

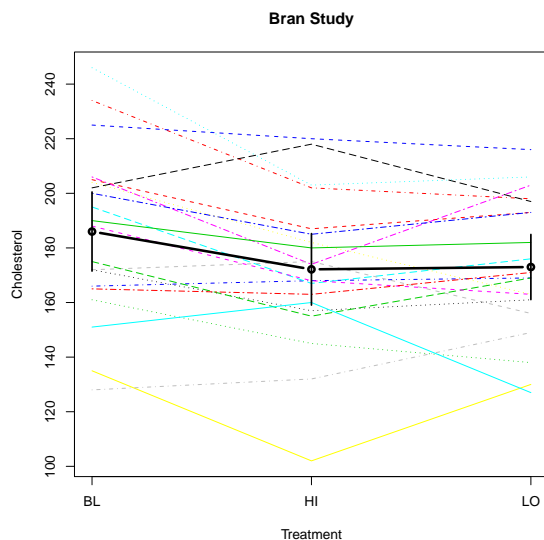
2/23/2010

## 36-402/608 ADA-II Breakout #12 Results

H. Seltman

The bran data (case1602.csv) are from 20 subjects studied in a randomized, double-blind diet manipulation trial. For each subject, cholesterol was measured 1) during a baseline period while the subject was on their normal diet, 2) at the end of the first 6 week diet trial, and 3) at the end of the second six week diet trial. (There was a return to normal diet for two weeks between the trials.) The two experimental diets were low-fiber (from wheat) and high-fiber (from oat bran). The order of the diets was randomly assigned. The scientific questions were whether or not the high fiber diet lowered cholesterol relative to baseline and whether the effect might be just due to reduced fat calories rather than the high fiber content of the oat bran.

```
# Raw bran data and means
plot(c(1,3), lim, type="n", xlab="Treatment", ylab="Cholesterol",
     main="Bran Study", axes=FALSE)
box() # bounding box for graph
axis(2) # normal y-axis
axis(1, at=1:3, labels=c("BL","HI","LO")) # custom x-axis
# individual data as lines:
for (i in 1:nrow(bran)) lines(1:3, bran[i,1:3], col=i+1, lty=i+1)
# Group means and SEs:
means = sapply(bran[,1:3],mean)
lines(1:3, means, lwd=3, type="b")
SE = sqrt(sapply(bran[,1:3],var)/nrow(bran)) # (allow non-constant variance)
multiplier = qt(0.975, nrow(bran)-1)
segments(1:3, means-SE*multiplier, 1:3, means+SE*multiplier, lwd=2)
```



**Question 1: What type of study is this? What should we test? How does the graphics code work?**

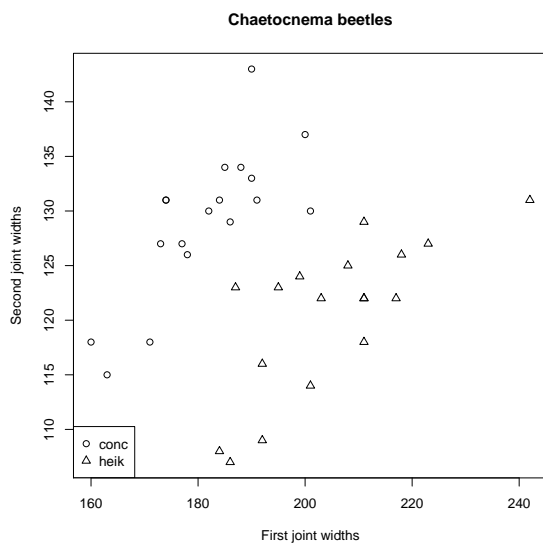
This is a crossover study because each experimental unit (person) is exposed to each treatment condition. As is typical, when possible, the order of treatments is randomized (counterbalanced) in an attempt to cancel out “carry-over” effects.

The questions asked suggest simply comparing the mean of baseline to hi-fiber and the mean of hi-fiber to lo-fiber. Both can handle the repeated measures aspect (correlation within subject) by using a t-test. Combining information from both questions could be done in a way that takes the three-way correlation into account.

The code uses the general `plot()` option `axes=FALSE` which inhibits the production of the four axes (bounding box) and the x and y tick marks and their labels. The `box()` function adds the bounding box. The `axis()` function draws any of the four axis (numbered counterclockwise starting with 1=x-axis) with appropriate default tick marks and labels or with custom ones using `at=` to set the locations of the tick marks, and `labels=` to set the tick labels.

Also, `segments()` has arguments, `x0`, `y0`, `x1`, `y1` and each set of 4 values represents one line segment.

The flea beetle data (`ex1613.csv`) consists of measurements of the first and second joint widths of the feet of flea beetles from two different species. These measurements are intended to distinguish between the two species more clearly than either individual measurement.



**Question 2: What type of study is this? How might you use these data to classify future beetles?**

This is a study with a multivariate (hence, probably correlated within beetle) outcome. It is neither longitudinal nor crossover. The main observation is that, focusing on either one measurement there is plenty of overlap between

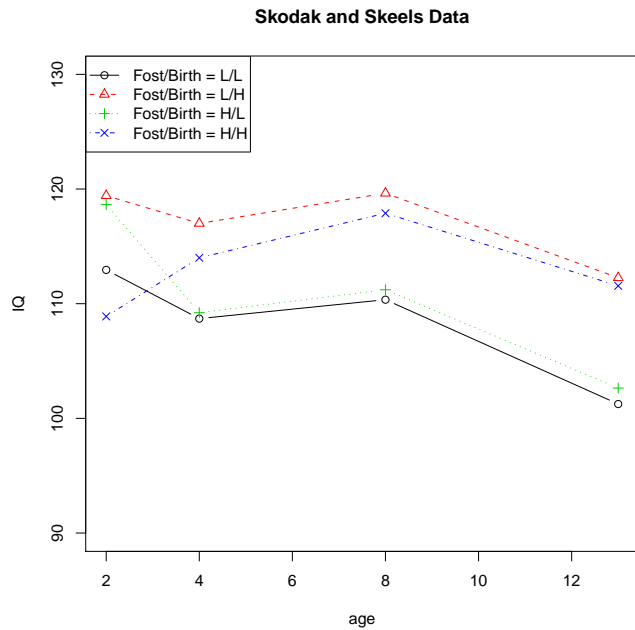
species, but as a correlated multivariate measurement, we can imagine tilted ellipses that completely separate the species.

The Skodak and Skeels data (ex1605.csv) looks at IQ in adopted children at ages 2, 4, 8 and 13. The foster mothers education (fmed) and the birth mothers IQ (tmiq) are also measured for each child.

	fmed	tmiq	age2iq	age4iq	age8iq	age13iq
1	10	100	120	115	109	106
2	10	71	131	109	113	95
3	14	89	126	115	113	90
...						
60	10	90	122	127	129	126
61	15	104	108	124	116	113
62	8	88	122	112	119	97

**Question 3: What type of study is this? What EDA would you produce? Consider what simplified data might be most appropriate, then choose what testing you might do to find out the effects of the two mothers on the child's IQ.**

This is a longitudinal study with two one-per-subject covariates. One possible EDA plot is to examine the mean change in IQ for, say, for groups defined by the median splits of the two maternal covariates:



Subject matter concerns suggest that the birth mother might have the greatest effect on early IQ and the foster mother on later IQ, so separate regressions with a single outcome per subject might be appropriate. More sophisticated repeated measures models could best detect the effects of mothers IQ/education on the pattern of change of IQ in the children.