1. Define both internal and external validity. Which type of validity is more important? Why? Using the Higgins and Marlatt study (drinking alcohol as a means of reducing tension) *and* the Argyle study (glasses and intelligence), tell me what Mook has to say about the importance of external validity. [10 pts]

Respond using Mook article notes.

2. Dr. Juan Moore decided to extend some of Darley and Latané's results by replicating their original experiment on bystander intervention with the addition of a couple of conditions. As you might remember, the original experiment involved 3 conditions (Group Size 2, Group Size 3, and Group Size 6). Dr. Moore decides to include Group Size 4 and Group Size 5 to see if he can determine the trend of delay in helping behavior. He decides to use 6 participants in each of his 5 conditions. Unlike the original experiment, Moore uses response time in *minutes* as his DV. Other than the additional conditions, all else remains the same (communication study in cubicles with different numbers of "participants," talking over telephone, "victim" sounds like he's having a seizure, measure how long it takes real participant to go to help the "victim"). Some incomplete information from the replication conducted by Dr. Moore is seen below.

a. Complete the source table and tell Dr. Moore what he can conclude on the basis of his experiment.

b. Suppose that Dr. Moore tells you that he wants to make his experiment more powerful by conducting a repeated measures design. What are two things that you might tell him about using that procedure in his quest for power for this study. [20 pts]

ANOVA Table for Time to Respond

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group Size	4	334.533	83.633	100.360	<.0001	401.440	1.000
Residual	25	20.833	.833				

Means Table for Time to Respond
Effect: Group Size

	Count	Mean	Std. Dev.	Std. Err.
2	6	1.667	.816	.333
3	6	3.500	1.049	.428
4	6	8.333	1.366	.558
5	6	9.667	.516	.211
6	6	9.667	.516	.211

You would reject H_0 [$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$], and conclude that there is a significant effect of group size, $F(4,25) = 100.36$, $MSE = .833$, $p < .001$. To determine which groups differ, you would need to compute Tukey's HSD:

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 4.16\sqrt{\frac{.833}{6}} = 1.55$$

Thus, any means that differed by 1.55 or more would lead you to conclude that the means were sampled from populations with different means.

	2	3	4	5	6
2	-----				
3	1.833	-----			
4	6.663	4.83	-----		
5	8.0	6.167	1.334	-----	
6	8.0	6.167	1.334	0	-----

We could conclude that respond significantly faster in Group Size 2 compared to all other group sizes. Furthermore, people in Group Size 3 respond significantly faster than groups of size 4, 5, or 6. Once the group reaches a certain size, there seem to be no additional increments in response times.

b. This study would not need the additional power of a repeated measures design because there are many significant effects. Moreover, given the deception involved, you would not be able to conduct this study with a repeated measures design (participants would catch on).

3. Dr. Kip Werkin is an industrial/organizational psychologist who is interested in the impact of environmental factors (such as noise) on productivity. He has a group of workers experience each of a set of background noise levels (70 dB, 80 dB, 90 dB, and 100 dB) as they work on a project that involves creating delicate instruments. The dependent variable is the number of errors made in the construction of the pieces. Complete the source table and tell Dr. Werkin what he should conclude from this study. [10 pts]

ANOVA Table for SPL

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	4.000	.444				
Category for SPL	3	13.900	4.633	22.339	<.0001	67.018	1.000
Category for SPL * Subject	27	5.600	.207				

Means Table for SPL
Effect: Category for SPL

	Count	Mean	Std. Dev.	Std. Err.
SPL 70 dB	10	.200	.422	.133
SPL 80 dB	10	.200	.422	.133
SPL 90 dB	10	1.000	.667	.211
SPL 100 dB	10	1.600	.516	.163

First of all, you would reject H_0 [$\mu_{70} = \mu_{80} = \mu_{90} = \mu_{100}$], concluding that the noise level had an impact on number of construction errors made, $F(3,27) = 22.339$, $MSE = .207$, $p < .001$. To determine which specific groups differed, you would compute Tukey's HSD:

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 3.86\sqrt{\frac{.207}{10}} = .555$$

Thus, the 100 dB group produced significantly more errors than all other groups. The 90 dB group made more errors than the 70 dB and the 80 dB groups.

3a. If the ***same*** data were analyzed with an independent groups design, what would the source table look like? Under which conditions would a repeated measures analysis of a data set not lead to a larger F-ratio than an independent groups analysis? **If the SS_{Subj} is relatively small, then the repeated measures ANOVA will not yield a larger F.** [5 pts]

Source	df	SS	MS	F
Treatment	3	13.9	4.633	17.37
Error	36	9.6	.267	
Total	39	23.5		

4. Dr. Bing Bada is convinced that sleep deprivation influences aggression. To test this assumption, volunteer participants are randomly assigned to sleep-deprivation periods of 0, 24, 28, or 72 hours and subsequently tested for aggressive behavior in a controlled social situation. The dependent variable (aggression) is operationally defined as the number of times the participant hits or curses at a confederate who attempts to annoy the participant. Complete the StatView output below and then tell Dr. Bada what to do next. [10 pts]

ANOVA Table for Agg Act

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Sleep Depriv	3	.275	.092	.264	.8509	.792	.095
Residual	36	12.500	.347				

Means Table for Agg Act
Effect: Sleep Depriv

	Count	Mean	Std. Dev.	Std. Err.
0	10	.200	.422	.133
24	10	.300	.483	.153
28	10	.400	.699	.221
72	10	.400	.699	.221

First, of course, you would retain H_0 because p > .05. Next, you would try to address the reason(s) that your study did not result in a significant effect. In other words, you will look for ways to increase power.

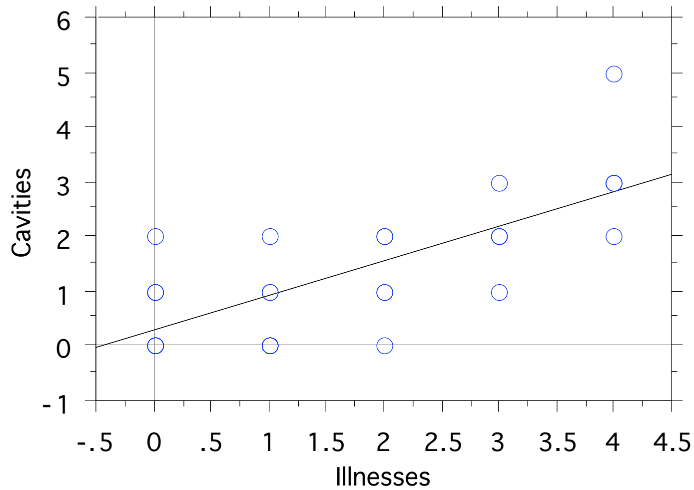
Dr. Bada used a no-deprivation control group, as well as a fairly wide range of periods of sleep deprivation. As a means of increasing the treatment effect, of course, you could always increase the range of sleep deprivation (96 hours, etc.). You might also look for a more sensitive DV than number of times that the participant hits or curses at someone else (e.g., include other forms of verbal abuse, or other forms of aggression). You should always consider increasing the sample size (only *n* = 10 here). Finally, you could reduce your error variability by reducing individual differences (e.g., use only males) or by reducing random variability (e.g., ensuring that instructions are clear, that testing is done in a quiet, private space).

5. Dr. Ginger Vitas is a health psychologist who is interested in the relationship between dental heath (operationally defined as number of cavities found in an annual checkup) and general health (operationally defined as the number of illnesses experienced in the preceding year). Analyze the output seen below as completely as you can. If a person had 3 illnesses in a year, how many cavities would you predict? If a person had 6 illnesses, how many cavities would you predict? What proportion of variability do these two variables share? [10 pts]

Regression Summary
Cavities vs. Illnesses

Count	30
Num. Missing	0
R	.738
R Squared	.545
Adjusted R Squared	.529
RMS Residual	.855

Regression Plot



Y = .276 + .629 * X; R^2 = .545

ANOVA Table
Cavities vs. Illnesses

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	24.501	24.501	33.521	<.0001
Residual	28	20.466	.731		
Total	29	44.967			

Regression Coefficients
Cavities vs. Illnesses

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	.276	.245	.276	1.128	.2691
Illnesses	.629	.109	.738	5.790	<.0001

There is a significant positive linear relationship between illnesses and number of cavities, *r*(28) = .738, *p* < .001. If a person had 3 illnesses, you would predict 2.16 (~2) cavities. If a person had 6 illnesses, you would not be able to make a safe prediction, given that you hadn't observed anyone with more than 4 illnesses. If the trend continued (a dicey prediction), you would predict 4.05 (~4) cavities. The two variables share .545 (54.5%) of their variability (r^2).

5.Conceptually, how does the error term (MS_{Error}) differ between an independent groups and a repeated measures ANOVA? Why do the error terms need to differ? [5 pts]

In a repeated measures ANOVA, the error term comes from random variability only, whereas in the independent groups ANOVA, the error term comes from individual differences and random variability. The reason for the difference is that in the repeated measures ANOVA, the numerator contains only treatment effects and random variability (no individual differences), so the appropriate error term needs to contain only random variability.

6. What is a manipulation check? Give an example of a situation in which you would want to conduct a manipulation check, including an illustration of what it would look like. [5 pts]

A manipulation check is used to determine if the manipulation has had its intended effect. You should consider a manipulation check when the impact of the independent variable is not readily observable (aside from its presumed impact on the dependent variable). For example, suppose that you wanted to study the effects of self-esteem (e.g., on people's willingness to engage in charitable behavior) in an experimental fashion (which may raise ethical issues, but let's leave those issues aside for this example). You could do something to your participants (e.g., give them false feedback from a self-esteem measure) that would place them into either the Low or High Self-Esteem group. The problem, of course, is that self-esteem in not readily observable, so you can't be sure that your false feedback actually influenced the participants' self-esteem. Under these circumstances, you may well use a manipulation check (at the end of the experiment) to determine if the self-esteem of the two groups actually differed in the predicted fashion.

If you were dealing with an independent variable that is less ambiguous, no manipulation check would be needed. For example, if you were interested in assessing the impact of lighting levels in a room (e.g., on reading speed), you can measure the lighting levels easily (and precisely) so no manipulation check would be needed.