**Reading statistical results: Narayanan, Tai & Kinias (2013)**

While or just after reading a research article, it is useful to make yourself a table answering the questions below about each study. If you can do it, then you really understand how they did the study and what the results are. It makes sense to re-read the Method and Results sections of the paper to complete the table.

|  |  |
| --- | --- |
| **Design** |  |
| Is this an experiment? (Why or why not?) |  |
| (If so) What is/are the independent variable(s) (IVs)? |  |
| How many levels does each IVs have, and what are they? |  |
| Are the IVs within- or between-participants? Are the factors crossed? |  |
| BONUS: How many numerator df are there total? For each possible effect? |  |
| **Measures** |  |
| What is the dependent variable?  |  |
| How was it measured? What is its possible range? |  |
| Were there any other measures? If so, how are they used (covariate, mediator…)? |  |
| **Analyses** |  |
| What kind of analysis did the researchers perform?  |  |
| What questions can that analysis answer? |  |
| BONUS: Was the analysis appropriate given the design and measures? |  |
| **Results** |  |
| What effects were reliable? |  |
| Are those effects consistent with the researchers’ hypotheses? (Are they in the right direction? Is the appropriate effects – contrasts, main effects or interactions, reliable?)  |  |
| BONUS: Do the results raise additional questions that must be answered with a new study? |  |

Here is how I would complete this table, going over the 4 experiments of Narayanan et al.

In addition to the information to know about each study in the table above, I will show you how to get a good start on understanding the studies by the most succinct kind of information the authors present.

Each experiment in the paper uses a different analysis method that we discussed when we went over inferential statistics, so this is a way to see those concepts in practice.

Experiment 1:

Design: Expt bc Ps are randomly assigned to condition. N = 45 men.

IV: 3 levels of power: low, high, control

Dependent measure: intention to connect with others (self-report scale, 1 to 12)

Other measure: Mood, to check whether manipulation of independent variable affected this and whether this corresponded to the dependent measure

Analysis: Used (one-way) analysis of variance (ANOVA) to compare the 3 conditions on both measures, separately.

ANOVA gives an overall (“omnibus”) test of whether there are any differences among the 3 conditions. So because k -1 = 2, there are two df in the numerator, so we can compare two sets of conditions orthogonally. A reliable F test (small p-value) does not say *which* conditions differ from others, so confidence intervals or simple effects tests to compare each pair of the 3 conditions are needed.

*Here is the first paragraph of the Results and Discussion section. See if you can identify where they present each effect I explained above, and the additional questions under Results in the table above:*



Answer next page.

You could take an educated guess about whether the results were as predicted by comparing the mean differences to the standard deviations by condition:

Condition Mean SD

High power 8.85 1.84 <- This mean is more than 1 SD bigger than the others.

}

Low power 6.40 2.38 The difference between these 2

Control 6.82 2.10 means is tiny compared to the SDs. *ns*

 *t*-tests always compare only 2 distributions. They usually have to be larger than about 2 to be reliably non-zero.

Experiment 2:

The pilot study of this experiment used a physical criterion to test a new self-report scale with multiple items about approach/avoidance. (In other words, scoring high on the scale means approach orientation and scoring low means avoidance orientation – they are measuring approach---avoid as if they oppose each other). The authors intended to use the new scale as a mediator between power condition and their DV, intention to connect with others, because that is a way to try to substantiate a process that is in between the IV and DV that might tell us *how* the outcome occurs.

The experimental design only has two conditions (low and high power) since a control was just like the low power condition in Study 1. In addition to measuring the DV, they have to measure the mediator in the same participants.

What they have labeled a “pilot” study is not actually piloting Experiment 2. It could be called a “validation study” because its purpose is to validate their new scale. They rely on the idea that people are used to using different arm muscles to push things away from them (avoiding) than they are to pull things towards them (approaching), and reason that if they have people do one of those two tasks, it will change the results of the scale. It did! Plus, they report an α, which is used to indicate how much the different items on the scale correlate with each other. Recall from measurement theory that if different things correlate, that is consistent with the idea that they are all driven by the same construct. The alpha is based on the average pair-wise correlations among all the items and can range from 0 to 1 where 1 is very high correlations (no one really should get that, but .75 to .95 are quite “good”).

The analyses for mediation involve three regression equations. The first uses just the IV on the DV. To replicate prior results (Study 1), they should get a reliable effect. The second regression equation tests whether the IV influences the hypothesized mediating variable. If this comes out null, then it wouldn’t make sense to think that the mediator is responsible for influence of the IV on the DV. The third (optional) regression equation tests whether the mediating variable does “influence” the dependent variable. One wouldn’t bother with this if the previous regression did not show reliable effects of the IV on the mediator. They did not report this one. The fourth regression is a *multiple* regression because it uses both the IV and the measured mediating variable to predict the DV. To demonstrate mediation, then the mediating variable should have a reliable effect on the DV, and in addition, having the mediating variable in the equation should reduce the effect of the IV compared to the first regression in which only the IV was allowed to predict the DV.

Below is the figure for Study 2 in which they show the results of the four regression equations. This is a standard way people show mediation tests. In this kind of figure, the outcome variable of a regression has an arrow (or more than one) pointing at it. The independent variable, labeled “Power” has no arrows pointing to it because it is independent (not contingent on anything except the experimenters’ random assignment). The dotted line indicates the first simple regression I described in which Power (only) predicts the DV (Intention to connect with others). The regression weight for this is 1.34 and is reliable as indicated by the asterix. The second regression equation I noted in which Power (IV) predicts the hypothesized mediator (Approach orientation) has a weight of .54, which is also reliable as the \* shows. As I mentioned, they don’t show us a regression with just the mediator predicting the DV, but they could have.

The fourth multiple (multiple) regression has both Power and Approach orientation predict the DV. What they should get if approach/avoid mediates the effect of Power on the DV (intention to connect) is that the mediator IS a reliable predictor of the DV, and the amount that the IV predicts the DV becomes weaker than in the first equation when just the IV predicts the DV. This is what they find. There is a reliable 1.47 regression weight from the mediator to the DV, and a B of 0.63 in parentheses from Power, with the latter one not reliable. This implies that (probably), we can accept the idea that approach orientation is responsible for much of the effect of power on the DV because the B for power on the DV was reduced substantially. (There actually is a separate test for comparing two B levels, which is not reported in this paper).

The idea here is *not* that magically Power doesn’t influence the DV any more (that would be inconsistent with the first regression results and the predictions), but rather that the effect of Power goes *through* the mediator. They postulate that the *way* that power changes the DV is because power changes approach/avoid orientation, and approach/avoid orientation is what changes the IV. In other words, approach orientation is the more proximal cause of intention to connect, and Power is the more distal cause.

So the logic of the study goes like this:

Researchers manipulated power, and then measured Approach and then Intention to Connect.

(They could have done ANOVA to do this because it is equivalent.) They then did regression and found that the power condition reliably affected the Approach measure. A second regression found that the power condition also reliably affected the Intention measures (B= 1.34), but notice that it was always completed AFTER the Approach measure was. So, we have to consider the *experimental condition + doing the approach measure* is the effect on the Intention measure. Another regression showed that including both power condition and the approach measure produced a big B for the approach measure, and reduced the B for experimental condition a lot compared to the regression that did not include the approach measure.



**Do mediations always have to “come out?”**

No (otherwise we wouldn’t have to test them).

Suppose we label the three variables IV, M, and DV (see my Figure 2, below). Here are results that could happen with this set-up that would not show mediation. Remember, the regressions are (Y on X):

(1) DV on IV

(2) M on IV

(3) DV on both IV and M

Figure 2. Generic Mediation Model

(3)

(2)

M

IV

DV

(1)/(3)

They way I’ve drawn the figure shows which equations produce each regression weight shown. We can call these “paths” because another name for this is “path analysis.” It just means a set of equations using the same big set of variables.

Ways mediation would not be indicated by the results:

1. It is possible that M separately predicts DV. So, we might get reliable effects in the multiple regression of both IV and M on DV, but this might not reduce the effect of IV on DV very much if IV and M are unrelated (i.e., path 2 is weak). If they are not related, then the regression of M on IV won’t be strong or perhaps reliable. *Tell yourself which paths would be non-zero and which would be reliable for this case.*
2. It is also possible that the IV predicts M, but M doesn’t predict the DV. That would not be mediation, but effects of the IV on different variables. *Tell yourself which paths would be non-zero and which would be reliably non-zero for this case.*
3. If the IV doesn’t predict the DV, then there is nothing to mediate. M could still predict the DV but we wouldn’t say it is mediating the IV because the IV doesn’t have any effects in the first place. *Tell yourself which paths would be non-zero and which would be reliably non-zero for this case.*

Experiment 3

Design: 2 x 2 cross factorial design. One factor is Power prime (high, low, so two levels) and the second factor is independent of the Power factor, and is whether the participant was included or excluded from cyberball. This is all between-Ps.

DV: Intention to connect with others.

Other measure: Mood scale.

Analysis:

Because the researchers are using the same DV as before, and one of the conditions is the same, this is study replicates Experiment 1 in some ways. However, the addition of the second factor (experience of inclusion/exclusion) means the study is extended. Possibly, the power main effect will replicate, but if it is *moderated* overall by the second factor, this might not come out.

To analyze both IVs at the same time, and to be able to test for their interaction, the researchers must use a 2-way ANOVA. The potential effects that analysis tests for are (1) Power main effect, (2) Experience of Inc/Exclusion main effect, (3) interaction of factors 1 and 2.

They predict that they will get high intention to connect *only* when power is high and the participant has been excluded. This means they predict an interaction with a particular pattern. (If they get a reliable F for the interaction term, but the pattern of means is not the particular one they predict, then that means their prediction was not upheld by the data).

Using how you read Experiment 1, you should be able to read the prose on p. 261 about how they did their analyses. They also illustrated the effects in Figure 2:



Instead of looking at numbers as digits, like the Table I made for Study 1, you can look at the bar graph above, and see whether the top of one bar falls within the little verticle I-beam for the other bars (that represents the SE, but I think it is just an estimate from their graph program and not the actual SE for each condition). The only bar whose top falls outside the I-beams is the 2nd from left, Exclude + High power condition.

If you visually average the Exclude bars and the Include bars, you can guess that there is not going to be a reliable main effect for that condition. Check the text they wrote to see if these interpretations of their figure are spot-on.

*Side Note:* A lovely thing these authors have done is to show the whole range of their scale on the Y axis. They could make the effect look a lot bigger if they truncated the Y axis to say, 6 to 9, but that is kind of misleading.

Experiment 4

*You should be able to read and interpret everything about Experiment 4 using what we did with the previous 3 studies, except for one thing:*

**Experiment 4 shows moderated mediation.**

Ugh – what is that?

Mediation was shown in Experiment 2. It means that the effect of the IV is transmitted to the mediator, and that effect is transmitted to the DV. Think of a train engine (IV) pulling car 1 which pulls car 2. The energy of the engine is transmitted to car 1 (M) and that is what transmits energy to car 2 (DV).

Moderation, remember, is an interaction. It means the effect of one factor is changed by the levels of a different factor. This means sometimes you see the effect, and sometimes you don’t, but you can tell why. Another way to think of moderation is that there is some boundary condition for an effect: across some line it doesn’t work the same way as inside that space.

Putting both together:

*Moderated mediation* means that mediation is going to be contingent on an IV. In other words, the size of the mediation effect is going to depend on the condition of the experiment – it won’t be the same in both.

Here is the figure in which they present the moderated mediation.



This looks basically like the IV🡪M🡪DV figure from Experiment 2 (with the triangle flattened)-- except there are two different paths from M to DV. Those different paths (arrows) are just showing you that there is one effect from M to DV (and it is reliable) for the excluded condition (top arrow), and there is a different effect for that path (and it is not reliable) for the included condition (bottom arrow). So, this is showing us that approach as a mediator works when people have been excluded.

A different way they could have presented their results would be like this:

Figure 3 Mediation for the Exclude Condition and the Include Condition

Exclude Condition: Include Condition:

.39

.43

Approach

Approach

-.02

.43

Power

Intention

Intention

Power

PUTTING ALL THE STUDIES TOGETHER – integrative review.

If you think back over the other experiments, each study showed a piece of the results of Study 4, but none of them showed all of it.

Expt 1 showed power IV to intention to connect (DV). (This is Path 1 on my Figure 2.)

Expt 2 showed the mediation of approach orientation between IV (power) and DV (intention to connect). (Combination of Paths 2 and 3 in my Figure 2.)

Expt 3 showed a contingency: that the IV (power) to DV (intention to connect) only happens for excluded participants. (This is like the red dashed line versus the green dashed line in my Figure 3).

So, Expt 4 is showing the combination of the results of Expt 2 and Expt 3. It does this with a different DV than those two studies did, so that is a nice extension.

When different studies, with different participants, and different manipulations of the IVs, and different measures of the DVs and/or mediators all find compatible results, that gives us confidence that the effects are robust.