Analysis of Covariance (ANCOVA)

- **Fertilizer Example.** An experiment was conducted to see the effects of two treatments, a slow-release fertilizer (s) and a fast-release fertilizer (f), on seed yield (grams) of peanut plants were compared with a control (c), a standard fertilizer. Ten replications of each treatment were to be grown in a greenhouse study. When setting up the experiment, the researcher recognized that the 30 peanut plants were not exactly at the same level of development or health. Consequently, the researcher recorded the height (cm) of the plant, a measure of plant development and health, at the start of the experiment. The results of the experiment are shown in the table below.

c		s		f	
Yield	Height	Yield	Height	Yield	Height
12.2	45	16.6	63	9.5	52
12.4	52	15.8	50	9.5	54
11.9	42	16.5	63	9.6	58
11.3	35	15.0	33	8.8	45
11.8	40	15.4	38	9.5	57
12.1	48	15.6	45	9.8	62
13.1	60	15.8	50	9.1	52
12.7	61	15.8	48	10.3	67
12.4	50	16.0	50	9.5	55
11.4	33	15.8	49	8.5	40

- **Covariance Model**

$$Y_{ij} = \mu_{.} + \tau_i + \gamma(X_{ij} - \bar{X}_{..}) + \epsilon_{ij}$$

where:

1. $\mu_{.}$ is an overall mean.

2. τ_i are the fixed treatment effects, subject to the restriction $\sum \tau_i = 0$.

3. γ is a regression coefficient for the relation between Y and X.

4. X are constants

5. ϵ_{ij} are independent $N(0, \sigma^2)$.

- **Analysis using R**

```
> data=read.csv("FertilizerYield.csv",header=T)
> attach(data)

> treatment
> as.numeric(treatment)

> plot(height,yield,pch=as.numeric(treatment))
> legend(60,15.5,c("Control","Slow","Fast"),pch=c(1,3,2))



# Check for constancy of slopes
> anova(aov(yield~height*treatment))
Response: yield
                 Df  Sum Sq Mean Sq   F value    Pr(>F)
height            1   0.472   0.472   31.7908 8.344e-06 ***
treatment         2 213.904 106.952 7201.3001 < 2.2e-16 ***
height:treatment  2   0.061   0.031    2.0627    0.1491
Residuals        24   0.356   0.015
```

```
# Check if the treatment has an effect on the covariate
> anova(aov(height~treatment))
Response: height
                  Df Sum Sq Mean Sq F value Pr(>F)
factor(treatment)  2  303.8 151.900  1.9086 0.1678
Residuals         27 2148.9  79.589

> height.centered=height-mean(height)
> result=lm(yield~treatment+height.centered)

# Check for constant variance
> plot(result$fit,result$res,xlab="Fitted Values",ylab="Residuals")

# Check for normality of error terms
> qqnorm(result$res); qqline(result$res)
> shapiro.test(result$res)
W = 0.9772, p-value = 0.7467

> anova(result)
Response: yield
                Df  Sum Sq Mean Sq F value    Pr(>F)
treatment        2 207.683 103.841 6463.47 < 2.2e-16 ***
height.centered  1   6.693   6.693  416.62 < 2.2e-16 ***
Residuals       26   0.418   0.016

> anova(aov(yield~treatment))
Response: yield
          Df  Sum Sq Mean Sq F value    Pr(>F)
treatment  2 207.683 103.841  394.28 < 2.2e-16 ***
Residuals 27   7.111   0.263

> summary(result)
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     12.314173   0.041085  299.72   <2e-16 ***
treatmentf      -3.144156   0.060374  -52.08   <2e-16 ***
treatments       3.571637   0.057033   62.62   <2e-16 ***
height.centered  0.055810   0.002734   20.41   <2e-16 ***

# Computing the adjusted means
ybar.control=12.31
ybar.fast=12.31-3.14=9.17
ybar.slow=12.31+3.57=15.88

> tapply(yield,treatment,mean)
    c     f     s
12.13  9.41 15.83
> tapply(height,treatment,mean)
   c    f    s
46.6 54.2 48.9

# 95% joint C.I.s for the (mu.f-mu.c) and (mu.s-mu.c) using Bonferroni adjustment
> confint(result,level=(1-.05/2))
```

- **Covariance Model for Two-Factor Studies**

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma(X_{ijk} - \bar{X}_{...}) + \epsilon_{ijk}$$

  where:

  1. $\mu_{..}$ is an overall mean.
  2. α_i are the fixed factor A effects, subject to the restriction $\sum_{i=1}^{a} \alpha_i = 0$.
  3. β_j are the fixed factor A effects, subject to the restriction $\sum_{j=1}^{b} \alpha_j = 0$.
  4. $(\alpha\beta)_{ij}$ are the interaction effects, subject to the restriction $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$.
  5. γ is a regression coefficient for the relation between Y and X.
  6. X are constants
  7. ϵ_{ij} are independent $N(0, \sigma^2)$.

- **Salable Flowers Example.** A horticulturist conducted an experiment to study the effects of flower variety (factor A: LP, WB) and moisture level (factor B: low, high) on yield of salable flowers (Y). Because the plots were not of the same size, the horticulturist wished to use plot size (X) as the covariate. Six replications were made for each treatment. The results of the experiment are stored in the excel file *SalableFlowers.csv.*

- **Analysis using R**

```
> data=read.csv("SalableFlowers.csv",header=T)
> attach(data)
> A=factor(variety)
> B=factor(moisture)
> size2=size-mean(size)

> result2=lm(yield~A*B+size2)

> plot(size,yield,pch=as.numeric(A:B))
> legend(12,60,c("trt11","trt21","trt12","trt22"),pch=c(1,2,3,4))

> anova(result2)
Response: yield
          Df Sum Sq Mean Sq  F value    Pr(>F)
A          1  486.0   486.0  77.2841 4.018e-08 ***
B          1  486.0   486.0  77.2841 4.018e-08 ***
size2      1 3978.5  3978.5 632.6608 4.760e-16 ***
A:B        1   16.0    16.0   2.5511    0.1267
Residuals 19  119.5     6.3

> summary(result2)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   76.542      1.028  74.429  < 2e-16 ***
A2            -5.723      1.454  -3.937 0.000885 ***
B2            -9.000      1.448  -6.216 5.69e-06 ***
size2          3.277      0.130  25.203 4.59e-16 ***
A2:B2          3.277      2.052   1.597 0.126719
```