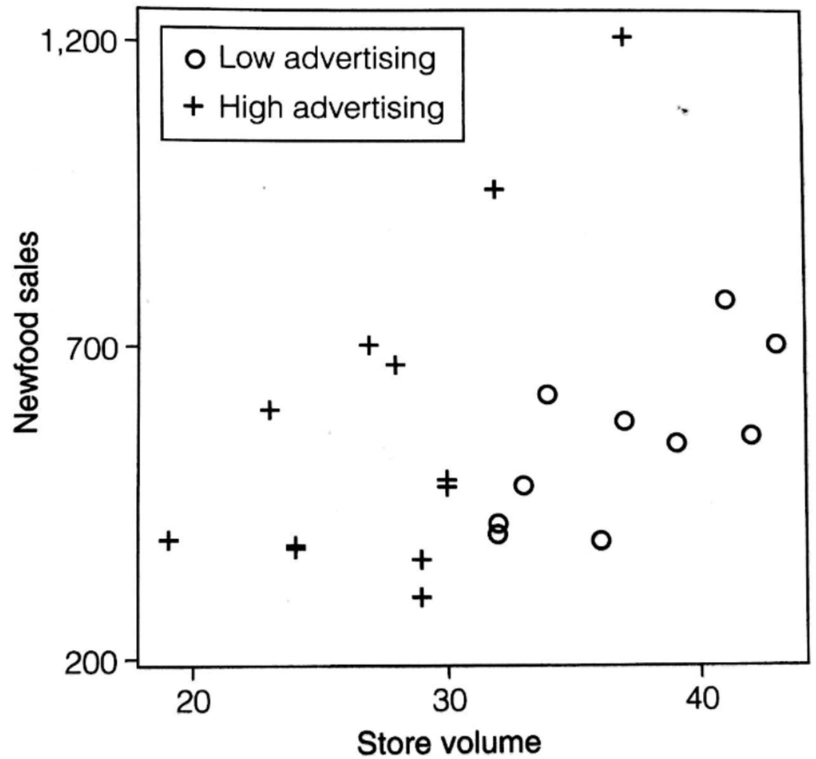
TABLE 11.8 Average store size in each of Newfood treatment groups

	Low Price	Medium Price	High Price
Low advertising	37.5	38	35.5
High advertising	30	26.5	26.5

Sample Problem:
Marketing a New Product

Controlling for Store Volume (ANCOVA)

FIGURE 11.11 Newfood sales versus store volume at low and high advertising



Sample Problem: Marketing a New Product

Controlling for Store Volume (ANCOVA)

TABLE 11.9 ANCOVA table for Newfood: Adjusted for store size (*VOL*)

Source	Sum of Squares	*df*	Mean Square	*F*-Value	Pr > *F*
PRICE	393082.79	2	196541.40	19.30	0.0001
ADV	113931.74	1	113931.74	11.19	0.0038
PRICE × *ADV*	45304.16	2	22652.08	2.22	0.1386
Error	173097.49	17	10182.21		
Adjusted total	725416.18	22			

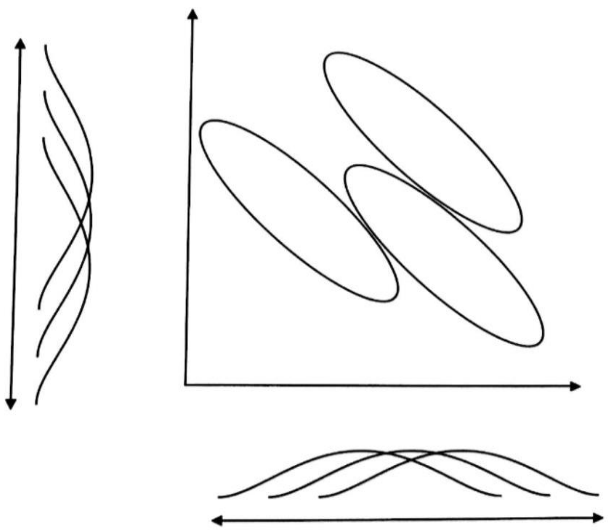
Least Squares Means

PRICE	*ADV*	Adjusted Mean Sales
Low	Low	581.97
Low	High	909.51
Med	Low	448.47
Med	High	642.03
High	Low	365.23
High	High	481.28

Multiple Analysis Of Variance(MANOVA)

- MANOVA: ANOVA with multiple dependent variables
- **Q:** Why not run a series of a simple ANOVAs, each with a different dependent variable?
  - **A:** The simple ANOVAs, run separately, cannot explain the pattern of covariation among the dependent measures

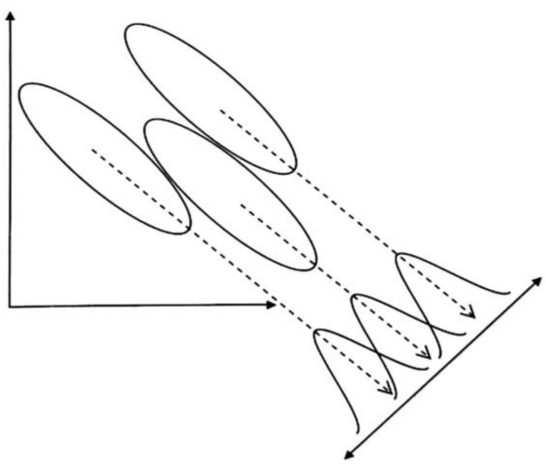
FIGURE 11.12 Marginal distributions of Y_1 and Y_2 and joint distribution of Y_1, Y_2



Multiple Analysis Of Variance(MANOVA)

- ANOVA looks at only one dependent variable at a time
- MANOVA look as both Y_1 and Y_2 at the same time
- Linear combination of Y_1 and Y_2 may reduce within-group variance relative to differences in group means

FIGURE 11.13 Linear combination of Y_1 and Y_2 reduces within-group variance relative to differences in group means



- Adding Y_1 and Y_2 together effectively may reduce within-group variation, making the group differences relatively more prominent

Multiple Analysis Of Variance(MANOVA)

- In two factor ANOVA

$$SS_T = SS(\alpha) + SS(\beta) + SS(\alpha\beta) + SS_W$$

- In two factor MANOVA, instead of partitioning a scalar sum of squares, we partition a sum of squares matrix as follows:

$$\boldsymbol{S}_T = \boldsymbol{S}_\alpha + \boldsymbol{S}_\beta + \boldsymbol{S}_{\alpha\beta} + \boldsymbol{S}_E$$

- For testing model, instead of *F*-test to compare two scalar estimates of σ^2 in ANOVA, we use **WILK's $\boldsymbol{\Lambda}$** as the following ratio:

$$\Lambda = \frac{|\boldsymbol{S}_E|}{|\boldsymbol{S}_T|} = \frac{\text{determinant of residual sum of squares matrix}}{\text{determinat of total sum of squares matrix}}$$

Multiple Analysis Of Variance(MANOVA)

- Hypothesis test (use Bartlett's Chi-Square Test/Rao's *F*-Test in Ch. 9)
  - Let $\boldsymbol{S}_H$ = the sum of squares matrix attributable to a particular hypothesis. For example, let $\boldsymbol{S}_H = \boldsymbol{S}_\alpha$ and then let's construct

$$\Lambda_H = \frac{|\boldsymbol{S}_E|}{|\boldsymbol{S}_H + \boldsymbol{S}_E|}$$

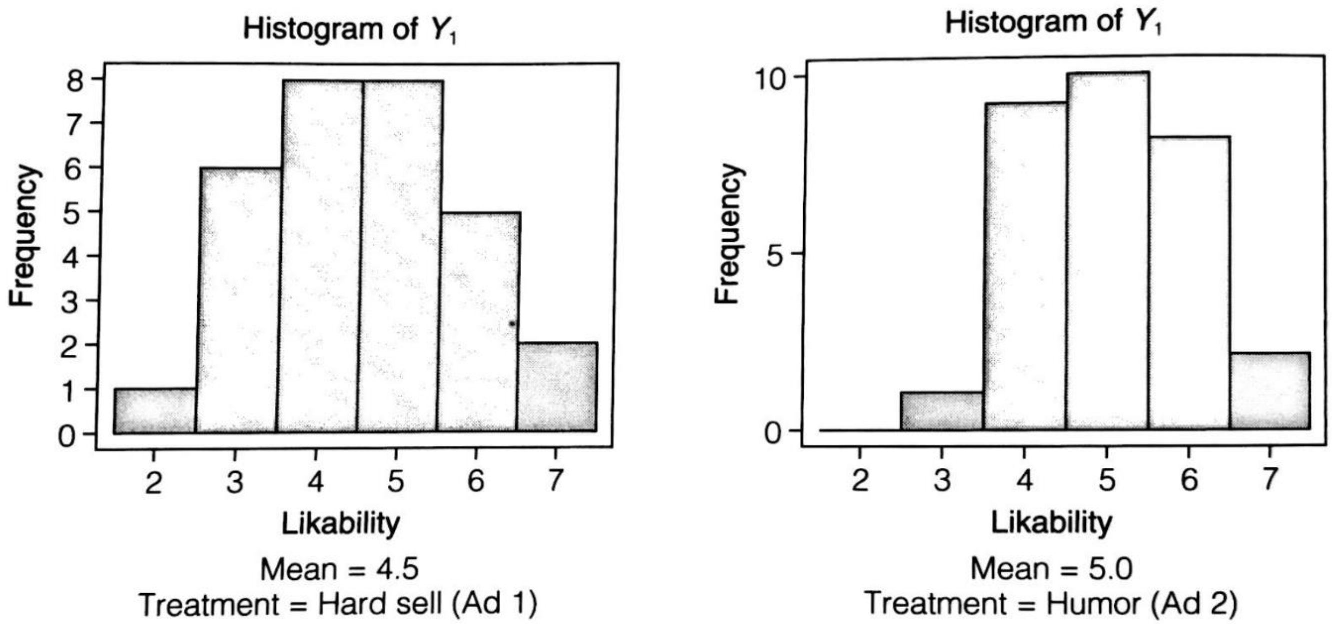
    which is a test statistic for $H_0\colon \ \alpha = 0$
  - The smaller Λ_H, the more significant the result.
- Just as ANOVA is essentially equivalent to multiple regression, MANOVA is essentially equivalent to canonical correlation.

Sample Problem (MANOVA): Testing Advertising Message Strategy

- Testing Advertising Message Strategy
  - Two responses Y_1 and Y_2 and One factor X
  - $Y_1 =$ "How much did you like the product?"
  - $Y_2 =$ "How likely are you to go out and buy this product?"
  - $X = \begin{cases} \text{Ad1: a strong persuasive approach ("hard sell")} \\ \text{Ad2: a light and humorous tone} \end{cases}$

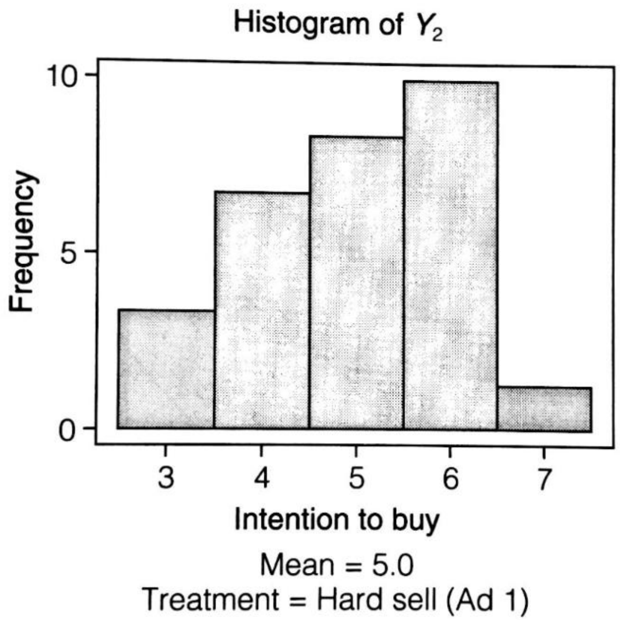
Sample Problem (MANOVA): Testing Advertising Message Strategy

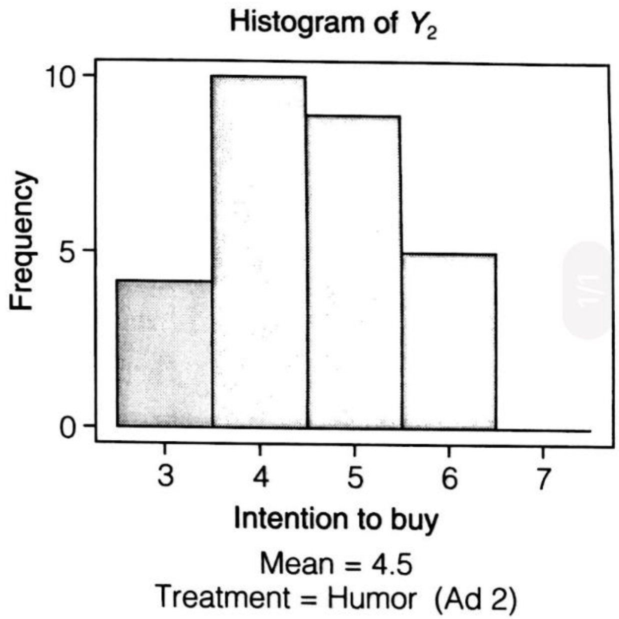
FIGURE 11.14 Distribution of Y_1 (likability) across ad-treatment groups



Sample Problem (MANOVA):
Testing Advertising Message Strategy

FIGURE 11.15
Distribution of Y_2 (intent) across ad-treatment groups





Sample Problem (MANOVA):
Testing Advertising Message Strategy

TABLE 11.10 Results from ANOVA of Y_1 (product likability)

Source	Sum of Squares	*df*	Mean Square	*F*-Value	Pr > *F*
Across	3.75	1	3.75	2.85	0.0970
Within	76.43	58	1.32		
Total	80.18	59			

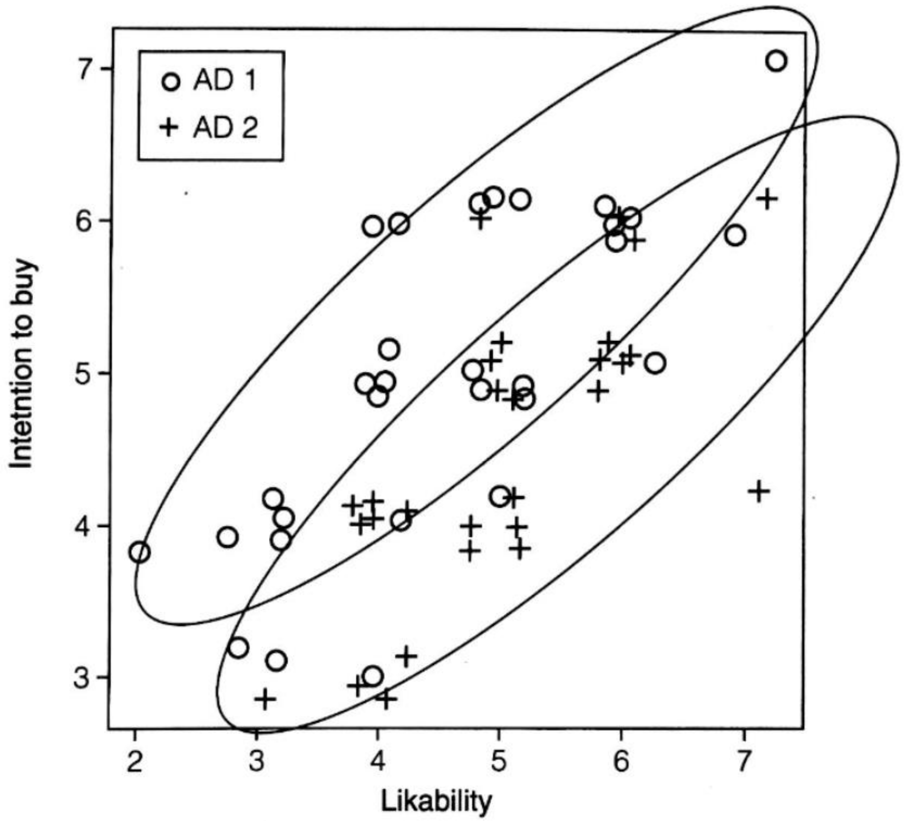
Sample Problem (MANOVA):
Testing Advertising Message Strategy

TABLE 11.11 Results from ANOVA of Y_2 (intention to buy)

Source	Sum of Squares	*df*	Mean Square	*F*-Value	Pr > *F*
Across	2.82	1	2.82	2.80	0.0999
Within	58.43	58	1.01		
Total	61.25	59			

Sample Problem (MANOVA): Testing Advertising Message Strategy

FIGURE 11.16 Joint distribution of Y_1 and Y_2 across ad-treatment groups



Sample Problem (MANOVA): Testing Advertising Message Strategy

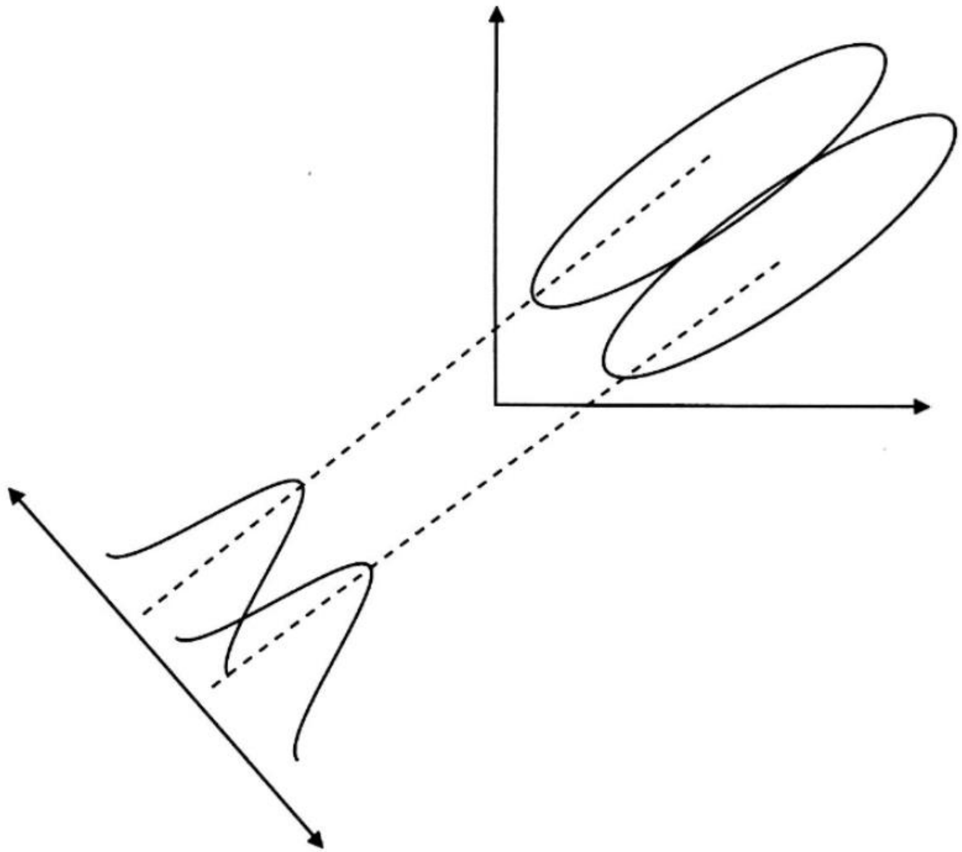
TABLE 11.12 Results from MANOVA of Y_1 (likability) and Y_2 (intention to buy) (Test for the hypothesis of no overall treatment effect)

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.7434	9.84	2	57	0.0002

		Y_1 (Likability)		Y_2 (Intention to Buy)	
Treatment	N	Mean	SD	Mean	SD
Ad 1	30	4.53	1.28	4.97	1.07
Ad 2	30	5.03	1.00	4.53	0.94

Sample Problem (MANOVA): Testing Advertising Message Strategy

FIGURE 11.17
Greater separation between groups with linear combination of Y_1 and Y_2



Testing Specific Contrasts in ANOVA

- Among μ_1, μ_2, μ_3, suppose we are interested in

$$H_0: \mu_1 = \mu_2 \quad \text{or} \quad \mu_1 - \mu_2 = 0$$

- With contrast vector $\mathbf{L}$ = [−1 1 0], the hypothesis is given by the linear combination $\mathbf{Lb} = \mathbf{0}$
- Test statistics:

$$(\mathbf{Lb})^{\mathbf{T}} \left(\mathbf{L}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{L}^{\mathbf{T}}\right)^{-\mathbf{1}} (\mathbf{Lb})$$

where $\mathbf{b} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{y}$

TABLE 11.13 Results of test comparing two treatment levels for Finnish liquor data

Contrast	Contrast SS	*df*	Mean Square	*F*-Value	Pr > *F*
Control vs package only	2.670	1	2.670	0.03	0.8701

Testing for Equal Within-Group Slopes in ANCOVA

- Compare a model of the same (common) slope

$$Y_{ij} = \mu + \tau_j + \beta Z_{ij} + \varepsilon_{ij} \quad \text{(Restricted model)}$$

with a model of different slopes within groups

$$Y_{ij} = \mu + \tau_j + \beta_j Z_{ij} + \varepsilon_{ij} \quad \text{(General model)}$$

- Test statistics:

$$F = \frac{(R_G^2 - R_R^2)/(df_G - df_R)}{(1 - R_G^2)/df_G} \sim F(df_G - df_R, df_G)$$

where R_G^2: R^2 for the general model

R_R^2: R^2 for the restricted model

df_G: degrees of freedom for general model

df_R: degrees of freedom for restricted model

Using MANOVA for Repeated Measures Designs

- Repeated-measures design
  - When multiple dependent variables exist because the same measure is taken from each subject at different points in time following the treatment.

Using MANOVA for Repeated Measures Designs

TABLE 11.14 Newfood sales data at three different points in time: First two months (Y_1), second two months (Y_2), and last two months (Y_3) after introduction

Price	Advertising	Y_1	Y_2	Y_3	Store Volume
Low	Low	225	190	205	34
Low	Low	323	210	241	41
Low	High	424	275	256	32
Low	High	268	200	201	28
Low	Low	224	190	209	34
Low	Low	331	178	267	41
Low	High	254	157	185	23
Low	High	492	351	365	37
Medium	Low	167	163	145	33
Medium	Low	226	148	170	39
Medium	High	210	134	128	30
Medium	High	289	212	200	27
Medium	Low	204	200	175	37
Medium	Low	288	171	247	43
Medium	High	245	120	117	30
Medium	High	161	116	111	19
High	Low	161	141	111	32
High	Low	246	126	184	42
High	High	128	83	83	29
High	High	154	122	102	24
High	Low	163	116	116	32
High	Low	151	112	119	36
High	High	180	100	75	29
High	High	150	122	101	24

Using MANOVA for Repeated Measures Designs: MANCOVA (Store volume)

TABLE 11.15 MANCOVA table for Newfood data

Test for the hypothesis of no overall *PRICE* effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.2677	4.66	6	30	0.0018

Test for the hypothesis of no overall *ADV* effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.2606	14.18	3	15	0.0001

Test for the hypothesis of no overall $PRICE \times ADV$ effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.5924	1.50	6	30	0.2133

Using MANOVA for Repeated Measures Designs

TABLE 11.16 Effects of time for Newfood data

Test for the hypothesis of no *TIME* effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.5235	7.28	2	16	0.0056

Manova test for the hypothesis of no *TIME* × *PRICE* effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.8105	0.89	4	32	0.4832

Manova test for the hypothesis of no *TIME* × *ADV* effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.2893	19.66	2	16	0.0001

Manova test for the hypothesis of no *TIME* × *PRICE* × *ADV* effect

Statistic	Value	F	Num df	Den df	$Pr > F$
Wilks's Λ	0.7003	1.56	4	32	0.2089

Using MANOVA for Repeated Measures Designs

FIGURE 11.18
Plot showing adjusted mean sales of Newfood for different time periods and different advertising-treatment levels