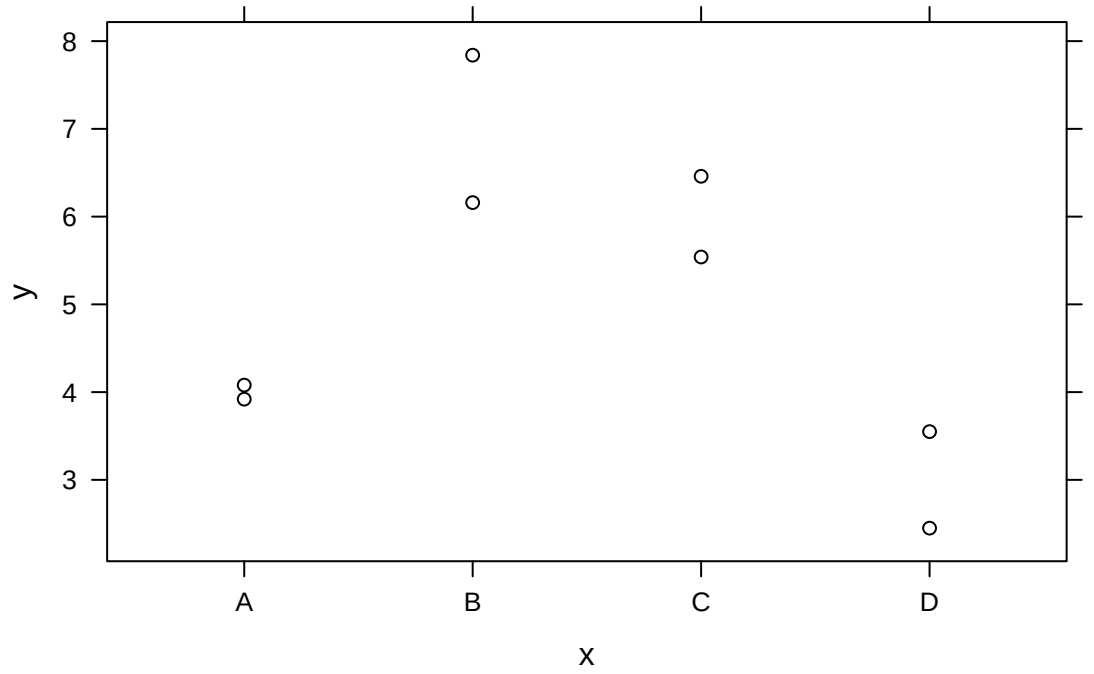
Some data I made up

```
  x    y
1 A 3.92
2 A 4.08
3 B 7.84
4 B 6.16
5 C 5.54
6 C 6.46
7 D 2.45
8 D 3.55
```



We're going to fit an ANOVA model to this data. How might we set up our multivariate normal model?

$$Y = X\alpha + \delta, \quad \text{where } \delta \sim N(0, \sigma^2) \text{ i.i.d}$$

Here are four possible ways of defining X and α: they all give the same predicted values!
It's important to know which one the computer used when interpreting the result.

For each, we'll discuss...

- What the columns of X mean?
- How do we interpret the corresponding α parameters?
- Do they all give us the same predicted values?

A "simple" parameterization

```
  xA xB xC xD
1  1  0  0  0
2  1  0  0  0
3  0  1  0  0
4  0  1  0  0
5  0  0  1  0
6  0  0  1  0
7  0  0  0  1
8  0  0  0  1
```

```
   Estimate Std. Error t value Pr(>|t|)
xA   4.0000     0.5536   7.225 0.001947
xB   7.0000     0.5536  12.643 0.000225
xC   6.0000     0.5536  10.837 0.000411
xD   3.0000     0.5536   5.419 0.005622
```

"first level baseline" style parameterization (the R default)

```
  (Intercept) xB xC xD
1           1  0  0  0
2           1  0  0  0
3           1  1  0  0
4           1  1  0  0
5           1  0  1  0
6           1  0  1  0
7           1  0  0  1
8           1  0  0  1
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.0000     0.5536   7.225  0.00195
xB            3.0000     0.7830   3.832  0.01859
xC            2.0000     0.7830   2.554  0.06301
xD           -1.0000     0.7830  -1.277  0.27064
```

"last level baseline" style parameterization (the R default)

```
  (Intercept) x1 x2 x3
1           1  1  0  0
2           1  1  0  0
3           1  0  1  0
4           1  0  1  0
5           1  0  0  1
6           1  0  0  1
7           1  0  0  0
8           1  0  0  0

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0000     0.5536   5.419  0.00562
x1            1.0000     0.7830   1.277  0.27064
x2            4.0000     0.7830   5.109  0.00694
x3            3.0000     0.7830   3.832  0.01859
```

"Sum" style parameterization

```
  (Intercept) x1 x2 x3
1           1  1  0  0
2           1  1  0  0
3           1  0  1  0
4           1  0  1  0
5           1  0  0  1
6           1  0  0  1
7           1 -1 -1 -1
8           1 -1 -1 -1

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0000     0.2768  18.062 5.52e-05
x1           -1.0000     0.4795  -2.086   0.1054
x2            2.0000     0.4795   4.171   0.0140
x3            1.0000     0.4795   2.086   0.1054
```

"Helmert" style parameterization

This contrasts the second level with the first, the third with the average of the first two, and so on.

```
  (Intercept) x1 x2 x3
1           1 -1 -1 -1
2           1 -1 -1 -1
3           1  1 -1 -1
4           1  1 -1 -1
5           1  0  2 -1
6           1  0  2 -1
7           1  0  0  3
8           1  0  0  3

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0000     0.2768  18.062 5.52e-05
x1            1.5000     0.3915   3.832   0.0186
x2            0.1667     0.2260   0.737   0.5018
x3           -0.6667     0.1598  -4.171   0.0140
```

Polynomial style parameterization

This assumes the predictor is ordered and looks for linear, quadratic, cubic, etc. effects.

```
  (Intercept)        x.L  x.Q        x.C
1           1 -0.6708204  0.5 -0.2236068
2           1 -0.6708204  0.5 -0.2236068
3           1 -0.2236068 -0.5  0.6708204
4           1 -0.2236068 -0.5  0.6708204
5           1  0.2236068 -0.5 -0.6708204
6           1  0.2236068 -0.5 -0.6708204
7           1  0.6708204  0.5  0.2236068
8           1  0.6708204  0.5  0.2236068

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0000     0.2768  18.062 5.52e-05
x.L          -0.8944     0.5536  -1.616  0.18150
x.Q          -3.0000     0.5536  -5.419  0.00562
x.C           0.4472     0.5536   0.808  0.46452
```

What if these four levels were really two, crossed?

```
  x    y x1 x2
1 A 3.92  A  C
2 A 4.08  A  C
3 B 7.84  B  C
4 B 6.16  B  C
5 C 5.54  A  D
6 C 6.46  A  D
7 D 2.45  B  D
8 D 3.55  B  D
```

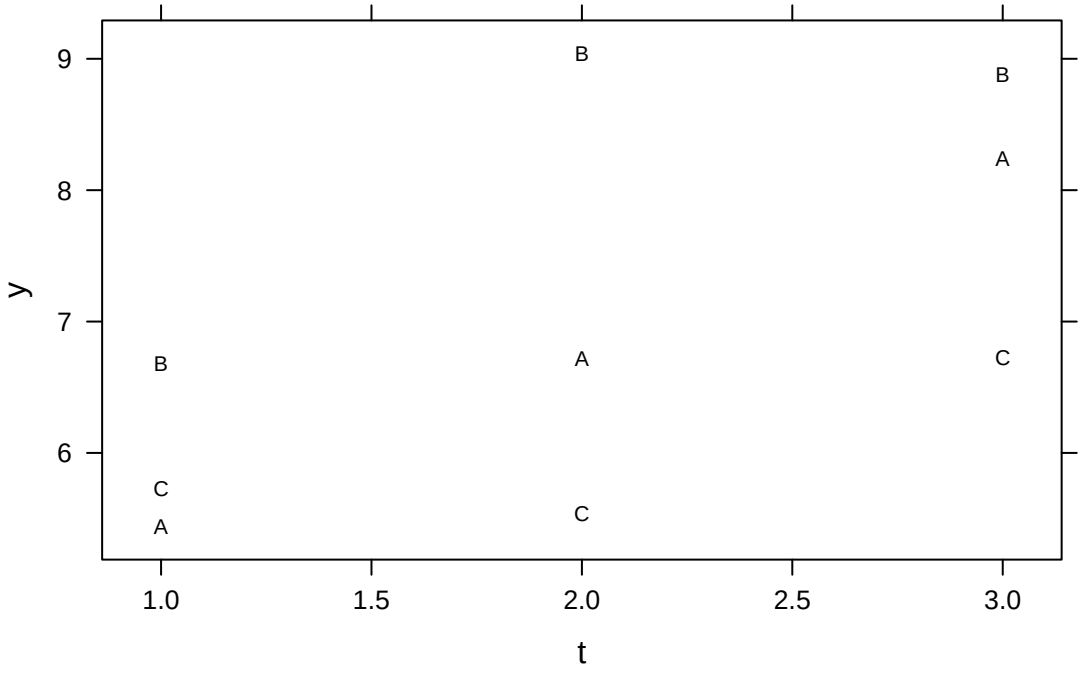
The first level as baseline parameterization: `R` default

```
  (Intercept) x1B x2D x1B:x2D
1           1   0   0       0
2           1   0   0       0
3           1   1   0       0
4           1   1   0       0
5           1   0   1       0
6           1   0   1       0
7           1   1   1       1
8           1   1   1       1

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.0000     0.5536   7.225  0.00195
x1B           3.0000     0.7830   3.832  0.01859
x2D           2.0000     0.7830   2.554  0.06301
x1B:x2D      -6.0000     1.1073  -5.419  0.00562
```

What if we have a continuous variable as well as a categorical one, and we want to fit separate slopes and intercepts?

```
  x t    y
1 A 1 5.44
2 A 2 6.72
3 A 3 8.24
4 B 1 6.68
5 B 2 9.04
6 B 3 8.88
7 C 1 5.73
8 C 2 5.54
9 C 3 6.73
```



A simple parameterization

```
     Estimate Std. Error t value Pr(>|t|)
xA     4.0000     1.0380   3.853   0.0309
xB     6.0000     1.0380   5.780   0.0103
xC     5.0000     1.0380   4.817   0.0170
xA:t   1.4000     0.4805   2.914   0.0618
xB:t   1.1000     0.4805   2.289   0.1060
xC:t   0.5000     0.4805   1.041   0.3746
```

```
  xA xB xC xA:t xB:t xC:t
1  1  0  0    1    0    0
2  1  0  0    2    0    0
3  1  0  0    3    0    0
4  0  1  0    0    1    0
5  0  1  0    0    2    0
6  0  1  0    0    3    0
7  0  0  1    0    0    1
8  0  0  1    0    0    2
9  0  0  1    0    0    3
```

Using baseline for intercept only

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.0000     1.0380   3.853   0.0309
xB            2.0000     1.4680   1.362   0.2663
xC            1.0000     1.4680   0.681   0.5446
xA:t          1.4000     0.4805   2.914   0.0618
xB:t          1.1000     0.4805   2.289   0.1060
xC:t          0.5000     0.4805   1.041   0.3746
```

```
  (Intercept) xB xC xA:t xB:t xC:t
1           1  0  0    1    0    0
2           1  0  0    2    0    0
3           1  0  0    3    0    0
4           1  1  0    0    1    0
5           1  1  0    0    2    0
6           1  1  0    0    3    0
7           1  0  1    0    0    1
8           1  0  1    0    0    2
9           1  0  1    0    0    3
```

Using baseline for intercept and slope

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.0000     1.0380   3.853   0.0309
xB            2.0000     1.4680   1.362   0.2663
xC            1.0000     1.4680   0.681   0.5446
t             1.4000     0.4805   2.914   0.0618
xB:t         -0.3000     0.6796  -0.441   0.6888
xC:t         -0.9000     0.6796  -1.324   0.2772
```

```
  (Intercept) xB xC t xB:t xC:t
1           1  0  0 1    0    0
2           1  0  0 2    0    0
3           1  0  0 3    0    0
4           1  1  0 1    1    0
5           1  1  0 2    2    0
6           1  1  0 3    3    0
7           1  0  1 1    0    1
8           1  0  1 2    0    2
9           1  0  1 3    0    3
```