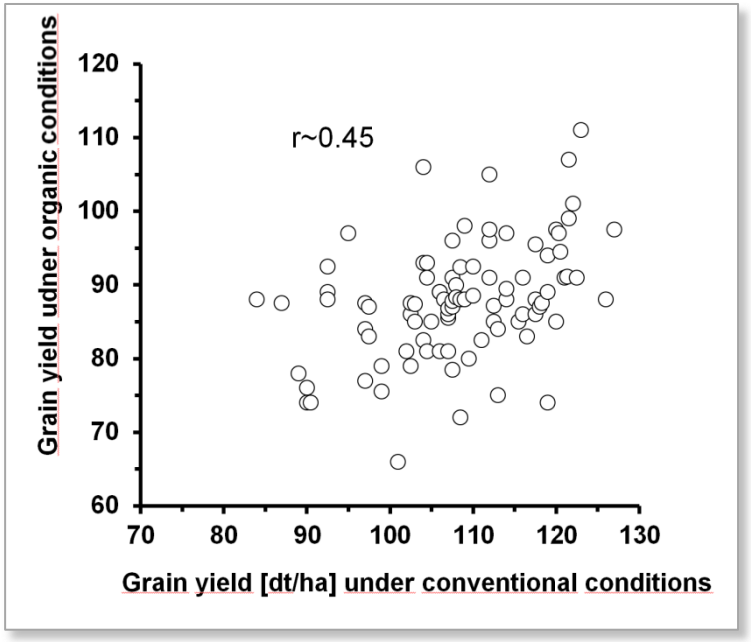
Mittwoch, 6. Juni 2018 - Thursday, June 14, 2018 wlink@gwdg.de Wolfgang Link, Universität Göttingen, DNPW, von Siebold 8, Göttingen **This document must not be changed, altered! Feed-back is very welcome.**



Often in breeding research we have Genotypes and two Treatment levels, such as

- conventional vs. organic
- fertilized vs. non-fertilized
- sprayed vs. non-sprayed
- vernalized vs. non-vernalized
- *per se* vs GCA
- diploid vs tetraploid
- last season vs current season
- pure stand vs. mixture
- and so on.

In such cases we may run ANOVA and find out whether $\sigma^2(G)$ and $\sigma^2(T)$ and $\sigma^2(GxT)$ are significant sources of variation. Usually we can even test even $\sigma^2(GxT)$ for significance - if we have a further layer such as replications R or environments E.

As diagram, we may visualize the data by plotting the genotypes (their means across replications or environments) at one treatment level against their results at the other treatment level, as shown here for 'conventional' vs. 'organic' (Becker, 2011). With a marked $\sigma^2(GxT)$ variance, we would expect a correlation between the two treatments of $r << 1$. For simplicity, I neglect here that $\sigma^2(GxT)$ may include shares of GxTxR, GxTxE and other higher interactions and error components.

Here I ask: somewhere in the quantity **"1-r"** we must find the σ²(GxT). **Where, how, inasmuch**?
This is my question.
I am not aware that this is algebraically available in the literature; at least I did not see it so far, although this may just be my mistake. I am currently studying **Yamada, 1962 (Jap. Jour. Genet. 37, 498-509)** to see inasmuch 'my' algebra was already presented there. Feedback is welcome!

σ² = variance; ω = covariance; here, for ease, I divide by N, not by (N-1); that does not matter, I verified this point.

$$\sigma^2_{G.i} = Variance\ of\ main\ effects\ of\ \boldsymbol{genotypes}$$

$$\sigma^2_{G_i xT1} = Var.of\ \ genotypes\ x\ treatment1\ \ interaction\ effects$$

$$\omega_{G.i \cdot GxT1} = Cov.(G_{.i})\ \ x\ (G \cdot T1)\ \ interaction\ effects$$

... this would be the covariance between the figures **8**, 4, (-12) and the figures **2**, 5, (-7) below here.

Illustration: Results of three genotypes (G) under two treatments (T), with zero main effect of treatments (just for ease of analysis; if we included a difference between the treatments' means, this would not make a difference for our purpose here**)**

	Treatments			
Genotypes	1	2	G Means	Main effects (G.i)
i=1	**110**	**106**	108	**8**
i=2	109	99	104	4
i=3	81	95	88	-12
T Means	**100**	**100**		**Grand mean μ=100**
σ²	**180.67**	**20.67**	74.67	74.67

	Interactions between G and T		
	T1	T2	Sum
G1	**2**	**-2**	**0**
G2	5	-5	0
G3	-7	7	0
Means	**0**	**0**	
σ²	**26.00**	**26.00**	

$r = \omega / (\sigma \cdot \sigma)$ I follow this algebra and follow it as good as I can, step by step.

Results of G1 and G2 and G3 in T1 is “G1T1” etc.; (these results are **110**, 109, 81).

Then, **G1T1** = μ + $G_{.1}$ + G1xT1-interaction; so, **G1T1** = 100 + **8** + **2** = **110**

The covariance ω between the genotypes’ performance in T1 and T2 is hence the covariance between the …

G1T1 = μ + $G_{.1}$ + G1xT1-interaction	.. with their counterparts …►	G1T2 = μ + $G_{.1}$ + G1xT2-interaction
G2T1 = μ + $G_{.2}$ + G2xT1-interaction		G2T2 = μ + $G_{.2}$ + G2xT2-interaction
G3T1 = μ + $G_{.3}$ + G3xT1-interaction		G3T2 = μ + $G_{.3}$ + G3xT2-interaction

This is, in the example, the covariance between

110	…►	**106**
109		99
81		95

In the illustration table above, we see that the interactions G1xT1 and G1xT2- etc. are symmetrical: Hence, their variance in T1 is the same as in T2 (in the example it is σ^2=26.00) and their covariance is minus-their-variance, which is ω = -26.00 in the example.

2	**-2**
5	-5
-7	7

So, back to the correlation.

The nominator, the covariance is, hence, ω ($G_{.i}$ + G_ixT1; $G_{.i}$ + G_ixT2)

We apply this rule: cov(X+Y, V+W) = cov(X,V) + cov(X,W) + cov(Y,V) + cov(Y,W)

So, ω ($G_{.i}$ + G_ixT1; $G_{.i}$ + G_ixT2) = ω ($G_{.i}$; $G_{.i}$) + ω ($G_{.i}$; G_ixT2) + ω (G_ixT1; G_i) + ω (G_ixT1; G_ixT2).

The part in yellow is zero because of the symmetry of the interactions. In the example, the covariance between the genotypes main effects G_i and the interactions is either 40.00 or minus 40.00, depending on whether we take the interactions with T1 or with T2.

The part in grey is $\sigma^2_{G.i}$. The part in blue is identical to "minus $\sigma^2_{G_ixT1}$" as we saw above. Therefore, the numerator of our sought-for correlation r, ω (G_i + G_ixT1; G_i + G_ixT2), is

ω = $\sigma^2_{G.i} - \sigma^2_{G_ixT1}$; this is 74.67 – 26 = 48.67 in our example data.

As next step we need the denominator of this correlation. Hence, we need the variance of the genotypes' results in T1, and the variance of the genotypes' results in T2.

We apply this rule: σ²(X+Y) = σ²(X) + σ²(W) + 2cov(Y,V)

So, for the variance in treatment T1, we get $\sigma^2_{G_iT1} = \sigma^2_{G.i} + \sigma^2_{G_ixT1} + 2\,\omega_{G.i;G_ixT1}$

And for the variance in treatment T2, we get $\sigma^2_{G_iT2} = \sigma^2_{G.i} + \sigma^2_{G_ixT2} + 2\,\omega_{G.i;G_ixT2}$

The variance in treatment T2 is as well: $\sigma^2_{G_iT2} = \sigma^2_{G.i} + \sigma^2_{G_ixT1} - 2\,\omega_{G.i;G_ixT1}$

This is because the covariance between genotypes' main effects G_i and their interaction effects G_ixT1 is "minus" the covariance between G_i and G_ixT2; as seen above.

For this part: $\sigma^2_{G.i} + \sigma^2_{G_ixT1}$ we write for a short while as an abbreviation $\sigma^2_{G_i+G_ixT1}$;
in our example: 74.67+26.00 = 100.67

This ("100.67") is NOT the variance of the genotypes in T1 (or in T2), but it is the variance of the genotypes in T1 or in T2 after eliminating the covariances between G.i and GixT1 or between G.i and GixT2 from the variances! Hm…

We re-write what we did above:

Variance in treatment T1: $\sigma^2_{G_iT1} = \sigma^2_{G_i+G_ixT1} + 2\,\omega_{G.i;GixT1}$ in our example: 100.67 + 2 · 40=180.67

Variance in treatment T2: $\sigma^2_{G_iT2} = \sigma^2_{G_i+G_ixT1} - 2\,\omega_{G.i;GixT1}$ in our example: 100.67 - 2 · 40 = 20.67

So, the correlation can meanwhile be written as:

$$r = {}^{\omega}/_{\sigma \cdot \sigma} = \frac{\sigma^2_{G.i} - \sigma^2_{G_ixT1}}{\{[\sigma^2_{G_i+G_ixT1} + 2\,\omega_{G.i;GixT1}]\,[\sigma^2_{G_i+G_iT1} - 2\,\omega_{G.i;GixT1}]\}^{0.5}}$$

$$r = {}^{\omega}/_{\sigma \cdot \sigma} = \frac{\sigma^2_{G.i} - \sigma^2_{G_ixT1}}{[(\sigma^2_{G_i+G_ixT1})^2 - (2\omega_{G.i;GixT1})^2]^{0.5}}$$

Nota bene: Two terms cancel out each other: these are the two products between the $(\sigma^2_{G_i+G_ixT1})$ and the covariances of the type G.i and GixT1.

For the $(2\omega_{G.i;GixT1})^2$, it is irrelevant whether we take the covariance with T1 (which is +40.00 in our example) or the one with T2 (-40.00 in our example), because we anyway take the square of it.

$$r = \frac{\sigma^2_{G.i} - \sigma^2_{G_ixT1}}{[(\sigma^2_{G_i+G_ixT1})^2 - (2\omega_{G.i;GixT1})^2]^{0.5}}$$

$$\boldsymbol{r = \frac{\sigma^2_{G.i} - \sigma^2_{G_ixT1}}{[(\sigma^2_{G.i} + \sigma^2_{G_ixT1})^2 - (2\omega_{G.i;GixT1})^2]^{0.5}}}$$

Numerator: 74.667 minus 26.000 = 48.667
Denominator: $[(74.667 \text{ plus } 26.000)^2 \text{ minus } (2\cdot 40)^2]^{½}$ = $(3733.845)^{½}$ = 61.105

Hence, the correlation between the genotypes' performance in the treatment T1 and the treatment T2 is

r = 48.667 / 61.105 = 0.796

A question is: Will the main effects of genotypes show a covariance with the genotypes' interactions with the one and the other treatment? Genotypes that show, on average across the two treatments, a high performance, may show mostly positive interactions with one treatment and negative interactions with the other treatments. It may indeed be that e.g. on-average higher yielding genotypes respond more marked on N-fertilization than on-average lower yielding genotypes. Or *vice versa*. It may be like that; but not necessarily so. Both cases lead to $(2\omega_{G.i;GixT1})^2$ being non-zero and positive. Yet, we anyway square to covariance. If non-zero, this covariance decreases the denominator and hence increases the correlation r. Both, negative and positive covariance between main effects and interactions increase the correlation between the two treatments.
It may as well be that there is no or a small covariance between the average performance of the genotypes and their interaction with the one (or other) treatment.
If this covariance between G_i and G_ixT1 was zero, then the composition of the correlation r is easier to write ☺.

Then: $r = \frac{\sigma^2_{G.i} - \sigma^2_{GxT}}{\sigma^2_{G.i} + \sigma^2_{GxT}}$

```
PLABSTAT   VERSION 3Awin  of 2004-11-29      Input:  3g2u.dat
 INSTITUT FUER PFLANZENZUECHTUNG, UNIVERSITAET HOHENHEIM
 At 2018/06/18   13h 06m 00s

  MEANS  GU  for X1
       I     U
 G     I      1      2 I   Mean
 ----------------------------
    1 I 110.00 106.00 I 108.00
    2 I 109.00  99.00 I 104.00
    3 I  81.00  95.00 I  88.00
 ----------------------------
 Mean I 100.00 100.00 I 100.00

 EFFECTS  GU  for X1
       I     U
 G     I      1      2 I   Mean
 ----------------------------
    1 I   2.00  -2.00 I   8.00
    2 I   5.00  -5.00 I   4.00
    3 I  -7.00   7.00 I -12.00
 ----------------------------
 Mean I   0.00   0.00 I 100.00

 Please check for outliers (test after ANSCOMBE and TUKEY)

Source   DF     SS      MS     Var.cp      F
U         1      0       0        -       -
G         2    448     224       73    2.87
GU        2    156      78       78   55.1543
Total     5    604
At 2018/06/18   13h 06m 00s
```

You see here the results of ANOVA with these data. What ANOVA does not give us is ($2\omega_{G.i\cdot GxT1}$); ω = 40.

The interaction variance $\sigma^2_{G_ixT1}$ is the variance between the figures 2, 5, (-7), which is $\sigma^2_{G_ixT1}$=26 if calculated with df=N and which is 39 if calculated with df=N-1=2. From the ANOVA we get σ²=78, this is double of 39. For the ANOVA, the interaction variance is calculated from these six data: 2; 5; -7; -2; -5; 7; with (2-1)·(3-1)=2 df (knowing only 2 of these 6 figures, such as 2 and 5, we can deduce all others: the -7 (because the columns' sum must be zero, and the second column is just the mirror picture of the first one anyway and hence does not add any df). The Sums of Squares of the first column (2, 5, -7) is 78, and dividing by df=2 we get σ²=39 (for the left as well as for the right column). The ANOVA takes the sums of squares of both columns, which is 156 (from six figures), and still divides only by df=2 as explained above. Thus, the $\sigma^2_{G_ixT1}$ variance (in the left column) as well as the $\sigma^2_{G_ixT2}$ in the right column are each half as large as the σ^2_{GxT} variance that is given by the ANOVA.

The direct correlation coefficient between the figure 110; 109; 81 and the figures 106; 99; 95 is r=0.79644789.

The composed correlation coefficient - if calculating the variances and covariances with (N-1) is
r=(73.0)/(91.6569692) = 0.79644789.
The composed correlation coefficient if calculating the variances and covariances with (N) is
r=(48.6666667)/(61.1046461) = 0.79644789. So, dividing by 'N' and by 'N-1' come to the same result.