13.2 : Analysis of variance : testing for the equality of K population means.

Analysis of variance can be used to test for the equality of K population means. The general form of the hypotheses tested is

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots\dots = \mu_K.$$

H_1 : Not all population means are equal.

Where

μ_j = mean of the jth population.

* We assume that a simple random sample of size n_j has been selected from each of the K populations or treatments. For the resulting sample data, let

i: الصف
j: العمود

x_{ij} = value of observation i for treatment j.

n_j = number of observations for treatment j.

$\bar{x}_j$ = sample mean for treatment j.

S_j^2 = sample variance for treatment j.

S_j = sample standard deviation for treatment j.

→ Testing for the equality of K population means sample mean for Treatment j :

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n_j}$$

→ sample variance for treatment j

$$S_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n_j - 1}$$

• The overall sample mean denoted $\bar{\bar{x}}$ is the sum of all the observations divided by the total number of observations, That is →

→ over sample mean :

$$\bar{\bar{X}} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} x_{ij}}{n_T}$$

Where $n_T = n_1 + n_2 + \dots + n_K$.

→ If the size of ach sample is n, $n_T = Kn$, in this case equation reduces to :

$$\bar{\bar{X}} = \frac{\sum_{j=1}^{n} \bar{X}_j}{K}$$

حالة خاصة

زي table 13.1

In the words, whenever the sample sizes are equal, the overall sample mean is just the average of the K-sample means.

→ Between-treatments estimate of population variance

we introduced the concept of a between-treatments estimate of δ^2 and showd how to compute it when the sample size ~~were~~ equal. This estimate of δ^2 is called the mean square due to treatments and is denoted MSTR. The general formula for computing MSTR is

$$\boxed{MSTR = \frac{\sum_{j=1}^{n} n_j(\bar{x}_j - \bar{\bar{x}})^2}{K-1}}$$

$\sum_{j=1}^{n} n_j(\bar{x}_j - \bar{\bar{x}})^2$: sum of squares due to treatments (denoted SSTR).

$K-1$: degrees of freedom associated with SSTR.

→ Mean square due to treatments

$$MSTR = \frac{SSTR}{K-1}$$

where $SSTR = \sum_{j=1}^{n} n_j (\bar{x}_j - \bar{\bar{x}})^2$.

Note :

If H_0 is true , MSTR provides an unbiased estimate of σ^2.

If the means of K populations are not equal , MSTR is not an unbiased estimate of σ^2.

→ Within-treatments estimate of population variance .

We introduced the concept of a within-treatments estimate of σ^2 and showed how to compute it when the sample sizes were equal. This estimate of σ^2 is called the mean square due to error and is denoted MSE. The general formula for computing MSE is :

$$\boxed{MSE = \frac{\sum_{j=1}^{n} (n_j - 1) s_j^2}{n_T - K}}$$

$\sum_{j=1}^{n} (n_j - 1) S_j^2$: sum of squares due to error and is denoted SSE .

$n_T - K$: degrees of freedom associated with SSE .

→ Mean square due to error :

$$MSE = \frac{SSE}{n_T - K} \quad \text{where} \quad SSE = \sum_{j=1}^{n} (n_j - 1) S_j^2 .$$

Note : MSE is based on the variation within each of the treatments ; it is not influenced by whether the null hypothesis is true. Thus , MSE always provieds an unbiased estimate of σ^2 .

→ Comparing the variance estimates : The F test .

- Test statistic for the equality of K population means

$$F = \frac{MSTR}{MSE}$$

The test statistic follows an F distribution with K-1 degrees of freedom in the numerator and $n_T - K$ degrees of freedom in the denominator.

df of MSTR = $K-1$

df of MSE = $n_T - K$

بتعامل معه كأنه upper tail test

→ Rejection Rule :

- Reject Ho if p-value ≤ α . (p-value approach).

By F-table

- Reject Ho if $F \geq F_\alpha$ (critical value approach).

→ ANOVA table :

source of variation	df	sum of squares	mean square	F
treatment [illegible] Treatments	$K-1$	SSTR	$MSTR = \frac{SSTR}{df}$	F statistic = $\frac{MSTR}{MSE}$
Error [illegible] Error	$n_T - K$	SSE	$MSE = \frac{SSE}{df}$	
Total	df associative with SST	SST		

ملخص لكل الشيء بكون في هاد الجدول

★ ANOVA table :

→ Total sums of squares .

$$SST = \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{\bar{x}})^2 .$$

→ partitioning of sum of squares = ضنی بعض

$$SST = SSTR + SSE .$$

→ ANOVA table of exp :

Source of var.	df	SS	ME	F
treatments	2	516	258	9
Error	15	430	28.67	
Total	17	946	—	—

reject if p-value < α يعني upper test دائماً في ANOVA

and reject if $F > F_\alpha$.

→ p-value < 0.01 so reject H_0 ($\alpha = 0.05$)

Not all μ_j are equal ($\alpha = 0.05$).

By F-table

→ $F_\alpha = F_{0.05} = 3.68$

$F > F_\alpha$ so reject H_0 ($\alpha = 0.05$)

Not all μ_j are equal ($\alpha = 0.05$).