Models for Repeated Measures

MA482/BE482 – Biostatistics

Summary: When the same response is collected multiple times on a single subject, the assumption of independence is often violated. The additional correlation we have introduced into the problem should be appropriately modeled. Understanding these models can lead to more powerfully designed studies.

1 Importance of Study Design

Summary: Study design is too often separated from the statistical analysis that follows. However, in addition to informing the conclusions we draw regarding the data, the study design helps in choosing an appropriate analysis to address the question of interest.

Example 1. *(Fecal Fat)* Lack of digestive enzymes in the intestine can cause bowel absorption problems. This will be indicated by excess fat in the feces. Pancreatic enzyme supplements can be given to ameliorate the problem. A study was conducted to determine if the form of the supplement makes a difference on the resulting fecal fat (g/day) present. The data are given below:

No Pill	44.5	33.0	19.1	9.4	71.3	51.2
Tablet	7.3	21.0	5.0	4.6	23.3	38.0

Is there evidence that receiving the supplement as a tablet results in a different amount of fecal fat produced, on average?

```r
# Construct Data
#  We can use a data.frame command to manually enter data when needed. While I
#  prefer having data stored as a csv file, this is sometimes useful.
#
#  Each column is separated by a comma, and a vector of data is read in for each
#  column, which is named prior to an equal sign. Here, we read all the data
#  into one column, with a second column containing the group. As we have seen,
#  it is preferred to organize the data this way as opposed to a different
#  column for each response.
fat1 <- data.frame(Fat = c(44.5, 33.0, 19.1, 9.4, 71.3, 51.2,
                           7.3, 21.0, 5.0, 4.6, 23.3, 38.0),
                   PillType = rep(c("No Pill", "Tablet"), each=6))



# Two-Sample T-test
#  One approach is to consider a two-sample t-test, allowing the variances to
#  differ between the two groups.
TwoSample(Fat ~ PillType, data=fat1, plots=FALSE)

## Two Sample Procedure using t-Distribution:
##   Difference in Means: -21.55
##   95% CI: (-46.078, 2.978)
##   Alternative Hypothesis: mu1 Not Equal mu2
##   P-Value: 0.0774
##   Degrees of Freedom: 8.127143
```

```r
# Regression Approach
#  This assumes the variance is the same in each of the two groups. We see
#  that the assumptions for this model are violated.
fit1.fat1 <- lm(Fat ~ PillType, data=fat1)
tidy(fit1.fat1)

##             term  estimate std.error statistic     p.value
## 1    (Intercept)  38.08333  7.541819  5.049622 0.000499524
```

```
## 2 PillTypeTablet -21.55000 10.665742 -2.020488 0.070922531
```

Example 2. *(Example 1 Cont.)* For the Fecal Fat example, suppose you are given the additional information that a cross-over study was conducted with a substantial wash-out period. The resulting data is now updated to include the subject identifier:

	Subject					
Pill Type	1	2	3	4	5	6
No Pill	44.5	33.0	19.1	9.4	71.3	51.2
Tablet	7.3	21.0	5.0	4.6	23.3	38.0

Given this additional information, how do your analysis and conclusions change, if at all?

```
# Paired Test
#  Add subject identifier
fat1$SubjectID <- c(1, 2, 3, 4, 5, 6)
PairedSample(Fat ~ PillType | SubjectID, data=fat1, plots=FALSE)

## Paired Sample Procedure using t-Distribution:
##   Mean Difference: -21.55
##   95% CI: (-39.369, -3.731)
##   Alternative Hypothesis: mu1 Not Equal mu2
##   P-Value: 0.0266
##   Degrees of Freedom: 5
```

1.1 Terminology

This paired t-test fits into a much larger framework for which we now introduce some terminology.

Definition 1. *(Repeated Measures)* Data for which the observed responses can be _____ based on some nuisance variable (typically the subject itself).

- Observations _____ a group tend to be more alike than observations _____ groups, leading to a violation of the independence assumption.

Definition 2. *(Correlation Structure)* This defines the direction and strength of the relationship between the observed _____.

- Ignoring the correlation structure does not tend to affect the _____, but can greatly affect the resulting _____ _____.