1. The topic that struck me as most important was this scientific obsession with dichotomous findings i.e. obtaining a p value <.05. Reality is not split into significant and non-significant findings, we (especially in neuroscience) have to make it acceptable to convey more sophisticated information than P(H0|data) instead of making it the punchline of any study. Indeed, I don’t have any issue with the logic of p<.05 (especially when the question is pre-registered) but this such a critical topic because of how it has guided scientific research. It provides incentives for researchers to ask questions that result in p<.05 even though that may not even be the true question of interest! This obsession introduces conscious (i.e. fraud) and unconscious (i.e. running more subjects until p<.05) biases into science that could be limited if we as a field accepted more continuous conclusions. During my time at Massachusetts General Hospital I remember calculating a p value of 0.07. The first question a post-doc had for me was ‘well why is your hypothesis false?’ I pushed back, arguing that there was still evidence that my Ha was true, I just didn’t have enough evidence to reject H0 in this specific context. This visceral experience eloquently captured the field’s view on p values and to me, drove home just how far we have to go was a field. It’s devastating to think of all of the promising lines of science that were abandoned just because of an arbitrary p-value that was above 05.

Appropriately, another topic that Rosnow and Rosenthal bring up that I believe is important is the danger of working with low power. I do agree that on a whole, type 1 errors can be more serious than type 2 but what is the purpose of conducting experiments if you’re working with a sample size that is probably too small to detect your Ha? The powerless plot shown in the lecture slides drives this point home: we need quite a large sample size to even *detect* (on the order of 40 plus for even robust effect sizes). I think researchers tend to value logistics or finances in terms of designing experiments when the mentality should be: if I’m spending this money anyways, I should design a study that’s powerful enough to answer my question of interest. Power analyses can help answer these logistical questions. I actually experienced nearly the very same Smith v. Jones example Rosnow and Rosenthal discuss; I was the one who tried replicating study A with much fewer mice and ultimately, my p value exceeded .2. However, I ended up calculating an effect size that was even larger than study A’s, indicating that my study was probably underpowered. It was a costly lesson for me to spend more time designing experiments and to confirm that I’m not handicapping myself before I’ve even started. Consequently, I think these two topics independently support the conclusion that we need to start including effect sizes and confidence intervals in our reports and recognizing how weak certain experimental designs are.

2a. Geometrically, orthogonality can be calculated if the product of two vectors is 0 i.e. they are perpendicular and have no overlap. In relevant statistical terms, orthogonality means that two vectors are independent from one another. This is the crux of multivariate statistics, if vectors were always independent from one another, univariate statistics could always be used.

2b. In the example of multiple regression, orthogonality between your predictors is extremely important. If your predictors covary (i.e. they are not independent/orthogonal), then it becomes very difficult to understand which one of the predictors actually contributed to the variance of the DV. This is referred to as a shared variance problem and must be controlled for by considering the covariances between predictors. Orthogonality is also a critical component of PCA. Simplistically, PCA takes R, a correlation matrix, and creates a new axis by using the off diagonals of R (covariances) to stretch the diagonals of R (variance of new axis) in a way that maximizes the variance of the original data. Once the covariance can’t be used to expand the new axis anymore, it moves on to expanding the next axis (another principal component). The inherent nature of this process means that the variance of one principal component is totally separate from the variance of another component which means that these principal components are now orthogonal to one another. This is a critical benefit of conducting PCA because your generated components are sure to be independent from one another. Orthogonality is also relevant when choosing which contrasts to run. Contrasts are essentially comparisons between levels in a factor that can indicate precisely which levels differ from others (thus contrasts are only used when there are more than 2 levels in a factor). It’s best to select orthogonal contrasts because they are non-overlapping comparisons. It allows you to run multiple comparisons within a factor and you don’t have a shared variance problem because the answer to one contrast is totally separate to the answer to another contrast (function of orthogonality). If you run non-orthogonal contrasts (which is possible) than it becomes difficult to separate the answers of each contrasts from one another.

3a. All three of these techniques deal with analyzing categorical outcomes and use maximum likelihood estimates. Logistic and logit techniques both involve predicting one categorical outcome (logit analysis can deal with having more than 2 levels for the DV – as opposed to logistic regression which has continuous predictors and would require something like softmax expansion to predict multiple levels of the outcome) while loglinear analysis can deal with multiple categorical outcome variables. This is precisely because loglinear analysis doesn’t formally characterize predictors and outcomes. Loglinear analysis involves testing your *model* of specified effect terms using the full data set with all terms. Loglinear and logit analyses are similar in that they both have to deal with categorical predictors and their mode-of-action involve dealing with the inherent unbalanced design of categorical DVs and IVs by log-transforming the categorical frequency and probability counts into linearized versions. Logit analysis however, doesn’t include the model terms that don’t include the DV and thus allows for analysis of the DV. Logistic regression also has a clear DV, but it calculates ‘standard’ residuals by fitting a curved regression line to the probability of the outcome event occurring. Thus, logistic regression can deal with categorical and continuous predictors.

3b. If you had data indicating whether someone voted from Trump or Clinton, whether they were male or female, whether they identified as liberal, conservative, or independent, and whether they were happy with their choice a year later (yes/no), you could use loglinear analysis to analyze your model terms of interest. This would allow you to test all your terms simultaneously (interaction and main effects of interest) to see, say, if happiness interacted with political orientation and/or that was affected by vote. A similar example could suffice for logit analysis, let’s say your only DV of interest was emotion and it was modified to include multiple categories of emotion. You could run a log linear analysis and exclude the terms that don’t include emotion to simultaneously calculate the main effect of voter choice and/or the interaction of voter choice and gender on emotion. A logistic regression would be a different paradigm, say you were interested in only predicting vote for Clinton (1 for yes, 0 for no). You could include both categorical and continuous predictors for this question (i.e. SES, political ideology, gender, age etc.) to understand which predictors significantly affect the *probability* of voting for Clinton.

4a. Simple effects refer to the question: does group membership in different levels of a factor matter? Simple effects only refer to one factor and importantly, simple effects collapse across everything else in the design. A simple effect would ask answer the question of whether one level of a factor is different than another level in that same factor. This can be answered via contrasts to identify precisely where differences lie (note: there are k-1 contrasts per main effect; k referring to the number of levels in the factor of interest). Interaction effects are *separate* from main effects and look at the comparison of comparisons. Interactions can also be calculated by specific interaction contrasts and precisely answers the question of: is the effect of one factor dependent on the presence of another factor? If so, an interaction is present. Another way to qualify an interaction is to say that two (or more) main effects interact in a non-additive way i.e. the outcome is different than you would expect if you just summed up the main effects.

4b. In the 2x3 design, if you’re just interested in the effect that the drug has (regardless of patient condition) than you would just run a factorial ANOVA to calculate this simple effect. You wouldn’t need to see if there’s an interaction because you could collapse across patient condition. However, if your research question was different: did you want to see if the drug was more or less effective on different patients? In that case, you would need to calculate the main effect of drug (collapsed across patient condition). You would also need to calculate the main effect of patient condition collapsed across drug and placebo, which is a very bizarre effect to collapse. Understanding this particular simple effect also requires 2 contrasts (ideally contrasts that are orthogonal) because there are 3 levels. The interaction term would be calculated by seeing if the effect of drug was different than just predicted by the summation of the main effects. In this case of analyzing the interaction, you would need to craft specific interaction contrasts (there would be 2 contrasts to test here) to determine exactly where the interaction is/answer your question of interest.

In the 2x2 design you would certainly be interested in the interaction term as well as the main effects in this factorial ANOVA. In this case, you would calculate the main effect of being married or not on childlessness (collapsing across orientation), the other main effect of orientation on childlessness (collapsing across marriage status), and the interaction term of does the effect of being married/unmarried *change* as a function of orientation and testing these effects against chance. Because is a 2x2 there is only one interaction term to interpret but all three terms (2 simple effects and 1 interaction effect) would be of interest.

5a.

|  |  |
| --- | --- |
|  | I used R to generate a pseudo-random data file that I’ll use for this analysis. A screenshot is shown here, with 300 participants. It has 7 variables (300 rows, each row representing a different person). Group: 0 if non-alcoholic group, 1 if low alcohol content and 2 for high alcohol content. Depr0, depr60, and depr120 refer to the ordinal depression score (for simplicity, I had depression be scored 1-10; 10 being more depressed) taken at the 3 time points. Prod0, prod60, and prod120 refer to the continuous productivity score (1-100; 100 = most productive) taken at the 3 time points. |

b. The study is a multi-factorial multiple-measure doubly-multivariate design with both within-subject and between-subject factors. The between-subject factor is a categorical ‘group’ variable with 3 levels analyzing differences between no alcohol, low alcohol, and high alcohol. The within-subject ‘time’ factor analyzes how individuals change over the 120 day period, also with 3 levels. There are two continuous outcome variables: depr (scored between 1-10) and prod (scored between 1-100).

c. The most appropriate analysis that considers all the data is to run a repeated-measure mixed(between-within) 3(group) X 3(time) X 2(prod and depr) ‘doubly-multivariate’ MANOVA. The possible effects here are a group main effect (there are two contrasts here because of the 3 group levels), a time main effect (also two contrasts here because of the 3 group levels), and an interaction effect (4 interaction contrasts here because of the 3x3 design). The number of these possible effects are doubled because they exist for depr and for prod. To consider all of the data at once, we can look at the generated discriminated functions that combine these DVs and account for covariance. In terms of choosing contrasts, to look at general shifts in group mean, specifically between baseline and everything else, I specified orthogonal Helmert contrasts for both the within-subject and between-subject factors (I could have also analyzed linear and polynomial contrasts for time, it depends on the specific hypotheses I want to test).

d. MANOVA depr0 depr60 depr120

prod0 prod60 prod120

BY group(0,2)

/WSFACTORS = time(3)

/CONTRAST(time) = HELMERT

/RENAME =constdepr firstHdepr secondHdepr constprod firstHprod secondHprod

/CONTRAST(group) = HELMERT

/print = transform

d. The question of whether wine intake showed any benefit can be captured by analyzing the main effect of time. We can collapse across placebo, low alcoholic content, and high alcoholic content (since we’re just interested in if wine showed any benefit, regardless of alcoholic content) and look at the contrasts of time across productivity and depression (since both outcomes would be considered to be beneficial). To just get a simple yes/no answer to the question (did drinking wine have any benefits across time?), I looked at the single discriminant function that combined the two within-subject contrasts for depr and two within-subject contrasts for prod into a single omnibus test. SPSS gives me an output of all of the different multivariate tests of significance which displays the main effect of time on the combination (discriminant function) of these dependent measures.