Independent Observations; with One Mean/Variance

Notation

Four ways of writing this model; variance-covariance matrix; identity matrix; i.i.d.

Estimation

Estimates and their notation (hats, tildes); using them to do inference (sd vs. se)es

Fitting in R

Could get mean and standard error using cast;

```
> sm <- read.delim("http://rem.ph.ucla.edu/rob/mld/data/tabdelimiteddata/smallmice.txt
> cast(day ~ ., value = "weight", data = sm, fun.aggregate = function(x) c(m = mean(x)
+     sd = sd(x), se = sd(x)/sqrt(length(x))))
```

```
  day        m        sd        se
1   2 206.2857  29.77203  7.956909
2   5 376.9286  47.67432 12.741498
3   8 545.1429  58.31461 15.585234
4  11 684.2857 102.53512 27.403663
5  14 801.7143 133.73132 35.741199
6  17 864.4286 137.24078 36.679143
7  20 945.2857 120.66765 32.249786
```

Could also get using `lm`.

```
> m2 <- lm(weight ~ 1, data = sm, subset = day == 2)
> summary(m2)
```

```
Call:
lm(formula = weight ~ 1, data = sm, subset = day == 2)

Residuals:
    Min      1Q  Median      3Q     Max
-65.286 -15.536  -0.786  12.464  71.714

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  206.286      7.957   25.93 1.41e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 29.77 on 13 degrees of freedom
```

```
> coef(m2)
```

```
(Intercept)
   206.2857
```

```
> vcov(m2)
```

```
            (Intercept)
(Intercept)     63.3124
```

```
> sqrt(diag(vcov(m2)))
```

```
(Intercept)
   7.956909
```

Repeated Observations; Mean/Variance Changes Over Time

Notation

Six ways of writing this model; writing the variance matrix; relationships between variance matrix and correlation matrix; what's i.i.d.

Estimation

The many parameters; how we estimate (in the balanced case) and use to do inference

Fitting in R

(This will NOT be the usual way of doing this...)

Using lm with multivariate data

```
> smw <- cast(id ~ day, value = "weight", data = sm)
> smw <- as.matrix(smw[, -1])
> colnames(smw) <- paste("day", c(2, 5, 8, 1, 14, 17, 20), sep = "")
> mm <- lm(smw ~ 1)
> coef(mm)

                day2     day5     day8     day1    day14    day17    day20
(Intercept) 206.2857 376.9286 545.1429 684.2857 801.7143 864.4286 945.2857

> vm <- vcov(mm)
> rownames(vm) <- colnames(vm) <- colnames(smw)
> sqrt(diag(vm))

     day2      day5      day8      day1     day14     day17     day20
 7.956909 12.741498 15.585234 27.403663 35.741199 36.679143 32.249786
```

Can compare the correlation using lm with correlation

```
> round(vm/sqrt(diag(vm)[row(vm)] * diag(vm)[col(vm)]), 2)

      day2 day5 day8 day1 day14 day17 day20
day2  1.00 0.92 0.57 0.36  0.23  0.23  0.38
day5  0.92 1.00 0.77 0.54  0.45  0.41  0.55
day8  0.57 0.77 1.00 0.86  0.80  0.76  0.81
day1  0.36 0.54 0.86 1.00  0.93  0.92  0.87
day14 0.23 0.45 0.80 0.93  1.00  0.96  0.89
day17 0.23 0.41 0.76 0.92  0.96  1.00  0.92
day20 0.38 0.55 0.81 0.87  0.89  0.92  1.00

> round(cor(smw), 2)

      day2 day5 day8 day1 day14 day17 day20
day2  1.00 0.92 0.57 0.36  0.23  0.23  0.38
day5  0.92 1.00 0.77 0.54  0.45  0.41  0.55
day8  0.57 0.77 1.00 0.86  0.80  0.76  0.81
day1  0.36 0.54 0.86 1.00  0.93  0.92  0.87
day14 0.23 0.45 0.80 0.93  1.00  0.96  0.89
day17 0.23 0.41 0.76 0.92  0.96  1.00  0.92
day20 0.38 0.55 0.81 0.87  0.89  0.92  1.00
```

Three Problems

1) Unbalanced or missing data The estimates shown here only work when the data is perfectly balanced; in cases where data is unbalanced or missing, there is not a closed form solution, but we can use iterative algorithms to get maximum likelihood estimates.

2) Too many covariance parameters It can be computationally difficult to fit that many covariance parameters; additionally, using data to fit them can reduce the precision of the mean estimates. The covariance matrix can be parameterized so that fewer parameters are needed.

How many parameters, AR(1), compound symmetry, different variances...

3) No covariates We'll add terms to our model to allow the mean to be different for each subject, depending on the values of covariates for that subject.

Notation with independent observations; using sum, vector, matrix

Notation with repeated measures; using sum, vector, matrix