# Analysis of Variance and Design of Experiments

## Experimental Design Models

:::

## Lecture 13

## One-Way Classification in Experimental Design Models

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Slides can be downloaded from http://home.iitk.ac.in/~shalab/sp

## One way classification: Model

**Let us now consider the analysis of variance model with an additional constraint. Let**

$$Y_{ij} = \beta_i + \varepsilon_{ij}, \ i = 1, 2, ..., p; \ j = 1, 2, ..., n_i$$
$$= \bar{\beta} + (\beta_i - \bar{\beta}) + \varepsilon_{ij}$$
$$= \mu + \alpha_i + \varepsilon_{ij}$$

**with**

$$\mu = \bar{\beta} = \frac{1}{p}\sum_{i=1}^{p}\beta_i, \quad \alpha_i = \beta_i - \bar{\beta}, \quad \sum_{i=1}^{p} n_i\alpha_i = 0, \quad n = \sum_{i=1}^{p} n_i,$$

**and $\varepsilon_{ij}$ 's are identically and independently distributed with mean 0 and variance $\sigma^2$.**

## One way classification: Null hypothesis

**The null hypothesis is** $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$

**and the alternative hypothesis is** $H_1$ **: atleast one** $\alpha_i \neq \alpha_j$ **for all** $i, j.$

**This model is a one-way layout in the sense that the observations** $y_{ij}$ **'s are assumed to be affected by only one treatment effect** $\alpha_i$**.**

**So the null hypothesis is equivalent to testing the equality of *p* population means or equivalently the equality of *p* treatment effects.**

## One way classification: Least squares estimation

**We use the principle of least squares to estimate the parameters $\mu, \alpha_1, \alpha_2, ... \alpha_p$ .**

**Minimize the error sum of squares**

$$E = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \varepsilon_{ij}^2 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i)^2$$

**with respect to** $\mu, \alpha_1, \alpha_2, ..., \alpha_p$.

## One way classification: Least squares estimation

**The normal equations are obtained as**

$$\frac{\partial E}{\partial \mu} = 0 \Rightarrow -2 \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i) = 0$$

**or** $$n\mu + \sum_{i=1}^{p} n_i \alpha_i = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij} \qquad (1)$$

$$\frac{\partial E}{\partial \alpha_i} = 0 \Rightarrow -2 \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i) = 0$$

**or** $$n_i \mu + n_i \alpha_i = \sum_{j=1}^{n_i} y_{ij} \qquad (i = 1, 2, ..., p). \qquad (2)$$

**Using** $\sum_{i=1}^{p} n_i \alpha_i = 0$ **in (1) gives** $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij} = \frac{G}{n} = \bar{y}_{oo}$

**where** $G = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}$ **is the grand total of all the observations.**

## One way classification: Least squares estimation

**Substituting $\hat{\mu}$ in (2) gives**

$$\hat{\alpha}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} - \hat{\mu}$$

$$= \frac{T_i}{n_i} - \hat{\mu}$$

$$= \bar{y}_{io} - \bar{y}_{oo}$$

**where** $T_i = \sum_{j=1}^{n_i} y_{ij}$ **is the treatment total due to *i*th effect** $\alpha_i$ **, i.e., a total of all the observations receiving the *ith* treatment and**

$$\bar{y}_{io} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}.$$

## One way classification: Sum of squares due to error

**Now the fitted model is** $y_{ij} = \hat{\mu} + \hat{\alpha}_i$ **and the error sum of squares after substituting** $\hat{\mu}$ **and** $\hat{\alpha}_i$ **in** ***E*** **becomes**

$$E = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left[ (y_{ij} - \bar{y}_{oo}) - (\bar{y}_{io} - \bar{y}_{oo}) \right]^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2 - \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\bar{y}_{io} - \bar{y}_{oo})^2$$

$$= \left( \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n} \right) - \left( \sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{G^2}{n} \right)$$

**where** $\frac{G^2}{n}$ **is called the correction factor (*CF*) and the total sum of squares *(TSS)*** $TSS = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n},$

# One way classification: Sum of squares

**To obtain a measure of variation due to treatments, let**

$$H_0 = \alpha_1 = \alpha_2 = ... = \alpha_p = 0$$

**be true. Then the model becomes**

$$Y_{ij} = \mu + \varepsilon_{ij}, \ i = 1, 2, ..., p; \ j = 1, 2, ..., n_i.$$

**Minimizing the error sum of squares** $E_1 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \mu)^2$ **with respect to** $\mu$**, the normal equation is obtained as**

$$\frac{\partial E_1}{\partial \mu} = 0 \Rightarrow -2 \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \mu) = 0$$

**or** $\hat{\mu} = \dfrac{G}{n} = \bar{y}_{oo}.$

**Substituting** $\hat{\mu}$ **in** $E_1$, **the error sum of squares becomes**

$$E_1 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu})^2 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}.$$

# One way classification: Sum of squares due to treatment

**Note that**

**$E_1$: Contains variation due to treatment and error both**

**$E$: Contains variation due to error only**

**So $E_1$- $E$ : contain variation due to treatment only.**

**The sum of squares due to treatment *( SSTr )* is given by**

$$SSTr = E_1 - E$$

$$SSTr = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\bar{y}_{io} - \bar{y}_{oo})^2$$

$$= \sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{G^2}{n}.$$

## One way classification: Sum of squares due to error

**The following quantity is called the error sum of squares or sum of squares due to error (SSE)**

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_{io})^2 .$$

**These sum of squares forms the basis for the development of tools in the analysis of variance. We can write**

$$TSS = SSTr + SSE.$$

**The distribution of degrees of freedom among these sum of squares is as follows:**

## One way classification: Degrees of freedom

**The distribution of degrees of freedom among these sum of squares is as follows:**

- **The total sum of squares is based on *n* quantities subject to the constraint that** $\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{oo})=0$**, so *TSS* carries (*n* - 1) degrees of freedom.**
- **The sum of squares due to the treatments is based on *p* quantities subject to the constraint** $\sum_{i=1}^{p} n_i(\bar{y}_{io}-\bar{y}_{oo})=0$ **, so *SSTr* has (*p* - 1) degrees of freedom.**
- **The sum of squares due to errors is based on *n* quantities subject to *p* constraints** $\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{io})=0,\ i=1,2,\ldots,p$ **so *SSE* carries (*n* - *p*) degrees of freedom.**

## One way classification: Distributions of sum of sqaures

**Also, note that**

$$TSS = SSTr + SSE,$$

**the *TSS* has been divided into two orthogonal components – *SSTr* and *SSE.***

**Moreover, all *TSS, SSTr* and *SSE* can be expressed in a quadratic form. Since $\varepsilon_{ij}$'s are assumed to be identically and independently distributed following $N(0,\sigma^2)$, so $y_{ij}$ 's are also independently distributed following $N(\mu+\alpha_i,\ \sigma^2)$.**

## Recall Theorems:

**Theorem 7:** **Let** $Y=(Y_1,Y_2,...,Y_n)'$ **follow a multivariate normal distribution** $N(\mu,\Sigma)$ **with mean vector** $\mu$ **and positive definite covariance matrix** $\Sigma$**.**

**Let** $Y'A_1Y \sim \chi^2(p_1, \mu' A_1 \mu)$ **and** $Y'A_2Y \sim \chi^2(p_2, \mu' A_2 \mu)$**.**

**Then** $Y'A_1Y$ **and** $Y'A_2Y$ **are independently distributed if** $A_1\Sigma A_2 = 0.$

## Recall Theorems:

**Theorem 9: Let** $Z = Y - X\beta^0$

$$Q_1 = Z'X(X'X)^{-1}X'Z$$

$$Q_2 = Z'[I - X(X'X)^{-1}X']Z.$$

**Then** $\frac{Q_1}{\sigma^2}$ **and** $\frac{Q_2}{\sigma^2}$ **are independently distributed.**

**Further, when** $H_0$ **is true, then**

$$\frac{Q_1}{\sigma^2} \sim \chi^2(p)$$

**and** $\frac{Q_2}{\sigma^2} \sim \chi^2(n-p)$

**where** $\chi^2(m)$ **denotes the** $\chi^2$ **distribution with '*m*' degrees of freedom.**

# One way classification: Distributions of sum of sqaures

**Now using the theorems 7 and 9 with** $q_1 = SSTr,\ q_2 = SSE,$

**we have under *H*$_0$,**

$$\frac{SSTr}{\sigma^2} \sim \chi^2(p-1)$$

**and**

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-p).$$

**Moreover, *SSTr* and *SSE* are independently distributed.**

**The mean squares is defined as the sum of squares divided by the degrees of freedom. So the**

- **mean square due to treatment is** $MSTr = \dfrac{SSTr}{p-1}$ **and**
- **mean square due to error is** $MSE = \dfrac{SSE}{n-p}.$

## One way classification: Test statistics

**Thus, under** $H_0$,

$$F = \frac{\left(\frac{MSTr}{\sigma^2}\right)}{\left(\frac{MSE}{\sigma^2}\right)} = \frac{MSTr}{MSE} \sim F(p-1,\ n-p).$$

**The decision rule is that reject** $H_0$ **at** $\alpha$ **% level of significance if**

$$F > F_{1-\alpha,\ p-1,\ n-p}$$

## One way classification: Test statistics

**If $H_0$ does not hold true, then**

$$\frac{MSTr}{MSE} \sim noncentral\ F(p-1,\ n-p,\ \delta)$$

**where** $\delta = \sum_{i=1}^{p} \frac{n_i \alpha_i^2}{\sigma^2}$ **is the non-centrality parameter.**

**Note that the test statistic** $\frac{MSTr}{MSE}$ **can also be obtained from the likelihood ratio test.**

# One way classification: ANOVA table

**The computations of this test of hypothesis can be represented in the form of an analysis of variance table.**

**ANOVA for testing** $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$

| Source of variation | Degrees of freedom | Sum of squares | Mean squares | *F*-value |
|---|---|---|---|---|
| Due to treatments | *p* – 1 | *SSTr* | *MStr* | $\frac{MSTr}{MSE}$ |
| Error | *n* – *p* | *SSE* | *MSE* | |
| Total | *n* – 1 | *TSS* | | |

## One way classification: Multiple comparison test

**If $H_0$ is rejected, then we go for multiple comparison tests and try to divide the population into several groups having the same effects.**

## One way classification: Expectation of SS due to treatment

**Now we find the expectations of *SSTr* and *SSE*.**

$$E(SSTr) = E\left[\sum_{i=1}^{p} n_i(\bar{y}_{io} - \bar{y}_{oo})^2\right] = E\left[\sum_{i=1}^{p} n_i\left\{(\mu + \alpha_i + \bar{\varepsilon}_{io}) - (\mu + \bar{\varepsilon}_{oo})\right\}^2\right]$$

**where**

$$\bar{\varepsilon}_{io} = \frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{ij}\ , \quad \bar{\varepsilon}_{oo} = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\varepsilon_{ij} \quad \text{and} \quad \sum_{i=1}^{p}\frac{n_i\alpha_i}{n} = 0.$$

$$E(SSTr) = E\left[\sum_{i=1}^{p} n_i\left\{\alpha_i + (\bar{\varepsilon}_{io} - \bar{\varepsilon}_{oo})\right\}^2\right]$$

$$= \sum_{i=1}^{p} n_i E(\alpha_i^2) + \sum_{i=1}^{p} n_i E(\bar{\varepsilon}_{io} - \bar{\varepsilon}_{oo})^2 + 0.$$

# One way classification: Expectation of SS due to treatment

$$E(\bar{\varepsilon}_{io}^2) = Var(\bar{\varepsilon}_{io}) = Var\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{ij}\right) = \frac{1}{n_i^2}n_i\sigma^2 = \frac{\sigma^2}{n_i}$$

$$E(\bar{\varepsilon}_{oo}^2) = Var(\bar{\varepsilon}_{oo}) = Var\left(\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\varepsilon_{ij}\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

$$E(\bar{\varepsilon}_{io}\bar{\varepsilon}_{oo}) = Cov(\bar{\varepsilon}_{io}, \bar{\varepsilon}_{oo}) = \frac{1}{n_i n}Cov\left(\sum_{j=1}^{n_i}\varepsilon_{ij}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\varepsilon_{ij}\right) = \frac{n_i\sigma^2}{n_i n} = \frac{\sigma^2}{n}.$$

# One way classification: Expectation of SS due to treatment

**Since**

$$E(SSTr) = \sum_{i=1}^{p} n_i \alpha_i^2 + \sigma^2 \sum_{i=1}^{p} n_i \left( \frac{1}{n_i} - \frac{1}{n} \right)$$

$$= \sum_{i=1}^{p} n_i \alpha_i^2 + (p-1)\sigma^2$$

**or** $$E\left( \frac{SSTr}{p-1} \right) = \sigma^2 + \frac{\sum_{i=1}^{p} n_i \alpha_i^2}{p-1}$$

**or** $$E(MSTr) = \sigma^2 + \frac{\sum_{i=1}^{p} n_i \alpha_i^2}{p-1}.$$

# One way classification: Expectation of SS due to error

**Next**

$$E(SSE) = E\left[\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{io})^2\right] = E\left[\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left\{(\mu+\alpha_i+\varepsilon_{ij})-(\mu+\alpha_i+\overline{\varepsilon}_{io})\right\}^2\right]$$

$$= E\left[\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\varepsilon_{ij}-\overline{\varepsilon}_{io})^2\right] = \sum_{i=1}^{p}\sum_{j=1}^{n_i}E(\varepsilon_{ij}^2+\overline{\varepsilon}_{io}^2-2\varepsilon_{ij}\overline{\varepsilon}_{io})$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\sigma^2+\frac{\sigma^2}{n_i}-\frac{2\sigma^2}{n_i}\right) = \sigma^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\frac{n_i-1}{n_i}\right)$$

$$= \sigma^2\sum_{i=1}^{p}\frac{n_i(n_i-1)}{n_i} = \sigma^2\sum_{i=1}^{p}(n_i-1) = (n-p)\sigma^2$$

**or** $$E\left(\frac{SSE}{n-p}\right) = \sigma^2$$

**or** $$E(MSE) = \sigma^2.$$

**Thus *MSE* is an unbiased estimator of** $\sigma^2$.