



Paired correlations problem

E-mail distributed on 08-04-2024

Dear all,

Some of you may be familiar with Fisher's test for comparing two independent correlations, that is, for comparing the same association (e.g., $\text{cor}(X,Y)$) measured in two separate groups. Much less known are tests for comparing two dependent correlations, that is, two associations taken from the same group. Dependent correlations may come in two varieties, either with or without an overlapping variable. All three types of correlation comparisons have been implemented in the wonderful R package [cocor](#), which is the subject of this month's stat support. Non-R-users will be happy to know, however, that the package also has an [online interface](#) to do the necessary calculations!

Comparing correlation coefficients

Consider the following data set of a study where researchers investigate the impact of LSD on creativity and well-being:

```
LSD <-  
read.csv("https://drive.switch.ch/index.php/s/TWkPWrFu8r6u6HE/download",  
as.is=FALSE)  
head(LSD)
```

```
> ID   Group Age Creativity_T0 Wellbeing_T0 Creativity_T1 Wellbeing_T1  
> 1    LSD  52         58          86          35          61  
> 2 Placebo 35         47          79          38          76  
> 3    LSD  28         72          64          81          62  
> 4 Placebo 31         85          43          93          36  
> 5    LSD  54         87          78          87          90  
> 6    LSD  45         59          40          48          46
```

The researchers hypothesize that after a 4-week LSD program, the association between creativity and well-being will be stronger than before the program, as compared to the placebo group. One way to analyze this would be to look at the pattern of the four relevant correlations:

```
cor(Creativity_T0[Group=="Placebo"], Wellbeing_T0[Group=="Placebo"])  
cor(Creativity_T0[Group=="LSD"], Wellbeing_T0[Group=="LSD"])  
cor(Creativity_T1[Group=="Placebo"], Wellbeing_T1[Group=="Placebo"])  
cor(Creativity_T1[Group=="LSD"], Wellbeing_T1[Group=="LSD"])
```

```
> -0.10 -0.08 -0.09 0.49
```

As expected, the T1 correlation for the LSD group appears to be quite strong. One could run correlation tests on each individual correlation (e.g., with `cor.test`) and establish that all correlations are non-significant, except for the LSD group at T1. However, this approach does not make a formal comparison between the correlation coefficients. To do this, there are two possible directions of testing, **(a)** compare coefficients between groups, within T0 and T1 separately, or **(b)** compare coefficients between times, within LSD and placebo separately. The first may be the more intuitive approach, since it involves comparing independent correlations. Using the `cocor` package, we could run:

```
cocor.indep.groups(-0.10,-0.08, n1=36,n2=48)
cocor.indep.groups(-0.09,0.49, n1=36,n2=48)
```

```
> fisher1925: Fisher's z (1925)
> z = -0.0612, p-value = 0.9512
> Null hypothesis retained
```

```
> fisher1925: Fisher's z (1925)
> z = -2.7240, p-value = 0.0065
> Null hypothesis rejected
```

Indeed it appears that the creativity-wellbeing association is significantly different at T1 for the LSD group compared to the placebo group, while at T0 the association was not different. The function returns the classical Fisher test but also a modern alternative by Zou (2007), based on a confidence interval. What about the other direction of testing? Within each group, this would involve comparing dependent correlations, with no overlapping variables, since $\text{cor}(\text{Creativity}_{T0}, \text{Well-being}_{T0})$ has different variables than $\text{cor}(\text{Creativity}_{T1}, \text{Well-being}_{T1})$. However, the relevant function in `cocor` will require *all* 6 correlations among the four measures to produce a result:

```
> Placebo
>
> Creativity_T0 Wellbeing_T0 Creativity_T1 Wellbeing_T1
> Creativity_T0      1.00
> Wellbeing_T0      -0.10      1.00
> Creativity_T1      0.63      -0.02      1.00
> Wellbeing_T1      -0.11      0.93      -0.09      1.00
```

```
> LSD
>
> Creativity_T0 Wellbeing_T0 Creativity_T1 Wellbeing_T1
> Creativity_T0      1.00
> Wellbeing_T0      -0.08      1.00
> Creativity_T1      0.73      -0.12      1.00
> Wellbeing_T1      0.32      0.72      0.49      1.00
```

```
cocor.dep.groups.nonoverlap(r.jk=-0.10,r.hm=-0.09,
  r.jh=0.63,r.jm=-0.11,r.kh=-0.02,r.km=0.93,n=36)
cocor.dep.groups.nonoverlap(r.jk=-0.08,r.hm=0.49,
  r.jh=0.73,r.jm=0.32,r.kh=-0.12,r.km=0.72,n=48)
```

This time we get results for 6 different methods to calculate a *p*-value, with the oldest dating to 1898! The package or reference paper does not provide any recommendation for which one should be

preferred, so we could opt for the most recent proposal by Silver, Hittner, and May (2004). In any case it appears that the middle four methods produce the same z - and p -values for these data:

```
> Placebo
> silver2004:
> z = -0.0605, p-value = 0.9517
> Null hypothesis retained
```

```
> LSD
> silver2004:
> z = -4.1519, p-value = 0.0000
> Null hypothesis rejected
```

This time we have established that the correlation coefficients are not significantly different between T0 and T1 for the placebo group, and significantly different between T0 and T1 for the LSD group.

Finally, we briefly examine the case of dependent correlations with an overlapping variable. Say the researchers on the LSD study are interested to compare correlations between participant age and baseline (a.k.a, T0) creativity and well-being, regardless of the group. This time, $\text{cor}(\text{Age}, \text{Creativity}_{T0})$ and $\text{cor}(\text{Age}, \text{Well-being}_{T0})$ share a variable, namely age. With `cocor`, we proceed as follows:

```
cocor.dep.groups.overlap(r_jk=-0.03, r_jh=-0.16, r_kh=-0.09, n=84)
```

```
> hittner2003:
> z = 0.7638, p-value = 0.4450
> Null hypothesis retained
```

Once again, we need to provide both the two target correlations and their intercorrelation, which reduces to only one extra coefficient when there is an overlapping variable. This time the output produces no less than 10 (!) different methods to construct the appropriate test. Once again, we choose a recent one (Hittner, May, & Silver, 2003), which indicates that the correlation between age and creativity is not significantly different from the correlation between age and well-being, at baseline.

Regression approach

The `cocor` tests will come in handy for many similar comparisons of correlation coefficients. For the specific scenario above, however, the `cocor` approach misses a rather crucial piece of information. That is, while one can establish that the LSD-placebo correlations differ at T1 but not T0, this does not correspond to the formal interaction test that would be required first, i.e., the difference of the difference in coefficients.

To achieve this, one could abandon the correlations and opt for a linear regression. Because we have within-subjects variables, we reformat the data to long-format and fit a multilevel model (loading packages `lme4`, `lmerTest`, `emmeans`, and `visreg`):

```
LSD2 <-
read.csv("https://drive.switch.ch/index.php/s/RS0mAsliP2egw6k/download",
  as.is=FALSE)
```

```

model <- lmer(Wellbeing~Group*Time*Creativity+(1|ID),data=LSD2)
anova(model,type=2)

> Type II Analysis of Variance Table with Satterthwaite's method
>
>          Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
> Group          363.36   363.36     1 151.202  16.291 8.601e-05 ***
> Time           334.57   334.57     1  82.732  15.001 0.0002141 ***
> Creativity     172.32   172.32     1 123.374   7.726 0.0062959 **
> Group:Time     391.48   391.48     1  82.616  17.552 6.962e-05 ***
> Group:Creativity 302.09   302.09     1 121.787  13.544 0.0003484 ***
> Time:Creativity 372.07   372.07     1  82.835  16.682 0.0001015 ***
> Group:Time:Creativity 422.58  422.58     1  82.679  18.946 3.822e-05 ***
> ---
> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The interest here is in the three-way interaction, which will tell us if the association between creativity and well-being is modified by levels of group and time combined. As expected, the test is highly significant. With `emmeans`, we can print and compare the creativity slopes in each condition:

```

emtrends(model, specs=c("Group", "Time"), var="Creativity")

>   Group   Time Creativity.trend      SE  df lower.CL upper.CL
> LSD     T0           0.0326 0.0926 120   -0.151  0.2159
> Placebo T0          -0.0411 0.0929 106   -0.225  0.1430
> LSD     T1           0.4518 0.0692 120    0.315  0.5889
> Placebo T1          -0.0811 0.0716 106   -0.223  0.0608

pairs(emtrends(model, specs=c("Group", "Time"), var="Creativity"), adjust="none")

> contrast          estimate      SE    df t.ratio p.value
> LSD T0 - Placebo T0    0.0737 0.1311 112.9   0.562  0.5752
> LSD T0 - LSD T1       -0.4192 0.0701  83.3  -5.978 <.0001
> LSD T0 - Placebo T1    0.1137 0.1170 114.6   0.972  0.3331
> Placebo T0 - LSD T1   -0.4929 0.1158 111.0  -4.255 <.0001
> Placebo T0 - Placebo T1 0.0400 0.0790  81.8   0.507  0.6136
> LSD T1 - Placebo T1    0.5329 0.0996 112.7   5.352 <.0001

```

In this table, the between-group comparisons are in blue, and a conceptual equivalent of `cocor.indep.groups`, while the within-group comparisons are in yellow, and a conceptual equivalent of `cocor.dep.groups.nonoverlap`. **CAUTION:** Be warned that these tests are not statistically equivalent to the `cocor` output! Crucially, the regression approach requires that the analyst chooses a direction of causality between creativity and well-being. We could have flipped this but when we do, mysteriously this happens:

```

model <- lmer(Creativity~Group*Time*Wellbeing+(1|ID),data=LSD2)
anova(model,type=2)

```

Type II Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Group	682.18	682.18	1	130.090	8.6794	0.003814	**
Time	313.42	313.42	1	80.388	3.9877	0.049219	*
Wellbeing	206.89	206.89	1	137.024	2.6322	0.107012	
Group:Time	123.52	123.52	1	80.138	1.5715	0.213634	
Group:Wellbeing	787.29	787.29	1	136.049	10.0167	0.001914	**
Time:Wellbeing	275.87	275.87	1	80.465	3.5100	0.064632	.
Group:Time:Wellbeing	162.80	162.80	1	80.426	2.0713	0.153977	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The reason is that, once a regression model contains more than one effect, bivariate relations are no longer symmetric when flipped. In the worst case, the two directions of causality may even disagree on their conclusion (see my earlier mail regarding [Berkson's paradox!](#)), and this is especially likely in models containing interactions. So which of these two models is more correct? The answer will depend on the specific data scenario but fortunately, most of the time, there will be a clear causal preference for what should be the independent and the dependent variable.

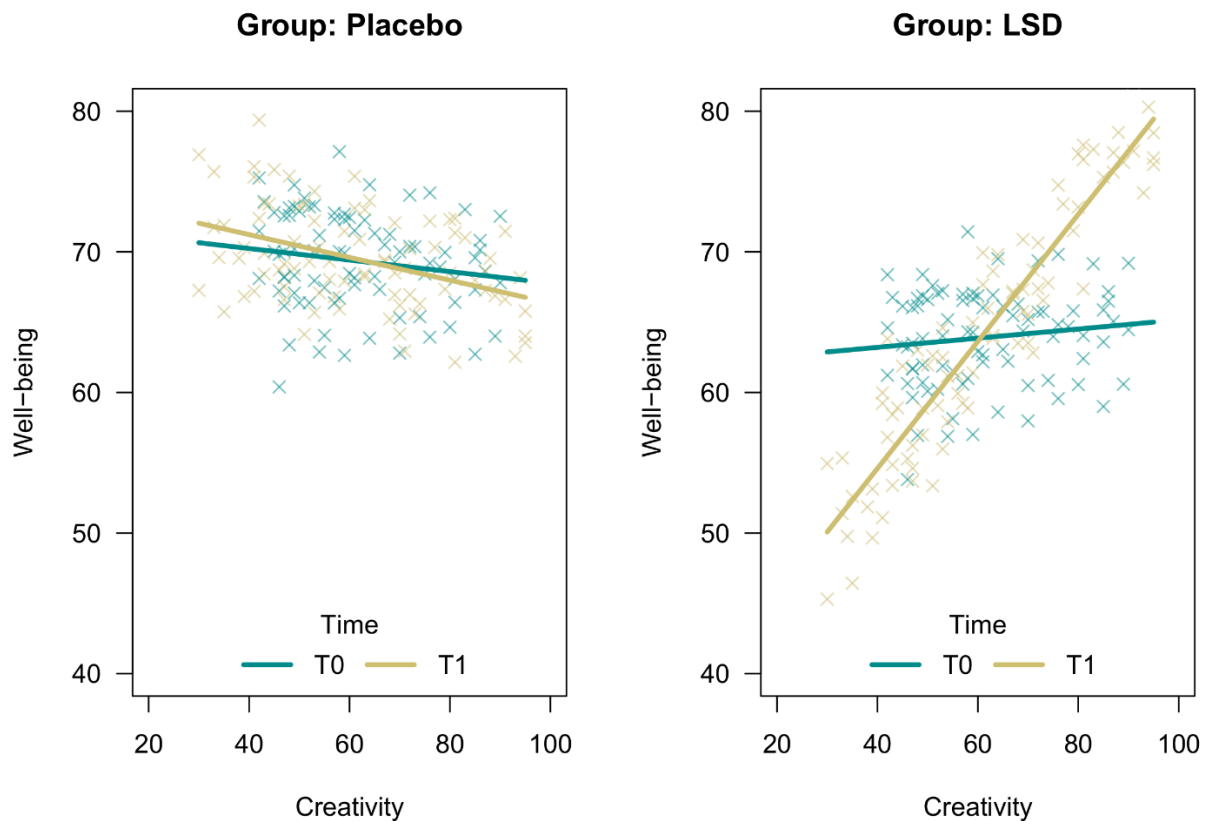


Figure 1. Creativity slopes for levels of group and time in a multilevel regression on well-being.

The `cocor` approach, on the other hand, has the advantage of not requiring a direction of causality, and is fully symmetric under any XY flip. The downside is that it cannot produce an interaction test of the type required for the above research question.

Best,
Ben

--

Ben Meuleman, Ph.D.

Statistician

Swiss Center for Affective Sciences

University of Geneva | Campus Biotech

Chemin des Mines 9 | CH-1202 Genève

ben.meuleman@unige.ch | +41 (0)22 379 09 79