Single Factor ANOVA

1 Introduction to Analysis of Variance

1.1 Using multiple pairwise comparisons to test multi-sample hypothesis → INVALID

- inflated risk of type I error: $P(\text{error}) = 1 - (1-\alpha)^z$ where z = no. of pairwise comparisons

1.2 Single Factor ANOVA Hypotheses:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$; k = # experimental groups (samples)

H_A: group means are not all equal

1.3 Notation

1.3.1 individual measurements: X_{ij} = j[th] measurement in i^{th} group

1[st] subscript = group # 2[nd] subscript = individual w/in group

1.3.2 sample sizes: n_i = # in group i (sample size of i^{th} group);

$N = \Sigma n_i$ → total # in expt

1.3.3 means: 2 kinds of means: group means (group i): $\bar{X}_i$ & grand mean: $\bar{X}$

1.3.4 Summation: $\sum_i X_i$ $\sum_i \sum_j X_{ij} = \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} X_{ij} \right)$

1.4 Linear model for single-factor ANOVA:

$X_{ij} = \mu + \tau_i + \varepsilon_{ij}$ where: μ is the grand mean

τ_i is the difference betw/ μ and i^{th} group mean

ε_{ij} is random error within i^{th} group

assumptions: (1) $\bar{\varepsilon}_{ij} = 0$ i.e., no bias w/in sample

(2) $\text{var}(\varepsilon_{ij}) = \sigma^2$ i.e., equal variances among groups

(3) ε_{ij} ~ normal

- reject H_0 if variation in X_{ij} influenced more by τ_i than by ε_{ij}
- do not reject H_0 if variation in X_{ij} influenced more by ε_{ij} than by τ_i

2 ANOVA Calculations

2.1 Partitioning variability & Sources of variation

- total variability
- within group variability ("error")
- among group variability

Deviation of each datum from grand mean partitioned into:

(1) deviation from group mean

(2) deviation of group mean from grand mean

$$(X_{ij} - \bar{X}) = (X_{ij} - \bar{X}_i) + (\bar{X}_i - \bar{X})$$

2.2 Sums of squares & degrees of freedom:

total: total $SS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$ total DF = $N - 1$

among-groups: among-groups $SS = \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2$ among-groups $DF = k - 1$

w/in-groups: within-groups $SS = \sum_{i=1}^{k} \left[\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \right]$

w/in-groups DF = total DF − groups DF = $N - k$

(continued)

2 ANOVA Calculations (continued)

2.3 *SS* & *DF* are additive: total *SS* = groups *SS* + error *SS*
total *DF* = groups *DF* + error *DF*

2.4 Mean Squared Deviation from mean: divide *SS* by *DF*

$$\text{groups MS} = \frac{\text{groups SS}}{\text{groups DF}} \qquad \text{error MS} = \frac{\text{error SS}}{\text{error DF}}$$

2.5 Summary of Calculations:

Source	Sum of squares (*SS*)	Degrees of freedom (*DF*)	Mean Square (*MS*)
Total	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2$	$N-1$	
among groups	$\sum_{i=1}^{k} n_i(\overline{X}_i-\overline{X})^2$	$k-1$	$\frac{\text{groups SS}}{\text{groups DF}}$
within groups	$\sum_{i=1}^{k}\left[\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2\right]$	$N-k$	$\frac{\text{error SS}}{\text{error DF}}$

- Summary of calculations w/ machine formulas − Zar, Table 10.2

3 Interpreting ANOVA

3.1 ANOVA Hypothesis Test

3.1.1 If H_o true, then groups MS & error MS each estimate common pop var (σ^2)
If pop means ≠, then groups MS > error MS
∴ test for equality of means using variance ratio test w/ groups MS in numerator

$$F = \frac{groups\ MS}{errorMS}$$

3.1.2 Critical value = $F_{\alpha(1),(k-1),(N-k)}$

3.1.3 If $F_{calc} \geq F_{crit}$, then reject H_0 => all *k* pop means not equal
→ to determine which means differ, use multiple comparisons test (e.g., Tukey test)

3.2 ANOVA with Means & Variances only (not original data)

3.2.1 First, determine SS_i

$$s_i^2 = (s_i)^2 = n_i(s_{\overline{X}_i})^2 = \frac{SS_i}{n_i - 1}$$

3.2.2 Next, calculate error *SS* & groups *SS*

$$\text{error SS} = \sum_{i=1}^{k} SS_i = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

$$\text{groups SS} = \sum_{i=1}^{k} n_i \overline{X}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \overline{X}_i\right)^2}{\sum_{i=1}^{k} n_i}$$

3.2.3 Complete ANOVA using groups SS & error SS (& groups DF & error DF)

3 Interpreting ANOVA (continued)

3.3 Two types of ANOVA

3.3.1 Fixed effects model (model I): Levels of factor (treatments) specifically chosen

Hypotheses: H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

3.3.2 Random effects model (model II): Levels of factor selected @ random

Null hypothesis: H_0: no difference in measurements among groups

– calculations identical to fixed effects model

3.3.3 Distinction very important for ANOVA involving ≥ 2 factors

3.4 Assumptions

3.4.1 Equal sample variances

– ANOVA fairly robust to violations, provided all n_i nearly equal

– if not, then different P(type I error):

→ if larger variances w/ larger samples, P(type I error) < α;

→ if larger variances w/ smaller samples, P(type I error) > α;

3.4.2 Samples from normal populations

– ANOVA fairly robust to violations (both skew & kurtosis)

4 Confidence Limits for Population Means

4.1 Confidence Interval calculation similar to two sample case

$1-\alpha$ confidence limits for μ_i ($\mu + \tau_i$): $\overline{X}_i \pm t_{\alpha(2),\nu}\sqrt{\frac{s^2}{n_i}}$

– where s^2 is error MS & ν is error DF (= $N-k$)

4.2 CI calculation appropriate only when know that μ_i different than other μ's

=> first must perform multiple comparison test (e.g., Tukey test)