Chapter 6: Tests of Statistical Hypotheses

1. Concepts of Hypothesis Testing

H_0: Null Hypothesis, H_1: Alternative Hypothesis

(1) Errors (Type I, Type II)

$$\begin{cases} \text{Type I error: reject a true } H_0 \\ \text{significance level: } \alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true}) \end{cases}$$

$$\begin{cases} \text{Type II error: Accept a false } H_0 \\ \beta = P(\text{Accept } H_0 \mid H_0 \text{ is false}) \\ \text{power} = 1-\beta \end{cases}$$

(2) $$\begin{cases} \text{Operating Characteristic Curve} \\ OC(\mu) = P(\text{accept } H_0 \mid \mu) \end{cases}$$

H_0: $\mu = \mu_0$ $\quad$ H_1: $\mu \neq \mu_0$ $\quad$ $(\mu_1 \neq \mu_0)$

$$\left.\begin{aligned} OC(\mu_0) &= P(\text{accept } H_0 \mid \mu=\mu_0) = 1-\alpha \\ OC(\mu_1) &= P(\text{accept } H_0 \mid \mu=\mu_1) = \beta \end{aligned}\right\} \Rightarrow \text{Determine sample size}$$

~~(3)~~

$H_0: \mu = \mu_0 \qquad H_1: \mu > \mu_0 \qquad (\mu_1 > \mu_0)$

Rejection region : $P\left\{ \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right\} = 1 - \alpha$

$$\beta = P\{\text{accept } H_0 \mid \mu = \mu_1\}$$

$$\beta = P\left\{ \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \le z_\alpha \;\middle|\; \mu = \mu_1 \right\}$$

$$\beta = P\left\{ \bar{x} \le z_\alpha \cdot \frac{\sigma}{\sqrt{n}} + \mu_0 \right\}$$

$$\beta = P\left\{ \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} \le \frac{z_\alpha \cdot \frac{\sigma}{\sqrt{n}} + \mu_0 - \mu_1}{\sigma/\sqrt{n}} \right\}$$

$$z_{1-\beta} \cdot \frac{\sigma}{\sqrt{n}} = z_\alpha \cdot \frac{\sigma}{\sqrt{n}} + \mu_0 - \mu_1$$

$$\mu_1 - \mu_0 = (z_\alpha - z_{1-\beta}) \frac{\sigma}{\sqrt{n}}$$

$$n = \left[\frac{(|z_\alpha| + |z_\beta|) \cdot \sigma}{\mu_1 - \mu_0} \right]^2 \qquad \begin{cases} z_{1-\beta} = |z_\beta| \\ z_\alpha = |z_\alpha| \end{cases} \text{ for small } \alpha \text{ and } \beta$$

5

- Test
  - paired t-test
  - χ^2-test
  - variance test
- ANOVA
  - linear Regression.
- sample size : OC (curve).

2. Tests of Population Mean and Population Proportion

(1) Z-test for population mean if σ^2 is known, $\begin{cases} \text{normal population} \\ n \geq 30 \end{cases}$

Statistic: $$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad H_0: \mu = \mu_0$$

Rejection region $\{ z_0 > Z_\alpha \}$ $\qquad H_1: \mu > \mu_1$

$\{ |z_0| > Z_{\frac{\alpha}{2}} \}$ $\qquad H_1: \mu \neq \mu_1$

$\{ z_0 < -Z_\alpha \}$ $\qquad H_1: \mu < \mu_1$

P-value: $P\{ Z > z_0 \}$ $\qquad H_1: \mu > \mu_1$

$2 \times P\{ Z > |z_0| \}$ $\qquad H_1: \mu \neq \mu_1$

$P\{ Z < z_0 \}$ $\qquad H_1: \mu < \mu_1$

(2) T-test if σ^2 is unknown, normal population

Statistic $$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \qquad H_0: \mu = \mu_0$$

Rejection region: $\{ t_0 > t(\alpha, n-1) \}$ $\qquad H_1: \mu > \mu_0$

P-value: $P\{ t(n-1) > t_0 \}$

(3) Z-test for population proportion $H_0: p = p_0$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1) \quad \text{under } H_0$$

3. Tests of Two Distributions

(1) Independent Samples : $H_0: \mu_1 = \mu_2$

a) Assume σ_1^2, σ_2^2 are known, normal or $n_1, n_2 \geq 20$

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad \text{under } H_0$$

b) ~~If $\sigma_1^2 = \sigma_2^2$~~ but unknown, with small sample size

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{S}_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2) \quad \text{under } H_0$$

where the pooled sample variance $\hat{S}_p^2 = \dfrac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$

(2) Paired Sample t-Test for paired data ~~H_0~~:

$H_0: \mu_X = \mu_Y \iff H_0: \mu_D = 0 \qquad \mu_D = \mu_X - \mu_Y$

$D_i = X_i - Y_i \qquad \bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$

$$t = \frac{\bar{D}}{S_D / \sqrt{n}} \sim t(n-1) \quad \text{under } H_0$$

(3) Test of two proportions $H_0: P_1 = P_2$.

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P} \cdot (1 - \hat{P}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \quad \text{under } H_0,$$

$\hat{P}_1 = \frac{y_1}{n_1} \qquad \hat{P}_2 = \frac{y_2}{n_2}$

$\hat{P} = \frac{y_1 + y_2}{n_1 + n_2}$.

(4) Test of variances $H_0: \sigma_1^2 = \sigma_2^2$ (normal populations)

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1) \quad \text{under } H_0$$

Rejection region $\{F > F(\alpha, n_1 - 1, n_2 - 1)\}$ for $H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1$

4. Chi-Square Tests

(1) χ^2 test for k Proportions (multinomial population)

H_0: $p_1 = p_{1,0}$, $p_2 = p_{2,0}$, $\cdots$, $p_k = p_{k,0}$ $\quad$ H_1: Not all p_i are equal

Statistic $$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - n\cdot p_{i,0})^2}{n\cdot p_{i,0}} \sim \chi^2(k-1) \text{ under } H_0$$

~~Rejection region~~ $\{ \chi^2 > \chi^2(\alpha, k-1) \}$

(2) χ^2 test for Independence

H_0: Two variables are independent, H_1: Variables are dependent.

statistic $$\chi^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{(y_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\cdot \hat{p}_{i\cdot}\cdot \hat{p}_{\cdot j}} \sim \chi^2((a-1)\cdot(b-1))$$

where $n\cdot \hat{p}_{i\cdot}\cdot \hat{p}_{\cdot j} = (y_{i\cdot}\cdot y_{\cdot j})/n = \left(\sum_{j=1}^{b} y_{ij}\right)\cdot\left(\sum_{i=1}^{a} y_{ij}\right)/n$

Rejection Region $\{ \chi^2 > \chi^2(\alpha, (a-1)(b-1)) \}$

(3) Goodness-of-fit Test

H_0: X follows a distribution (normal, poisson, exponential.)

H_1: X does not follow the specified distribution

a) Divide Sample to k disjoint cells, frequency y_i

b) Find expected probability p_i of ith cell after the parameters are estimated.

c) $$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i} \sim \chi^2(k-1-h)$$ where h is the number of parameters

Rejection region $\{ \chi^2 > \chi^2(\alpha, k-1-h) \}$

Chap 8. Regression Analysis

1. Simple Linear Regression Model

Model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i=1,\dots,n$

where error $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

(1) Least Square Estimator $(\hat{\beta}_0, \hat{\beta}_1)$:

$$\{\hat{\beta}_0, \hat{\beta}_1\} = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \{Y_i - \beta_0 - \beta_1 x_i\}^2 \right\}$$

$$\hat{\beta}_1 = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

(2) Fitted Value: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Residual $e_i = y_i - \hat{y}_i \quad \Rightarrow \sum_{i=1}^n e_i = 0, \quad \sum_{i=1}^n e_i x_i = 0$

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2 \sum_i x_i^2}{n \sum_i (x_i-\bar{x})^2}\right), \quad \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_i (x_i-\bar{x})^2}\right)$$

$\hat{\beta}_0, \hat{\beta}_1 \perp SSE$

$$\frac{\sum_{i=1}^n e_i^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2(n-2), \quad MSE = \frac{SSE}{n-2} \; (= \hat{\sigma}^2)$$

(3) ANOVA

Source	DF	SS	MS	F
Regression	1	$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \cdot \sum_{i=1}^n (x_i - \bar{x})^2$	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Error	$n-2$	$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$MSE = \frac{SSE}{n-2}$	
Total	$n-1$	$SSTO = \sum_{i=1}^n (y_i - \bar{y})^2$		

$H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ $\quad F \overset{H_0}{\sim} F(1, n-2)$

$R^2 = \frac{SSR}{SSTO}$ coefficient of determination

Standard error: $s(\hat{\beta}_1) = \sqrt{MSE / \sum_{i=1}^n (x_i - \bar{x})^2}$, $\quad t_{\hat{\beta}_1} = \hat{\beta}_1 / s(\hat{\beta}_1) \overset{H_0}{\sim} t(n-2)$

C.I. for β_1: $\hat{\beta}_1 \pm t_{\frac{\alpha}{2}}(n-2) \cdot s(\hat{\beta}_1)$

2. Multiple Linear Regression

Model: $Y = X \cdot \beta + \varepsilon$ $\quad \beta = (\beta_0, \beta_1, \cdots, \beta_p)'$ $\quad X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{pmatrix}_{n \times (p+1)}$

$Y = (Y_1, \cdots, Y_n)'$ $\quad \varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, i.e. $\varepsilon \sim N_n(0, \sigma^2 I_n)$

(1) $\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \underbrace{(Y - X\beta)'(Y - X\beta)}_{S(\beta)}$

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'Y + 2X'X\beta = 0 \quad \Rightarrow \quad (X'X) \cdot \beta = X'Y$$

$$\hat{\beta}_{LS} = (X'X)^{-1} X'Y \quad \sim N_{p+1}(\beta, \sigma^2 (X'X)^{-1})$$

(2) $\hat{Y} = X \cdot \hat{\beta}_{LS} = \underbrace{X(X'X)^{-1}X'}_{H} Y = HY$

Residual $\hat{\varepsilon} = Y - \hat{Y} = (I - H)Y \sim N_{(p+1)}(0, \sigma^2(I-H))$

$E(SSE) = E(\hat{\varepsilon}' \cdot \hat{\varepsilon}) = \sigma^2(n-p-1)$

where $SSE = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{\varepsilon}' \cdot \hat{\varepsilon} = \varepsilon' \cdot (I - H) \cdot \varepsilon$

$Cov(\hat{\beta}, \hat{\varepsilon}) = 0$

(3) ANOVA Table

Source	SS	DF	MS	F
Reg.	$SSR = \sum_i (\hat{y}_i - \bar{y})^2$	p	$MSR = SSR/p$	$F = MSR/MSE$
Error	$SSE = \sum_i (y_i - \hat{y}_i)^2$	$n-p-1$	$MSE = SSE/(n-p-1)$	
Total	$SSTO = \sum_i (y_i - \bar{y})^2$	$n-1$		

$H_0: \beta_1 = \cdots = \beta_p = 0$ vs H_1: not all β_i's are 0s.

$$F = \frac{MSR}{MSE} \overset{H_0}{\sim} F(p, n-p-1)$$

Partial t-test: $\quad t_{\hat{\beta}_k} = \dfrac{\hat{\beta}_k}{S(\hat{\beta}_k)} \overset{H_0: \beta_k = 0}{\sim} t(n-p-1)$ $\quad S^2(\hat{\beta}_k) = MSE \cdot C_{k+1}$

$k = 1, \cdots, p$

C_{k+1} is the $(k+1)$-th diagonal element of $(X'X)^{-1}$.

C.I. for β_k: $\quad \hat{\beta}_k \pm t_{\frac{\alpha}{2}}(n-p-1) \cdot S(\hat{\beta}_k)$

Chap 6. Experiment with one factor

1. Completely Randomized Design (CRD) with one fixed factor

Model: $Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i=1,\cdots,k$; $j=1,\cdots,n_i$.

μ_i : i-th trt mean τ_i : i-th trt effect, $\sum_{i=1}^{k} n_i\tau_i = 0$

$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

$\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ $\bar{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$, $N = \sum_{i=1}^{k} n_i$

(1) ANOVA Table

Source	SS	DF	MS	F
Treatment	$SSTR = \sum_{i=1}^{k} n_i\cdot(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$k-1$	$MSTR = \frac{SSTR}{k-1}$	$F = MSTR/MSE$
Error	$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$N-k$	$MSE = \frac{SSE}{N-k}$	
Total	$SSTO = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$N-1$		

$E[MSE] = \sigma^2$, $E[MSTR] = \sigma^2 + \frac{1}{k-1}\sum_{i=1}^{k} n_i\cdot\tau_i^2$

$H_0: \tau_1 = \cdots = \tau_k = 0$ vs H_1: at least one $\tau_i \neq 0$. $F = \frac{MSTR}{MSE} \overset{H_0}{\sim} F(k-1, N-k)$

(2) Estimation of Means: $\hat{\mu}_i = \bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$

$\hat{\sigma}^2 = MSE$.
$Var(Y_{ij}) = \sigma^2$
$E(Y_{ij}) = \mu + \tau_i$

C.I. for μ_i: $\bar{Y}_{i\cdot} \pm t_{\frac{\alpha}{2}}(N-k)\cdot\sqrt{\frac{MSE}{n_i}}$

C.I. for $(\mu_i - \mu_j)$: $(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) \pm t_{\frac{\alpha}{2}}(N-k)\cdot\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})}$

(3) Simultaneous C.I. for $(\mu_i - \mu_j)$, $\forall(i,j)$.

Tukey's: $(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) \pm q(\alpha, k, N-k)\cdot\sqrt{\frac{MSE}{2}(\frac{1}{n_i} + \frac{1}{n_j})}$

Bonferroni's: $(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) \pm t_{\frac{\alpha}{k\cdot(k-1)}}(N-k)\cdot\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})}$

2. CRD with one random factor (balanced)

model: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $\quad i = 1, \dots, k$, $\; j = 1, \dots, n$, $\quad N = n.k$

where $\tau_i \overset{iid}{\sim} N(0, \sigma_\tau^2)$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $\{\tau_i\}$ independent of $\{\varepsilon_{ij}\}$

(SSTO, SSTR, SSE, same as those in CRD with one fixed factor)

(1) ANOVA

Source	SS	DF	MS	F
Treatment	SSTR	$k-1$	$MSTR = SSTR/(k-1)$	$F = MSTR/MSE$
Error	SSE	$N-k$	$MSE = SSE/(N-k)$	
Total	SSTO	$N-1$		

$E[MSE] = \sigma^2$, $\quad E[MSTR] = \sigma^2 + n\sigma_\tau^2$

$H_0: \sigma_\tau^2 = 0$ vs $H_1: \sigma_\tau^2 > 0$, $\quad F = \frac{MSTR}{MSE} \overset{H_0}{\sim} F(k-1, N-k)$

$\hat{\sigma}^2 = MSE$, $\quad \hat{\sigma}_\tau^2 = \frac{MSTR - MSE}{n}$

(2) $Cov(Y_{ij}, Y_{i'j'}) = \begin{cases} 0 & i \neq i' \\ \sigma_\tau^2 & i = i',\ j \neq j' \\ \sigma^2 + \sigma_\tau^2 & \text{if } i = i',\ j = j' \end{cases}$, $\quad Y_{ij} \sim N(\mu, \sigma^2 + \sigma_\tau^2)$

(3) Estimation of μ: $\quad \hat{\mu} = \bar{Y}_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} Y_{ij}$

$Var(\hat{\mu}) = \frac{1}{N}(\sigma^2 + n\sigma_\tau^2) = \frac{1}{N} E[MSTR]$

C.I. for μ:

$$\bar{Y}_{\cdot\cdot} \pm t_{\frac{\alpha}{2}}(k-1) \cdot \sqrt{\frac{MSTR}{N}}$$

3. Randomized Complete Block Design (RCBD, fixed)

Model: $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ $\quad i=1,\cdots,k \quad j=1,\cdots,b \qquad N = k\cdot b$

where $\sum_{i=1}^{k} \tau_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

$\bar{Y}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b} Y_{ij} \qquad \bar{Y}_{\cdot j} = \frac{1}{k}\sum_{i=1}^{k} Y_{ij} \qquad \bar{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{b} Y_{ij}$

$E(Y_{ij}) = \mu + \tau_i + \beta_j$, $\quad Var(Y_{ij}) = \sigma^2$

ANOVA

Source	SS	DF	MS	F
Treatment	$SSTR = b\cdot\sum_{i=1}^{k}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$k-1$	$MSTR = \frac{SSTR}{k-1}$	$F_{TR} = \frac{MSTR}{MSE}$
Block	$SSB = k\cdot\sum_{j=1}^{b}(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$	$b-1$	$MSB = \frac{SSB}{b-1}$	$F_B = \frac{MSB}{MSE}$
Error	$SSE = \sum_{i=1}^{k}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2$	$(k-1)\cdot(b-1)$	$MSE = \frac{SSE}{(k-1)(b-1)}$	
Total	$SSTO = \sum_{i=1}^{k}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$k\cdot b-1$		

$E[MSTR] = \sigma^2 + \frac{b}{k-1}\sum_{i=1}^{k}\tau_i^2$, $\quad E[MSB] = \sigma^2 + \frac{k}{b-1}\sum_{j=1}^{b}\beta_j^2$

H_0: $\tau_i = 0$, $\forall i=1,\cdots,k$ vs H_1: at least one $\tau_i \neq 0$

$$F_{TR} = \frac{MSTR}{MSE} \overset{H_0}{\sim} F(k-1, (k-1)(b-1))$$

H_0: $\beta_j = 0$, $\forall j=1\cdots b$ vs H_1: at least one $\beta_j \neq 0$

$$F_B = \frac{MSB}{MSE} \overset{H_0}{\sim} F(b-1, (k-1)(b-1))$$

Tukey's Simultaneous C.I. for treatment means $(\bar{\mu}_{i\cdot} - \bar{\mu}_{j\cdot})$:

$$(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) \pm q(\alpha, k, (k-1)(b-1)) \cdot \sqrt{\frac{MSE}{b}}$$

Chap 7. Experiments with Two Factors

1. Completed Crossed and ~~Bla~~ Balanced Design (Factorial Design)

Model: $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, $i=1,\cdots,a;\ j=1,\cdots,b$, $k=1,\cdots,n$

Effect Model: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, $\varepsilon_{ijk} \overset{iid}{\sim} N(0,\sigma^2)$

main effects: α_i, β_j interaction effects: $(\alpha\beta)_{ij}$

Factor A and Factor B both fixed

$$\sum_{i=1}^{a}\alpha_i = 0,\quad \sum_{j=1}^{b}\beta_j = 0,\quad \sum_{i=1}^{a}(\alpha\beta)_{ij} = \sum_{j=1}^{b}(\alpha\beta)_{ij} = 0$$

Factor A and Factor B both random

$$\alpha_i \sim N(0,\sigma_\alpha^2),\quad \beta_j \sim N(0,\sigma_\beta^2),\quad (\alpha\beta)_{ij} \overset{iid}{\sim} N(0,\sigma_{\alpha\beta}^2)$$

$$\{\alpha_i\} \perp \{\beta_j\} \perp \{(\alpha\beta)_{ij}\} \perp \{\varepsilon_{ijk}\}$$

Factor A fixed, Factor B random (mixed effects)

$$\sum_{i=1}^{k}\alpha_i = 0,\quad \beta_j \sim N(0,\sigma_\beta^2),\quad (\alpha\beta)_{ij} \sim N\left(0, \frac{(a-1)\sigma_{\alpha\beta}^2}{a}\right),\quad \sum_i(\alpha\beta)_{ij} = 0,\quad Cov((\alpha\beta)_{ij}, (\alpha\beta)_{i'j}) = -\frac{1}{a}\sigma_{\alpha\beta}^2$$

	$E(Y_{ijk})$	$Var(Y_{ijk})$
A, B fixed	$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$	σ^2
A fixed, B random	$\mu + \alpha_i$	$\sigma^2 + \sigma_\beta^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2$
A, B random	μ	$\sigma^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2$

$$SSA = \sum_i\sum_j\sum_k (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSB = \sum_i\sum_j\sum_k (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSAB = \sum_i\sum_j\sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSTR = SSA + SSB + SSAB = \sum_i\sum_j\sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSTO = \sum_i\sum_j\sum_k (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2,\quad SSE = \sum_i\sum_j\sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2$$

Source	DF	MS	E(MS·): A, B Fixed	E(MS·): A Fixed B Random	E(MS·): A, B Random
A	$(a-1)$	$MSA = \frac{SSA}{(a-1)}$	$\sigma^2 + nb \cdot \frac{\sum_{i=1}^{a} \alpha_i^2}{a-1}$	$\sigma^2 + nb \frac{\sum \alpha_i^2}{a-1} + n\sigma^2_{\alpha\beta}$	$\sigma^2 + nb\sigma_\alpha^2 + n\sigma^2_{\alpha\beta}$
B	$(b-1)$	$MSB = \frac{SSB}{(b-1)}$	$\sigma^2 + na \cdot \frac{\sum_{j=1}^{b} \beta_j^2}{b-a}$	$\sigma^2 + na\sigma_\beta^2$	$\sigma^2 + na\sigma_\beta^2 + n\sigma^2_{\alpha\beta}$
AB	$(a-1)\cdot(b-1)$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$\sigma^2 + \frac{n\sum_i\sum_j(\alpha\beta)^2_{ij}}{(a-1)(b-1)}$	$\sigma^2 + n\sigma^2_{\alpha\beta}$	$\sigma^2 + n\sigma^2_{\alpha\beta}$
Error	$ab(n-1)$	$MSE = \frac{SSE}{ab(n-1)}$	σ^2	σ^2	σ^2
Total	$abn-1$		H_0: $\alpha_i=0$ — MSA/MSE; $\beta_j=0$ — MSB/MSE; $(\alpha\beta)_{ij}=0$ — $MSAB/MSE$	H_0: $\alpha_i=0$, $(\alpha\beta)_{ij}=0$; $\sigma_\beta^2=0$, $\sigma^2_{\alpha\beta}=0$; $\alpha_i=0$ — $MSA/MSAB$; $\sigma_\beta^2=0$ — MSB/MSE	H_0: $\sigma_\alpha^2=0$ — $MSA/MSAB$; $\sigma_\beta^2=0$ — $MSB/MSAB$; $\sigma^2_{\alpha\beta}=0$ — $MSAB/MSE$

φ. (cont.) Estimation of Means for 2-way ANOVA Factorial Design

① A, B both fixed: $\hat{\mu}_{ij} = \bar{Y}_{ij\cdot} \sim N(\mu_{ij}, \frac{\sigma^2}{n})$,

$$\hat{\bar{\mu}}_{i\cdot} = \bar{Y}_{i\cdot\cdot} \sim N(\bar{\mu}_{i\cdot}, \frac{\sigma^2}{bn}), \quad \hat{\bar{\mu}}_{\cdot j} = \bar{Y}_{\cdot j\cdot} \sim N(\bar{\mu}_{\cdot j}, \frac{\sigma^2}{an})$$

C.I. for $(\bar{\mu}_{i\cdot} - \bar{\mu}_{i'\cdot})$: $(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) \pm t_{\frac{\alpha}{2}}(ab(n-1)) \cdot \sqrt{\frac{2MSE}{bn}}$

$(\bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot j'})$: $(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot j'\cdot}) \pm t_{\frac{\alpha}{2}}(ab(n-1)) \cdot \sqrt{\frac{2MSE}{an}}$

Tukey's C.I. for $(\bar{\mu}_{i\cdot} - \bar{\mu}_{i'\cdot})$, $\forall (i, i')$

$$(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) \pm q(\alpha, a, ab(n-1)) \cdot \sqrt{\frac{MSE}{bn}}$$

Tukey's C.I. for $(\bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot j'})$, $\forall (j, j')$

$$(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot j'\cdot}) \pm q(\alpha, b, ab(n-1)) \cdot \sqrt{\frac{MSE}{an}}$$

② A fixed B random: $\hat{\bar{\mu}}_{i\cdot} = \bar{Y}_{i\cdot\cdot}$, $\hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}$

C.I. for α_i: $(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) \pm t_{\frac{\alpha}{2}}((a-1)(b-1)) \cdot \sqrt{\frac{MSAB}{bn}}$

C.I. for $(\alpha_i - \alpha_{i'})$: $(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) \pm t_{\frac{\alpha}{2}}((a-1)(b-1)) \sqrt{\frac{2MSAB}{bn}}$

Tukey's C.I. for $(\alpha_i - \alpha_{i'})$, $\forall (i, i')$

$$(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) \pm q(\alpha, a, (a-1)(b-1)) \cdot \sqrt{\frac{MSAB}{bn}}$$

2. Nested Design

Model: $Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$, $i=1 \cdots a$; $j=1 \cdots b$; $k=1 \cdots n$

SSTO, SSA, $SSB(A) = \sum_i \sum_j \sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2$, SSE

Source	DF	MS	E(MS·) A,B fixed	E(MS·) A fixed, B random	E(MS·) A,B random
A	a	$MSA = \frac{SSA}{a-1}$	$\sigma^2 + bn \cdot \frac{\sum_{i=1}^{a} \alpha_i^2}{a-1}$	$\sigma^2 + bn \frac{\sum_i \alpha_i^2}{a-1} + n\sigma_\beta^2$	$\sigma^2 + bn\sigma_\alpha^2 + n\sigma_\beta^2$
B(A)	$a(b-1)$	$MSB(A) = \frac{SSB(A)}{a(b-1)}$	$\sigma^2 + n \frac{\sum_i \sum_j \beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
Error	$ab(n-1)$	$MSE = \frac{SSE}{ab(n-1)}$	σ^2	σ^2	σ^2
Total	$abn-1$				

A,B fixed:

$H_0: \alpha_i = 0$ MSA/MSE

$\beta_{j(i)} = 0$ $MSB(A)/MSE$

C.I. for $\mu_{i\cdot}$

$\bar{Y}_{i\cdot\cdot} \pm t_{\frac{\alpha}{2}}(ab(n-1)) \cdot \sqrt{\frac{MSE}{bn}}$

A fixed, B random:

$H_0: \alpha_i = 0$ $MSA/MSB(A)$,

$\sigma_\beta^2 = 0$ $MSB(A)/MSE$

$\bar{Y}_{i\cdot\cdot} \pm t_{\frac{\alpha}{2}}(a(b-1)) \cdot \sqrt{\frac{MSB(A)}{bn}}$

A,B random:

$H_0: \sigma_\alpha^2 = 0$ $MSA/MSB(A)$

$\sigma_\beta^2 = 0$ $MSB(A)/MSE$

3. 2^k - factorial design

Coded variable $x_1, x_2, \cdots x_k$, for factor A, B, ⋯, (k factors) with two levels $(+,-)$

Design matrix includes 2^k runs, complete combined treatment

Run	x_1	x_2	x_1x_2	obs.
1	−	−	+	$\bar{Y}_{1\cdot}$
2	+	−	−	$\bar{Y}_{2\cdot}$
3	−	+	−	$\bar{Y}_{3\cdot}$
4	+	+	+	$\bar{Y}_{4\cdot}$

n replications of 2^k runs

$\text{effect}(1) = (\bar{Y}_2 - \bar{Y}_1 + \bar{Y}_3 - \bar{Y}_4)/4$ ⋯

$\widehat{\text{effect}} = \frac{1}{2^k} \sum_{i=1}^{2^k} C_i \bar{Y}_i$, $C_i = \pm 1$

$Var(\text{effect}) = \frac{1}{n \cdot 2^k} Var(Y)$

$\widehat{Var}(\text{effect}) = \frac{1}{2^k \cdot n} \cdot S^2$, where S^2 is the pooled sample variance

and $S^2 = \frac{1}{2^k} \sum_{i=1}^{2^k} S_i^2 = \frac{1}{2^k} \sum_{i=1}^{2^k} \left(\frac{1}{(n-1)} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_i)^2 \right)$

C.I. for effect: $\widehat{\text{effect}} \pm 2 \cdot S(\widehat{\text{effect}})$, standard error $S(\widehat{\text{effect}}) = \sqrt{\widehat{Var}(\text{effect})}$