整理（除了 Experimental methods）：

Data normalization and Statistical Analysis:

Ethics:

Group Settings and controls 与方法对应（参考 20Q4）：

Nuisance Variables/Biases:

- **对照组设置**：通过对比验证假设（如空白对照、阳性/阴性对照）。
- **重复与随机化**：
  - *重复（Replication）*：减少偶然误差，提高结果可靠性。
  - *随机化（Randomisation）*：避免系统性偏差（如样本分组随机）。
- **分层（Stratification）**：按特征分组（如性别、年龄）以提高组间均衡性。

How to Minimize Bias?

Key strategies highlighted in the PPT:

1. **Clearly define the study population**: Establish strict **inclusion/exclusion criteria** (e.g., excluding individuals with peanut allergies in a dietary study).
2. **Randomization & Stratification**: Use **random assignment** to ensure baseline characteristics (e.g., age, sex) are balanced between intervention and control groups. Apply **stratification** for critical variables (e.g., grouping participants by exercise experience before randomization).
3. **Blinding (Masking)**: **Single-blind**: Participants are unaware of their group assignment. **Double-blind**: Both participants and researchers are unaware of group assignments (e.g., using **coded labels** instead of explicit group names).
4. **Pilot testing 预实验**: Conduct small-scale preliminary studies to identify potential issues (e.g., participant dropout patterns or measurement inconsistencies).
5. **Standardized protocols**: Ensure consistency in data collection (e.g., measuring cortisol levels at fixed times of day to control for circadian rhythm effects).

CRD **(Completely Randomized Design)**：one-way ANOVA，比较处理间差异。

RCBD **(Randomized Complete Block Design)**：Two-way ANOVA（处理 + 区组），分离区组效应。Separates variability into **treatment effect**, **block effect** (e.g., soil differences, time batches), and random error.

Common Statistical Methods in Biological Experimental Design

1. **t-test**
   - **Purpose**: Compare means between two groups.
   - **Application**: Determine if a drug treatment significantly alters protein expression levels (e.g., treated vs. control groups).
   - **Types**: Independent (unpaired) t-test for distinct groups; Paired t-test for repeated measurements (e.g., pre- vs. post-treatment).
2. **Wilcoxon Signed-Rank Test**
   - **Purpose**: Non-parametric alternative to the paired t-test for non-normally

distributed data.
    - **Application**: Compare ranked data, such as tumor size reduction scores before and after therapy when data are skewed.
3. **Mann-Whitney U Test**
    - **Purpose**: Non-parametric alternative to the independent t-test.
    - **Application**: Compare gene expression levels between two species (e.g., wild-type vs. mutant) without assuming normality.
4. **ANOVA (Analysis of Variance)**
    - **Purpose**: Compare means across three or more groups.
    - **Application**: Test if nutrient levels (low, medium, high) differentially affect plant growth rates.
    - **Variants**: One-way ANOVA (single factor); Two-way ANOVA (multiple factors, e.g., diet and genotype interactions).
5. **Post-hoc Tests (e.g., Tukey's HSD, Bonferroni)**
    - **Purpose**: Identify specific group differences after ANOVA detects significance.
    - **Application**: Determine which drug doses (low, medium, high) differ significantly in reducing viral load.
6. **Kruskal-Wallis Test**
    - **Purpose**: Non-parametric alternative to ANOVA for non-normal data.
    - **Application**: Compare median survival times across multiple cancer subtypes.
7. **Chi-square Test**
    - **Purpose**: Assess associations between categorical variables.
    - **Application**: Test if genotype frequencies (AA, Aa, aa) deviate from Hardy-Weinberg equilibrium in a population.
8. **Fisher's Exact Test**
    - **Purpose**: S**mall-sample categorical data (alternative to Chi-square)**.
    - **Application**: Determine if a rare mutation is significantly associated with disease status in a case-control study.
9. **Pearson/Spearman Correlation**
    - **Purpose**: Measure **linear (Pearson 线性 参数)** or **monotonic (Spearman 指数等非线性单调 非参数，抗异常值强)** relationships between continuous variables.
    - **Application**: Evaluate if mRNA expression correlates with protein abundance (Pearson) or ranked metabolic activity (Spearman).

生物学研究实例

- **药物疗效试验**:
    - Log-rank 检验比较治疗组与安慰剂组生存差异 → 若显著，用 Cox 模型调整患者年龄、并发症等，确认药物独立效应。
- **癌症预后模型**:
    - Kaplan-Meier 展示不同分期的生存率 →Logistic 回归预测早期转移风险 → Cox 模型整合基因组数据优化预后分层。

场景选择指南

1. **需要比较组间生存率?**
    - 单变量：Log-rank 检验 + Kaplan-Meier 曲线。

- 多变量：Cox 模型。

2. **预测是否发生某事件（非时间相关）？**
   - 使用 Logistic 回归。
3. **描述生存趋势并可视化？**
   - 选择 Kaplan-Meier 分析。
4. **量化多个因素对生存风险的影响？**
   - 采用 Cox 模型，验证比例风险假设。

10. **Logistic Regression**

目的：预测二元事件（是/否）发生的概率，基于预测变量（连续或分类）。

适用数据：仅适用于静态结局，无法分析时间依赖性或删失数据（如患者失访）

- **结果变量**：二分类（如肿瘤复发 vs. 未复发）。
- **预测变量**：临床参数（年龄、性别）、基因表达量等。
  - **Purpose**: Model **binary outcomes** based on p**redictor variables.**
  - **Application**: Predict the likelihood of tumor recurrence based on genetic markers and clinical parameters.

11. **Kaplan-Meier Survival Analysis**

- **目的**：估计生存概率随时间变化的曲线，比较不同组间的生存差异。
  适用数据：仅描述生存分布，无法量化协变量（如年龄、治疗）对风险的直接影响。
- **时间-事件数据**（如生存时间、复发时间）。
- **删失数据**（如研究结束时患者仍存活或失访）。
  - **Purpose**: Estimate survival probabilities over time.
  - **Application**: Compare survival curves **between patients receiving different therapies.**

12. Log-rank Test

Purpose: Compare survival distributions between two or more groups.

Experimental Design:

- **Example**: Testing if a new drug (Group A) extends survival compared to standard therapy (Group B) in a randomized clinical trial.
- **Steps**:
  1. Define the event (e.g., death, relapse) and follow-up time.
  2. Record event times and censoring status (e.g., patients lost to follow-up).
  3. Use the log-rank test to calculate a p-value for differences in survival curves.

13. **Cox Proportional Hazards Model (week 4)**

目的：分析时间-事件数据中多个协变量对事件风险的动态影响。

适用数据：包含时间、事件状态（发生/删失）及协变量（如年龄、治疗分组）。

关键特点：

- **半参数模型**：不指定基准风险函数，仅假设协变量效应为乘法关系（比例风险）。
- **风险比（HR）**：解释变量对事件风险的瞬时影响（HR=1.5 表示某因素使风险增加 50%）。
  - **Purpose**: Analyze time-to-event data **with covariates.**
  - **Application**: Assess how age, gender, and treatment influence relapse risk in a cohort.

14. **Multiple Testing Correction (e.g., Benjamini-Hochberg FDR)**
    - **Purpose**: Control false positives in high-throughput experiments (e.g., genomics).
    - **Application**: Adjust p-values when testing thousands of genes for differential expression.
15. **Repeated Measures ANOVA**
    - **Purpose**: Compare groups with longitudinal/repeated observations.
    - **Application**: Analyze blood pressure changes in the same subjects over multiple time points.
16. **Linear Regression**
    - **Purpose**: Model relationships between dependent and independent variables.
    - **Application**: Predict enzyme activity based on substrate concentration and pH level.

Key Considerations:

- **Parametric vs. Non-parametric**: Use t-test/ANOVA for normally distributed data; opt for Wilcoxon/Kruskal-Wallis for skewed distributions.
- **Data Type**: Match methods to data (continuous: t-test/ANOVA; categorical: Chi-square/Fisher's).
- **Experimental Design**: Account for repeated measures, multiple factors, or hierarchical structures (e.g., nested ANOVA for batch effects).

统计方法选择

- **参数检验（如 t 检验）**：要求数据正态分布；
- **非参数检验（如 Mann-Whitney U, Wilcoxon）**：适用于非正态数据。
- **多重检验校正**（如 Bonferroni）：避免假阳性。

一、参数情况（数据符合正态分布）

1. One-way ANOVA 的 Post-hoc 检验

- **方法**：
    - **Tukey HSD**：控制整体误差率，适用于所有组间两两比较。
    - **Bonferroni adjusted**：通过调整显著性水平（如 α/比较次数）减少假阳性。
    - **Scheffé 检验**：灵活处理任意复杂比较（如组 A vs 组 B+组 C 的平均值）。
- **目的**：仅针对单一因素的多个水平进行均值差异比较（如比较三种药物的效果）。

2. Two-way ANOVA 的 Post-hoc 检验

- **方法**：
    - **Main Effects 主效应分析**：若某个因素（如药物类型）的主效应显著，需像 One-way ANOVA 一样比较其水平（如药物 A vs 药物 B vs 对照）。
    - **Interaction Effects 交互作用分析**: 若交互作用显著(如药物×剂量), analyze **simple effects**：
        1. **按一个因素分层**(如固定剂量水平), 在每层内进行 One-way ANOVA

的 post-hoc 比较（如比较药物 A vs 药物 B 在低剂量下的差异）。
    2. **多因素组合比较**（如药物 A 高剂量 vs 药物 B 低剂量），使用 **adjusted tests**（e.g., Sidak, Bonferroni）。
- **复杂性**：需同时处理主效应和交互作用，导致更多比较次数，需更严格的多重校正（如分阶段控制误差率）。

二、非参数情况（数据不符合正态分布）

1. One-way ANOVA 的非参数替代与 Post-hoc 检验

- **替代方法**：
    - **Kruskal-Wallis 检验**（non-parametric One-way ANOVA）。
- **Post-hoc 检验**：
    - **Dunn 检验**：控制整体误差率，基于秩次进行两两比较，需 Bonferroni 或 Holm 校正。
    - **Conover-Iman 检验**：允许更多灵活的比较，但需多重校正。

2. Two-way ANOVA 的非参数替代与 Post-hoc 检验

- **替代方法**：
    - **Friedman 检验**（适用于重复测量或区组设计，非参数版 Two-way ANOVA）。
    - **Aligned Ranks Test** 或 **Quade 检验**：处理更一般的双因素设计。
- **Post-hoc 检验**：
    - **主效应分析**：若因素主效应显著，使用非参数两两比较（如 Wilcoxon test + Bonferron adjusted）。
    - **交互作用分析**：需分解数据：
        1. **按一个因素分层**（如固定时间点），在每层内进行 Kruskal-Wallis 或 Wilcoxon 检验。
        2. **交互组合比较**（如治疗组早晨 vs 对照组晚上），使用非参数检验并严格校正（如 Holm-Sidak 方法）。
- **挑战**：非参数方法对交互作用的分解更复杂，且 statistical power 通常低于参数方法。

1. 动物实验（Animal Experiments）

Ethics Statement

All animal experiments were conducted in compliance with the guidelines established by the **Institutional Animal Care and Use Committee (IACUC)**. Procedures were designed to minimize animal suffering and distress, including the use of humane endpoints, appropriate anesthesia (e.g., isoflurane inhalation), and analgesia (e.g., buprenorphine for post-operative pain). Animal housing conditions (temperature: 22±1°C, 12h light/dark cycle) and enrichment (nesting materials, social grouping) were maintained to ensure welfare. The principles of the **3Rs (Replacement, Reduction, Refinement) were strictly followed**.

---

2. 人类实验（Human Studies）

Ethics Statement

All human studies were approved by **the Institutional Review Board (IRB)** (Protocol No. XXX-2023) and conducted in accordance with the Declaration of Helsinki. **Written informed consent was obtained from all participants prior to enrollment**. For minors or vulnerable populations, consent was provided by legal guardians with additional assent from participants

aged ⩾12 years. Personal data were anonymized and **stored securely in compliance with GDPR (General Data Protection Regulation)** and local **privacy laws**. Risks (e.g., blood sampling) were minimized through standardized protocols, and **participants were free to withdraw at any stage** without penalty.

---

3. 细胞系实验（Cell Line Experiments）

Ethics Statement

All cell lines used in this study were obtained from commercial repositories (e.g., ATCC) or collaborators under material transfer agreements (MTAs). The origin and ethical sourcing of cell lines were verified:

- **Commercial cell lines**: Certified by suppliers with documentation confirming compliance with ethical standards (e.g., no use of embryonic tissues without explicit consent).
- **Patient-derived cell lines**: Originally collected under IRB-approved protocols (No. XXX-2010) with donor consent for research use.
  Cell line identities were authenticated via STR profiling, and mycoplasma contamination was routinely tested (e.g., PCR-based assays).

---

4. 人原代细胞实验（Human Primary Cell Experiments）

Ethics Statement

Human primary cells (e.g., peripheral blood mononuclear cells, PBMCs) were isolated from donor samples collected under IRB-approved protocols (No. XXX-2022). Written informed consent was obtained from all donors, specifying the scope of research use and anonymization of data. Tissue samples (e.g., surgical waste) were acquired through collaborations with certified biobanks, adhering to the Human Tissue Act (HTA) and local regulations. Experimental procedures excluded commercial exploitation of donor materials, and all data were de-identified to protect privacy.

5. 人类医疗数据收集（如年龄、病史等）

Ethics Statement

All procedures involving the collection of human medical data (e.g., age, diagnosis, treatment history) were reviewed and approved by the **Institutional Review Board (IRB)** (Protocol No. XXX-2023). **Written informed consent** was obtained from all participants, explicitly stating the purpose of data collection, potential risks (e.g., privacy breaches), and data usage scope (e.g., academic research only). **Personal identifiers (e.g., names, ID numbers) were removed**, and data were pseudonymized using unique codes. **Access** to **raw data was restricted to authorized researcher**s, and all records were stored in encrypted databases compliant with GDPR and HIPAA regulations.

核心要求

- 知情同意（明确数据用途与风险）
- 匿名化/假名化处理
- 数据加密与访问权限控制

---

6. 人类影像学数据（如 MRI、CT、X 光）

Ethics Statement

The acquisition and analysis of human imaging data (e.g., MRI, CT scans) were conducted under IRB approval (No. XXX-2024) with explicit participant consent for image storage and secondary research use. To **prevent re-identification**, all metadata (e.g., acquisition date, hospital codes) were removed, and **facial features in head scans were anonymized** using defacing tools (e.g., *Quickshear*). Images were **stored in a secure PACS (Picture Archiving and Communication System)** with access logs audited quarterly. For public sharing, images were converted to non-DICOM formats (e.g., NIfTI) and verified against the *BIDS* (Brain Imaging Data Structure) standard to ensure privacy compliance.

核心要求

- 影像匿名化（去面部特征、去元数据）
- 安全存储系统（PACS/BIDS 标准）
- 二次使用需重新确认伦理许可

---

7. 公共数据使用（如开放数据库、政府公开数据）

Ethics Statement

Public datasets (e.g., UK Biobank, NHANES) used in this study were accessed under data use agreements (DUA No. XXX-2023) and **complied with** the **original ethical approvals granted to the data providers**. All data were **de-identified prior to public release**, as confirmed by the source institutions. Researchers adhered to the terms of **use (e.g., no attempt to re-identify individuals, non-commercial purposes) and cited data sources according to FAIR principles (Findable, Accessible, Interoperable, Reusable**). Local IRB exemption (No. XXX-2023) was obtained for secondary analysis of anonymized public data.

核心要求

- 遵守原始数据提供方的伦理条款
- 禁止再识别（Re-identification）尝试
- 引用与合规声明