ANOVA (Analysis of variance)

group id	Sample			sample size	population mean		sample mean	sample sd
group 1:	y_{11}, y_{12},	$\cdots$	y_{1n_1}	n_1	μ_1	σ	$\bar{y}_1$	S_1
group 2:	y_{21}, y_{22},	$\cdots$	y_{2n_2}	n_2	μ_2	σ	$\bar{y}_2$	S_2
$\vdots$				$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
group t:	y_{t1}, y_{t2},	$\cdots$	y_{tn_t}	n_t	μ_t	σ	$\bar{y}_t$	S_t

$$\sum_{i=1}^{t} n_i = N \quad \leftarrow \text{ total sample size}$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad \leftarrow \text{ sample mean of } i\text{-th group}$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{t} \sum_{j=1}^{n_i} y_{ij} \quad \leftarrow \text{ sample mean of whole sample}$$

$$S_i^2 = \frac{1}{n_i - 1} \sum_{i=1}^{n_i} \left(y_{ij} - \bar{y}_i \right)^2 \quad \leftarrow \text{ sample sd of } i\text{-th grou}$$

$$SS Total = SS Trt + SSE$$

$$\sum_i \sum_j \left(y_{ij} - \bar{y} \right)^2 = \sum_i n_i \left(\bar{y}_i - \bar{y} \right)^2 + \sum_i \sum_j \left(y_{ij} - \bar{y}_i \right)^2$$

Source of Variation	df	Sum sq	MS
Treatment	$t-1$	$SSTrt = \sum_{i=1}^{t} n_i \left(\bar{y}_i - \bar{y} \right)^2$	$MSTrt = SSTrt / (t-1)$
Error	$n-t$	$SSE = \sum_{i=1}^{t} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i \right)^2$	$MSE = SSE / (n-t)$
Total	$n-1$	$SSTotal = \sum_{i=1}^{t} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y} \right)^2$	$SSTotal / (n-1)$

$$\cancel{\text{X}} \quad MSE = \frac{SSE}{n-t} = \frac{(n_1 - 1)S_1^2 + \cdots + (n_t - 1)S_t^2}{n-t} \quad \leftarrow \text{ pooled estimated variance.}$$

$H_0: M_1 = M_2 = \cdots = M_t$ $H_1:$ otherwise

$$F = \frac{MSTrt}{MSE} = \frac{SSTrt / (t-1)}{SSE / (n-t)} \qquad \overset{H_0}{\sim} \quad F_{t-1, \, n-t}$$

F_{obs} is the observed F stat we compute from sample.
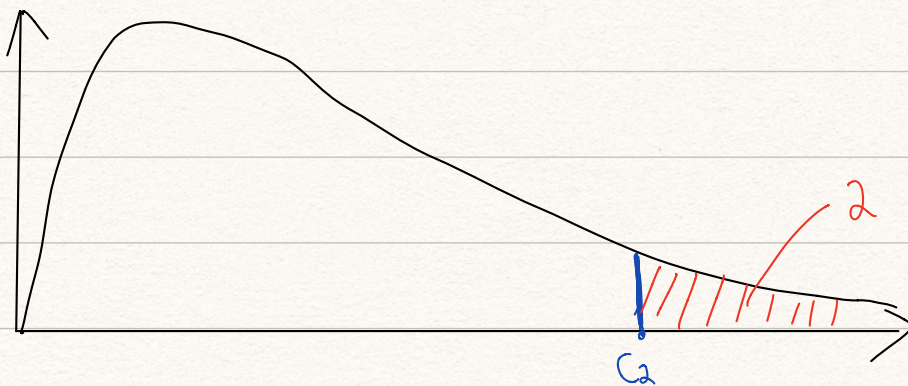
Significance level : α

Suppose F is a random variable with distribution $F_{t-1, \, n-t}$

① reject region

$$P(F > C_\alpha) = \alpha$$

↓
want to find such C_α

$$C_\alpha = qf(1-\alpha, \; df1 = t-1, \; df2 = n-t)$$

reject H_0 if $F_{obs} > C_\alpha$

② p-value

$$p\text{-value} = P(F > F_{obs})$$

$$= pf(F_{obs}, df1 = t-1, df2 = n-t)$$

reject H_0 if p-value $< \alpha$

② off - the - shelf method:

group : $\overbrace{1 \cdots 1}^{n_1}$, $2 \cdots 2$ $\cdots$, $\overbrace{t \cdots t}^{n_t}$

y : $y_{11} \cdots y_{1n_1}$, $y_{21} \cdots y_{2n_2}$, $\cdots$, $y_{t1} \cdots y_{tn_t}$

- mod = aov(y ~ group)
- anova(mod)