Reduced / Means Model

Factor Effect Model

multiple t-tests → type I error

ANOVA

purpose: same as t-test, method is different

ANOVA: analysis of variance → misnomer, actually analyzes significant differences in means across 3+ groups using variance

1. → variation in mean across groups
2. → variation in each group
3. → insight if groups have different means

factor: independent var or dependent?

Z-test vs t-test vs ANOVA

Z-test

- large sample size ($n>30$) [or] population σ is known
- Assumes underlying population is normal

→ one sample (one sample mean)

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$ (two tailed)

$\mu > \mu_0$ (right tailed)

$\mu < \mu_0$ (left tailed)

→ two sample (two sample means)

2 independent samples from populations with same means

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$\mu_1 > \mu_2$ (right-tailed)

$\mu_1 < \mu_2$ (left-tailed)

t-test

- Pop σ not known / small sample ($n<30$)
- Shape of distribution depends on dof = $n-1$

→ One sample t-test

Sample μ differs significantly from pop. μ

→ Two sample t-test

Compares means of independent groups

→ Paired t-test

formula different

Issue: Check normality (Shapiro-Wilk test)

When ANOVA?

1. Multiple groups (3+)
2. Dependent var continuous
3. Independent var categorical → leads to groups
4. Groups should be independent
5. Normal distribution
6. Homoscedasticity: groups should have similar variances → if not ratio messes up

minor deviation is fine

some commonality?

One-way ANOVA = compare means of two or more groups (independent var) on one dependent variable to see if group means differ

→ assumes means across distributions → same

→ uses variance to test H_0: no difference in distributions

if equal variance b/w groups, feature has no impact on response

one-way ANOVA = impact of one single factor on particular outcome | two way = two factors impact, and whether those two factors interact

→ used when categorical variable has at least 3 groups

Hypothesis

H_0: all group means are the same

$\mu_1 = \mu_2 = \cdots = \mu_n$

H_1: at least one is different → we don't know which is different

$$F\text{-statistic} = \frac{MSB}{MSW}$$

MSB: Mean Square Between groups

MSW: " " Within groups

Applications

Anything with 3 or more groups like t-test

- Quant trading
- Credit risk

var1 = loan purpose category

var2 = credit risk (FICO) category

dependent = interest rate for borrowing

MANOVA

dep var → two or more continuous variables

indep var → categorical var with 3 or more levels

Example

Spearman rank corr vs one way ANOVA

no dependent independent

Regression vs ANOVA

continuous predictors — categorical predictors

Setting up ANOVA (two way)

$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \varepsilon_{ijk}$

α_i = effect (effect) of factor A ($i = 1 \ldots a$)

β_j = " " " " B ($j = 1 \ldots b$)

$\alpha_i\beta_j$ = interaction of ith level of factor A, j^{th} level of factor B

ε_{ijk} = error associated with k^{th} obs. at level i of factor A, level j of factor B

Test of effect of A

$H_0: \alpha_i = 0$

H_A: Not all $\alpha_i = 0$

Test of effect of B

$H_0: \beta_i = 0$

H_A: Not all $\beta_i = 0$

Test of interaction effect

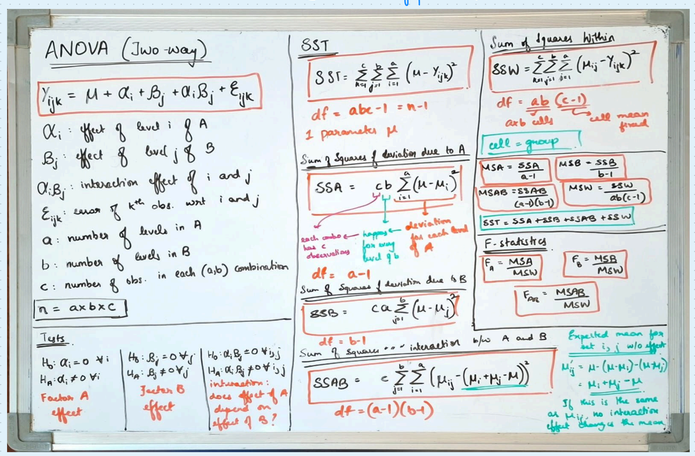
$H_0: \alpha_i\beta_j = 0$

H_A: Not all $\alpha_i\beta_j = 0$

$SST = SSA + SSB + SSAB + SSW$

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(Y_{ijk} - \mu)^2$$

c → no of observations in each group (assume each grp → same)



Setting up ANOVA (one-way)

k levels → k groups

k = no. of groups / sample

n_i = no. of observations in group i

n = total obs.

Y_{ij} = Observation j of group i

μ

μ_i

total variation

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \mu)^2$$

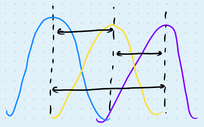
dof for SST is n-1

$$MST = \frac{SST}{n-1}$$

$$SSB = \sum_{i=1}^{k} n_i \times (\mu_i - \mu)^2$$

dof for SSB is k-1

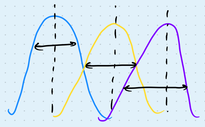
$$MSB = \frac{SSB}{k-1}$$



$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \mu_i)^2$$

dof for SSW is n-k

$$MSW = \frac{SSW}{n-k}$$



$SST = SSB + SSW$

Cochran's Theorem

$\frac{SSB}{\sigma^2}, \frac{SSW}{\sigma^2}$ are χ^2 distributed — quadratic forms LA

F-test

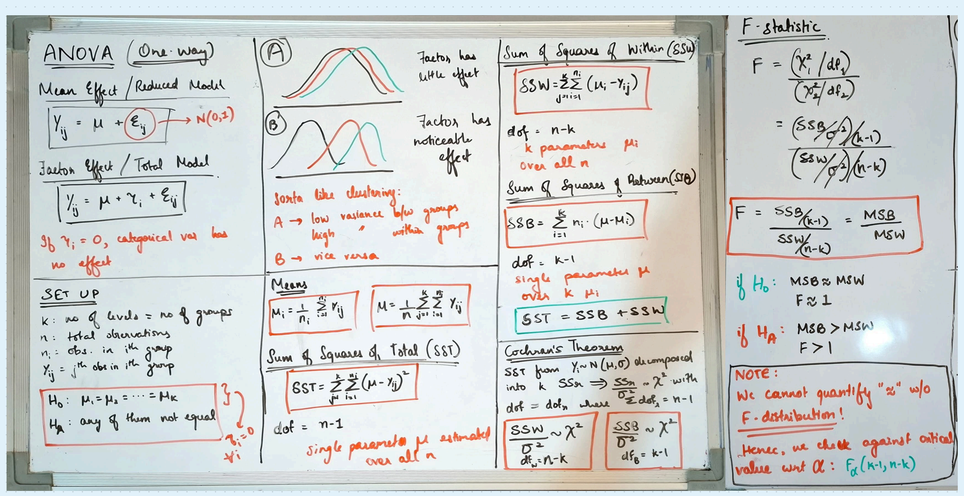
if H_0 true → means are same → MSB = MSW

if H_0 true → " " different → MSB > MSW

$$F = \frac{SSB/k-1}{SSW/n-k} = \frac{MSB}{MSW} \approx 1 \quad \text{if means same}$$

$$\frac{MSB}{MSW} > 1 \quad \text{if means different}$$

test statistic → right tailed → only checking if variation b/w groups greater than within

NOTE: two dofs to note in numerator and denominator

$$F_{a,b} = \frac{MSB}{MSW}$$