One-Way Analysis of Variance using Regression

Recall Example: A group of 32 rats were randomly assigned to each of 4 diets labelled (A,B,C,and D). The response is the liver weight as a percentage of body weight. Two rats escaped and another died, resulting in the following data

A	B	C	D
3.42	3.17	3.34	3.65
3.96	3.63	3.72	3.93
3.87	3.38	3.81	3.77
4.19	3.47	3.66	4.18
3.58	3.39	3.55	4.21
3.76	3.41	3.51	3.88
3.84	3.55		3.96
	3.44		3.91

Anova Model:

$$X_{ij} = \mu_i + e_{ij}$$

Assumptions e_{ij} are independent $N(0, \sigma^2)$, $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, n_i$

Indicator Variables

Define $a - 1$ indicator variables to identify the a groups

Index the subjects using a single index i, $i = 1, \ldots, n$

Let

$X_{i1} = 1$ if subject i is in group 1, and 0 otherwise.

$X_{i2} = 1$ if subject i is in group 2, and 0 otherwise.

$X_{i3} = 1$ if subject i is in group 3, and 0 otherwise.

...

$X_{i,a-1} = 1$ if subject i is in group $a - 1$, and 0 otherwise.

Regression Model :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{a-1} X_{i,a-1} + \epsilon_i$$

Table of Means

Group	ANOVA mean	Regression mean
1	μ_1	$\beta_0 + \beta_1$
2	μ_2	$\beta_0 + \beta_2$
	...	
$a - 1$	μ_{a-1}	$\beta_0 + \beta_{a-1}$
a	μ_a	β_0

Equivalent ANOVA and regression hypotheses.

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_a$$

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{a-1}$$

```
MTB > oneway c1 c2
One-way ANOVA: C1 versus C2

Source  DF      SS      MS      F      P
C2       3  1.1649  0.3883  10.84  0.000
Error   25  0.8954  0.0358
Total   28  2.0603

S = 0.1893   R-Sq = 56.54%   R-Sq(adj) = 51.32%

MTB > indicator c2 c3-c6
MTB > print c2 c3-c6
Data Display

Row  C2  C3  C4  C5  C6
  1   1   1   0   0   0
  2   1   1   0   0   0
  3   1   1   0   0   0
...
  7   1   1   0   0   0
  8   2   0   1   0   0
  9   2   0   1   0   0
 10   2   0   1   0   0
...
 18   3   0   0   1   0
 19   3   0   0   1   0
 20   3   0   0   1   0
...
 28   4   0   0   0   1
 29   4   0   0   0   1

MTB > regress c1 3 c3-c5

Regression Analysis: C1 versus C3, C4, C5

The regression equation is
C1 = 3.94 - 0.133 C3 - 0.506 C4 - 0.338 C5

Predictor      Coef  SE Coef      T      P
Constant    3.93625  0.06691  58.83  0.000
C3         -0.13339  0.09795  -1.36  0.185
C4         -0.50625  0.09463  -5.35  0.000
C5          -0.3379   0.1022  -3.31  0.003

S = 0.189253   R-Sq = 56.5%   R-Sq(adj) = 51.3%

Analysis of Variance

Source          DF       SS       MS      F      P
Regression       3  1.16490  0.38830  10.84  0.000
Residual Error  25  0.89541  0.03582
Total           28  2.06032
```