1. (From G&W) In order to study cardiovascular responses to embarrassment, Harris (2001) had people sing the *Star Spangled Banner* in front of a video camera while she recorded their heart rate and blood pressure. She found that blood pressure increases steadily for two minutes before gradually returning to normal. What about the heart rate data? Below is a partially completed source table for these heart-rate data. Complete the table and analyze/interpret the results as completely as possible. Is the pattern for heart rate similar to that for blood pressure? [15 pts]

Descriptive Statistics

	Mean	Std. Deviation	N
Baseline Heart Rate	76.9167	1.72986	12
Heart Rate at 1 Min	89.2500	1.86474	12
Heart Rate at 2 Min	78.0833	1.78164	12

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
time	Sphericity Assumed	1112.667	2	556.333	333.800	.000	.968	667.600	1.000
Error(time)	Sphericity Assumed	36.667	22	1.667					

H_0: $\mu_{Baseline} = \mu_{1Min} = \mu_{2Min}$ **H_1: Not H_0**

Decision: Reject H_0, $p < .001$

Post Hoc Test: $HSD = 3.55\sqrt{\frac{1.667}{12}} = 1.32$

	Base	1Min	2Min
Base	------		
1Min	12.3	------	
2Min	1.2	11.2	------

There is a significant effect of time of assessing heart rate, $F(2,22) = 333.8$, $MSE = 1.667$, $p < .001$, $\eta^2 = .968$. Post hoc tests using Tukey's HSD indicate that heart rate is significantly higher at one minute ($M = 89.25$) compared to Baseline ($M = 76.917$) or two minutes ($M = 78.083$). Thus, the data aren't similar to the blood pressure data, because the hear rate data have almost returned to baseline.

Even though this is a repeated measures design, no counterbalancing is possible. Briefly explain why not.

Because you're interested in the effects of time, you cannot counterbalance that variable.

2. Mook argues that external validity is not always the purpose behind psychological research. For each of the studies below, indicate why the study is not externally valid, then why it's not a concern, given the intentions of the researcher(s). [10 pts] **Answer each question using the Mook article.**

Study	Why not externally valid	Why lack of EV is not a concern
Argyle (glasses and intelligence)		
Harlow (infant monkeys and drive reduction theory)		
Hecht (dark adaptation)		
Brown & Hanlon (parental role in grammar acquisition)		

3. In the first lab, we collected a number of different academic measures from members of both sections of PS 306. Below are the results from a correlation analysis of two different SAT scores (Math and Verbal/Critical Reading). First of all, tell me what you could conclude from these results. Then, given an SAT-V score of 600, what SAT-M score would you predict using the regression equation? Given the observed correlation, if a person studied and raised her or his SAT-V score, would you expect that person's SAT-M score to increase as well? What would you propose as the most likely source of the observed relationship? [10 pts]

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.438[a]	.192	.173	59.47261

a. Predictors: (Constant), satv

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	36109.377	1	36109.377	10.209	.003[a]
	Residual	152090.623	43	3536.991		
	Total	188200.000	44			

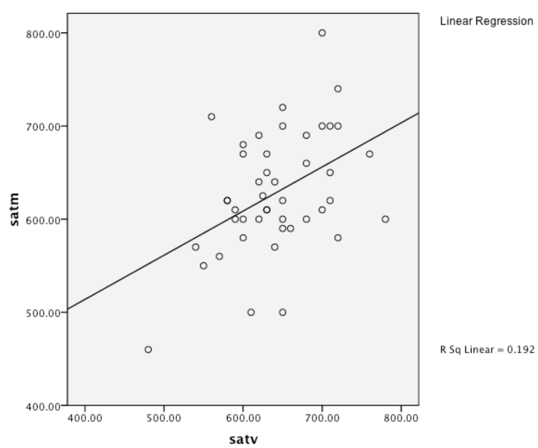
a. Predictors: (Constant), satv

b. Dependent Variable: satm

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	324.301	95.566		3.393	.001
	satv	.474	.148	.438	3.195	.003

a. Dependent Variable: satm



There is a significant positive linear relationship between SAT-M and SAT-V, *r*(43) = .438, *p* = .003. The effect size (coefficient of determination) is r^2 = .192.

SAT-M = (.474)(600) + 324.3 = 608.7

No, raising SAT-V would not necessarily have an impact on SAT-M. The relationship is not a causal one.

A third variable, such as intelligence (IQ) or educational experience may well lead to the scores on both SAT-M and SAT-V.

4. (From G&W) Intelligence is offered as one possible explanation for why some birds migrate and others maintain year-round residency in a single location. That is, small bird brains (relative to body size) don't have enough computational power to allow the bird to find food during the winter, so they must migrate to warmer climates where more food is available (Sol, Lefebvre, & Rodriguez-Tejeiro, 2005). On the other hand, large bird brains (relative to body size) produce sufficient computational power that their owners are more creative and can find food even when the weather turns harsh. Below is a partially completed source table consistent with the actual research results. The numbers represent the relative brain size for the individual birds in each sample. Complete the table and analyze/interpret the results as completely as possible. [15 pts]

Descriptives

Relative Brain Size

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Non-Migrating	15	15.0000	5.35857	1.38358	12.0325	17.9675	8.00	23.00
Short-distance Migrant	15	10.0000	5.11301	1.32017	7.1685	12.8315	5.00	16.00
Long-distance Migrant	15	5.0000	4.72077	1.21890	2.3857	7.6143	2.00	14.00
Total	45	10.0000	6.44910	.96138	8.0625	11.9375	2.00	23.00

Test of Homogeneity of Variances

Relative Brain Size

Levene Statistic	df1	df2	Sig.
.942	2	42	.398

ANOVA

Relative Brain Size

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	750.000	2	375.000	14.583	.000
Within Groups	1080.000	42	25.714		
Total	1830.000	44			

H_0: $\mu_{NonMigrating} = \mu_{Short\ Mig} = \mu_{Long\ Mig}$ **H_1: Not H_0**

Decision: Reject H_0, because $p < .001$

Post Hoc Test: $HSD = 3.44\sqrt{\frac{25.71}{15}} = 4.50$

	NonMig	Short Mig	Long Mig
NonMig	-----		
Short Mig	5	-----	
Long Mig	10	5	-----

There is a significant effect of type of bird on brain size, F(2,42) = 14.583, MSE = 25.714, p < .001. {Note, however, that the birds are not randomly assigned to condition (type of bird is a non-manipulated characteristic of the subjects), so one cannot make a causal claim.} Post hoc tests using Tukey's HSD indicate that the brains of non-migrating birds are significantly higher (*M* = 15) than those of birds with short (*M* = 10) or long migration patterns (*M* = 5). The short-distance migration birds have brains that are larger than those that migrate for long distances.

5. (From G&W) There is some evidence to suggest that high school students justify cheating in class on the basis of the teacher's skills or stated concern about cheating (Murdock, Miller, & Kohlhardt, 2004). Thus, students appear to rationalize their illicit behavior on perceptions of how their teachers view cheating. Poor teachers are thought not to know or care whether or not students cheat, so cheating in their classes is viewed as acceptable. Good teachers, on the other hand, do care and are alert to cheating, so students tend not to cheat in their classes. Below is a partially completed source table and summary statistics that are consistent with the findings of Murdock et al. The scores represent judgments of the acceptability of cheating for students in each sample. Complete the source table below and interpret the data as completely as you can. What is your best estimate of the population variance (σ^2)? [10 pts]

Descriptives

Acceptability of Cheating

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Good Teacher	20	2.1000	.85224	.19057	1.7011	2.4989	1.00	4.00
Poor Teacher	20	6.0500	1.27630	.28539	5.4527	6.6473	3.00	8.00
Total	40	4.0750	2.26894	.35875	3.3494	4.8006	1.00	8.00

Test of Homogeneity of Variances

Acceptability of Cheating

Levene Statistic	df1	df2	Sig.
2.147	1	38	.151

ANOVA

Acceptability of Cheating

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	156.025	1	156.025	132.491	.000
Within Groups	44.750	38	1.178		
Total	200.775	39			

In order to complete this problem, you need to compute MS_{Within} from the supplied standard deviations. First, you need to square the standard deviations to turn them into variances (.726 and 1.629), then you need to compute the mean of the variances (1.178), which is MS_{Within}. That is also the best estimate of the population variance (σ^2).

There is a significant effect of type of teacher on acceptability of cheating, $F(1,38) = 132.491$, $MSE = 1.178$, $p < .001$. Student in a class with a teacher rated as "good" view cheating as less acceptable ($M = 2.1$) than students in a class with a teacher rated as "bad" ($M = 6.05$). Note, however, that this study is not an experiment, with no random assignment to conditions.

6. Briefly define the following terms and explain why each term is important to experimental design. [10 pts]

floor effect

The task is so difficult that all scores are low (near minimum), so that any differences among groups would be obscured. Thus, you would want to be sure that you avoid a floor effect.

running in randomized replications

When using an independent groups design, it's important to ensure that you run a person in each condition before running another person in any one condition. The order in which

you run the people would be randomized, hence running in randomized replications. In a sense, it's much like counterbalancing for a repeated measures design. Doing so protects against problems that occur over time (e.g., changes in the experimenter's demeanor, equipment changes, world events). Thus, if you had four conditions (A, B, C, D), you would run your first four Ss as, for example, S_1 = A, S_2 = B, S_3 = C, S_4 = D. Then for the next four Ss: S_5 = C, S_6 = B, S_7 = A, S_8 = D. And so forth…

random assignment to conditions

In a true experiment for independent groups, it's vital that you randomly assign participants to conditions. In so doing, you should expect that any individual differences would be roughly equivalent among your conditions (especially with a sufficiently large sample size). Thus, any differences that emerge among your groups would be due to the treatment and not due to any differences among the participants in the groups.

counterbalancing

In order to distribute any order or carry-over effects evenly among the conditions, it's crucial that you run your participants in different orders. For relatively small numbers of treatment conditions (e.g., five or fewer), you would likely use complete counterbalancing. Using complete counterbalancing, with *k* treatments, you'd have *k*! orders. Using the digram-balanced approach to incomplete counterbalancing, with an even number of *k* treatments, you'd have *k* orders. With an odd number of *k* treatments, you'd have 2*k* orders.

reliability

There are actually many different types of reliability. In general, what we mean by reliability is the extent to which a measure taken on a participant would be consistent over time, presuming that the characteristic being measured doesn't change. That is, if intelligence is a consistent trait of an individual, an intelligence test would be considered reliable were it to produce the same measure of intelligence if administered over a number of different times.

7. Psychologists often use the recognition paradigm to study memory. Thus, a list of words would be presented and then there would typically be a distractor phase (count backwards by 3 from 275) and then a test that contains both new items (not in the original list) as well as old items (from the original list). Suppose that you were interested in testing the extent to which the frequency of the word in the English language had an impact on recognition memory. You acquire four lists of words that vary in frequency (from low to high frequency, labeled I, II, III, IV), with 20 words in each list. You want to use a repeated measures design in which you present the words in four blocks (all the items in each block from the same frequency group, I, II, III, or IV). Briefly describe how you would design your study and how many participants you would use. [10 pts]

First of all, you'd surely want to use counterbalanced blocks. With only four levels of word frequency, you'd use complete counterbalancing. Thus, there would be 4!, or 24 orders. For example, here are some of the orders:

I, II, III, IV	**I, II, IV, III**	**I, III, II, IV**	**I, III, IV, II**
I, IV, II, III	**I, IV, III, II**	**II, I, III, IV**	**II, I, IV, III**
II, III, I, IV	**II, III, IV, I**	**II, IV, I, III**	**II, IV, III, I**

As a result, you'd need to run in multiples of 24 participants.

Given the number of participants you propose, complete the following source table and tell me what you could conclude from his study. (You don't need to know F_{Crit}…right? And you can ignore the Subject line if you prefer.)

Source	***SS***	***df***	***MS***	***F***
Subject		23		
Treatment	40.0	3	13.3	1.0
Error	917.7	69	13.3	

This source table assumes that I used 24 participants.

Even though the means for each condition are not provided to you here, what can you tell me about the means of the four conditions?

The means would be quite similar (near identical), in order to produce such a small *F*.