Correlation Examples

What we did in Class that was quite incorrect.

$$r_{xy} = \frac{\text{cov}}{\textit{total}\ \text{var}} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}}$$

$$= \frac{5376 - \frac{(76)(710)}{13}}{\sqrt{546 - \frac{(76)^2}{13}}\sqrt{50058 - \frac{(710)^2}{13}}} = \frac{5376 - \frac{5375}{13}}{\sqrt{546 - \frac{5776}{13}}\sqrt{50058 - \frac{504100}{13}}} =$$

$$= \frac{5376 - 4150.7692}{\sqrt{546 - 444.3077}\sqrt{50058 - 38776.9231}} = \frac{1224.2308}{\sqrt{101.6923}\sqrt{11281.0769}} =$$

$$= \frac{1224.2308}{(10.0842)(106.2124)} = \frac{1224.2308}{1071.0671} = 1.1430(\textit{wrong})$$

What we should have done:
Our mistake was using 5376 for the sum of X times Y.
It should have been 5015

Thus

$$r_{xy} = \frac{\text{cov}}{\textit{total}\ \text{var}} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}}$$

$$= \frac{5015 - \frac{(76)(710)}{13}}{\sqrt{546 - \frac{(76)^2}{13}}\sqrt{50058 - \frac{(710)^2}{13}}} = \frac{5015 - \frac{5375}{13}}{\sqrt{546 - \frac{5776}{13}}\sqrt{50058 - \frac{504100}{13}}} =$$

$$= \frac{5015 - 4150.7692}{\sqrt{546 - 444.3077}\sqrt{50058 - 38776.9231}} = \frac{864.2308}{\sqrt{101.6923}\sqrt{11281.0769}} =$$

$$= \frac{864.2308}{(10.0842)(106.2124)} = \frac{864.2308}{1071.0671} = .8069$$

To test the significance of this correlation we need to use Table D (p 335 of WEC)
The df for pearson's *r* is –2.
In this case we have 13-2 = 11 degrees of freedom.
The *r* critical at the *p* .05, .02, and .01 levels (two tailed) are .553, 634, and .684, respectively.
Thus we can reject the null hypothesis (Ho : $r = 0$), fail to reject the alternative hypothesis (Ha: *r* not equal to 0) and conclude that the correlation between hotness of the sauce and the distance that 3rd graders spew the sauce is significant at least at the .01 level.

In APA format this would be reported as

$r(11)= .81, p < .01$.

Further we can conclude that 65.38% of the variance in Spewing distance is accounted for by the hotness of the sauce. ($r^2= .6538$).
Also, we can conclude that 34.62 % of the variance in Spewing distance is unaccounted for and likely due to other factors (e.g. spew experience, developmental differences, experience with hot food, etc).
We know this by finding $1-r^2$.
This is referred to as residual variance or error variance.
In this case $1-r^2 = 1 - .6538 = .3462$