overall F-test in LM → F-Test

noncentral parameter, $\sum_i n_i \tau_i / N$

$\frac{1}{k-1}\left[\|P_{\Omega\omega}\mu\|^2+(k-1)\sigma^2\right]$ ← ANOVA Table ← By (P4) in LNp.2-36 → $\sigma^2+\frac{\sum_i n_i(\tau_i-\bar{\tau})^2}{k-1}$

Source	Degrees of Freedom (df)	Sum of Squares ($RSS_\omega - RSS_\Omega = \|P_{\Omega\omega} y\|^2$)	Mean Squares	Expected MS
treatment (between)	$k-1$	$SSTr=\sum_{i=1}^k n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2$	$MSTr=SSTr/df$	$E_\Omega(MSTr)$
residual (within)	$N-k$	$SSE=\sum_{i=1}^k\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2$ (RSS_Ω)	$MSE=SSE/df$ ($=\hat{\sigma}^2$)	$E_\Omega(MSE)$
total	$N-1$	$\sum_{i=1}^k\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\cdot\cdot})^2$ (RSS_ω)		

$N-k$: μ_i's; $N-1$: $\sum n_i$, η. $MSTr \div MSE$ → F. By (P5) in LNp.2-36 $E_\Omega(MSE) = \sigma^2$.

noncentral parameter: $\|P_{\Omega\omega}(\mu)\|^2=\sum n_i(\mu_i-\bar{\mu})^2$, $\mu = \begin{pmatrix}\mu_1\\ \vdots\\ \mu_1\\ \mu_2\\ \vdots\\ \mu_2\\ \vdots\\ \mu_k\\ \vdots\\ \mu_k\end{pmatrix} \in \Omega$, $\mu - \frac{\mu^T \mathbf{1}}{\|\mathbf{1}\|^2}\mathbf{1}$, $\frac{\sum n_i\mu_i}{N}=\bar{\mu}$

The F statistic for the null hypothesis that there is no difference between the treatments, i.e.,

non-central parameter $=0$ under H_0: $\mu_1=\cdots=\mu_k$ ($\mu_i=\eta+\tau_i$)

$$H_0:\tau_1=\cdots=\tau_k, =\tau$$

$\Omega: y_{ij}=\eta+\tau_i+\varepsilon_{ij} \Rightarrow df_\Omega=N-k$

$\omega: y_{ij}=(\eta+\tau)+\varepsilon_{ij} \Rightarrow df_\omega=N-1$

is

Intuition: graphs in LNp.3-2~3

$$F=\frac{\sum_{i=1}^k n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^k\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2/(N-k)}=\frac{MSTr}{MSE},$$

$\leftarrow F=\frac{(RSS_\omega-RSS_\Omega)/df_\omega-df_\Omega}{RSS_\Omega/df_\Omega}$

It follows a noncentral chi-square dist. under $\Omega\backslash\omega$ (LM, LNp.4-7) & (LNp 2-35)

$RSS_\Omega \sim \sigma^2\chi^2_{df_\Omega}$ under Ω & ω

which has an F distribution with parameters $k-1$ and $N-k$. (under ω)

ANOVA for Pulp Experiment

Source	Degrees of Freedom (df)	Sum of Squares	Mean Squares	F
operator	3	1.34	0.447	4.20
residual	16	1.70	0.106	
total	19	3.04		

qualitative factor

0.106: $\hat{\sigma}^2$; 19: 20 obs

- $Prob(F_{3,16}>4.20)=0.02=$ p-value, thus declaring a significant operator-to-operator difference at level 0.02.

  Note. ANOVA answers the question: "whether there exist difference between treatment means?"

- Further question: among the $6=\binom{4}{2}$ pairs of operators, what pairs show significant difference?

  Answer: Need to use multiple comparisons. → answer "how different?" question.

Identifiability (LM, LNp 5-10~11)

Constraint on the Parameters τ_i's

- The model in LNp.3-4 has k distinct levels, but k+1 regression parameters

  $\because$ ω, Ω not change under the over-parameterized model

  $\Rightarrow$ over-parameterized $\Rightarrow X^TX$ singular $\Rightarrow$ unidentifiable

  $\Rightarrow$ cannot estimate parameters (**Q**: but why can do overall F-test?)

  $\because \hat{\tau}_i$'s, $\hat{\eta}$ have infinite many solutions

- Some common constraint on τ_i's (Ω: $y_{ij} = \underline{\eta + \tau_i} + \varepsilon_{ij} = \underline{\mu_i} + \varepsilon_{ij}$)

  - $\sum_{i=1}^k \tau_i = 0 \Rightarrow$ dummy variables: sum coding ← LM, LNp.8-11~16

    $\tau_1+\tau_2+\tau_3+\tau_4=0 \Rightarrow \tau_4 = -(\tau_1+\tau_2+\tau_3) = \mu_4 - \bar{\mu}$

    $$\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{bmatrix} = \eta\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \tau_1\begin{bmatrix}1\\0\\0\\0\end{bmatrix} + \tau_2\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \tau_3\begin{bmatrix}0\\0\\1\\0\end{bmatrix} + \tau_4\begin{bmatrix}0\\0\\0\\1\end{bmatrix} \cdots (*)$$

    $$\Rightarrow \eta\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \tau_1\begin{bmatrix}1\\0\\0\\-1\end{bmatrix} + \tau_2\begin{bmatrix}0\\1\\0\\-1\end{bmatrix} + \tau_3\begin{bmatrix}0\\0\\1\\-1\end{bmatrix}$$

    $\eta = \bar{\mu} = (\mu_1 + \cdots + \mu_4)/4$, $\tau_1 = \mu_1 - \bar{\mu}$, $\tau_2 = \mu_2 - \bar{\mu}$, $\tau_3 = \mu_3 - \bar{\mu}$ ← Interpretation of τ_i's, η

    When $n_1 = \cdots = n_k$ the sum codings are orthogonal to $\mathbb{1}$.

  - $\tau_1 = 0 \Rightarrow$ dummy variables: treatment coding

    $$(*) = \eta\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \tau_2\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \tau_3\begin{bmatrix}0\\0\\1\\0\end{bmatrix} + \tau_4\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

    $\eta = \mu_1$, $\tau_2 = \mu_2 - \mu_1$, $\tau_3 = \mu_3 - \mu_1$, $\tau_4 = \mu_4 - \mu_1$

  - $\eta = 0$

    $$(*) = \tau_1\begin{bmatrix}1\\0\\0\\0\end{bmatrix} + \tau_2\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \tau_3\begin{bmatrix}0\\0\\1\\0\end{bmatrix} + \tau_4\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

    $\tau_i = \mu_i$

Recall. ANOVA decomposition in LNp.3-5 use the concept of sum codings, $\because$

$\Omega \ominus \omega = \text{span}\{\text{sum codings}\}$

ANOVA $\Rightarrow H_0: \tau_1 = \cdots = \tau_k$

$H_0: \tau_1 = \cdots = \tau_k = 0$ ← after the first 2 constraints on τ_i's are added

❖ **Reading**: textbook, 2.1 ← check LNp.2-34

Multiple Comparisons

→ answer "how different" problem: $\mu_i \overset{?}{=} \mu_j$, $\forall i<j$

- Consider the full model $y_{ij} = \underbrace{\eta + \tau_i}_{\mu_i} + \varepsilon_{ij}$. For one pair, say (i, j), of treatments, test $H_0^{ij}: \tau_i = \tau_j$ against $H_A^{ij}: \tau_i \neq \tau_j$.

  $\Leftrightarrow \mu_i = \mu_j$ ($\mu_i - \mu_j = 0$)

  decided before exp't, not related to the observed y

Ω: $\; y = \eta\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} + \cdots + \tau_i\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ \vdots\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix} + \cdots + \tau_j\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ \vdots\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix} + \cdots + \varepsilon$ (ith, jth)

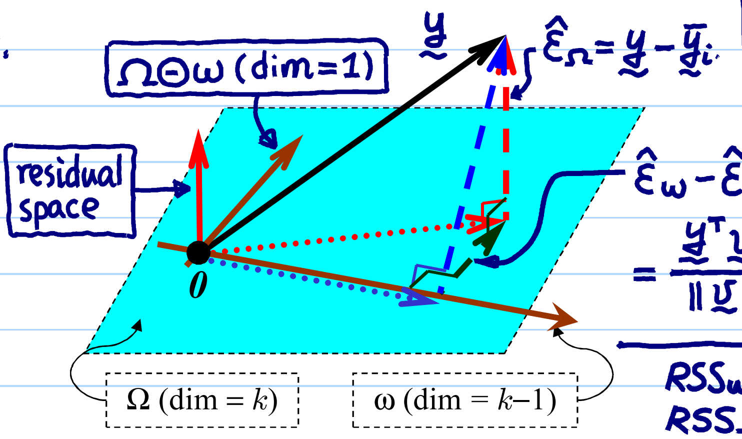
ω: $\; y = \eta\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} + \cdots + \tau_i\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ \vdots\\ 1\\ 0\\ \vdots\\ 0\\ 1\\ \vdots\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix} + \cdots + \cancel{\otimes} + \cdots + \varepsilon$ (ith, jth)

$F_{ij} = \dfrac{RSS_\omega - RSS_\Omega / 1}{RSS_\Omega / df_\Omega} \sim F_{1, df_\Omega}$ (under ω)

$\Omega \ominus \omega = \text{span}\{v\}$, where

$$v = \begin{bmatrix}0\\ \vdots\\ 0\\ +1/n_i\\ \vdots\\ +1/n_i\\ 0\\ \vdots\\ 0\\ -1/n_j\\ \vdots\\ -1/n_j\\ 0\\ \vdots\\ 0\end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow i\text{th} \\ \\ \\ \\ \\ \\ \leftarrow j\text{th} \\ \\ \\ \\ \\ \end{matrix}$$

How can we derive it?



$$\hat{\varepsilon}_\omega - \hat{\varepsilon}_\Omega = P_{\Omega\ominus\omega}\, y = \frac{y^T v}{\|v\|^2} v = \frac{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}{\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \cdot \frac{v}{\|v\|}$$

$RSS_\omega - RSS_\Omega = \|\hat{\varepsilon}_\omega - \hat{\varepsilon}_\Omega\|^2$

$RSS_\Omega = \|\hat{\varepsilon}_\Omega\|^2$

- It is common to use the t-test and the t-statistic

$t_{ij}^2 = F_{ij}$

$$t_{ij} = \frac{\boxed{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}} - 0}{\hat{\sigma}\boxed{\sqrt{1/n_i + 1/n_j}}},$$

μ_i μ_j $\bar{y}_{i\cdot}$ $\bar{y}_{j\cdot}$

$Var(\bar{y}_{i\cdot} - \bar{y}_{j\cdot})$
$= Var(\bar{y}_{i\cdot}) + Var(\bar{y}_{j\cdot})$
$= \sigma^2/n_i + \sigma^2/n_j$

where n_i = number of observations for treatment i,
$\hat{\sigma}^2 = RSS_\Omega/df_\Omega$ in ANOVA; declare "treatments i and j different at level α" if

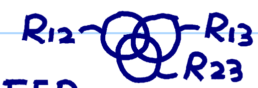
$$|t_{ij}| > t_{N-k,\frac{\alpha}{2}}.$$

$F_{1,df_\Omega} \to$ null dist.

df_Ω

null of ANOVA in LNp.3-6

multiple testing



$R_{ij} \equiv$ rejection region of H_0^{ij}
$P(R_{ij} \mid \mu_i = \mu_j \, (\tau_i = \tau_j)) = \alpha$

EER $= P(\bigcup R_{ij} \mid \mu_1 = \cdots = \mu_k)$
$> \alpha$ ← why not good?
usually

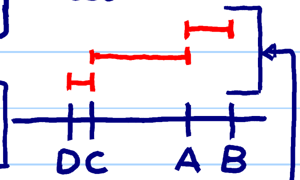
- Suppose k' tests are performed to test $H_0 : \tau_1 = \cdots = \tau_k$.

e.g. $\binom{k}{2}$

Experimentwise error rate (EER) = Probability of declaring at least one pair of treatments significantly different under H_0. Need to use multiple comparisons to control EER.

$\bigcap_{i,j} H_0^{ij}$ ← union-intersection test

H_1: $\bigcup_{i,j} H_1^{ij}$



Note. deductive logic does not hold (Why?)

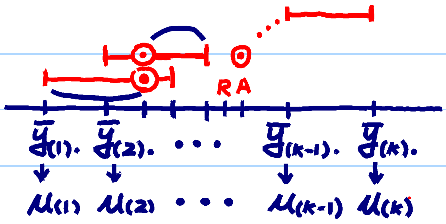
$t_{16,\frac{0.05}{2}} = 2.12$

t_{ij}

A vs. B	A vs. C	A vs. D	B vs. C	B vs. D	C vs. D
−0.87	1.85	2.14	2.72	3.01	0.29

$|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| > \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \cdot t_{N-k,\frac{\alpha}{2}}$

a scale



$\mu_{(1)}, \cdots, \mu_{(k)}$: sorted $\mu_1, \cdots, \mu_k$
$(\mu_{(1)} \le \mu_{(2)} \le \cdots \le \mu_{(k)})$

$\bar{y}_{(1)\cdot}, \cdots, \bar{y}_{(k)\cdot}$: order statistics of $\bar{y}_{1\cdot}, \cdots, \bar{y}_{k\cdot}$.

Bonferroni Method

- Declare "τ_i different from τ_j at level α" if $|t_{ij}| > t_{N-k,\frac{\alpha}{2k'}}$, where k' = number of tests.

cf. $\frac{\alpha}{2}$

EER $\le \alpha$

very conservative when k' is large

- For one-way layout with k treatments, $k' = \binom{k}{2} = \frac{1}{2}k(k-1)$, as k increases, k' increases, and the critical value $t_{N-k,\frac{\alpha}{2k'}}$ gets bigger (i.e., method less powerful in detecting differences).

some indep. assumptions btwn the test stat. of these tests, e.g., indep. $\bar{y}_i$'s

- Advantage: It works without requiring independence assumption.

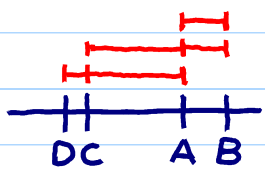
can be widely applied

- For pulp experiment, take $\alpha = 0.05$, $k = 4$, $k' = 6$, $t_{16,0.05/12} = 3.008$. Among the 6 t_{ij}-values (see LNp.3-10), only the t-value for B-vs-D, 3.01, is larger. Declare "B and D different at level 0.05".

2×6

$$P(\bigcup R_{ij} \mid \bigcap H_0^{ij}) \le \sum_{i,j} P(R_{ij} \mid H_0^{ij}) = k' \cdot \alpha' = \alpha$$

$$\Rightarrow \alpha' = P(R_{ij} \mid H_0^{ij}) = \alpha/k'$$



For simplicity, assume $n_1=\cdots=n_k=n$

Tukey Method

Same procedure can be applied to unequal sample size case → Tukey-Kramer test

- Declare "τ_i different from τ_j at level α" if

$$|t_{ij}| > \frac{1}{\sqrt{2}} q_{k,N-k,\alpha},$$

EER ≤ α

∵ $\sqrt{1/n+1/n} = \sqrt{2}/\sqrt{n}$

LNp.3-10

where $q_{k,N-k,\alpha}$ is the upper α-quantile of the **Studentized range** (**SR**) distribution with parameter k and $N-k$ degrees of freedom. (see distribution table on LNp.3-13)

- For pulp experiment,

LNp.3-13

$$\frac{1}{\sqrt{2}} q_{k,N-k,0.05} = \frac{1}{\sqrt{2}} q_{4,16,0.05} = \frac{4.05}{\sqrt{2}} = 2.86.$$

Again only B-vs-D has larger t_{ij}-value than 2.86 (See LNp.3-10). Tukey method is more powerful than Bonferroni method because 2.86 is smaller than 3.01 (why?)

- compare $\bar{y}_{i\cdot}-\bar{y}_{j\cdot}$, $\forall (i,j)$ ← check LNp.3-10
- For $1 \le i < j \le k$, $|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}| \le \bar{y}_{(k)\cdot}-\bar{y}_{(1)\cdot}$.
- $P(\cup R_{ij} \mid \cap H_0^{ij})$ ← EER

$= P($at least one $|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}|$ larger than $C^* \mid \cap H_0^{ij})$

$= P(\bar{y}_{(k)\cdot}-\bar{y}_{(1)\cdot} > C^* \mid \cap H_0^{ij})$

$= \alpha$

- Under $\cap H_0^{ij}$, $\bar{y}_{1\cdot},\ldots,\bar{y}_{k\cdot} \overset{\text{indep.}}{\sim} N(\mu, \sigma^2/n) \Rightarrow$

$$\frac{\sqrt{n}(\bar{y}_{(k)\cdot}-\bar{y}_{(1)\cdot})}{\hat{\sigma}} \sim SR_{k,\nu}$$

$\hat{\sigma}$: $\sqrt{RSS_\Omega/df_\Omega}$ (LNp.3-9)

where $\nu\hat{\sigma}^2 \sim \sigma^2\chi^2_\nu$ and indep. of $\bar{y}_{1\cdot},\ldots,\bar{y}_{k\cdot}$.

$\Rightarrow C^* = \hat{\sigma}/\sqrt{n} \cdot SR_{k,\nu,\alpha}$

$\Rightarrow \sqrt{n}|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}|/\hat{\sigma} > SR_{k,\nu,\alpha}$

Q: For the case in LNp.3-10, after getting $\bar{y}_{i\cdot}$'s, want to test $H_0^{ij}: \mu_D=\mu_B$ (only one test). Which critical value should we use? $H_0: \mu_{(1)}=\mu_{(k)}$?

Q: Among the 3 critical values of $|t_{ij}|$, which one is the smallest? the largest?

Selected values of $q_{k,\nu,\alpha}$ for $\alpha = 0.05$

ν: N-k

increasing (why?)

ν \ k	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59	51.96	53.20	54.33	55.36
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.15	10.35	10.52
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65
16	3.00	3.65	$\boxed{4.05}$	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59

α=upper tail probability, ν=degrees of freedom, k=number of treatments

decreasing (Why?)

For complete tables corresponding to various values of α refer to Appendix E.

❖ **Reading**: textbook, 2.2

One-Way ANOVA with a Quantitative Factor

cf. Qualitative

★ response : bonding strength

★ (treament) factor : power (quantitative)

3 levels – 40, 50, 60 (equally spaced)

★ Exp'tal units : one composite

9 EUs

★ Each treatment repeats 3 times (3 replicates)

- **Data** :

Design matrix

power	strength
40	25.66
40	28.00
40	20.65
50	⋮
50	⋮
50	⋮
60	
60	
60	35.66

y = bonding strength of composite material,

x = laser power at 40, 50, 60 watt.

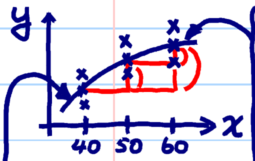
homogeneous → EU1 · · · EU9

$\frac{9!}{3!3!3!}$ ← completely randomized → r 40, 50, 60 r

Table 2: Strength Data, Composite Experiment

Laser Power (watts)		
40	50	60
25.66	29.15	35.73
28.00	35.09	39.56
20.65	29.79	35.66

Initial Data Analysis : scatter plot



the line is meaningful only when the factor is quantitative

major difference between quantitative & qualitative factors

conceptual model

$\mu_x = E(y_x)$

$y_x = \mu_x + \varepsilon$

$\mu_x = \beta_0 + \beta_1 x$

★ $\mu_x = \beta_0 + \beta_1 x + \beta_2 x^2$

$\mu_x = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

$\mu_x = \beta_0 + \beta_1 x \log x + \beta_2 e^x$

⋮