Lab 2: RCT with Two Endpoints

Grading: Turn in your first attempt at the tasks for a binary "fair attempt or not" grade on Canvas. That is, your first attempt need not be neat or correct.

Done early? Come discuss your work with me and I'll give you one suggestion on how to improve what you've got in the remaining time. Iterate if necessary.

Scientific context

You are a mathematical statistician at the FDA tasked with determining the FDA policy on approving experimental interventions on the basis of multiple outcomes.

For concreteness, you and your clinical colleagues consider a case study of an RCT designed to evaluate a new treatment for cardiovascular disease where both LDL cholesterol and systolic blood pressure are measured for each subject, with the analysis plan of reporting the point estimate, p-value, and 95% confidence interval from a two-sample t-test with unequal variances.

For computational efficiency, in today's lab, don't bother computing or saving the 95% confidence intervals.

Task 1

Imagine that your clinical colleagues tell you that the treatment should only be considered effective if the mean systolic blood pressure **and** the mean LDL cholesterol is higher in the treatment group.

Recall the "naive" approval strategy discussed in class: approve the treatment if the p-value for total cholesterol is ≤ 0.05 and the p-value for LDL cholesterol is ≤ 0.05.

Simulate to investigate:

- The probability of approving under this strategy when the treatment is ineffective

- The probability of approving under this strategy when the treatment is ineffective

Summarize the results from your investigation.

Task 2

Imagine that your clinical colleagues tell you that the treatment should only be considered effective if the mean systolic blood pressure **or** the mean LDL cholesterol is higher in the treatment group.

Recall the "naive" approval strategy discussed in class: approve the treatment if the p-value for total cholesterol is ≤ 0.05 or the p-value for LDL cholesterol is ≤ 0.05.

Simulate to investigate the probability of approving under this strategy when the treatment is ineffective. Summarize the results from your investigation.

Code

Data generation

The following function generates hypothetical data from a single RCT where the population distribution of blood pressure and cholesterol is multivariate normal. In the function below, `rho` is the correlation between the two cholesterol measurements and `sigma` is a vector of length 2 containing the standard deviation for blood pressure and cholesterol, respectively.

```r
generate_trial_data <- function(n, mu_treat, mu_control, sig, rho) {
    Sigma <- rbind(c(sig[1]^2, sig[1]*sig[2]*rho), c(sig[1]*sig[2]*rho, sig[2]^2))

    treat_data <- t(t(matrix(rnorm(n*2, 0, 1), n, 2, byrow=TRUE)%*%chol(Sigma)) + mu_trea

    control_data <- t(t(matrix(rnorm(n*2, 0, 1), n, 2, byrow=TRUE)%*%chol(Sigma)) + mu_co

    data <- data.frame(sbp = c(treat_data[, 1], control_data[, 1]),
                                    ldl = c(treat_data[, 2], control_data[, 2]),
                                    group = c(rep("T", n), rep("C", n)))

    data
}

n <- 50
mu_control <- mu_treat <- c(140, 160)
sig <- sqrt(c(40, 25))
rho <- 0

set.seed(123)
ex_data <- generate_trial_data(n, mu_treat, mu_control, sig, rho)
head(ex_data)
```

```
      sbp      ldl group
1 136.4552 158.8491     T
2 149.8581 160.3525     T
3 140.8177 168.5753     T
4 142.9151 153.6747     T
5 135.6560 157.7717     T
6 147.7418 161.7991     T
```

Data analysis

The following function takes in data of the format generated by `generate_trial_data()`, tests for no difference in blood pressure, tests for no difference in cholesterol, and returns the p-values, the point estimates for the differences in means, and 95% confidence intervals for the differences in means.

```
analyze_trial_data <- function(data) {

    sbp_results <- t.test(data[data$group == "T", "sbp"],
                          data[data$group == "C", "sbp"],
                          alternative = "two.sided")

    ldl_results <- t.test(data[data$group == "T", "ldl"],
                          data[data$group == "C", "ldl"],
                          alternative = "two.sided")

    output <- c(sbp_results$estimate,
                          sbp_results$p.value,
                          ldl_results$estimate,
                          ldl_results$p.value)

    names(output) <- c("sbp_treat_mean", "sbp_control_mean", "sbp_pval",
                          "ldl_treat_mean", "ldl_control_mean", "ldl_pval")

    output
}

(ex_analysis <- analyze_trial_data(ex_data))
```

```
  sbp_treat_mean sbp_control_mean          sbp_pval   ldl_treat_mean
     140.6988408      139.3746075         0.2465396      160.3515769
ldl_control_mean         ldl_pval
     159.4189482        0.3483850
```