Contents

1 Lab 2: Observational and Interventional studies

Fill the blanks in the following table:

Term	Description
Descriptional study	
Analytical study	
Observational study	
Interventional study	
Retrospective study	
Prospective study	
Ecological study	
Ecological fallacy	
Prevalence	
Incidence	
Cumulative incidence rate	
Incidence rate	
Case report	
Case series	
Cross-sectional study	
Ecological study	
Case-control study	
Cohort study	
Bias	
Selection bias	
Recall bias	
Randomized controlled trial	
Randomized clinical trial	A study where participants receive or or more treatments to answer questions a
Parallel design	A trial design where each subject is assigned to either experimental treatment c
Cossover design	A trial design where every subject serves as his/her own control
Factorial design	
Randomization	
Simple randomization	
Complete randomization	

Term	Description
Efron's biased coin randomization	
Wei's urn randomization	
Cluster randomization	
Stratified randomization	
Minimization randomization	A dynamic randomization strategy in a clinical trial to balance the assignment
Superiority design	A trial design whose aim is to show that the efficacy of the experimental treatm
Equivalence design	A trial design whose aim is to show that the results of experimental and control
Non-inferiority design	A trial whose objective is to validate that the results of a treatment are not mu
Blinding	A strategy in a clinical trial where one or more parties involved (e.g., participan
Open-label	A strategy in a clinical trial where all parties know the treatment assignment.
Single-blinded	A blinding strategy in a clinical trial where only the participants have no idea c
Double blinded	A blinding strategy in a clinical trial where neither participants nor clinicians k
Triple blinded	A blinding strategy in a clinical trial where participants, clinicians and statistici
PROBE	Prospective, randomized, open-label
Protocol	A document describing the objectives, design, statistical consideration etc. of a
Intention-to-treat analysis	An analysis that is analyzing every randomized subject as assigned to their rand
Per-protocol analysis	Choose only the participants who perfectly follow the protocol (excluding the p
As-treatment analysis	Analyzing the results based on the participants real treatment.
t-test	A statistical testing method to compare the means for two groups with normal
Paired t-test	A statistical testing method for comparing the means of two matched paired gr
Wilcoxon rank-sum test	A nonparametric testing method for comparing two medians of two independen
Wilcoxon signed rank test	A nonparametric testing method for comparing two medians of two correated g
Chi-squared test	
Fisher's exact test	
McNemar's test	
Analysis of variance (ANOVA)	
Analysis of covariance (ANCOVA)	
Repeated measures ANOVA	
Friedman's test	
Cochrane's Q test	
Missing data imputation	
Dropouts	
Fixed-value imputation	An imputation strategy that substitutes each missing or dropout value with a fi
Multiple imputation	
Last observation carried forward (LOCF)	One fixed-value imputation strategy by filling the missing data with the last no

1.1 Exercises

1. Use R to draw the density curve of $X \sim N(0, 5)$ and $Y \sim N(2, 5)$, and mark the type I error $\alpha = 0.05$ and the corresponding Type II error β.
2. A randomized clinical trial is designed to evaluate the efficacy of a newly developed drug to reduce pain in patients after joint replacement surgery by comparing with the standard care. 100 patients were assigned to receive either the new drug or the standard care. The primary outcome was a reduction of 3 or 3+ scale points (clinically meaningful reduction). The data are summarized in the following table:

Treatment | n | #patients with 3+ reduction | proportion |
— | — | — | — |
New drug | 50 | 23 | 0.46 |
Stanard care | 50 | 11 | 0.22 | - How would you analyze the data for this superiority design? Write down the R code.

3. A small randomized clinical trial was conducted to test whether treatment A (new drug) was effective in lowering DBP as compared to B (standard) and to describe changes in DBP across times at which it was measured (`DBP.dat`).

- Are the baseline (DBP1) and the potential confounding factors balanced in these two groups? How to analyze? Write down the R code and also the results and conclusion.
- Is the treatment A more effective in lowering the DBP than B? Use the parametric methods, adjusted by the confounders.
- Can you analyze the data using the nonparametric methods?
- Or permutation-based method? Or even bootstrap-based methods?

4. Bioequivalence, crossover clinical trial In this exercise we will use a dataset from a bioequivalence clinical trial described in Chow and Liu (2009). The trial utilized a standard two-sequence (i.e., 1=RT and 2=TR), two-period, two-formulation (T=Test; R=Reference) (i.e., $2 \times 2 \times 2$) crossover design to compare two oral formulations of a drug, and was conducted with 26 healthy volunteers (subjects). Subjects were randomized to either five 50mg tablets (T) or 5 mL of a suspension (R) at the first period baseline, and then crossed over to the alternative formulation at the second period baseline. And the bioavailability outcome is the area under the concentration-by-time curve (AUC) over the interval from 0 to 48 hours. The data file is `ChowLiu2009data.csv`.

- Write an R code to compute the area under the concentration-by-time curve (AUC):

$$AUC = \sum_{\tau=1}^{k} \frac{(c_\tau + c_{\tau-1}) \times (t_\tau - t_{\tau-1})}{2}$$

  where t_τ is the τ-th time point of blood sample collection and c_τ is the τ-th blood or plasma concentration and $\tau = 0, 1, 2, \ldots, k$.

- Test for the carryover effect
- Compute the subject totals across two periods:

$$U_{ik} = Y_{i1k} + Y_{i2k}$$

  where

  - $k = 1, 2$: the sequence
  - $i = 1, \ldots, n_k$: the subject in each sequence k

- Y_{ijk}: the AUC for subject i in sequence k and period j.

- Calculate the sample mean across all the subjects in each sequence:

$$\overline{U_{*k}} = \frac{1}{n_k}\sum_{i=1}^{n_k} U_{ik}, k = 1, 2$$

- Compute the differential carryover effect C:

$$\hat{C} = \overline{U_{*2}} - \overline{U_{*1}}$$

- $\hat{C}$ is normally distributed with mean C and variance:

$$\widehat{Var}(\hat{C}) = \hat{\sigma_u^2}(\frac{1}{n_1} + \frac{1}{n_2})$$

and

$$\hat{\sigma_u^2} = \frac{1}{n_1 + n_2 - 2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(U_{ik} - \overline{U_{*k}})^2$$

- Compute the statistic:

$$T = \frac{\hat{C}}{\sqrt{\widehat{Var}(\hat{C})}} \sim t(n_1 + n_2 - 2)$$

- Compute the p-value, and draw the conclusion.

- Test for direct formulation effect:

- Compute the difference in periods for each subject within each sequence:

$$d_{ik} = \frac{1}{2}(Y_{i2k} - Y_{i1k}), i = 1, \ldots, n_k; k = 1, 2$$

- Compute the sample means for the period differences for each sequence:

$$\overline{d_{*k}} = \frac{1}{n_k}\sum_{i=1}^{n_k} d_{ik}$$

- Compute the direct differential formulation effect:

$$\hat{F} = \overline{d_{*1}} - \overline{d_{*2}}$$

- If no carryover effect, $\hat{F} \sim N(F, \widehat{Var}(\hat{F}))$, where

  - (1) $\widehat{Var}(\hat{F}) = \hat{\sigma}_d^2(\frac{1}{n_1} + \frac{1}{n_2})$
  - (2) $\hat{\sigma}_d^2 = \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(d_{ik} - \overline{d}_{*k})^2$

- Similarly, compute the t-statistic:

$$T_F = \frac{\hat{F}}{\sqrt{\widehat{Var}(\hat{F})}}$$

- Compute the p-value and reach the conclusion.

- Analysis of variance (ANOVA) "‘ Data <- data.frame(subj = as.factor(dat$subj$), $formu$ = $as.factor(dat$formulation), seq = as.factor(datseq), $prd = as.factor(dat$prd), auc = dat$auc)

summary(aov(auc ~ seq*formu + Error(subj), data=Data)) "‘

- Two one-sided t-test

FDA has specified a decision criterion for concluding bioequivalence of a test formulation (T) to a reference formulation (R): T is bioequivalent to R if the 90%CI on the ratio of the mean of T to the mean of R is between 80% and 125% for bioequivalent outcome AUC. * Compute the mean AUC for each formulation * Determine the decision CI (θ_L, θ_R) for the difference in means calculated using the mean of the reference formulation (R). * Use two one-sided t-test to validate the bioequivalence of the two formulations.

$$T_L = \frac{\overline{Y}_T - \overline{Y}_R - \theta_L}{\sqrt{\hat{\sigma}_d^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$T_U = \frac{\overline{Y}_T - \overline{Y}_R - \theta_U}{\sqrt{\hat{\sigma}_d^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$