

1. Dr. Melody Brooks is interested in assessing whether or not people eat more when they are anxious. She brings her participants into the lab one at a time and places a bowl of 30 salted peanuts in front of them, explaining that the participant should feel free to eat as many peanuts as she or he would like. Dr. Brooks then has half the participants fill out an ominous-looking consent form, absolving her of all liability should the experiment result in permanent damage to the participant. She then tells these participants that they will receive a series of painful electric shocks (though there will be “no permanent tissue damage”), to determine their tolerance for pain. The other half of the participants is given a standard consent form that simply describes the experiment and asks for the participant’s cooperation. Dr. Brooks then tells this group that they will be serving as a control group for an experiment and will have to wear a shock electrode and receive a mild electric shock, but that it will only feel like a mild tingle. (Dr. Brooks might well have seen the same movie you saw at the beginning of the semester.) Neither group actually receives a shock, with the threat of the different shock levels intended to create different anxiety levels. At the conclusion of the experiment, Dr. Brooks is surprised to see that there is no difference between the two groups in terms of peanuts eaten (M for the anxious group = 29.8, M for the non-anxious group = 29.5). What advice would you give Dr. Brooks about her experiment? Because you want to make the experiment more powerful, would you suggest using a repeated measures (within subjects) design? [10 pts]

First of all, given the deceptive nature of this study, I don’t think that one could readily change it to a repeated measures design. That is, given that the participant is not actually shocked, you wouldn’t be able to deceive them when introducing the second condition (this time I’m really, really, really going to shock you!). Next, you should note that the means don’t differ much and that they are both quite high (people typically eating all of the nuts), which typifies a ceiling effect. Thus, you should probably run the experiment again, but with more than 30 nuts in the bowl. I think that you should also consider a control group in which the participants experience no undue anxiety (i.e., no threat of shock at all).

2. Dr. Nick O. Thyme was interested in studying the effects of practice on performance on video games that involve a great deal of eye-hand coordination. He decides to conduct a two-factor mixed design, with amount of practice (3 hours, 6 hours, 12 hours) as the between (independent groups) factor and input device (keyboard, mouse, joystick) as the within-subjects (repeated measures) factor. He uses the score on a novel video game as the dependent variable. [15 pts]

a. Given the design specified above, if Dr. Thyme wants to have a minimum of 20 scores per cell (condition), tell him how many participants he would need to run.

This study would involve a 3x3 mixed design. To completely counterbalance the repeated factor (3 levels of input device) would require 6 orders. Because 24 is the first multiple of 6 to exceed 20, I would need to run 24 participants in each of the three levels of the between factor (practice), for a total of $3 \times 24 = 72$ participants.

b. Briefly describe a good procedure that Dr. Thyme might use to conduct his study.

First of all, I would want to include a 0 hours of practice control group (to determine how well people would perform if they had no practice at all). If I were constrained to 3 levels of the practice factor, I would use 0, 6, and 12 hours of practice. I would randomly assign participants to the amount of practice factor. Because those practice times are too great to administer all at once, I would have participants practice in blocks of .5 hour (for example) on each input device, with a

10-minute break between practice sessions and with 3 sessions per day (one session on each input device). I would also use random orders of devices for each practice session. That would mean that the participants in the 12-hour practice group, for example, would have to practice over a period of 8 days.

At the end of the practice period, I would test each participant on each input device, using one of the 6 orders: KMJ, KJM, JKM, JMK, MJK, MKJ. Each order would be used equally often, so each order would be used 4 times within each of the levels of the practice factor.

c. Suppose that he runs the study and finds no main effects or interaction. What would you advise him to change in his study to make it more powerful? Do you think it would help if he were to run the study as a completely within (repeated measures) design?

Of course, one approach would be to increase n (sample size). Beyond that approach, I would consider increasing treatment effects by using even greater practice times (0, 12, 24, 48 hours) and possibly other input devices. I may also look to see if there are ceiling or floor effects, which might lead me to use a more difficult game or an easier game. I would also work to decrease variability, possibly by using more similar participants (all with very little video game experience, or all with a great deal of video game experience, or all athletes with good eye-hand coordination, etc.). I would also work to ensure that the testing room was well isolated to minimize distractions, make sure that the instructions were very clear, etc.

3. A major research technique in the field of behavioral genetics is to breed animals selectively on the basis of particular characteristics exhibited by the animals and then to observe the relative performance of the offspring. Suppose that an experiment is conducted in which three strains of rats are to be compared. One strain was obtained by selectively breeding rats who performed exceptionally well in a maze-learning task (the “bright” rats); a second strain was obtained by selectively breeding rats who performed quite poorly on the same task (the “dull” rats); and a third strain consisted of rats who were bred without regard for maze-learning performance (the “mixed” rats). One group from each strain was raised under “enriched” conditions, and a second group was raised under “impoverished” conditions. The enriched environment consisted of a large cage containing objects for the animals to play with; the impoverished environment consisted of a similar cage containing nothing but the bare essentials of rat life (food and water dispensers). Following six months of exposure to one of the two environments, all of the rats were tested in a standard laboratory maze. There were eight rats randomly assigned to each of the six groups. The learning scores (trials needed to learn the maze) are presented below. Analyze the results as completely as you can, then provide a description of the findings. [15 pts]

ANOVA Table for Learning Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Environment	1	154.083	154.083	23.092	<.0001	23.092	.999
Breed	2	175.792	87.896	13.173	<.0001	26.345	.998
Environment * Breed	2	17.792	8.896	1.333	.2746	2.666	.262
Residual	42	280.250	6.673				

Means Table for Learning Score
Effect: Environment * Breed

	Count	Mean	Std. Dev.	Std. Err.
Enriched, Bright	8	2.750	1.165	.412
Enriched, Dull	8	8.875	2.167	.766
Enriched, Mixed	8	6.125	2.232	.789
Impoverished, Bright	8	8.000	2.449	.866
Impoverished, Dull	8	11.250	3.955	1.398
Impoverished, Mixed	8	9.250	2.712	.959

Because the interaction is not significant, I would focus on the two main effects (both with P-Values < .0001). For the Environment factor there are only two levels, so no post hoc test is needed. I would be able to conclude that the rats in the Enriched environment needed significantly fewer trials to learn the maze ($M = 5.92$) compared to rats in the Impoverished environment ($M = 9.5$). For the Breed factor, I would need to compute a post hoc test:

$$HSD = 3.44 \sqrt{\frac{6.67}{16}} = 2.22$$

Thus, I would be able to conclude that the Bright rats needed significantly fewer trials to learn the maze ($M = 5.375$) than Mixed ($M = 7.69$) or Dull ($M = 10.06$) rats. Moreover, Mixed rats needed fewer trials to learn the maze ($M = 7.69$) than Dull rats ($M = 10.06$). You should also note that rats could be randomly assigned to environments, but not to Breed, which is a non-manipulated variable.

4a. 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 test the possibility that all clinical groups would respond similarly.

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

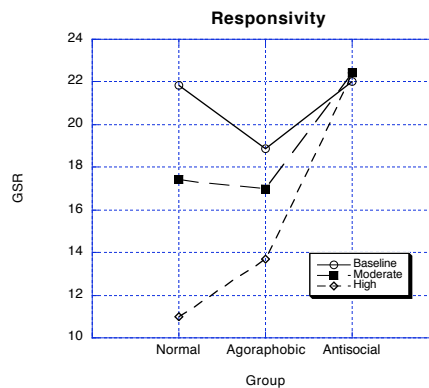
	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group	2	414.0	207.0	34.5	<.0001	69.076	1.000
Stress	2	319.2	159.6	26.6	<.0001	53.222	1.000
Group * Stress	4	223.2	55.8	9.3	<.0001	37.311	1.000
Residual	54	324.0	6.0				

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

Given the significant interaction ($p < .0001$), I would focus my attention on explaining that effect. First, I draw myself a figure and look at it to see what the interaction looks like, then I would compute a post hoc test:



$$HSD = 4.6 \sqrt{\frac{6}{7}} = 4.26$$

Thus, for the Antisocial people, there is no difference in GSR for Baseline, Moderate, and High Stress. As predicted, they tend to show little difference in response under the three different levels of stress. On the other hand, Normal people showed higher GSR response for Baseline than Moderate Stress and higher GSR response for Moderate stress than High stress. These results are consistent with greater physiological response to greater stress. Finally, the Antisocial people showed significantly higher GSR under Baseline conditions compared to the High stress condition, but Moderate Stress led to GSR levels that didn't differ from Baseline or High stress.

4b. What is your best estimate of the population variance (σ^2) from which these scores were selected? [2 pts]

MS_{Error} (6.0)

4c. Suppose that you were to compute a one-way analysis of variance on these same data, looking only at the group factor. Complete the source table below to show what the one-way ANOVA would look like. [5 pts]

Source	df	SS	MS	F
Group	2	414	207	14.375
Error	60	866.4	14.4	
Total	62	1280.4		

4d. Compare your F-ratios for Group in the Two-Way ANOVA and in the One-Way ANOVA. Under which conditions would your F-ratio be larger? Under which conditions would your F-ratio be smaller? [3 pts]

The F-ratio for the One-Way ANOVA would be greater when MS_{Error} gets smaller in the One-Way ANOVA (compared to the Two-Way ANOVA). (Note that the MS_{Group} would not change from the One-Way to the Two-Way ANOVA.) That situation will emerge when there is not a lot of variability in the second factor or the interaction. That is, SS_{Stress} and SS_{GroupxStress} would be small.

The F-ratio for the One-Way ANOVA would be smaller then MS_{Error} gets larger in the One-Way ANOVA (compared to the Two-Way ANOVA). That situation will emerge when there is a great deal of variability in the second factor or the interaction. That is, SS_{Stress} and/or SS_{GroupxStress} would be large.

5. Compare a 3x5 completely between (independent groups) and a 3x5 completely within subjects (repeated measures) designs to illustrate the efficiency of a repeated measures design. [5 pts]

In a 3x5 repeated measures design, I would need to counterbalance, which would lead me to use incomplete counterbalancing. Thus, I would need to run in multiples of 30 participants. Let's assume that I use just 30 participants. Doing so would produce 450 pieces of data in this study. In a 3x5 independent groups design, on the other hand, I would need one participant for each piece of data. Thus, in order to generate 30 pieces of data in each cell, I would need to run 450 people. The repeated measures design is a good deal more efficient, needing only 30 people to generate 450 pieces of data, whereas the independent groups design would require 450 people.