

1. Suppose that you were interested in studying the impact of a treatment that you thought would have only a small effect (e.g., Drug A has a major effect in alleviating depression, but you think that combining it with Drug B may provide a small additional positive impact that may be useful, so you do a study to look at DrugA + Placebo vs. Drug A + Drug B). What aspect of an experiment must one seek to enhance when studying such small treatment effects? How would you design your study to implement the desired outcome? [5 pts]

When you expect that the treatment effect is small, you need lots of power. Drug A has already been found to be effective, so there seems little point in including a Placebo control group. Given that it's a drug study, I would be inclined to use an independent groups design. I would probably include four groups (Drug A alone [serving as a control group], Drug B alone [to see how effective Drug B is on its own], Drug A + Drug B [to test the combined effects of the two drugs], and Drug A + Placebo [to ensure that the combined effects of the two drugs isn't due to a belief that more drugs is better]). To enhance the treatment effects, I would use fairly large (but not ridiculous) doses of Drug B (and normal doses of Drug A). I would use a fairly homogenous group of participants to minimize the error term. And I would use a very large sample size to detect the small incremental effects of Drug B.

2. Suppose that you are interested in studying the impact of a drug on maze learning in rats. Because you are unsure of the level at which the drug might be most effective, you decide to use 4 different levels of the drug. First of all, tell me (in very general terms) how you would determine the 4 levels that you would use in your experiment. You want to avoid carry-over effects of the drug, so each rat will be exposed to only one level of drug. Because you think that the drug may lead to better performance on some mazes than on other mazes, you want to run each of your rats through three different mazes (Easy, Moderate, and Difficult). Thus, this experiment would be a 4x3 mixed design. In *very explicit fashion*, tell me how you would run this experiment, including the minimum number of rats you'd need for your study and how many you'd actually use, the procedure you'd use, etc. [20 pts]

	No Drug	Low Drug	Mod Drug	High Drug
Easy Maze	n = 30	n = 30	n = 30	n = 30
Mod. Maze				
Diff. Maze				

In order to counterbalance the orders of the repeated factor (Maze Difficulty), I would need 6 orders, which means that I need to run in multiples of 6 rats. The six orders are:

EMD, EDM, MED, MDE, DEM, DME

To achieve reasonable power, I might then use 30 rats at each level of Drug. Thus, I would need a total of 120 rats. The minimum I could get away with would be 24.

Let's assume that the drug is administered through a syringe. I would need to inject all the rats, but those in the No Drug condition would be injected with a saline solution. I would then wait the appropriate time for the drug to take effect and then run the rat through the first maze. Presuming that the drug remained active for a

sufficient duration, I would then run the rat through the next two mazes. I would use time to complete the maze as the DV. If I could not run the rat through all three mazes in one day, I would space the running out over three days, but would then inject the rat in the same fashion prior to each run of the maze.

3. The Ross, et al. article on the effectiveness of debriefing provides us with some cautionary data. How do the two studies reported in the article inform us about the effectiveness of debriefing. [10 pts]

You need to carefully articulate the findings of the Ross, et al. study.

4. We talked about the Doob & Wood (1972) study (Catharsis and aggression: Effects of annoyance and retaliation on aggressive behavior) when we discussed two-factor designs. However, we didn't really talk about it when discussing ethics. I'd like you to use the APA guidelines to evaluate the extent to which the Doob & Wood study is ethical. In essence, I'm asking you to "pretend" that you're a member of an IRB and to evaluate whether or not you would approve the Doob & Wood study. [10 pts]

Here's a brief recapitulation of the study.

Time 1: Participant is in the waiting room and a confederate of the experimenter comes into the room and either annoys the participant (Annoy) or doesn't annoy (No Annoy) the participant.

Time 2: In the "learning phase" of the experiment, and similar to the Milgram study, the confederate is "randomly" chosen to be the learner and subsequently makes errors. He is either shocked by the experimenter (Experimenter Shocks), shocked by the participant (Participant Shocks), or receives no punishment for incorrect answers (No Shock).

Time 3: In the "creativity phase" of the experiment participants are asked to judge the creativity of the confederate's responses in a free-association task (participant says a word and confederate gives a verbal response, which the participant judges by shock level). The participant was asked to give no shock if the response was creative, but to give a shock if the response was not creative and to use the duration of the shock to indicate the degree to which the response lacked creativity (longer shocks for less creative responses). [The confederate is never actually shocked at any point in the study.]

Time 4: The experimenter fully explains the study to the participants.

The article isn't completely clear about informed consent, but let's presume that it occurs prior to the "actual" experiment (that is, during Time 2) and tells the participant that it's a learning experiment.

As always, what's important is how well you make use of the APA ethical guidelines in assessing this study.

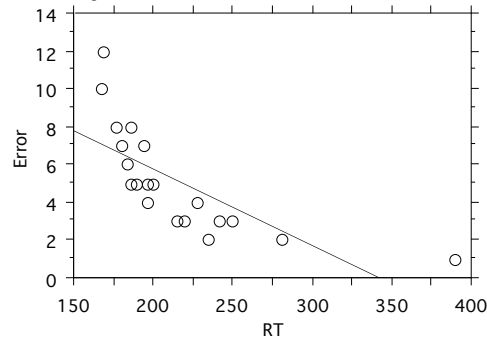
5. Researchers who measure reaction time (RT) for human participants often observe a relationship between the RTs and the number of errors that participants commit. This relationship is known as the *speed-accuracy trade-off*. In the data below, 20 participants were shown computer screens filled with the letter E. Some of the letters were in the usual orientation and some of the letters were backward. Participants are told to assess the number of backward E's as quickly as possible. The average number of backward letters missed over 10 trials constituted the mean number of errors. The average response time on each trial (in milliseconds) constituted the other variable. Interpret the results of the study as completely as you can. How many errors would you predict a person would make who responded with a RT of 200 milliseconds? How many errors would you predict a person would make who responded with a RT of 500 milliseconds? How much of the variability in errors is shared with RT? [10 pts]

Regression Summary

Error vs. RT

Count	20
Num. Missing	0
R	.724
R Squared	.524
Adjusted R Squared	.498
RMS Residual	2.007

Regression Plot



ANOVA Table

Error vs. RT

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	80.012	80.012	19.855	.0003
Residual	18	72.538	4.030		
Total	19	152.550			

Regression Coefficients

Error vs. RT

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	13.774	1.987	13.774	6.933	<.0001
RT	-.040	.009	-.724	-4.456	.0003

There is a significant negative linear relationship between Errors and RT, $r = -.724$, $p = .0003$. If a person responded in 200 ms, the number of predicted errors would be:

$$\hat{Y} = 13.774 - (.04)(200) = 5.77$$

If a person responded in 500 ms, you would not be able to predict the number of errors, because you did not observe anyone with a RT that large. The shared variability between these two variables would be $r^2 = .524$.

You might also note that the relationship appears to be curvilinear and that one point may be something of an outlier.

6. Dr. Luke Attem was interested in factors that influence memory for faces. During the acquisition phase, participants were shown a series of 60 computer-generated male faces one at a time for 30 seconds each. Fifteen of the faces wore sunglasses, fifteen of the faces had full beards, fifteen of the faces wore hats, and fifteen of the faces were unadorned (no glasses, beard, hat, etc.). Each type of face occurred equally often within portions of the acquisition phase, to control for any position bias. (Note that the random ordering of the faces serves the same function as counterbalancing.) At test, 120 “unadorned” faces were presented, 60 new faces and the 60 original faces. That is, regardless of how the face was seen at acquisition, it was seen unadorned at test (no sunglasses, beard, or hat). The dependent variable was the percentage of faces of each type correctly recognized (100% indicating perfect recognition). Complete the analysis below and interpret the results of this experiment as completely as you can. [15 pts]

ANOVA Table for Acq Face

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	23	7041.490	306.152				
Category for Acq Face	3	5554.281	1851.427	263.688	<.0001	791.063	1.000
Category for Acq Face * Subject	69	484.469	7.021				

Means Table for Acq Face

Effect: Category for Acq Face

	Count	Mean	Std. Dev.	Std. Err.
Unadorned	24	60.167	9.867	2.014
Sunglasses	24	41.417	9.007	1.839
Beard	24	44.583	8.423	1.719
Hat	24	55.042	8.819	1.800

Decision: Reject H_0 , because $p < .05$ ($p < .0001$).

Post Hoc Test:

$$HSD = 3.7 \sqrt{\frac{7.0}{24}} = 2.0$$

Conclusion:

When people wore glasses, they were recognized significantly less often than if they wore beards or hats or if their faces were unadorned. When people wore beards, they were recognized significantly less often than when they wore hats or were unadorned. When people wore hats, they were recognized significantly less often than when their faces were unadorned.

7. Individuals who are identified as having an antisocial personality disorder also tend to have reduced physiological responses to painful or anxiety-provoking stimuli. In everyday terms, these individuals show a limited physical response to fear, guilt, or anxiety. One way of measuring this response is with the galvanic skin response (GSR). With GSR, higher scores indicate lower responsivity and lower GSR scores indicate greater responsivity. In the study summarized below, three groups of individuals were tested: Normal Personality, Antisocial Personality, and Agoraphobics. First, briefly tell me why a group of Agoraphobics (or some other clinically diagnosed group) would be included in such a study:

To determine if the results are specific to antisocial people or to any clinical group.

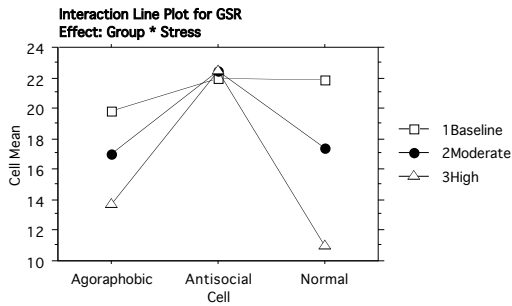
As you can see, a third of each group is given the GSR under ordinary circumstances (baseline), a third is given a moderately stressful situation, and a third is given a highly stressful situation. Complete the source table and interpret the results of this study as completely as you can. [20 pts]

ANOVA Table for GSR

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group	2	419.937	209.968	34.538	<.0001	69.076	1.000
Stress	2	323.556	161.778	26.611	<.0001	53.222	1.000
Group * Stress	4	226.825	56.706	9.328	<.0001	37.311	1.000
Residual	54	328.286	6.079				

Means Table for GSR
Effect: Group * Stress

	Count	Mean	Std. Dev.	Std. Err.
Agoraphobic, 1Baseline	7	19.857	2.116	.800
Agoraphobic, 2Moderate	7	17.000	1.633	.617
Agoraphobic, 3High	7	13.714	1.113	.421
Antisocial, 1Baseline	7	22.000	3.109	1.175
Antisocial, 2Moderate	7	22.429	2.637	.997
Antisocial, 3High	7	22.429	2.225	.841
Normal, 1Baseline	7	21.857	2.410	.911
Normal, 2Moderate	7	17.429	4.117	1.556
Normal, 3High	7	11.000	1.414	.535



There is a significant interaction, as well as two significant main effects. However, you would focus your attention on the interaction.

Compute HSD. $HSD = 4.58\sqrt{\frac{6}{7}} = 4.24.$

Thus, Antisocial people show no difference between Baseline, Moderate, or High Stress. However, Agoraphobics and Normal people show a different pattern. Agoraphobics have higher GSR for Baseline than High Stress levels, but GSR levels for Baseline and Moderate Stress are equal and GSR levels for Medium and High Stress are equal. For Normal people, however, GSR levels are significantly higher for Baseline than Moderate and High Stress levels. GSR levels are also significantly higher for Moderate Stress than High Stress.