small,scriptsize,tiny,p0.5linewidth

Simple linear

$S_{xx} \sum_{i=1}^{10}(x_i - \bar{x})^2 = 472, S_{yy} = \sum_{i=1}^{10}(y_i - \bar{y})^2 = 731, S_{xy} \sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = 274$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{274}{472} = 0.5805; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \bar{y} - 0.5805\bar{x}$

$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{1}{8}(S_{yy} - \frac{S_{xy}^2}{S_{xx}}) = \frac{1}{8}(731 - \frac{274^2}{472}) = 71.4926$

$\hat{\beta}_1 \sim N(\beta_1, \frac{\hat{\sigma}^2}{S_{xx}}); \hat{\beta}_0 \sim N(\beta_0, \hat{\sigma}^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}]);$

$Var(\hat{\beta}_1) = \frac{71.4926}{472} = 0.1515; Var(\hat{\beta}_1) = 71.4926(\frac{1}{10} + \frac{\bar{x}^2}{472})$

$H_0 : \hat{\beta}_1 = 0; t_0 = \frac{\hat{\beta}_1 - 0}{\sqrt{Var(\hat{\beta}_1)}} = \frac{0.5805085}{\sqrt{0.1514673}} = 1.491589 < t_{0.05,n-2} = 2.31$ Fail to reject

$\hat{\beta}_1 \pm t_{\frac{0.05}{k+1},n-2}se(\hat{\beta}_1) = 0.5805085 \pm 2.31\sqrt{0.1514673}, (-0.3185158, 1.479533)$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_0 = 168.3609; \bar{x} = 147.6923, S_{XX} = \sum_{i=1}^{13}(x_i - \bar{x})^2 = 6230.769$

Confidence interval: $se(y_0) = \sqrt{MSE(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}})} = $

$\sqrt{275.06005(\frac{1}{13} + \frac{(170-147.6923)^2}{6230.769})} = 6.567093$

$\hat{y} \pm t_{n-2,0.025}se(y_0) = 168.3609 \pm 2.200985 * 6.567093, (153.9068, 182.815)$

Prediction interval: $se(y_0) = \sqrt{MSE(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}})} = 17.83779; 168.3609 \pm 2.200985 * 17.83779; (129.1002, 207.6216)$

Multi linear

• least-squares estimators

$SSE = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$

$\frac{\partial SSE}{\partial \beta_{0,1,2}} = 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)(-1, -x_i, -x_i^2) \overset{set}{=} 0$

• Regression with indicator

$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + \beta_3 w_i x_i + \varepsilon_i, w_i = \begin{cases} 0 & 1 \le i \le \frac{n}{2} \\ 1 & \frac{n}{2}+1 \le i \le n \end{cases}$

$\mathbf{Y_{n \times 1}} = \mathbf{X_{n \times (k+1)}} \mathbf{fi_{(k+1) \times 1}} + \mathbf{"_{n \times 1}}$

equality of slopes $H_0 : \beta_3 = 0, k = 3, r = 1$

$dfE_{Ful} = n - (k+1), dfE_{Red} - dfE_{Ful} = r$

• Partitioned regression

(df)	SS_F	$-SS_{3,4,5}$	$SS_{1,2}$	$-SS_{2,4}$	$SS_{1,3,5}$
R	$\sum_{i=1}^{5} = 367(5)$	$-\sum_{i=3}^{5}$	$\sum_{i=1}^{2} = 271(2)$	-153(2)	214(3)
E	T-R=336(69)	+96(3)	432(72)	+153(2)	489(71)
T	703(74)				

$H_0 : \beta_3 = \beta_4 = \beta_5 = 0; r = 3; H_0' : \beta_2 = \beta_4 = 0; r = 2;$

$F_{p,3,69} = \frac{96/3}{336/69} = 6.5714; F' = \frac{153/2}{336/69} = 15.7098$ $p < 0.05$ reject H_0 at 0.05 level of significance

$R^2 = \frac{SSR}{SST} = \frac{271}{703} = 0.3855$ coefficient of determination is the proportion of variation explained by the regressor x

$R_{adj}^2 = 1 - \frac{SSE/dfE}{SST/dfT} = 1 - \frac{432/72}{703/74} = 0.3684$

CRD

$y_{ijkl} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, i = 1,2; j = 1,2,3; k = 1,2,3,4; l = 1,2, [a = 2, b = 3, n = 4]$ where μ overall mean

τ_i is fixed main effect of ith level of Factor A;

β_j is fixed main effect of jth level of Factor B;

$(\tau\beta)_{ij}$ is fixed interaction effect of ith level of Factor A and j^{th} level of Factor B;

ε_{ijk} is random error for the kth replicate EU when i^{th} level of Factor A and j^{th} level of Factor B are applied; y_{ijkl} is response for the;

Assumptions: $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$ (constant variance, zero mean, independent);

$\sum_i^2 \tau_i = 0; \sum_j^3 \beta_j = 0; \sum_i^2 (\tau\beta)_{ij} = 0; \sum_j^3 (\tau\beta)_{ij} = 0$

term	i(f)	j(f)	k(r)	df	SS	EMS
$A\tau_i$f	0	b	n	a-1	$bn\sum^a(\bar{y}_{i..} - \bar{y}_{...})^2; \frac{\sum^a y_{i..}^2}{bn} - \frac{y_{...}^2}{abn};96;96$	$\sigma^2 + \frac{b\sum\tau_i^2}{a-1}$
$B\beta_j$f	a	0	n	b-1	$an\sum^b(\bar{y}_{.j.} - \bar{y}_{...})^2; \frac{\sum^b y_{.j.}^2}{an} - \frac{y_{...}^2}{abn};208;104$	$\sigma^2 + \frac{a\sum\beta_j^2}{b-1}$
AB $(\tau\beta)_{ij}$f	0	0	n	(a-1)(b-1)	$n\sum^a\sum^b(y_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2;112;56$ $n\sum\sum y_{ij.}^2 - \frac{1}{abn}y_{...}^2 - SS_A - SS_B$	$\sigma^2 + \frac{\sum\sum(\tau\beta)_{ij}}{(a-1)(b-1)}$
$E\varepsilon_{ijk}$r	1	1	1	ab(n-1)	$SST - \sum SS;(n-1)\sum^a\sum^b S_{ij}^2;126;7$	σ^2
Total				abn-1	$\sum\sum\sum(y_{ijk} - \bar{y}_{...})^2 \sum\sum y_{ijk}^2 - \frac{y_{...}^2}{abn};542$	

$H_0 : (\tau\beta)_{ij} = 0 \forall i, j; F_{p,2,18} \frac{MS_{AB}}{MSE} = \frac{56}{7} = 8; F_{0.05,2,8} = 3.55.$ There is enough evidence to reject H_0. The model may not be reduced, as the interaction effects is significant at 5% significance level.

$\bar{y}_{12.} - \bar{y}_{22.} \pm t_{\frac{\alpha}{2},18}\sqrt{\frac{2MSE}{n}} = 14 - 8 \pm 2.1\sqrt{\frac{2*7}{4}} = 6 \pm 3.9287; [2.0713, 9.9287]$

Latin Square

$y_{ijk} = \mu + \tau_i + \alpha_j + \beta_k + \varepsilon_{ijk}, i, j, k = 1, .., 6;$ where μ overall mean

τ_i is effect of ith treatment; α_j is effect of jth block of factor R; β_k effect of kth block of factor C;

ε_{ijkl} is random error when i^{th} treatment is applied at j^{th} block of factor R and k^{th} block of factor C; y_{ijkl} is response ;

Assumptions: $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$. Further assumptions would be based on whether the treatment and blocking factors are random or fixed.

The Latin-Squre design can only use 2 blocking factors, as we distribute the levels of the treatment factor on a table with rows of one blocking factor (each row is fore one block) and columns of the other blocking factor (each column is one block)

You can test three blocking factors by turning the Latin-square design into a Graeco-Latin square design, which allows to add a Greek letter to each entry in the table, where each Greek letter stands for a block of the factor G.

Relative Efficiency $\frac{(df_{E(LS)}+1)(df_{E(CRD)}+3)}{(df_{E(LS)}+3)(df_{E(CRD)}+1)} = 2.3$

$df_{E(LS)} = (p-1)(p-2) = 20, (df_{E(GS)} = (p-1)(p-3) = 15, df_{E(CRD)} = a(n-1)$

BIBD

Trt$a$ = 7, Rep$r$ = 15, Blk$b$ = 21, Size$k$ = 5, $ar = bk = 105$, replications of each pair$\lambda = \frac{(k-1)}{a-1}r = \frac{k(k-1)}{a(a-1)}b = 10$

$\lambda = \frac{2}{3}r \in \mathbf{N}^+$ the smallest number of observations per treatment is a multiple of 3 and 5

Factorial model

k factors, p generators; 2^p blocks/fraction; 2^{k-p}Run, Blk size;

$2^p - 1$ alias; $2^p - p - 1$ auto confounded;

I=ABC=BCD=AD

AD+,ABC-,BCD-;(1),bc,abd,acd;ABC+BCD+;b,c,ad,abcd

AD-,ABC-,BCD+;ab,ac,bcd,d;ABC+BCD+;a,abc,bd,cd

ANOVA-df: T= $2^k n - 1$;Blk= $2^p - 1$;E= $2^k(n-1)$;others=1

2^{k-p}	2^k	2^p	2^{k-p}	Generator		
2_{III}^{3-1}	8	2	4	C=AB		
2_{IV}^{4-1}	16	2	8	D=ABC		
2_V^{5-1}	32	2	16	E=ABCD		
2_{III}^{5-2}	32	4	8	D=AB;E=AC		
2_{VI}^{6-1}	64	2	32	F=ABCDE		
2_{IV}^{6-2}	64	4	16	E=ABC;F=BCD	A-F;AB;AC;AD;**AE**;AF;BD;BF;**ABD**	
2_{III}^{6-3}	64	8	8	D=AB;E=AC;F=BC		
2_{VII}^{7-1}	128	2	64	G=ABCDEF		
2_{IV}^{7-2}	128	4	32	F=ABCD;G=ABDE	I=ABCDF=ABDEG=CEFG;CE=FG,CF=EG	
2_{IV}^{7-3}	128	8	16	E=ABC;F=BCD;G=ACD	$2^p - p - 1 = 4$ generalized interactions	
2_{III}^{7-4}	128	16	8	D=AB;E=AC;F=BC;G=ABC		

Fixed v.s. random; crossed v.s. nested

(a)Fixed effects are constant across individuals, random effects vary. Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small part of the population the corresponding variable is random. If an effect is assumed to be a realized value of a random variable, it is called a random effect. Fixed effects are estimated using least squares and random effects are estimated with shrinkage.

Example: A lab want to test the quality of products. There are only two test machines. Each machine assigned two operators who randomly sellected from a large amont of operators. Each poerator test the products in three specific temperature. In this test, the factor mathine and temperature have fixed effects. The factor operator has random effects.

(b)Two factors are crossed when every category of one factor co-occurs in the design with every category of the other factor. There is at least one observation in every combination of categories for the two factors.

A factor is nested within another factor when each category of the first factor co-occurs with only one category of the other. An observation has to be within one category of Factor 2 in order to have a specific category of Factor 1. All combinations of categories are not represented.

You can calculate an interaction between two crossed factors. If they are nested, you cannot because you do not have every combination of one factor along with every combination of the other.

In **Example** (a), Three temperature are applied on each machine. Thus, temperature and machine have crossed effect. The operators were assigned to each machine are different. The operators are nested in the levels of factor machine. Thus, the effects of the factor operator are nested effect.

Mixed model

$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}; \sum_{i=1}^a \alpha_i = 0, \beta_{ij} \sim N(0, \sigma_\beta^2), \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

$Cov(y_{111}, y_{112}) = Cov(\beta_{11} + \varepsilon_{111}, \beta_{11} + \varepsilon_{112}) = Var(\beta_{11}) + Cov(\varepsilon_{111}, \varepsilon_{112}) = \sigma_\beta^2$

$Cov(y_{111}, y_{121}) = Cov(\beta_{11} + \varepsilon_{111}, \beta_{12} + \varepsilon_{121}) = Cov(\beta_{11}, \beta_{12}) + Cov(\varepsilon_{111}, \varepsilon_{121}) = 0$

$Cov(y_{111}, y_{211}) = Cov(\beta_{11} + \varepsilon_{111}, \beta_{21} + \varepsilon_{211}) = Cov(\beta_{11}, \beta_{21}) + Cov(\varepsilon_{111}, \varepsilon_{211}) = 0$

$Var(y_{111}) = \sigma_\beta^2 + \sigma_\varepsilon^2 = Var(y_{112}); Cor(y_{111}, y_{112}) = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_\varepsilon^2} Cor(y_{111}, y_{121}) = Cor(y_{111}, y_{211}) = 0$

A linear combination of normal distributed random variables and constants are normal distributed.

$E[\bar{y}_{ij.}] = E[\frac{1}{n}\sum_{k=1}^n y_{ijk}] = \frac{\sum_{k=1}^n}{n}E[\mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}] = \mu + \alpha_i, \forall i = 1, .., a; j = 1, .., b$

$Var[\bar{y}_{ij.}] = Var[\frac{1}{n}\sum_{k=1}^n y_{ijk}] = \frac{\sum_{k=1}^n}{n^2}Var[\mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}] = \frac{1}{n}(\sigma_\beta^2 + \sigma_\varepsilon^2), \forall i = 1, .., a; j = 1, .., b$

$f(\bar{y}_{i1.}, ..\bar{y}_{ib.}) = \prod_{j=1}^b f(\bar{y}_{ij.}) = (2\pi\frac{\sigma_\beta^2 + \sigma_\varepsilon^2}{n})^{-\frac{b}{2}}\exp[\frac{-n}{2(\sigma_\beta^2 + \sigma_\varepsilon^2)}\sum_{j=1}^b(\bar{y}_{ij.} - \mu - \alpha_i)^2]$

$\hat{\alpha}_1 - \hat{\alpha}_2 \sim N(\alpha_1 - \alpha_2, \frac{2}{b}\sigma_\beta^2 + \frac{2}{bn}\sigma_\varepsilon^2)$

$\frac{\partial SSE}{\partial \alpha_i} = 2\sum_{j=1}^b\sum_{k=1}^n(y_{ijk} - \mu - \alpha_i - \beta_{ij})(-1) \overset{set}{=} 0; \hat{\alpha}_i = \frac{1}{bn}\sum_{j=1}^b\sum_{k=1}^n y_{ijk} - \mu - \frac{1}{b}\sum_{j=1}^b \beta_{ij} = \bar{y}_{i..} - \mu$

$\hat{\alpha}_1 - \hat{\alpha}_2 = \bar{y}_{1..} - \mu - (\bar{y}_{2..} - \mu) = \frac{\sum_{j=1}^b\sum_{k=1}^n(y_{1jk} - y_{2jk})}{bn} = \frac{\sum_{j=1}^b\sum_{k=1}^n(\alpha_1 - \alpha_2 + \beta_{1j} - \beta_{2j} + \varepsilon_{1jk} - \varepsilon_{2jk})}{bn} = \alpha_1 - \alpha_2 + \bar{\beta}_{1.} - \bar{\beta}_{2.} + \bar{\varepsilon}_{1..} - \bar{\varepsilon}_{2..}$

$Var[\hat{\alpha}_1 - \hat{\alpha}_2] = Var[\bar{\beta}_{1.} - \bar{\beta}_{2.} + \bar{\varepsilon}_{1..} - \bar{\varepsilon}_{2..}] = \frac{\sum_{j=1}^b}{b^2}(Var[\beta_{1.}] + Var[\beta_{2.}]) + \frac{\sum_{j=1}^b\sum_{k=1}^n}{b^2 n^2}(Var[\varepsilon_{1..}] + Var[\varepsilon_{2..}]) = \frac{2}{b}\sigma_\beta^2 + \frac{2}{bn}\sigma_\varepsilon^2$

$E[\hat{\alpha}_1 - \hat{\alpha}_2] = \alpha_1 - \alpha_2; \quad SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \mu - \alpha_i - \beta_{ij})^2$

$\bar{y}_{ij.} - \bar{y}_{i..} = \mu + \alpha_i + \beta_{ij} + \bar{\varepsilon}_{ij.} - (\mu + \alpha_i + \bar{\beta}_{i.} + \bar{\varepsilon}_{i..}) = \beta_{ij} - \bar{\beta}_{i.} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}$

$E[\bar{y}_{ij.} - \bar{y}_{i..}] = E[\beta_{ij} - \bar{\beta}_{i.} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}] = 0$

$Cov(\beta_{ij}, \bar{\beta}_{i.}) = \frac{1}{b}Cov(\beta_{ij}, \sum_{j=1}^{b}\beta_{ij}) = \frac{1}{b}[1 \cdot \sigma_\beta^2 + (b-1)\cdot 0]$

$Cov(\bar{\varepsilon}_{ij.}, \bar{\varepsilon}_{i..}) = Cov(\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{ijk}, \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n}\varepsilon_{ijk}) = \frac{1}{b}\frac{\sum_{k=1}^{n}}{n^2}Cov(\varepsilon_{ijk}, \sum_{j=1}^{b}\varepsilon_{ijk}) = \frac{1}{bn}\sigma_\varepsilon^2$

$Var[\bar{y}_{ij.} - \bar{y}_{i..}] = Var[\beta_{ij} - \bar{\beta}_{i.} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}] = Var[\beta_{ij} - \bar{\beta}_{i.}] + Var[\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}] = Var[\beta_{ij}] + Var[\bar{\beta}_{i.}] - 2Cov(\beta_{ij}, \bar{\beta}_{i.}) + Var[\bar{\varepsilon}_{ij.}] + Var[\bar{\varepsilon}_{i..}] - 2Cov(\bar{\varepsilon}_{ij.}, \bar{\varepsilon}_{i..}) = \sigma_\beta^2 + \frac{1}{b}\sigma_\beta^2 - \frac{2}{b}\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2 + \frac{1}{bn}\sigma_\varepsilon^2 - \frac{2}{bn}\sigma_\varepsilon^2 = \frac{b-1}{b}(\sigma_\varepsilon^2 + \frac{1}{n}\sigma_\varepsilon^2)$

$E[\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..})^2] = \sum_{i=1}^{a}\sum_{j=1}^{b}(Var[\bar{y}_{ij.} - \bar{y}_{i..}] + E[\bar{y}_{ij.} - \bar{y}_{i..}]^2)$

$\sum_{i=1}^{a}\sum_{j=1}^{b}[\frac{b-1}{b}(\sigma_\varepsilon^2 + \frac{1}{n}\sigma_\varepsilon^2) + 0] = a(b-1)(\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2)$

term	i(f)	j(r)	k(r)	df	EMS	F
α_i f	0	b	n	a-1	$\frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i + n\sigma_\beta^2 + \sigma_\varepsilon^2$	$\frac{MS_A}{MS_{AB}}$
β_{ij} r	0	1	n	a(b-1)	$n\sigma_\beta^2 + \sigma_\varepsilon^2$	$\frac{MS_{AB}}{MS_E}$
ε_{ijk} r	1	1	1	ab(n-1)	σ_ε^2	
Total				abn-1		

$\hat{\sigma}^2 = \frac{MS_{AB} - MS_E}{n}; \quad E[\hat{\sigma}^2] = \frac{1}{n}(n\sigma_\beta^2 + \sigma_\varepsilon^2 - \sigma_\varepsilon^2) = \sigma_\beta^2$

Nested model

term	i(f)	j(r)	k(f)	l(r)	df	EMS	F
$A\tau_i$ f	0	b	c	n	a-1	$\frac{bcn\sum_{i=1}^{a}\tau_i^2}{a-1} + cn\sigma_\beta^2 + \sigma^2$	$\frac{MS_A}{MS_{B(A)}}$
$B(A)\beta_{j(i)}$ r	1	1	c	n	a(b-1)	$cn\sigma_\beta^2 + \sigma^2$	$\frac{MS_{B(A)}}{MS_E}$
$C(\gamma)_k$ f	a	b	0	n	c-1	$\frac{abn\sum_{k=1}^{c}\gamma_k^2}{c-1} + bn\sigma_\gamma^2 + n\sigma_{\gamma\beta}^2 + \sigma^2$	$\frac{MS_C}{MS_{CB(A)}}$
$AC(\tau\gamma)_{ik}$ f	0	b	0	n	(a-1)(c-1)	$\frac{bn\sum_{i=1}^{a}\sum_{k=1}^{c}(\tau\gamma)_{ik}^2}{(a-1)(c-1)} + n\sigma_{\gamma\beta}^2 + \sigma^2$	$\frac{MS_{AC}}{MS_{CB(A)}}$
$CB(A)(\gamma\beta)_{kj(i)}$ r	1	1	0	n	a(b-1)(c-1)	$n\sigma_{\gamma\beta}^2 + \sigma^2$	$\frac{MS_{CB(A)}}{MS_E}$
$E\varepsilon_{(ijk)l}$	1	1	1	1	abc(n-1)	σ^2	
T					abcn-1		