Notes on contrasts for linear trend:

Let y_i = mean for i_{th} level of a factor (i=1,…p), with p = # levels in factor. Assume levels are coded as $x_i = i$. So, the y's represent the p coefficients that would be estimated in a regression model (without an intercept) including the p conventional indicator variables for factor levels.

Slope for the linear relationship between y & x is derived as the solution to the least squares estimator for β:

$$\sum_{i=1}^{p}(y_i - \beta x_i)^2$$

The estimate is obtained by setting the 1st derivative of the least squares estimator to zero and solving for β:

$$0 = -2\sum_{i=1}^{p}(x_i - \bar{x})[(y_i - \bar{y}) - \beta(x_i - \bar{x})]$$

$$\beta = \frac{\sum_{i=1}^{p}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{p}(x_i - \bar{x})^2}$$

The last equation equals zero when the numerator=0 (i.e. this motivates the contrast for the y_i's which represent the level-specific means, or, equivalently the coefficients of interest). Note that if the y's were all equal (i.e. under the null hypothesis of no trend), the values (x_i - $\bar{x}$) would have to sum to zero for β =0 to hold.

Example: x = (1, 2, 3, 4, 5). From the numerator of the expression for β above, the coefficient for the i^{th} value of x is: (x_i - $\bar{x}$).

For example, for x_1 = 1:

$$(1 - \frac{(1+2+3+4+5}{4})) = (1-3) = -2$$

For x_4 = 4:

$$(4 - \frac{(1+2+3+4+5}{4})) = (4-3) = 1$$

So, the complete contrast coefficients are (-2, -1, 0, 1, 2).

For predictors with even numbers of categories the same approach works, but the coefficients are scaled to integer values by multiplying by 2. In the case of four levels (1, 2, 3, 4) here is the value for the 4th coefficient

$$(4 - \frac{(1+2+3+4)}{4}) = (4 - 2.5) = 1.5$$

All four are given by (-1.5, -0.5, 0.5, 1.5). Scaling by a factor of 2 gives: (-3, -1, 1, 3). (Note that using the contrast based on the unscaled coefficients gives the same result.)