The Analysis of Variance

Compare means of more than two populations

Examples:

1. To compare the effectiveness of three methods of teaching the programming of a certain computer

method A: 73, 77, 67, 71
method B: 91, 81, 87, 85
method C: 72, 77, 76, 79

2. To compare the weight losses of certain machine parts due to friction when three different lubricants were used

lub A: 12.2, 11.8, 13.1, 11.0
lub B: 10.9, 5.7, 13.5, 9.4
lub C: 12.7, 19.9, 13.6, 18.3

basic framework and assumptions

Treatment1: $X_{11}, X_{12}, \ldots, X_{1J} \sim N(\mu_1, \sigma_1)$
Treatment2: $X_{21}, X_{22}, \ldots, X_{2J} \sim N(\mu_2, \sigma_2)$

Treatment$I$: $X_{I1}, X_{I2}, \ldots, X_{IJ} \sim N(\mu_I, \sigma_I)$

1. Each sample is a random sample
2. Samples are independent with each other
3. $\sigma_1 = \sigma_2 = \ldots = \sigma_I$

Hypotheses: $H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$ vs
H_a : at least two of the μ_i's are different

Test statistic??
Without H_0:

With H_0:

Mean square for treatment(MSTr)

$$\text{MSTr} = \frac{J}{I-1} \sum_{i=1}^{I} (\bar{X}_{i.} - \bar{X}_{..})^2$$

Mean square for error(MSE)

$$\text{MSE} = \frac{S_1^2 + S_2^2 + \ldots + S_I^2}{I}$$

Test statistic

$$F = \frac{\text{MSTr}}{\text{MSE}}$$

Theoretical results
1.

$$E(\text{MSE}) = \sigma^2$$

$$E(\text{MSTr}) = \sigma^2 + \frac{J}{I-1} \sum_{i=1}^{I} (\mu_i - \mu)^2$$

where

$$\mu = \frac{\sum \mu_i}{I}$$

Remark: MSTr tends to overestimate σ, because it is biased

2. When H_0 is right

$$F = \frac{\text{MSTr}}{\text{MSE}} \sim F(I-1, IJ-I)$$

Rejection region: $(F_{\alpha,I-1,IJ-i}, +\infty)$

Computational Formulas and ANOVA table

Total sum of squares:

$$\text{SST} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = \sum_i \sum_j x_{ij}^2 - \frac{1}{IJ} x_{..}^2$$

where $x_{..} = \Sigma_i \Sigma_j x_{ij}$.

Treatment Sum of Sqares

$$\text{SSTr} = \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..}^2) \frac{1}{J} \sum_{i.}^{2} - n \frac{1}{IJ} x_{..}^2$$

where $x_{i.} = \Sigma_j x_{ij}$.

Error Sum of Squares

$$\text{SSE} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2$$

An important indentity

$$\text{SST} = \text{SSTr} + \text{SSE}$$

ANOVA table

Sou. of Var.	df	Sum of Sqar.	Mean Squares	F
Treatments	$I-1$	SSTr	MSTr=SSTr/$(I-1)$	F=MSTr/MSE
Error	$IJ-I$	SSE	MSE=SSE/$(IJ-I)$	
Total	$IJ-1$	SST		

Multiple comparisons in ANOVA

Tukey's procedure

With probability $1-\alpha$

$$\bar{X}_{i1.} - \bar{X}_{i2.} - Q_{\alpha,I,I(J-1)}\sqrt{\text{MSE}/J}$$
$$\leq \mu_{i1} - \mu_{i2} \leq$$
$$\bar{X}_{i1.} - \bar{X}_{i2.} + Q_{\alpha,I,I(J-1)}\sqrt{\text{MSE}/J}$$

for every $i1$ and $i2$ ($i1 = 1, \ldots, I$ and $i2 = 1, 2, \ldots, I$) with $i1 \neq i2$

1. $Q_{\alpha,I,IJ-I}$: the upper-tail α critical value of the sudentized range distribution.
2. Simultaneous confidence statement and confidence level
3. Indentify different population means