

1. First, some random questions about topics we covered this semester. [10 pts]

a. In a repeated measures design, what is the effect of counterbalancing on order or carry-over effects?

Counterbalancing distributes any order or carry-over effects equally among the conditions (it doesn't eliminate the effects).

b. In a repeated measures design with 7 levels of the factor, how many participants would you need to run if you used incomplete counterbalancing and wanted to have a minimum of 40 scores per cell/condition?

You'd need 42 participants, with 14 different orders used three times each.

c. Both the standard deviation of a sample (s) and the standard error (s_M or $s_{\bar{X}}$) are standard deviations. How do they differ?

The standard deviation for a sample measures the variability of raw scores. It serves as an estimate of the population standard deviation. The standard error measures the variability of sample means in the sampling distribution of the mean. The standard error will always be less than the population standard deviation, as seen in the formula $\sigma_{\bar{X}} = \sigma / \sqrt{n}$.

d. Provide one example of a parametric statistic and one example of a nonparametric statistic.

The F ratio is a parametric statistic, as is r . Spearman's rho is a nonparametric statistic.

e. You're interested in testing $H_0: \mu = 10$. Under which circumstances would you compute a z -score and under which circumstances would you compute a t -test?

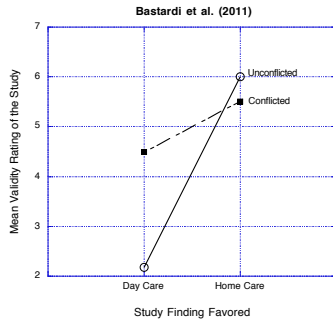
In the extraordinary situation that you would know σ , you could compute a z to test H_0 . However, the much more likely scenario is that you would not know σ , but would estimate it using s and then compute a t .

2. Bastardi et al. (2011) published a paper titled "Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence." (I'll simplify the study a bit.) In their study, participants read a scientific article. For half of the people, the "scientific article" provided results *favoring day care*. The other half of the people read a "scientific article" that produced results *favoring home care*. Before the study (i.e., a pre-test), *all* participants in this study had expressed a preference for home care for their children. However, half of the people actually were going to provide home care for their children (so they were Unconflicted) and the other half were going to place their children in day care (so they were Conflicted). The dependent variable was a rating of how valid the participants rated the studies to be, ranging from 1 (*studies were extremely unconvincing/invalid*) to 7 (*studies were extremely convincing/valid*). Below are data that are roughly consistent with the results of the Bastardi et al. study. Analyze these data and interpret the results as completely as you can. [25 pts]

	The Study Favored Day Care		The Study Favored Home Care	
	Conflicted	Unconflicted	Conflicted	Unconflicted
	5	3	5	6
	4	2	6	7
	6	1	5	5
	3	3	6	6
	5	2	6	6
	4	2	5	6
$\Sigma X (T)$	27	13	33	36
$\bar{X} (M)$	4.5	2.17	5.5	6.0
SS	5.5	2.83	1.5	2.0
s	1.049	.753	.548	.632
ΣX^2	127	31	183	218

Source	SS	df	MS	F
Conflicted	5.1	1	5.1	8.6
Study	35.1	1	35.1	59.5
Conflicted x Study	11.97	1	11.97	20.3
Within	11.83	20	.59	
Total	64	23		

$F_{Max} = 5.5/1.5 = 3.7$ and $F_{MaxCritical} = 13.7$, so no concern about heterogeneity of variance and I'd proceed with $\alpha = .05$. Thus, $F_{Critical}(1,20) = 4.35$.



$$HSD = 3.96 \sqrt{\frac{.59}{6}} = 1.24$$

There was a significant main effect for whether or not the parent participant was conflicted or not, $F(1,20) = 8.6$, $MSE = .59$, $p < .05$, $\eta^2 = .30$. There was a significant main effect of the finding of the study, $F(1,20) = 59.5$, $p < .05$, $\eta^2 = .75$. There was also a significant interaction between the participant's conflict and the finding of the study, $F(1,20) = 20.3$, $p < .05$, $\eta^2 = .50$. As seen in the figure above, and confirmed by Tukey's HSD, when the study outcome was in favor of home care, whether or not participant's were conflicted, they rated the article as highly (and equally) valid. However, when the study was in favor of day care, those who were conflicted rated the study as more valid than did those who were unconflicted.

3. Tybur et al. (2011) published a paper titled "Smells Like Safe Sex: Olfactory Pathogen Primes Increase Intentions to Use Condoms." We can think of the study as a 2x2 independent groups design. Let's say that the participants were 80 undergraduates. Half of the participants were Sexually Active (had sex with at least one person within the past year) and half were Not Sexually Active (no sex in the past year). Before filling out a bunch of questionnaires, and after a saliva test, each participant left the testing room to use a water fountain.

"In the pathogen-prime condition, while participants were out of the room, the experimenter sprayed the wall of the room with a single pump of 'Liquid ASS,' a novelty odor liquid that smells strongly of common bacterial threats (e.g., feces). In the control condition, the experimenter did not administer the spray."

Thus, there were two levels to the Odor condition (Odor vs. No Odor). The DV is the intention to use a condom (1 = "not at all likely" to 7 = "extremely likely"). Complete the source table below, then interpret the data as completely as you can. [10 pts]

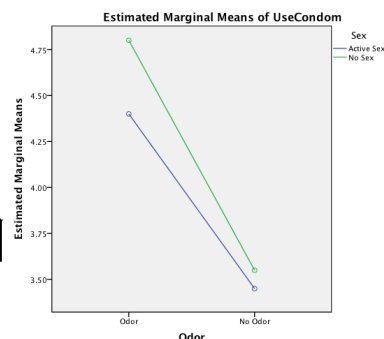
Descriptive Statistics
Dependent Variable: UseCondom

Odor	Sex	Mean	Std. Deviation	N
Odor	Active Sex	4.4000	.68056	20
	No Sex	4.8000	.89443	20
	Total	4.6000	.81019	40
No Odor	Active Sex	3.4500	.51042	20
	No Sex	3.5500	.68633	20
	Total	3.5000	.59914	40
Total	Active Sex	3.9250	.76418	40
	No Sex	4.1750	1.00989	40
	Total	4.0500	.89866	80

Levene's Test of Equality of Error Variances^a
Dependent Variable: UseCondom

F	df1	df2	Sig.
2.036	3	76	.116

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Odor + Sex + Odor * Sex



Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power ^b
Odor	24.200	1	24.2	48.4	.000	.390	1.000
Sex	1.250	1	1.25	2.5	.118	.032	.346
Odor * Sex	.43	1	.43	.86	.345	.012	.155
Error	37.92	76	.499				
Corrected Total	63.800	79					

At the outset, you might well consider the confound presented by the lack of a control group that experienced a non-feces odor. That is, as currently designed, one cannot be certain that it's the feces odor that is affecting the participants or any odor at all.

The simplest way to complete the ANOVA is to realize that MS_{Error} is the average of the condition variances, which you could determine by squaring the four sample standard deviations and taking their mean. You could also compute the SS for the main effects by multiplying the means by the appropriate n to determine the T values.

Given the Levene test, there is no reason to be concerned about violating the homogeneity of variance assumption. You'd reach the same conclusion had you chosen to compute Hartley's $F_{\text{Max}} = 3.08$. And, of course, given that the only significant effect had a p -value of $< .001$, your effect would be significant even had the homogeneity of variance assumption have been violated.

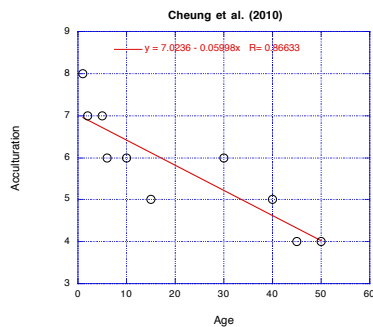
There was no main effect for whether or not the participants were sexually active, $F(1,76) = 2.5$, $MSE = .499$, $p > .05$, $\eta^2 = .032$. There was no interaction between sexual activity level and whether or not the odor was present, $F(1,76) = .86$, $p > .05$, $\eta^2 = .012$. However, there was a main effect for the type of odor, $F(1,76) = 48.4$, $p < .05$, $\eta^2 = .39$. Participants who were exposed to the Liquid Ass odor reported a significantly higher likelihood of using a condom ($M = 4.6$) than those who were not exposed to the feces odor ($M = 3.5$).

4. If the above ANOVA had been computed as a one-way ANOVA on the odor in the room (Odor), how would the source table change (illustrate below)? [5 pts]

Source	SS	df	MS	F
Odor	24.2	1	24.2	47.45
Within	39.6	78	.51	
Total	63.8	79		

5. Cheung et al. (2010) published a paper titled “Evidence for a sensitive period of acculturation: Younger immigrants report acculturating at a faster rate.” In one of their analyses, they looked at the relationship between age at immigration to Canada for a number of people from Hong Kong and a measure of how acculturated to Canadian culture they believed themselves to be. They had a lot of data, but I’ll give you a more limited data set to analyze as completely as you can. [15 pts]

	Age at Immigration (in Years)	Acculturation Score	XY
	1	8	8
	50	4	200
	30	6	180
	10	6	60
	15	5	75
	2	7	14
	40	5	200
	6	6	36
	5	7	35
	45	4	180
ΣX	204	58	988
$\bar{X} (M)$	20.4	5.8	
SS	3254.4	15.6	
s	19.02	1.32	



$$r = \frac{988 - \frac{204 * 58}{10}}{\sqrt{3254.4 * 15.6}} = -.87$$

Coefficient of determination, $r^2 = .757$

$$r_{\text{Critical}}(8) = .632$$

There is a significant negative linear relationship between Acculturation and Age, $r(8) = -.87, p < .05$.

Given the significant relationship, it’s appropriate to compute a regression equation: $\hat{y} = -.06x + 7.02$

The Standard Error of Estimate would be: $SEE = \sqrt{\frac{3.8}{8}} = .69$

6. Helmrich, Aronson, and LeFan (1970) examined the effects of seeing a person commit a social blunder. The participants were divided into three conditions on the basis of their self-esteem scores (Low, Medium, High). Some participants in each self-esteem group saw a competent person accidentally spill a cup of coffee on the floor (Spill). The other participants saw the competent person in the same situation but not spilling the coffee (No Spill). All participants were asked to indicate how much they liked the person on a 20-point scale (20 = like a lot). Below is a partially completed source table consistent with results from this study. Complete the source table and then interpret the results as completely as you can. [15 pts]

Descriptive Statistics

Dependent Variable: Liking

SelfEsteem	CoffeeSpill	Mean	Std. Deviation	N
Low	No Spill	12.2000	2.04396	10
	Spill	5.0000	1.24722	10
	Total	8.6000	4.04449	20
Medium	No Spill	8.0000	1.76383	10
	Spill	13.8000	1.31656	10
	Total	10.9000	3.33877	20
High	No Spill	13.3000	1.49443	10
	Spill	10.3000	1.56702	10
	Total	11.8000	2.14231	20
Total	No Spill	11.1667	2.88974	30
	Spill	9.7000	3.91417	30
	Total	10.4333	3.49026	60

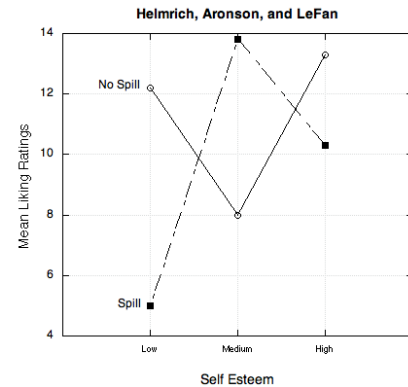
Levene's Test of Equality of Error Variances^a

Dependent Variable: Liking

F	df1	df2	Sig.
.651	5	54	.662

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + SelfEsteem + CoffeeSpill + SelfEsteem * CoffeeSpill



Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power ^b
SelfEsteem	108.7	2	54.4	21.4	.000	.442	1.000
CoffeeSpill	32.3	1	32.3	12.7	.001	.190	.938
SelfEstm * CoffSpill	439.4	2	219.7	86.5	.000	.762	1.000
Error	137.4	54	2.54				
Corrected Total	717.8	59					

Given the Levene test ($p = .662$), I wouldn't be concerned about having violated the homogeneity of variance assumption, so I'd use $\alpha = .05$.

$$HSD = 4.2 \sqrt{\frac{2.54}{10}} = 2.12$$

There was a significant main effect of self esteem, $F(2,54) = 21.4$, $MSE = 2.54$, $p < .001$, $\eta^2 = .442$. There was also a main effect of whether or not they observed the person spill coffee, $F(1,54) = 12.7$, $p = .001$, $\eta^2 = .19$. The interaction between self esteem and whether or not the person spilled coffee was also significant, $F(2,54) = 86.5$, $p < .001$, $\eta^2 = .762$. As seen in the figure, and confirmed by Tukey's HSD, for people with low and high self esteem, the liking rating for the person was significantly higher when the person had not spilled coffee than when the person had spilled coffee. However, for people with medium self esteem, they rated the person who spilled coffee as higher than the person who had not spilled coffee.

7. Whitchurch et al. (2011) published an article titled “He loves me, he loves me not...”: Uncertainty can increase romantic attraction.” I’ll modify their study slightly, while keeping to the basic results. The participants were women who were told that some men had read their Facebook profile. They were then shown the Facebook profile for three of the men who had seen their profile. They were told that one of the men “thought that he would like you the best.” They were told that another of the men “thought that he would like you about average.” Finally, they were told that the third man had read their profile, but had not made any response to their profile. Thus, we can consider the men as three levels of the factor: Best, Average, Uncertain. The women participants read each of the men’s profiles one at a time and then rated each guy on a scale of 1-8 (1 = “not attracted” to 8 “extremely attracted”). Complete the analysis below and interpret the results as completely as you can. [10 pts]

Descriptive Statistics

	Mean	Std. Deviation	N
Best	5.0000	1.00000	15
Average	3.5333	1.06010	15
Uncertain	6.8000	1.01419	15

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power ^a
Liking	Sphericity Assumed	80.311	2	40.15	81.9	.000	.854	1.000
Error(Liking)	Sphericity Assumed	13.689	28	.49				

First, you should note that there is likely a counterbalancing problem. That is, with three levels of the repeated factor, one would use complete counterbalancing, which leads to six orders (and multiples of six participants).

$$HSD = 3.5 \sqrt{\frac{.49}{15}} = .63$$

The type of man responding to the Facebook profile had a significant effect on ratings of attractiveness of the man, $F(2,28) = 81.9$, $MSE = .49$, $p < .001$, $\eta^2 = .854$. Tukey’s HSD indicated that women found the men most attractive when the women had no information about how the men rated them ($M = 6.8$), compared to men who liked them best ($M = 5.0$) and men who rated them as average ($M = 3.533$). The men who liked them best were rated as more attractive than the men who liked them an average amount.