

Statistical Theory

Daniel Mao

Contents

1	Descriptive Statistics	1
1.1	Population	1
1.2	Order, Rank, and Quantiles	3
1.3	Statistical Distance	4
2	Sampling Theory	5
2.1	Sampling Distribution	5
2.2	Inclusion Probability	7
2.3	Survey Errors	7
3	Experimental Design	9
3.1	Completely Random Design	10
3.2	Randomized Block Design	16
3.3	Two-Way Factorial Design	23
3.4	Two-Level Factorial Design	26
4	Estimation Theory	27
4.1	Maximum Likelihood Estimation	27
4.2	Examples	28
4.3	Confidence Intervals	29
4.4	Simple Random Sampling Without Replacement	29
5	Hypothesis Testing	31
5.1	p -values	31
6	Asymptotic Theory	33
6.1	Law of Large Numbers	34
6.2	Central Limit Theorem	34
7	Power Transformation	37

Chapter 1

Descriptive Statistics

1.1 Population

DEFINITION 1.1 (Population). We define a **population**, denoted by \mathcal{P} , to be a finite set of elements.

DEFINITION 1.2 (Unit). We define a **unit**, denoted by u , to be an element of some population.

DEFINITION 1.3 (Invariants and Equivariants). Let \mathcal{P} be a population. Let a be an attribute. We say that the attribute a is

1. **location invariant** if

$$\forall b \in \mathbb{R}, a(\mathcal{P} + b) = a(\mathcal{P}).$$

2. **location equivariant** if

$$\forall b \in \mathbb{R}, a(\mathcal{P} + b) = a(\mathcal{P}) + b.$$

3. **scale invariant** if

$$\forall m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P}) = a(\mathcal{P}).$$

4. **scale equivariant** if

$$\forall m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P}) = m \times a(\mathcal{P}).$$

5. **location-scale invariant** if

$$\forall b \in \mathbb{R}, m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P} + b) = a(\mathcal{P}).$$

6. **location-scale equivariant** if

$$\forall b \in \mathbb{R}, m \in \mathbb{R}^{\geq 0}, a(m \times \mathcal{P} + b) = m \times a(\mathcal{P}) + b.$$

7. **replication invariant** if

$$\forall k \in \mathbb{N}, a(\mathcal{P}^k) = a(\mathcal{P}).$$

8. **replication equivariant** if

$$\forall k \in \mathbb{N}, a(\mathcal{P}^k) = k \times a(\mathcal{P}).$$

PROPOSITION 1.4. The population average is location-scale equivariant and replication invariant.

1.2 Order, Rank, and Quantiles

DEFINITION 1.5 (Order Statistic). Let $\mathcal{P} = \{y_1, \dots, y_N\}$ be a population. We define the **order statistic**, denoted by $y_{(1)}, \dots, y_{(N)}$, to be the ordered values of the variate values from \mathcal{P} such that

$$y_{(1)} \leq \dots \leq y_{(N)}.$$

DEFINITION 1.6 (Rank Statistic). Let $\mathcal{P} = \{y_1, \dots, y_N\}$ be a population. We define the **rank statistic**, denoted by r_1, \dots, r_N , to be the ranks of the variate values from \mathcal{P} .

So $\forall u \in \mathcal{P}, y_u = y_{(r_u)}$.

DEFINITION 1.7 (Quantiles). **Quantiles that Measure Center:**

- Median: $Q_y(1/2)$.

- Mid-hinge: $\frac{Q_y(1/4) + Q_y(3/4)}{2}$.
- Mid-range: $\frac{Q_y(1/N) + Q_y(1)}{2}$.
- Trimean: $\frac{Q_y(1/4) + Q_y(1/2) + Q_y(3/4)}{3}$.

Quantiles that Measure Spread:

- Range: $Q_y(1) - Q_y(1/N)$.
- Inter-Quartile Range: $Q_y(3/4) - Q_y(1/4)$.

1.3 Statistical Distance

DEFINITION 1.8 (Kullback–Leibler (KL) Divergence).

- Discrete Case:

Let \mathcal{X} be a probability space. Let P and Q be discrete probability distributions over \mathcal{X} . We define the **Kullback–Leibler divergence** from Q to P , denoted by $D_{KL}(P||Q)$, to be the sum given by

$$D_{KL}(P||Q) := \mathbb{E}_{X \sim P} \left[\log \frac{P(X)}{Q(X)} \right] = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

- Continuous Case:

Let P and Q be continuous probability distributions. We define the **Kullback–Leibler divergence** from Q to P , denoted by $D_{KL}(P||Q)$, to be the integral given by

$$D_{KL}(P||Q) := \mathbb{E}_{X \sim P} \left[\log \frac{P(X)}{Q(X)} \right] = \int_{\mathbb{R}} P(x) \log \frac{P(x)}{Q(x)} dx.$$

Chapter 2

Sampling Theory

2.1 Sampling Distribution

PROPOSITION 2.1. Let $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$. Then

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

Proof. Part 1: show that

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)s^2 + n(\bar{X} - \mu)^2.$$

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\ &= \sum_{i=1}^n [(X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(X_i - \bar{X})(\bar{X} - \mu)] \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 + \sum_{i=1}^n 2(X_i - \bar{X})(\bar{X} - \mu) \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X}) \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n X_i - 2(\bar{X} - \mu) \sum_{i=1}^n \bar{X} \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu)n\bar{X} - 2(\bar{X} - \mu)n\bar{X} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\
&= (n-1) \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\
&= (n-1)s^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

That is,

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)s^2 + n(\bar{X} - \mu)^2,$$

as desired.

Part 2

Since

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)s^2 + n(\bar{X} - \mu)^2,$$

we get

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{(n-1)s^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2.$$

Since $X_i \sim N(\mu, \sigma^2)$, we get

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1).$$

So

$$\left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(1)$$

and hence

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

Since $\bar{X} \sim N(\mu, \sigma^2/n)$, we get

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

So

$$\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1).$$

Notice $\frac{(n-1)s^2}{\sigma^2}$ and $\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$ are independent, it follows that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

□

2.2 Inclusion Probability

Simple Random Sampling Without Replacement

- inclusion probability: $\pi_u = \frac{n}{N}$.

Derivation:

$$\begin{aligned}
 \pi_u &= \sum_{S \in \mathcal{P}_S} P(S) I(u \in S) \\
 &= \frac{1}{\binom{N}{n}} \binom{N-1}{n-1} \\
 &= \frac{1}{\frac{N!}{n!(N-n)!}} \frac{(N-1)!}{(n-1)!(N-n)!} \\
 &= \frac{n}{N}.
 \end{aligned}$$

- joint inclusion probability: $\pi_{uv} = \frac{n(n-1)}{N(N-1)}$

Derivation:

$$\begin{aligned}
 \pi_{uv} &= \sum_{S \in \mathcal{P}_S} P(S) I(u, v \in S) \\
 &= \frac{1}{\binom{N}{n}} \binom{N-2}{n-2} \\
 &= \frac{1}{\frac{N!}{n!(N-n)!}} \frac{(N-2)!}{(n-2)!(N-n)!} \\
 &= \frac{n(n-1)}{N(N-1)}.
 \end{aligned}$$

2.3 Survey Errors

DEFINITION 2.2 (Response Rate). Define the following notations:

- I : complete interviews;
- P : partial interview;
- R : refusals or break-offs;
- NC : non-contacts;
- O : other eligible non-respondents;
-

DEFINITION 2.3 (Non-Response Bias). Let 1 denote the respondent group. Let 2 denote the non-respondent group. Let \bar{Y}_1 and \bar{Y}_2 denote the population means of group 1 and group 2, respectively. We define the **non-response bias**

Chapter 3

Experimental Design

Contents

3.1	Completely Random Design	10
3.1.1	Estimation of Mean	10
3.1.2	Estimation of Variance	11
3.1.3	Hypothesis Testing	15
3.2	Randomized Block Design	16
3.2.1	Estimation of Mean	16
3.2.2	Estimation of Variance	18
3.2.3	Hypothesis Testing	22
3.3	Two-Way Factorial Design	23
3.3.1	Estimation of Mean	23
3.3.2	Estimation of Variance	23
3.3.3	Hypothesis Testing	25
3.4	Two-Level Factorial Design	26

3.1 Completely Random Design

DEFINITION 3.1 (Completely Random Design). Let k denote the number of treatments. Let n_i denote the number of units that receive the i -th treatment. We model the population as

$$y_{ij} = \mu_i + e_{ij}, \text{ for } i \in \{1, \dots, k\} \text{ and } j \in \{1, \dots, n_i\}$$

where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ are assumed to be independent. The total number of parameters in this model is $k + 1$

3.1.1 Estimation of Mean

PROPOSITION 3.2. Let y_{ij} for $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, n_i\}$ be given. Consider the following optimization problem:

$$\begin{aligned} \text{(P)} \quad & \min \quad f(\mu) := \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 \\ & \text{subject to: } \mu \in \mathbb{R}^k. \end{aligned}$$

Then the minimizer $\hat{\mu} \in \mathbb{R}^k$ of (P) is given by

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \text{ for } i \in \{1, \dots, k\}.$$

Proof. Let $p \in \{1, \dots, k\}$ be arbitrary. Then

$$\begin{aligned} \frac{\partial}{\partial \mu_p} f(\mu) &= \frac{\partial}{\partial \mu_p} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 = \sum_{j=1}^{n_p} \frac{\partial}{\partial \mu_p} (y_{pj} - \mu_p)^2 = -2 \sum_{j=1}^{n_p} (y_{pj} - \mu_p), \text{ and} \\ \frac{\partial^2}{\partial \mu_p^2} f(\mu) &= \frac{\partial}{\partial \mu_p} \left[-2 \sum_{j=1}^{n_p} (y_{pj} - \mu_p) \right] = 2n_p > 0. \end{aligned}$$

Suppose $\hat{\mu} \in \mathbb{R}^k$ is a minimizer of f . Then we have $\nabla f(\hat{\mu}) = 0 \in \mathbb{R}^k$. So

$$\frac{\partial}{\partial \mu_i} f(\hat{\mu}) = 0 \implies \hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.$$

Testing the Hessian of f at point $\hat{\mu} \in \mathbb{R}^k$ confirms that it is indeed a minimizer of f . \square

PROPOSITION 3.3 (Mean of the Mean Estimator). We have

$$\mathbb{E}(\hat{\mu}) = \mu \in \mathbb{R}^k.$$

i.e., $\hat{\mu}$ is an unbiased estimator for μ .

Proof. Recall that $\forall i \in \{1, \dots, k\}$, $\forall j \in \{1, \dots, n_i\}$, we have $y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$. So $\forall i \in \{1, \dots, k\}$, $\forall j \in \{1, \dots, n_i\}$, $\mathbb{E}(y_{ij}) = \mu_i$. Now we can compute

$$\mathbb{E}(\hat{\mu}_i) = \mathbb{E}\left(\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}\right) = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{E}(y_{ij}) = \frac{1}{n_i} \sum_{j=1}^{n_i} \mu_i = \mu_i.$$

That is, $\mathbb{E}(\hat{\mu}) = \mu$, as desired. \square

PROPOSITION 3.4 (Variance of the Mean Estimator). We have

$$\forall i \in \{1, \dots, k\}, \quad \mathbb{V}(\hat{\mu}_i) = \frac{\sigma^2}{n_i}.$$

Proof. Recall that $\forall i \in \{1, \dots, k\}$, $\forall j \in \{1, \dots, n_i\}$, we have $y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$. So $\forall i \in \{1, \dots, k\}$, $\forall j \in \{1, \dots, n_i\}$, $\mathbb{V}(y_{ij}) = \sigma^2$. Now we can compute

$$\mathbb{V}(\hat{\mu}_i) = \mathbb{V}\left(\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}\right) = \sum_{j=1}^{n_i} \frac{1}{n_i^2} \mathbb{V}(y_{ij}) = \sum_{j=1}^{n_i} \frac{1}{n_i^2} \sigma^2 = \frac{\sigma^2}{n_i}.$$

\square

3.1.2 Estimation of Variance

In this subsection, we assume that $\forall i \in \{1, \dots, k\}$, $n_i = n$ for some $n \in \mathbb{Z}_{++}$.

DEFINITION 3.5 (Sum of Squares). We define the following terms:

$$\begin{aligned} \text{SS}_{\text{trt}} &:= n \sum_{i=1}^k (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2, \\ \text{SS}_{\text{err}} &:= \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i\cdot})^2, \end{aligned}$$

$$SS_{\text{tot}} := \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2.$$

PROPOSITION 3.6 (Decomposition of SS_{tot}). We have

$$SS_{\text{tot}} = SS_{\text{trt}} + SS_{\text{err}}.$$

Proof.

$$\begin{aligned} SS_{\text{tot}} &= \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^n \left[(y_{ij} - \bar{y}_{i.})^2 + (\bar{y}_{i.} - \bar{y}_{..})^2 + 2(y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}_{..}) \right] \\ &= SS_{\text{trt}} + SS_{\text{err}} + \sum_{i=1}^k \sum_{j=1}^n 2(y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}_{..}) \\ &= SS_{\text{trt}} + SS_{\text{err}} + 2 \sum_{i=1}^k \sum_{j=1}^n y_{ij} \bar{y}_{i.} - 2 \sum_{i=1}^k \sum_{j=1}^n y_{ij} \bar{y}_{..} - 2 \sum_{i=1}^k \sum_{j=1}^n \bar{y}_{i.}^2 + 2 \sum_{i=1}^k \sum_{j=1}^n \bar{y}_{i.} \bar{y}_{..} \\ &= SS_{\text{trt}} + SS_{\text{err}} + 2n \sum_{i=1}^k \bar{y}_{i.}^2 - 2n \bar{y}_{..} \sum_{i=1}^k \bar{y}_{i.} - 2n \sum_{i=1}^k \bar{y}_{i.}^2 + 2n \bar{y}_{..} \sum_{i=1}^k \bar{y}_{i.} \\ &= SS_{\text{trt}} + SS_{\text{err}} + 0 = SS_{\text{trt}} + SS_{\text{err}}. \end{aligned}$$

□

DEFINITION 3.7 (Mean Squares). We define the following estimators for the variance σ^2 .

$$\begin{aligned} MS_{\text{trt}} &:= SS_{\text{trt}} / (k - 1), \\ MS_{\text{err}} &:= SS_{\text{err}} / (k(n - 1)). \end{aligned}$$

PROPOSITION 3.8 (Mean of MS_{err}). We have

$$\mathbb{E}(MS_{\text{err}}) = \sigma^2.$$

i.e., MS_{err} is an unbiased estimator for σ^2 .

Proof. Recall that $\forall i \in \{1, \dots, k\}$, $\forall j \in \{1, \dots, n\}$, we have $y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$. So

$$\bar{y}_{i\cdot} = \frac{1}{n} \sum_{j=1}^n y_{ij} \sim \mathcal{N}(\mu_i, \frac{\sigma^2}{n}), \quad \forall i \in \{1, \dots, k\}.$$

So

$$\begin{aligned} \mathbb{E}(y_{ij}^2) &= \mathbb{V}(y_{ij}) + \mathbb{E}^2(y_{ij}) = \sigma^2 + \mu_i^2, \quad \forall i, j, \text{ and} \\ \mathbb{E}(\bar{y}_{i\cdot}^2) &= \mathbb{V}(\bar{y}_{i\cdot}) + \mathbb{E}^2(\bar{y}_{i\cdot}) = \frac{\sigma^2}{n} + \mu_i^2, \quad \forall i. \end{aligned}$$

Now we can compute

$$\begin{aligned} \mathbb{E}(\text{MS}_{\text{err}}) &= \mathbb{E}(\text{SS}_{\text{err}}/(k(n-1))) = \mathbb{E}\left(\frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i\cdot})^2\right) \\ &= \mathbb{E}\left(\frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (y_{ij}^2 + \bar{y}_{i\cdot}^2 - 2y_{ij}\bar{y}_{i\cdot})\right) \\ &= \mathbb{E}\left(\frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 + n \sum_{i=1}^k \bar{y}_{i\cdot}^2 - 2 \sum_{i=1}^k \bar{y}_{i\cdot} \sum_{j=1}^n y_{ij} \right]\right) \\ &= \mathbb{E}\left(\frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 + n \sum_{i=1}^k \bar{y}_{i\cdot}^2 - 2n \sum_{i=1}^k \bar{y}_{i\cdot}^2 \right]\right) \\ &= \mathbb{E}\left(\frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - n \sum_{i=1}^k \bar{y}_{i\cdot}^2 \right]\right) \\ &= \frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n \mathbb{E}(y_{ij}^2) - n \sum_{i=1}^k \mathbb{E}(\bar{y}_{i\cdot}^2) \right], \text{ by linearity} \\ &= \frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n (\sigma^2 + \mu_i^2) - n \sum_{i=1}^k \left(\frac{\sigma^2}{n} + \mu_i^2 \right) \right], \text{ by above} \\ &= \frac{1}{k(n-1)} \left[(kn - k)\sigma^2 + n \sum_{i=1}^k (\mu_i^2 - \mu_i^2) \right] = \sigma^2. \end{aligned}$$

That is, $\mathbb{E}(\text{MS}_{\text{err}}) = \sigma^2$, as desired. \square

PROPOSITION 3.9 (Mean of MS_{trt}). Under the assumption that $\mu = \mathbb{1}\mu_0 \in \mathbb{R}^k$ for

some $\mu_0 \in \mathbb{R}$, we have

$$\mathbb{E}(\text{MS}_{\text{trt}}) = \sigma^2.$$

i.e., MS_{trt} is an unbiased estimator for σ^2 given that $\mu = \mathbb{1}\mu_0$ for some μ_0 .

Proof. Assume that $\mu = \mathbb{1}\mu_0 \in \mathbb{R}^k$ for some $\mu_0 \in \mathbb{R}$. Then $\forall i \in \{1, \dots, k\}$, $\forall j \in \{1, \dots, n\}$, we have $y_{ij} \sim \mathcal{N}(\mu_0, \sigma^2)$. So

$$\bar{y}_{i\cdot} = \frac{1}{n} \sum_{j=1}^n y_{ij} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{n}), \quad \forall i \in \{1, \dots, k\}, \text{ and}$$

$$\bar{y}_{\cdot\cdot} = \frac{1}{Kn} \sum_{i=1}^k \sum_{j=1}^n y_{ij} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{Kn}).$$

So

$$\mathbb{E}(\bar{y}_{i\cdot}^2) = \mathbb{V}(\bar{y}_{i\cdot}) + \mathbb{E}^2(\bar{y}_{i\cdot}) = \frac{\sigma^2}{n} + \mu_0^2, \quad \forall i, \text{ and}$$

$$\mathbb{E}(\bar{y}_{\cdot\cdot}^2) = \mathbb{V}(\bar{y}_{\cdot\cdot}) + \mathbb{E}^2(\bar{y}_{\cdot\cdot}) = \frac{\sigma^2}{Kn} + \mu_0^2.$$

Now we can compute

$$\begin{aligned} \mathbb{E}(\text{MS}_{\text{trt}}) &= \mathbb{E}(\text{SS}_{\text{trt}}/(k-1)) = \mathbb{E}\left(\frac{n}{k-1} \sum_{i=1}^k (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2\right) \\ &= \mathbb{E}\left(\frac{n}{k-1} \sum_{i=1}^k (\bar{y}_{i\cdot}^2 + \bar{y}_{\cdot\cdot}^2 - 2\bar{y}_{i\cdot}\bar{y}_{\cdot\cdot})\right) \\ &= \mathbb{E}\left(\frac{n}{k-1} \left[\sum_{i=1}^k \bar{y}_{i\cdot}^2 + k\bar{y}_{\cdot\cdot}^2 - 2\bar{y}_{\cdot\cdot} \sum_{i=1}^k \bar{y}_{i\cdot} \right]\right) \\ &= \mathbb{E}\left(\frac{n}{k-1} \left[\sum_{i=1}^k \bar{y}_{i\cdot}^2 + k\bar{y}_{\cdot\cdot}^2 - 2K\bar{y}_{\cdot\cdot}^2 \right]\right) \\ &= \mathbb{E}\left(\frac{n}{k-1} \left[\sum_{i=1}^k \bar{y}_{i\cdot}^2 - k\bar{y}_{\cdot\cdot}^2 \right]\right) \\ &= \frac{n}{k-1} \left[\sum_{i=1}^k \mathbb{E}(\bar{y}_{i\cdot}^2) - k\mathbb{E}(\bar{y}_{\cdot\cdot}^2) \right], \text{ by linearity} \\ &= \frac{n}{k-1} \left[\sum_{i=1}^k \left(\frac{\sigma^2}{n} + \mu_0^2 \right) - k \left(\frac{\sigma^2}{Kn} + \mu_0^2 \right) \right], \text{ by above} \\ &= \frac{n}{k-1} \left[\left(\frac{k}{n} - \frac{1}{n} \right) \sigma^2 + (k-k) \mu_0^2 \right] = \sigma^2. \end{aligned}$$

That is, $\mathbb{E}(\text{MS}_{\text{trt}}) = \sigma^2$, as desired. \square

3.1.3 Hypothesis Testing

DEFINITION 3.10 (ANOVA Table).

Table 3.1: ANOVA Table for Completely Randomized Design

	Sum of Squares	Degree of Freedom	Mean Squares	F_0
Treatment	SS_{trt}	$k - 1$	MS_{trt}	$MS_{\text{trt}}/MS_{\text{err}}$
Error	SS_{err}	$k(n - 1)$	MS_{err}	
Total	SS_{tot}	$kn - 1$		

3.2 Randomized Block Design

DEFINITION 3.11 (Randomized Block Design). Let $a \in \mathbb{Z}_{++}$ denote the number of treatments. Let $b \in \mathbb{Z}_{++}$ denote the number of blocks. We model the population as

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \text{ for } i \in \{1, \dots, a\} \text{ and } j \in \{1, \dots, b\}$$

with constraints $\mathbf{1}^\top \alpha = 0$ and $\mathbf{1}^\top \beta = 0$, and $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ are assumed to be independent. The total number of parameters in this model is $2 + a + b$.

3.2.1 Estimation of Mean

PROPOSITION 3.12. Let y_{ij} for $i \in \{1, \dots, a\}$ and $j \in \{1, \dots, b\}$ be given. Consider the following optimization problem:

$$\begin{aligned} \text{(P)} \quad & \min \quad f(\mu, \alpha, \beta) := \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j)^2 \\ & \text{subject to: } \mu \in \mathbb{R}, \alpha \in \mathbb{R}^a, \beta \in \mathbb{R}^b, \\ & \quad \mathbf{1}^\top \alpha = 0, \mathbf{1}^\top \beta = 0. \end{aligned}$$

Then the minimizer $(\hat{\mu}, \hat{\alpha}, \hat{\beta}) \in \mathbb{R} \oplus \mathbb{R}^a \oplus \mathbb{R}^b$ of (P) is given by

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij}, \\ \hat{\alpha}_i &= \bar{y}_{i.} - \bar{y}_{..} = \frac{1}{b} \sum_{j=1}^b y_{ij} - \bar{y}_{..}, \text{ for } i \in \{1, \dots, a\}, \\ \hat{\beta}_j &= \bar{y}_{.j} - \bar{y}_{..} = \frac{1}{a} \sum_{i=1}^a y_{ij} - \bar{y}_{..}, \text{ for } j \in \{1, \dots, b\}. \end{aligned}$$

Proof. Form the Lagrangian function $\mathcal{L} : \mathbb{R} \oplus \mathbb{R}^a \oplus \mathbb{R}^b \oplus \mathbb{R} \oplus \mathbb{R}$

$$\mathcal{L}(\mu, \alpha, \beta, \xi, \eta) := f(\mu, \alpha, \beta) - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta.$$

Compute the derivatives:

$$\begin{aligned} \frac{\partial}{\partial \mu} \mathcal{L}(\mu, \alpha, \beta, \xi, \eta) &= \frac{\partial}{\partial \mu} \left[f(\mu, \alpha, \beta) - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] \\ &= \frac{\partial}{\partial \mu} \left[\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j)^2 - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^a \sum_{j=1}^b \frac{\partial}{\partial \mu} (y_{ij} - \mu - \alpha_i - \beta_j)^2 \\
&= -2 \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j), \\
\frac{\partial}{\partial \alpha_p} \mathcal{L}(\mu, \alpha, \beta, \xi, \eta) &= \frac{\partial}{\partial \alpha_p} \left[f(\mu, \alpha, \beta) - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] \\
&= \frac{\partial}{\partial \alpha_p} \left[\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j)^2 - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] \\
&= \sum_{i=1}^a \sum_{j=1}^b \frac{\partial}{\partial \alpha_p} (y_{ij} - \mu - \alpha_i - \beta_j)^2 - \xi \frac{\partial}{\partial \alpha_p} \mathbf{1}^\top \alpha \\
&= -2 \sum_{j=1}^b (y_{pj} - \mu - \alpha_p - \beta_j) - \xi, \\
\frac{\partial}{\partial \beta_q} \mathcal{L}(\mu, \alpha, \beta, \xi, \eta) &= \frac{\partial}{\partial \beta_q} \left[f(\mu, \alpha, \beta) - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] \\
&= \frac{\partial}{\partial \beta_q} \left[\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j)^2 - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] \\
&= \sum_{i=1}^a \frac{\partial}{\partial \beta_q} (y_{iq} - \mu - \alpha_i - \beta_q)^2 - \eta \frac{\partial}{\partial \beta_q} \mathbf{1}^\top \beta \\
&= -2 \sum_{i=1}^a (y_{iq} - \mu - \alpha_i - \beta_q) - \eta, \\
\frac{\partial}{\partial \xi} \mathcal{L}(\mu, \alpha, \beta, \xi, \eta) &= \frac{\partial}{\partial \xi} \left[f(\mu, \alpha, \beta) - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] = -\mathbf{1}^\top \alpha, \\
\frac{\partial}{\partial \eta} \mathcal{L}(\mu, \alpha, \beta, \xi, \eta) &= \frac{\partial}{\partial \eta} \left[f(\mu, \alpha, \beta) - \xi \mathbf{1}^\top \alpha - \eta \mathbf{1}^\top \beta \right] = -\mathbf{1}^\top \beta.
\end{aligned}$$

Let $(\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\xi}, \hat{\eta}) \in \mathbb{R} \oplus \mathbb{R}^a \oplus \mathbb{R}^b \oplus \mathbb{R} \oplus \mathbb{R}$ be such that $\nabla \mathcal{L}(\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\xi}, \hat{\eta}) = \mathbf{0} \in \mathbb{R}^{a+b+3}$. Then we get the following system of equations:

$$\left\{ \begin{array}{l} -2 \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j) = 0 \\ -2 \sum_{j=1}^b (y_{pj} - \hat{\mu} - \hat{\alpha}_p - \hat{\beta}_j) - \hat{\xi} = 0, \forall p \\ -2 \sum_{i=1}^a (y_{iq} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_q) - \hat{\eta} = 0, \forall q \\ -\mathbf{1}^\top \hat{\alpha} = 0 \\ -\mathbf{1}^\top \hat{\beta} = 0 \end{array} \right. \implies \left\{ \begin{array}{l} \hat{\mu} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij} \\ \hat{\alpha}_i = \bar{y}_{i\cdot} - \hat{\mu}, \forall i \\ \hat{\beta}_j = \bar{y}_{\cdot j} - \hat{\mu}, \forall j \\ \hat{\xi} = 0 \\ \hat{\eta} = 0. \end{array} \right.$$

Testing the Hessian of \mathcal{L} at point $(\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\xi}, \hat{\eta}) \in \mathbb{R} \oplus \mathbb{R}^a \oplus \mathbb{R}^b \oplus \mathbb{R} \oplus \mathbb{R}$ confirms that it is indeed a minimizer of \mathcal{L} . \square

PROPOSITION 3.13 (Mean of the Mean Estimator). We have

$$\mathbb{E}(\hat{\mu}) = \mu \in \mathbb{R}, \quad \mathbb{E}(\hat{\alpha}) = \alpha \in \mathbb{R}^a, \quad \text{and} \quad \mathbb{E}(\hat{\beta}) = \beta \in \mathbb{R}^b.$$

i.e., $\hat{\mu}$ is an unbiased estimator for μ , $\hat{\alpha} \in \mathbb{R}^a$ is an unbiased estimator for $\alpha \in \mathbb{R}^a$, and $\hat{\beta} \in \mathbb{R}^b$ is an unbiased estimator for $\beta \in \mathbb{R}^b$.

Proof. Recall that $\forall i \in \{1, \dots, a\}, \forall j \in \{1, \dots, b\}$, we have $y_{ij} \sim \mathcal{N}(\mu + \alpha_i + \beta_j, \sigma^2)$. So

$$\begin{aligned} \bar{y}_{i\cdot} &= \frac{1}{b} \sum_{j=1}^b y_{ij} \sim \mathcal{N}\left(\mu + \alpha_i, \frac{\sigma^2}{b}\right), \quad \forall i \in \{1, \dots, a\}, \\ \bar{y}_{\cdot j} &= \frac{1}{a} \sum_{i=1}^a y_{ij} \sim \mathcal{N}\left(\mu + \beta_j, \frac{\sigma^2}{a}\right), \quad \forall j \in \{1, \dots, b\}, \\ \bar{y}_{\cdot\cdot} &= \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{ab}\right). \end{aligned}$$

Now we can compute

$$\begin{aligned} \mathbb{E}(\hat{\mu}) &= \mathbb{E}\left(\frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij}\right) = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \mathbb{E}(y_{ij}) = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b (\mu + \alpha_i + \beta_j) \\ &= \mu + \frac{1}{a} \sum_{i=1}^a \alpha_i + \frac{1}{b} \sum_{j=1}^b \beta_j = \mu, \\ \mathbb{E}(\hat{\alpha}_i) &= \mathbb{E}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) = \mathbb{E}(\bar{y}_{i\cdot}) - \mathbb{E}(\bar{y}_{\cdot\cdot}) = \mu + \alpha_i - \mu = \alpha_i, \quad \forall i, \\ \mathbb{E}(\hat{\beta}_j) &= \mathbb{E}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) = \mathbb{E}(\bar{y}_{\cdot j}) - \mathbb{E}(\bar{y}_{\cdot\cdot}) = \mu + \beta_j - \mu = \beta_j, \quad \forall j. \end{aligned}$$

That is, $\mathbb{E}(\hat{\mu}) = \mu$, $\mathbb{E}(\hat{\alpha}) = \alpha$, and $\mathbb{E}(\hat{\beta}) = \beta$, as desired. \square

PROPOSITION 3.14 (Variance of the Mean Estimator).

3.2.2 Estimation of Variance

DEFINITION 3.15 (Sum of Squares). We define the following terms:

$$\begin{aligned} \text{SS}_{\text{trt}} &:= b \sum_{i=1}^a (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2, \\ \text{SS}_{\text{blk}} &:= a \sum_{j=1}^b (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2, \\ \text{SS}_{\text{err}} &:= \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2, \\ \text{SS}_{\text{tot}} &:= \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{\cdot\cdot})^2. \end{aligned}$$

PROPOSITION 3.16 (Decomposition of SS_{tot}). We have

$$\text{SS}_{\text{tot}} = \text{SS}_{\text{trt}} + \text{SS}_{\text{blk}} + \text{SS}_{\text{err}}.$$

DEFINITION 3.17 (Mean Squares). We define the following estimators for the variance σ^2 .

$$\begin{aligned} \text{MS}_{\text{trt}} &:= \text{SS}_{\text{trt}} / (a - 1), \\ \text{MS}_{\text{err}} &:= \text{SS}_{\text{err}} / ((a - 1)(b - 1)). \end{aligned}$$

PROPOSITION 3.18. We have

$$\mathbb{E}(\text{MS}_{\text{err}}) = \sigma^2.$$

i.e., MS_{err} is an unbiased estimator for σ^2 .

Proof. Recall that $\forall i \in \{1, \dots, a\}$, $\forall j \in \{1, \dots, b\}$, we have $y_{ij} \sim \mathcal{N}(\mu + \alpha_i + \beta_j, \sigma^2)$. So

$$\begin{aligned} \bar{y}_{i\cdot} &= \frac{1}{b} \sum_{j=1}^b y_{ij} \sim \mathcal{N}\left(\mu + \alpha_i, \frac{\sigma^2}{b}\right), \quad \forall i \in \{1, \dots, a\}, \\ \bar{y}_{\cdot j} &= \frac{1}{a} \sum_{i=1}^a y_{ij} \sim \mathcal{N}\left(\mu + \beta_j, \frac{\sigma^2}{a}\right), \quad \forall j \in \{1, \dots, b\}, \text{ and} \end{aligned}$$

$$\bar{y}_{..} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij} \sim \mathcal{N}(\mu, \frac{\sigma^2}{ab}).$$

So

$$\begin{aligned} \mathbb{E}(y_{ij}^2) &= \mathbb{V}(y_{ij}) + \mathbb{E}^2(y_{ij}) = \sigma^2 + (\mu + \alpha_i + \beta_j)^2, \quad \forall i, j, \\ \mathbb{E}(\bar{y}_{i.}^2) &= \mathbb{V}(\bar{y}_{i.}) + \mathbb{E}^2(\bar{y}_{i.}) = \frac{\sigma^2}{b} + (\mu + \alpha_i)^2, \quad \forall i, \\ \mathbb{E}(\bar{y}_{.j}^2) &= \mathbb{V}(\bar{y}_{.j}) + \mathbb{E}^2(\bar{y}_{.j}) = \frac{\sigma^2}{a} + (\mu + \beta_j)^2, \quad \forall j, \text{ and} \\ \mathbb{E}(\bar{y}_{..}^2) &= \mathbb{V}(\bar{y}_{..}) + \mathbb{E}^2(\bar{y}_{..}) = \frac{\sigma^2}{ab} + \mu^2. \end{aligned}$$

Now we can compute

$$\begin{aligned} \mathbb{E}(\text{MS}_{\text{err}}) &= \mathbb{E}(\text{SS}_{\text{err}} / ((a-1)(b-1))) = \mathbb{E}\left(\frac{1}{(a-1)(b-1)} \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2\right) \\ &= \frac{1}{(a-1)(b-1)} \mathbb{E}\left(\left[\sum_{i=1}^a \sum_{j=1}^b \begin{aligned} &+ y_{ij}^2 & - y_{ij}\bar{y}_{i.} & - y_{ij}\bar{y}_{.j} & + y_{ij}\bar{y}_{..} \\ &- \bar{y}_{i.}y_{ij} & + \bar{y}_{i.}^2 & + \bar{y}_{i.}\bar{y}_{.j} & - \bar{y}_{i.}\bar{y}_{..} \\ &- \bar{y}_{.j}y_{ij} & + \bar{y}_{.j}\bar{y}_{i.} & + \bar{y}_{.j}^2 & - \bar{y}_{.j}\bar{y}_{..} \\ &+ \bar{y}_{..}y_{ij} & - \bar{y}_{..}\bar{y}_{i.} & - \bar{y}_{..}\bar{y}_{.j} & + \bar{y}_{..}^2 \end{aligned} \right]\right) \\ &= \frac{1}{(a-1)(b-1)} \mathbb{E}\left(\begin{aligned} &+ \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 & - \sum_{i=1}^a \sum_{j=1}^b y_{ij}\bar{y}_{i.} & - \sum_{i=1}^a \sum_{j=1}^b y_{ij}\bar{y}_{.j} & + \sum_{i=1}^a \sum_{j=1}^b y_{ij}\bar{y}_{..} \\ &- \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{i.}y_{ij} & + \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{i.}^2 & + \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{i.}\bar{y}_{.j} & - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{i.}\bar{y}_{..} \\ &- \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j}y_{ij} & + \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j}\bar{y}_{i.} & + \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j}^2 & - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j}\bar{y}_{..} \\ &+ \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{..}y_{ij} & - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{..}\bar{y}_{i.} & - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{..}\bar{y}_{.j} & + \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{..}^2 \end{aligned}\right) \\ &= \frac{1}{(a-1)(b-1)} \mathbb{E}\left(\begin{aligned} &+ \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 & - b \sum_{i=1}^a \bar{y}_{i.}^2 & - a \sum_{j=1}^b \bar{y}_{.j}^2 & + ab\bar{y}_{..}^2 \\ &- b \sum_{i=1}^a \bar{y}_{i.}^2 & + b \sum_{i=1}^a \bar{y}_{i.}^2 & + ab\bar{y}_{..}^2 & - ab\bar{y}_{..}^2 \\ &- a \sum_{j=1}^b \bar{y}_{.j}^2 & + ab\bar{y}_{..}^2 & + a \sum_{j=1}^b \bar{y}_{.j}^2 & - ab\bar{y}_{..}^2 \\ &+ ab\bar{y}_{..}^2 & - ab\bar{y}_{..}^2 & - ab\bar{y}_{..}^2 & + ab\bar{y}_{..}^2 \end{aligned}\right) \\ &= \frac{1}{(a-1)(b-1)} \mathbb{E}\left(\sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - a \sum_{j=1}^b \bar{y}_{.j}^2 - b \sum_{i=1}^a \bar{y}_{i.}^2 + ab\bar{y}_{..}^2\right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{(a-1)(b-1)} \left[\sum_{i=1}^a \sum_{j=1}^b \mathbb{E}(y_{ij}^2) - a \sum_{j=1}^b \mathbb{E}(\bar{y}_{\cdot j}^2) - b \sum_{i=1}^a \mathbb{E}(\bar{y}_{i \cdot}^2) + ab \mathbb{E}(\bar{y}_{\cdot \cdot}^2) \right] \\
&= \frac{1}{(a-1)(b-1)} \left[\sum_{i=1}^a \sum_{j=1}^b (\sigma^2 + (\mu + \alpha_i + \beta_j)^2) + ab \left(\frac{\sigma^2}{ab} + \mu^2 \right) \right. \\
&\quad \left. - a \sum_{j=1}^b \left(\frac{\sigma^2}{a} + (\mu + \beta_j)^2 \right) - b \sum_{i=1}^a \left(\frac{\sigma^2}{b} + (\mu + \alpha_i)^2 \right) \right] \\
&= \frac{1}{(a-1)(b-1)} \left[(ab + 1 - a - b)\sigma^2 + (ab + ab - ab - ab)\mu^2 \right. \\
&\quad \left. + 0\mu + (b-b) \sum_{i=1}^a \alpha_i^2 + (a-a) \sum_{j=1}^b \beta_j^2 + \sum_{i=1}^a \sum_{j=1}^b \alpha_i \beta_j \right] \\
&= \frac{1}{(a-1)(b-1)} (ab + 1 - a - b)\sigma^2 = \sigma^2.
\end{aligned}$$

That is, $\mathbb{E}(\text{MS}_{\text{err}}) = \sigma^2$, as desired. \square

PROPOSITION 3.19. Under the assumption that $\alpha = \mathbf{0} \in \mathbb{R}^a$, we have

$$\mathbb{E}(\text{MS}_{\text{trt}}) = \sigma^2.$$

i.e., MS_{trt} is an unbiased estimator for σ^2 given that $\alpha = \mathbf{0}$.

Proof. Under the assumption that $\alpha = \mathbf{0} \in \mathbb{R}^a$, we have $\forall i \in \{1, \dots, a\}$, $\forall j \in \{1, \dots, b\}$, $y_{ij} \sim \mathcal{N}(\mu + \beta_j, \sigma^2)$. So

$$\begin{aligned}
\bar{y}_{i \cdot} &= \frac{1}{b} \sum_{j=1}^b y_{ij} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{b}\right), \quad \forall i \in \{1, \dots, a\}, \text{ and} \\
\bar{y}_{\cdot \cdot} &= \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{ab}\right).
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{E}(\bar{y}_{i \cdot}^2) &= \mathbb{V}(\bar{y}_{i \cdot}) + \mathbb{E}^2(\bar{y}_{i \cdot}) = \frac{\sigma^2}{b} + \mu^2, \quad \forall i, \text{ and} \\
\mathbb{E}(\bar{y}_{\cdot \cdot}^2) &= \mathbb{V}(\bar{y}_{\cdot \cdot}) + \mathbb{E}^2(\bar{y}_{\cdot \cdot}) = \frac{\sigma^2}{ab} + \mu^2.
\end{aligned}$$

Now we can compute

$$\mathbb{E}(\text{MS}_{\text{trt}}) = \mathbb{E}(\text{SS}_{\text{trt}}/(a-1)) = \mathbb{E}\left(\frac{b}{a-1} \sum_{i=1}^a (\bar{y}_{i \cdot} - \bar{y}_{\cdot \cdot})^2\right)$$

$$\begin{aligned}
&= \frac{b}{a-1} \mathbb{E} \left(\sum_{i=1}^a (\bar{y}_{i.}^2 + \bar{y}_{..}^2 - 2\bar{y}_{i.}\bar{y}_{..}) \right) = \frac{b}{a-1} \mathbb{E} \left(\sum_{i=1}^a \bar{y}_{i.}^2 + a\bar{y}_{..}^2 - 2a\bar{y}_{..}^2 \right) \\
&= \frac{b}{a-1} \mathbb{E} \left(\sum_{i=1}^a \bar{y}_{i.}^2 - a\bar{y}_{..}^2 \right) = \frac{b}{a-1} \left[\sum_{i=1}^a \mathbb{E}(\bar{y}_{i.}^2) - a\mathbb{E}(\bar{y}_{..}^2) \right] \\
&= \frac{b}{a-1} \left[\sum_{i=1}^a \left(\frac{\sigma^2}{b} + \mu^2 \right) - a \left(\frac{\sigma^2}{ab} + \mu^2 \right) \right] \\
&= \frac{b}{a-1} \left[\left(\frac{a}{b} - \frac{1}{b} \right) \sigma^2 + (a-a)\mu^2 \right] = \sigma^2.
\end{aligned}$$

That is, $\mathbb{E}(\text{MS}_{\text{trt}}) = \sigma^2$, as desired. \square

3.2.3 Hypothesis Testing

We are interested in testing the following hypothesis:

- $H_0 : \alpha = \mathbf{0} \in \mathbb{R}^a$ vs $H_1 : \alpha \neq \mathbf{0} \in \mathbb{R}^a$.

DEFINITION 3.20. We define the F -statistic as

$$F_0 := \frac{\text{MS}_{\text{trt}}}{\text{MS}_{\text{err}}} \sim \mathcal{F}(a-1, (a-1)(b-1)).$$

DEFINITION 3.21 (ANOVA Table).

Table 3.2: ANOVA Table for Randomized Block Design

	Sum of Squares	Degrees of Freedom	Mean Squares	F_0
Treatment	SS_{trt}	$a-1$	MS_{trt}	$\text{MS}_{\text{trt}}/\text{MS}_{\text{err}}$
Block	SS_{blk}	$b-1$	MS_{blk}	
Error	SS_{err}	$(a-1)(b-1)$	MS_{err}	
Total	SS_{tot}	$ab-1$		

3.3 Two-Way Factorial Design

DEFINITION 3.22. Let $a \in \mathbb{Z}_{++}$ denote the number of treatments of factor A . Let $b \in \mathbb{Z}_{++}$ denote the number of treatments of factor B . Let $n \in \mathbb{Z}_{++}$ denote the number of repetitions for each combination of treatments. Let $\mu \in \mathbb{R}$ denote the overall mean. We model the population as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \text{ for } i \in \{1, \dots, a\}, j \in \{1, \dots, b\}, k \in \{1, \dots, n\}$$

with constraints $\mathbf{1}^\top \alpha = 0$, $\mathbf{1}^\top \beta = 0$, $\gamma^\top \mathbf{1} = \mathbf{0}$, and $\gamma \mathbf{1} = \mathbf{0}$, and $e_{ijk} \sim \mathcal{N}(0, \sigma^2)$ are assumed to be independent. The total number of parameters in this model is $2 + a + b + ab$.

3.3.1 Estimation of Mean

3.3.2 Estimation of Variance

DEFINITION 3.23 (Sum of Squared Errors). We define the following terms:

$$\begin{aligned} \text{SS}_A &:= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 \\ \text{SS}_B &:= an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ \text{SS}_{AB} &:= n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \\ \text{SS}_{\text{err}} &:= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ijk} - \bar{y}_{ij.})^2 \\ \text{SS}_{\text{tot}} &:= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ijk} - \bar{y}_{...})^2. \end{aligned}$$

PROPOSITION 3.24 (Decomposition of SS_{tot}). We have

$$\text{SS}_{\text{tot}} = \text{SS}_A + \text{SS}_B + \text{SS}_{AB} + \text{SS}_{\text{err}}.$$

DEFINITION 3.25 (Variance Estimator). We define the following estimators for the variance σ^2 .

$$\begin{aligned} \text{MS}_A &:= \text{SS}_A / (a - 1), \\ \text{MS}_B &:= \text{SS}_B / (b - 1), \\ \text{MS}_{AB} &:= \text{SS}_{AB} / ((a - 1)(b - 1)), \\ \text{MS}_{\text{err}} &:= \text{SS}_{\text{err}} / (ab(n - 1)). \end{aligned}$$

PROPOSITION 3.26. We have

$$\mathbb{E}(\text{MS}_{\text{err}}) = \sigma^2.$$

i.e., MS_{err} is an unbiased estimator for σ^2 .

PROPOSITION 3.27. Under the assumption that $\alpha = \mathbf{0} \in \mathbb{R}^a$, we have

$$\mathbb{E}(\text{MS}_A) = \sigma^2.$$

i.e., MS_A is an unbiased estimator for σ^2 given that $\alpha = \mathbf{0}$.

PROPOSITION 3.28. Under the assumption that $\beta = \mathbf{0} \in \mathbb{R}^b$, we have

$$\mathbb{E}(\text{MS}_B) = \sigma^2.$$

i.e., MS_B is an unbiased estimator for σ^2 given that $\beta = \mathbf{0}$.

PROPOSITION 3.29. Under the assumption that $\gamma = \mathbf{0} \in \mathbb{R}^{a \times b}$, we have

$$\mathbb{E}(\text{MS}_{AB}) = \sigma^2.$$

i.e., MS_{AB} is an unbiased estimator for σ^2 given that $\gamma = \mathbf{0}$.

3.3.3 Hypothesis Testing

We are interested in testing the following hypothesis

- $H_0 : \alpha = \mathbf{0} \in \mathbb{R}^a$ vs $H_1 : \alpha \neq \mathbf{0}$.
- $H_0 : \beta = \mathbf{0} \in \mathbb{R}^b$ vs $H_1 : \beta \neq \mathbf{0}$.
- $H_0 : \gamma = \mathbf{0} \in \mathbb{R}^{a \times b}$ vs $H_1 : \gamma \neq \mathbf{0}$.

DEFINITION 3.30 (ANOVA Table).

Table 3.3: ANOVA Table for Two-Way Factorial Design

	Sum of Squares	Degrees of Freedom	Mean Squares	F_0
A	SS_A	$a - 1$	MS_A	MS_A/MS_{err}
B	SS_B	$b - 1$	MS_B	MS_B/MS_{err}
AB	SS_{AB}	$(a - 1)(b - 1)$	MS_{AB}	MS_{AB}/MS_{err}
Error	SS_{err}	$ab(n - 1)$	MS_{err}	
Total	SS_{tot}	$abn - 1$		

3.4 Two-Level Factorial Design

Chapter 4

Estimation Theory

4.1 Maximum Likelihood Estimation

DEFINITION 4.1 (Likelihood Function). Let Ω be a parameter space. Let θ be an unknown parameter in Ω . Let \mathbf{Y} be a random variable and \mathbf{y} be an observed sample. Let f denote the probability function of Y . We define the **likelihood function** for θ , denoted by \mathcal{L} , to be a function from Ω to \mathbb{R} given by

$$\mathcal{L}(\theta; \mathbf{y}) := f(\mathbf{y}; \theta).$$

DEFINITION 4.2 (Maximum Likelihood Estimate). Let Ω be a parameter space. Let θ be an unknown parameter in Ω . Let \mathbf{Y} be a random variable and \mathbf{y} be an observed sample. Let \mathcal{L} be the likelihood function for θ . We define the **maximum likelihood estimate** for θ , denoted by $\hat{\theta}$, to be a parameter given by

$$\hat{\theta} := \operatorname{argmax}_{\theta \in \Omega} \mathcal{L}(\theta; \mathbf{y}).$$

DEFINITION 4.3 (Relative Likelihood Function). Let Ω be a parameter space. Let θ be an unknown parameter in Ω . Let \mathbf{Y} be a random variable and \mathbf{y} be an observed sample. Let \mathcal{L} be the likelihood function for θ . We define the **relative likelihood**

function for θ , denoted by \mathcal{R} , to be a function from Ω to \mathbb{R} given by

$$\mathcal{R}(\theta; \mathbf{y}) := \frac{\mathcal{L}(\theta; \mathbf{y})}{\mathcal{L}(\hat{\theta}; \mathbf{y})}.$$

DEFINITION 4.4 (Log Likelihood Function). Let Ω be a parameter space. Let θ be an unknown parameter in Ω . Let \mathbf{Y} be a random variable and \mathbf{y} be an observed sample. Let \mathcal{L} be the likelihood function for θ . We define the **log likelihood function** for θ , denoted by ℓ , to be a function from Ω to \mathbb{R} given by

$$\ell(\theta; \mathbf{y}) := \ln \mathcal{L}(\theta; \mathbf{y}).$$

4.2 Examples

PROPOSITION 4.5 (MLE of Binomial Distribution). Let $n \in \mathbb{N}$. Let θ be an unknown parameter. Let y be an observed value from a $\text{Binomial}(n, \theta)$ distribution. Then the likelihood function for θ is

$$\mathcal{L}(\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}.$$

PROPOSITION 4.6 (MLE of Poisson Distribution). Let θ be an unknown parameter. Let y_1, \dots, y_n be an observed random sample from a $\text{Poisson}(\theta)$ distribution. Then the likelihood function for θ is

$$\mathcal{L}(\theta) = \theta^{n\bar{y}} e^{-n\theta} \text{ for } \theta \in \mathbb{R}_+,$$

and the maximum likelihood estimation of θ is

$$\hat{\theta} = \bar{y}.$$

PROPOSITION 4.7 (MLE of Exponential Distribution). Let θ be an unknown parameter. Let y_1, \dots, y_n be an observed random sample from a $\text{Exponential}(\frac{1}{\theta})$ distribution.

bution. Then the likelihood function for θ is

$$\mathcal{L}(\theta) = \theta^{-n} e^{-n\bar{y}/\theta},$$

and the maximum likelihood estimate $\hat{\theta}$ of θ is

$$\hat{\theta} = \bar{y}.$$

4.3 Confidence Intervals

DEFINITION 4.8. Let X be a random variable from a probability distribution with statistical parameter $\theta \in \Omega$. We define a **confidence interval** for the parameter θ with confidence level γ to be an interval with random endpoints $u(X)$ and $v(X)$ such that

$$\forall \theta \in \Omega, \quad \mathbb{P}_{\theta}(u(X) < \theta < v(X)) = \gamma.$$

4.4 Simple Random Sampling Without Replacement

4.4.1 Population Mean Estimator

Chapter 5

Hypothesis Testing

5.1 p -values

DEFINITION 5.1. Let H_0 be a null hypothesis. Let D be a discrepancy measure. Let \mathbf{y} be an observed value of the data. We define the p -**value** of the test of the hypothesis H_0 , using the discrepancy measure D , to be the probability given by

$$p\text{-value} := \mathbb{P}(D(\mathbf{Y}) \geq D(\mathbf{y}); H_0).$$

Chapter 6

Asymptotic Theory

In statistics: asymptotic theory, or large sample theory, is a framework for assessing properties of estimators and statistical tests. Within this framework, it is often assumed that the sample size n may grow indefinitely; the properties of estimators and tests are then evaluated under the limit of $n \rightarrow \infty$. In practice, a limit evaluation is considered to be approximately valid for large finite sample sizes too. (Wikipedia, 2021-07-11)

In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed. The LLN is important because it guarantees stable long-term results for the averages of some random events. For example, while a casino may lose money in a single spin of the roulette wheel, its earnings will tend towards a predictable percentage over a large number of spins. Any winning streak by a player will eventually be overcome by the parameters of the game. It is important to remember that the law only applies (as the name indicates) when a large number of observations is considered. There is no principle that a small number of observations will coincide with the expected value or that a streak of one value will immediately be “balanced” by the others (see the gambler’s fallacy). (Wikipedia, 2021-07-11)

6.1 Law of Large Numbers

THEOREM 6.1 (The Weak Law of Large Numbers (Khinchin's Law)). Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables. Suppose that they all have finite expected value and standard variation. Let μ denote their common expected value. Then the sample average converges weakly (in probability) to the common expected value:

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{weakly}} \mu.$$

Proof. Note that

$$\begin{aligned} \mathbb{E}[\bar{X}_n] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu \text{ and} \\ \text{var}[\bar{X}_n] &= \text{var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{var}[X_i] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

So by the Chebyshev's inequality, we get

$$\Pr \left[|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \frac{1}{n} \varepsilon \right] \leq \frac{n^2 \sigma^2}{\varepsilon^2}.$$

□

THEOREM 6.2 (The Strong Law of Large Numbers (Kolmogorov's Law)). Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables. Suppose that they all have finite expected value and standard variation. Let μ denote their common expected value. Then the sample average converges strongly (almost surely) to the common expected value:

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{strongly}} \mu.$$

6.2 Central Limit Theorem

THEOREM 6.3 (Central Limit Theorem). Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed Lebesgue integrable random variables. Let μ denote their

common expectation and σ denote their common standard deviation. Then we have

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. When MGF's exist:

$$\begin{aligned} M_n(t) &= \mathbb{E} \left[\exp \left\{ t \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right\} \right] \\ &= \mathbb{E} \left[\exp \left\{ \frac{t}{\sqrt{n}} \sum_{i=1}^n Y_i \right\} \right], \quad Y_i := \frac{X_i - \mu}{\sigma} \\ &= \prod_{i=1}^n \mathbb{E} \left[\exp \left\{ \frac{t Y_i}{\sqrt{n}} \right\} \right] \\ &= \left[M_Y \left(\frac{t}{\sqrt{n}} \right) \right]^n \\ &= \left[M_Y(0) + M_Y'(0) \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} M_Y''(0) \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\ &= \left[1 + \mathbb{E}[Y_i] \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} \mathbb{E}[Y_i^2] \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\ &= \left[1 + \mathbb{E}[Y_i] \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} (\text{var}[Y_i] + (\mathbb{E}[Y_i])^2) \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\ &= \left[1 + 0 \cdot \left(\frac{t}{\sqrt{n}} \right) + \frac{1}{2} (1 + 0) \left(\frac{t}{\sqrt{n}} \right)^2 + o \left(\frac{t}{\sqrt{n}} \right)^2 \right]^n \\ &= \left[1 + \frac{t^2/2}{n} + o \left(\frac{t^2}{n} \right) \right]^n \\ &\rightarrow \exp \left\{ \frac{t^2}{2} \right\}. \end{aligned}$$

□

Chapter 7

Power Transformation

DEFINITION 7.1 (Box-Cox Transformation). We define the one-parameter **Box-Cox transformation** of y , denoted by $T_\lambda(y)$, to be a function given by

$$T_\lambda(y) := \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \ln y, & \text{if } \lambda = 0. \end{cases}$$

PROPOSITION 7.2. The Box-Cox transformation is monotonic increasing.