PSY 216 Exam 4 Review

Multi-level experiment – an experiment with a single IV that has more than 2 levels

- level – a value that an IV can take on
- as the number of levels increases, the number of comparisons needed to compare all levels to all other levels increases very rapidly: $(n^2 – n) / 2$ where n is the number of levels
- as the number of comparisons increases, the probability of making at least 1 Type-I error increases rapidly: $\alpha_{\text{family-wise}} = 1 – (1 - \alpha_{\text{comparison wise}})^N$
  - $\alpha_{\text{family-wise}}$ = probability of making at least 1 Type-I error
  - $\alpha_{\text{comparison-wise}}$ = probability of making exactly 1 Type-I error in a single comparison
  - N = number of comparisons being made
  - Type – I error: incorrectly rejecting H_0 when H_0 is true
- Do not use multiple t-tests to analyze multi-level experiments as the $\alpha_{\text{family-wise}}$ will be large and you cannot determine which comparison(s) suffer from a Type-I error

Analysis of Variance (ANOVA) is an inferential statistic that can be used to analyze the results of multi-level experiments when the DV is interval or ratio scaled

- H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$
  - The omnibus null hypothesis – covers all aspects
  - Always 2-tailed
- H_1: not H_0
- ANOVA provides two estimates of the variability in the data
  - Between-groups estimate of variance – how different the distributions for each condition are from each other
    - Caused by the effect of the IV (treatment) on the DV – the larger the effect, the larger between-groups variance will be
    - Caused by sampling error – even if the IV has no effect on the DV, we would not expect the means of the conditions to be exactly equal
    - To reject H_0, between-groups variance should be as large as possible
  - Within-groups estimate of variance – how different the scores are from each other within a given condition
    - Caused by things that were not, or could not be controlled for in the study, such as individual differences
    - To reject H_0, within-groups variance should be as small as possible
- F = between-groups variance / within-groups variance
  - If H_0 is true, the expected value of F is 1 (effect of IV on DV is 0, leaving error / error = 1)
  - If H_0 is not true, the expected value of F is larger than 1 (effect of IV on DV is greater than 0, leaving (greater than 0 + error) / error
- Reject H_0 when the observed F is at least as large as the critical F
  - Size of critical F increases as α decreases
  - Size of critical F decreases as sample size increases
  - As long as the within-groups degrees of freedom is larger than 2 (which is almost always true in a well designed study), the size of critical F decreases as the number of levels of the IV increases
  - Size of observed F tends to increase as the difference between the levels of the IV increases
  - Size of observed F tends to increase as experimental control increases
- ANOVA summary table summarize the key results of ANOVA
  - Source column – the type of variance: between-groups, within-groups, total
  - SS – sum of squares – the sum of the squared deviate scores – numerator of variance estimate

- df – degrees of freedom – the number of scores free to take on any value after restrictions have been placed on the data set – denominator of variance estimate
    - $df_{\text{between-groups}}$ = number of levels of IV – 1
    - $df_{\text{within-groups}} = \Sigma$(number of participants in a condition – 1) summed across conditions
  - MS – mean square – SS / df = variance estimate
  - F – the observed F ratio = $MS_{\text{between-groups}} / MS_{\text{within-groups}}$ – reject H_0 when observed F ≥ critical F
  - p – the p value or significance level – reject H_0 when $p \le \alpha$
- ANOVA assumptions
  - Sampling error is normal (or Gaussian) in shape; centered around mean of distribution
  - Homogeneity of variance – variance in all conditions approximately equal
  - Independence of observation – knowing one score tells you nothing about the other scores
  - ANOVA is robust to violations of the assumptions of normality and homogeneity of variance as long as sample size is large and equal across conditions
- ANOVA only tells us that it is probably the case that not all the means are equal. It does not tell us which means are likely to be different from which other means
- Different ANOVAs for between-subjects (different people in each condition) and within-subjects (same people in all conditions) designs

Multiple comparison (MC) techniques tell us which means are likely to be different from which other means
- Similar to performing multiple t-tests, but with additional protection from making Type-I errors
- Many types of MCs – Tukey, Newman-Keuls, etc. – they differ in the amount of protection from Type-I error and statistical power (which trade off)
- Perform Tukey tests only if the IV has more than 2 levels *and* the corresponding main effect of that IV is statistically significant
  - H0: $\mu_1 = \mu_2$
  - H1: $\mu_1 \neq \mu_2$
  - HSD = $q_{\alpha,k,df_{\text{within-groups}}}\sqrt{\frac{MS_{\text{within-groups}}}{n}}$, q = tabled q value with specified α level, k = number of levels of the IV, $df_{\text{within-groups}}$ = degrees of freedom of within-groups estimate of variance, $MS_{\text{within-groups}}$ = mean square within-groups, n = number of participants in each condition
  - Take the difference of the means; if the difference is at least as large as the HSD, reject H_0
  - Repeat as necessary for other comparisons
- A-posteriori (post-hoc) tests (such as Tukeys) do not specify which comparisons will be made until after the data have been collected and partially analyzed; performed only after ANOVA
- A-priori tests specify which comparisons will be made prior to data collection; tend to have greater statistical power than a-posteriori comparisons

Factorial designs are experiments in which there are at least two IVs and all possible combinations of the levels of each IV occur
- Factor ≈ IV
- n X m design has two IVs, first with n levels and the second with m levels; there are n X m conditions
- n X m Factorial designs answer three questions
  - Is there a main effect of the first IV (does the first IV influence the DV)?
  - Is there a main effect of the second IV (does the second IV influence the DV)?
  - Are the effects of the two IVs independent of each other (do the two IVs interact)?

- A main effect occurs when an IV influences the DV (the mean value of the DV for all conditions at one level of the IV is sufficiently different from the mean value of the DV for all conditions at some other level of the IV)
    - Can be determined either from factorial designs or multi-level experiments with a single IV
- Interaction effects occur when the main effects of two IVs are not independent of each other
    - Interaction occurs when the simple main effect of 1 IV is different depending on the level of the other IV
    - Interaction occurs when the effects of the IVs are not additive (you cannot add the effects of the IVs to one condition to predict the value of all other conditions)
    - Interaction occurs when the lines on a graph of the results are not parallel

Factorial ANOVA answers the questions of the n X m factorial design by providing three F ratios

- For the main effects: H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$ (subscripts represent level of the IV) H_1: not H_0
- For the interaction: H0: $\mu_{1,1} - \mu_{2,1} = \mu_{2,1} - \mu_{2,2}$ (first subscript represents level of first IV, second subscript represents level of second IV) H_1: not H_0
    - States that the distance between pairs of points are equal; i.e. lines are parallel
- ANOVA summary table as before, but with three between-groups estimates of variance – one for each of the two main effects and one for the interaction
- APA style: $\underline{F}(df_{between\text{-}groups}, df_{within\text{-}groups})$ = F value, $\underline{p}$ = p value, $\underline{MS}_{error}$ = MSwithin-groups, $\alpha = .05$
- Three types of ANOVA for n X m designs
    - Between-subjects – all IVs are manipulated between-subjects (different participants in all conditions)
        - Only 1 estimate of $MS_{within\text{-}groups}$
        - Lowest statistical power
    - Mixed-design – at least 1 IV is between-subjects and at least 1 IV is within-subjects
        - 2 estimates of $MS_{within\text{-}groups}$ – one for the main effect of the between-subjects IV, one for the main effect of the within-subjects IV and the interaction
        - Intermediate statistical power
    - Within-subjects – all IVs are manipulated within-subjects (same participants in all conditions)
        - 3 estimates of $MS_{within\text{-}groups}$ -- One for main effect of first IV, one for main effect of second IV, one for the interaction
        - Highest statistical power
    - Always use the appropriate ANOVA for a given design
- ANOVA can be used with any number of IVs, but 3 or 4 is usually an upper limit due to complexity of design
    - Higher-order interactions – interaction of interactions – nature of a lower-order interaction depends on the level of third (or higher) IV

Parametric statistics, such as t-tests and ANOVA make assumptions about the distribution of the data

- usually the assumption of normality
- increases statistical power
- cannot be used with nominal (and technically ordinally) scaled variables

Non-parametric statistics do not make assumptions about the distribution of the data set

- have lower statistical power
- can be used with any level of measure (especially if other assumptions of the parametric test have been violated)

Binomial test is a non-parametric test used with nominally scaled (or above) data that can have only two levels or values

- Answers the question: In a sample, is the proportion of observations in one category different from a given proportion?
- H_0: P ≤ .5 H_0: P = .5 H_0: P ≥ .5 (three different examples of H_0)
- Compare the number of observations in the category to the critical binomial value (in a table: column corresponds to the proportion in question, row corresponds to sample size and α); if observed ≥ critical, reject H_0
- If sample size ≥ 50 or if N * P * (1 – P) ≥ 9 then use normal approximation to binomial test

χ^2 test is a non-parametric test used with nominally scaled (or above) data that can have any number of levels or values

- Answers the question: Is the observed number of items in each category different from a theoretically expected number of observations?
- H_0: $\Sigma(O - E)^2 = 0$ The sum of the square differences between the observed (O) and expected (E) number of observations in each category equals 0
- observed $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$
- Determine critical χ^2 from table with df = number of categories – 1 and α
- If observed observed χ^2 ≥ critical χ^2, then reject H_0 -- it is unlikely that the observed and expected values are the same

χ^2 Test of Independence is a non-parametric test used with nominally scaled (or above) data

- Answers the questions: are the responses to one question independent of the responses to another question?
- H_0: $\Sigma\Sigma\ (O - E)^2 = 0$ Summing over both variables
- Expected values are given by $E_{ij} = \frac{r_i c_j}{T}$ where E_{ij} is the expected frequency for cell at row i and column j, r_i is the total number of observations in row i, c_j is the total number of observations in column j and T is the total number of observations
- observed $\chi^2 = \sum_{i=1}^{number\_of\_rows} \sum_{j=1}^{number\_of\_columns} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
- Determine critical χ^2 from table with df = (number of rows – 1) X (number of columns – 1) and α
- If observed χ^2 ≥ critical χ^2 , then reject H_0 – it is unlikely that the responses to one questions are independent of the responses to the other question

χ^2 Assumptions

- Independence – each observation must be unique; an individual cannot be contained in more than one category or counted in a given category more than once
  - Violations of this assumption greatly increase the probability of making a Type-I error
- Frequencies – the date must correspond to frequencies and not percentages or proportions
- Sufficient sample size – expected frequencies in each cell must be at least 0.33 – larger is better
  - Violations of this assumption increase the probability of making a Type-I error