We've introduced the multivariate normal model,

$$Y = X\alpha + \delta, \text{ where } \delta \sim N(0, \Sigma)$$

and (briefly) the version including random effects,

$$Y = X\alpha + X\beta + \delta,$$

which also induces correlation between the responses.

There are two big questions we'll be discussing for most of the rest of the semester.

- How do we make inferences about Y and X (the fixed effects)?

- What correlation structure Σ best fits the data?

The primary tool we will use to is **likelihood**. In particular, we'll use the likelihood ratio test and various information criteria; they both involve the idea of likelihood.

Likelihood

The likelihood of a set of population parameters is the probability of getting the sample data that we got, assuming that set of population parameters really is the truth.

Example: We are interested in the probability that a Hershey's kiss will lands on its base. Might it be 0.5? We flip ten kisses and three land on the base. We'll compute the likelihood of 0.5. The likelihood is

$$\binom{10}{3}(0.5)^3(0.5)^7 = 0.117.$$

Likelihood can be hard to interpret for just one set of parameters; instead it's used to compare several possible sets of parameters.

What's the likelihood of 0.4?

$$\binom{10}{3}(0.5)^3(0.5)^7 = 0.215.$$

It's more likely than 0.5, but perhaps we can do better.

For a given probability p, the likelihood is

$$\binom{10}{3} p^3 p^7$$

What value of p maximizes this?

We call this value the **Maximum Likelihood Estimate**, or, the **MLE**.

An aside: why use least squares for regression? .

.

Pediatric Pain Data   Here's a model for the pediatric pain data, with no fixed effect and a random intercept for each subject.

```
> pp <- read.delim("http://rem.ph.ucla.edu/rob/mld/data/tabdelimiteddata/pain.txt")
> pp <- subset(pp, !is.na(l2paintol))
> m3 <- lme(l2paintol ~ cs, random = ~1 | id, data = pp, method = "ML")
```

Here are the maximum likelihood estimates of the parameters:

```
> summary(m3)

...
Random effects:
 Formula: ~1 | id
        (Intercept)  Residual
StdDev:   0.8966126 0.5980688

Fixed effects: l2paintol ~ cs
                 Value Std.Error  DF   t-value p-value
(Intercept)  4.535189 0.1687056 181 26.882274  0.0000
csdistracter 0.455631 0.2383229  62  1.911821  0.0605
...
```

And here's the log likelihood.

```
> logLik(m3)

'log Lik.' -293.6357 (df=4)
```

Comparing Likelihoods

We usually have several models in mind, where we use different parameters, and want to compare these models.

For example, in the Hershey's kiss example, we've looked at two models.

1. The probability it lands on its base is exactly 0.5 (the parameter is fixed).

2. The probability it lands on its base is something other than 0.5 (the parameter will be estimated).

The first model has no parameters to be estimated; the second has one.

Adding parameters to a model will *always* increase the likelihood. Why?
When two models are related in that way they are called "nested."

So, why not add in as many parameters as we can??

More pediatric pain models

Are these nested? Which is the best, taking into account only likelihood?

No predictors

```
> m1 <- lme(l2paintol ~ 1, random = ~1 | id, data = pp, method = "ML")
> logLik(m1)

'log Lik.' -295.4209 (df=3)
```

With cs

```
> m2 <- lme(l2paintol ~ cs, random = ~1 | id, data = pp, method = "ML")
> logLik(m2)

'log Lik.' -293.6357 (df=4)
```

With trial

```
> m3 <- lme(l2paintol ~ factor(trial), random = ~1 | id, data = pp,
+     method = "ML")
> logLik(m3)

'log Lik.' -293.89 (df=6)
```

With cs and trial

```
> m4 <- lme(l2paintol ~ cs + factor(trial), random = ~1 | id, data = pp,
+     method = "ML")
> logLik(m4)

'log Lik.' -292.0698 (df=7)
```

With cs and trial and interaction

```
> m5 <- lme(l2paintol ~ cs * factor(trial), random = ~1 | id, data = pp,
+     method = "ML")
> logLik(m5)

'log Lik.' -291.6393 (df=10)
```

AR(1) model with cs

```
> m6 <- gls(l2paintol ~ cs, correlation = corAR1(form = ~1 | id),
+     data = pp, method = "ML")
> logLik(m6)

'log Lik.' -301.5021 (df=4)
```

The Likelihood Ratio Test

If two models are *nested*, the value

$$-2\log\left(\frac{\text{Likelihood of smaller model}}{\text{Likelihood of larger model}}\right)$$

is distributed like a χ^2 distribution, with degrees of freedom equal to the change in the number of parameters between the two models.

Test two models for the Hershey's kiss .

Test some models for the pediatric pain data

```
> anova(m1, m5)

   Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m1     1  3 596.8418 607.3456 -295.4209
m5     2 10 603.2785 638.2911 -291.6393 1 vs 2 7.563334  0.3727


> anova(m4, m5)

   Model df      AIC      BIC    logLik   Test   L.Ratio p-value
m4     1  7 598.1395 622.6483 -292.0697
m5     2 10 603.2785 638.2911 -291.6393 1 vs 2 0.8609908  0.8348
```

Information Criteria

```
> AIC(m1, m2, m3, m4, m5, m6)

    df       AIC
m1   3 596.8418
m2   4 595.2714
m3   6 599.7799
m4   7 598.1395
m5  10 603.2785
m6   4 611.0042
```