Why 'N' How seminar series - ANOVA and General Linear Models

Hari Bharadwaj

Research Affiliate, MGH Neurology

December 17, 2009

Outline

1. Detection and Hypothesis testing
   - Basic problem
2. ANOVA revisited
   - 2 sample t-test
   - One-way ANOVA
   - Correlation
3. The General Linear Model (GLM) perspective
   - GLM and Design matrices
   - The most general GLM
   - Contrasts and Interactions
4. Non-parametric approaches
5. Multiple Comparisons and Topological Inference

2 Alternatives discrimination problem

$\mathcal{H}_1$: **There is an 'effect'**
$\mathcal{H}_0$: **There is no 'effect'**

- Example: $\mathcal{H}_1$: Average IQ of Group1 subjects $<$ Group2 subjects
  $\mathcal{H}_0$: Average IQ of Group1 subjects $=$ Group2 subjects
- Given data we wish to probabilistically test out the hypotheses

Universal Frequentist Recipe

ALL univariate statistical tests entail the following:

1. Construct $\mathcal{H}_0$ and $\mathcal{H}_1$, could be competing models

Universal Frequentist Recipe

ALL univariate statistical tests entail the following:

1. Construct $\mathcal{H}_0$ and $\mathcal{H}_1$, could be competing models
2. Calculate a statistic, a scalar (T), that summarizes the effect you are trying to capture (example: difference in mean IQs of 2 groups)

Universal Frequentist Recipe

ALL univariate statistical tests entail the following:

1. Construct $\mathcal{H}_0$ and $\mathcal{H}_1$, could be competing models
2. Calculate a statistic, a scalar (T), that summarizes the effect you are trying to capture (example: difference in mean IQs of 2 groups)
3. Determine the distribution of T when $\mathcal{H}_0$ is true (Here is where usually many assumptions come in)

Universal Frequentist Recipe

ALL univariate statistical tests entail the following:

1. Construct $\mathcal{H}_0$ and $\mathcal{H}_1$, could be competing models
2. Calculate a statistic, a scalar (T), that summarizes the effect you are trying to capture (example: difference in mean IQs of 2 groups)
3. Determine the distribution of T when $\mathcal{H}_0$ is true (Here is where usually many assumptions come in)
4. If $p(T|\mathcal{H}_0) < 0.05$ or any other *ad hoc* threshold, reject $\mathcal{H}_0$ (This doesn't necessarily mean we have evidence for $\mathcal{H}_1$)

Where we are...

2 group t-test as a linear regression

- $y_i = \mu + \mu_{group} + \epsilon_i$
- Example: A given Group1 IQ $= \mu_{population} + \mu_{Group1} + \epsilon_{measurement}$
- Fitting IQ data to groups is like forming a linear regression btween IQ (independent variable) and Group (The explanatory variable)
- Asking is IQs are different for the 2 groups is like asking if the slope of the regression between y (IQ) and x (group in this case) is not 0
- This generalises to multiple group tests (F-tests) or ANOVA
- This perspective also gives us ways to assess significance of covariance

GLM with 1 factor (group) $\rightarrow$ One-Way ANOVA with 2 levels

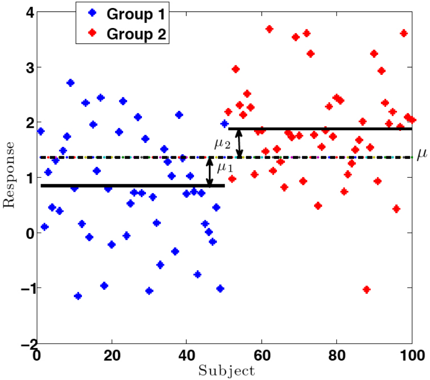


Figure: The data y is explained by 1 factor, namely 'group' $x = 0$ or 1 denoting Group1 or Group2 for example. Does regressing y as a linear function of x help explain the variance in y better than when not modeled as a function of x?

F test instead of t test

- Is $y = \mu_{group} + \mu + \epsilon$ a better model than $y = \mu + \epsilon$?
- Fit both models and see if gives one smaller errors ϵ than the other
- Let S_1 be the sum of squared errors after fitting model 1 and S_2 the sum of squared errors after fitting model 2 (null model): Is S_1 significantly smaller than S_2? (Use our **recipe**)
- $F = \frac{(S_2 - S_1)/(k_2 - k_1)}{S_1/k_1} = \frac{ExplainedVariance}{UnexplainedVariance} = \frac{BetweenGroupVariability}{WithinGroupVariability}$
- Under $\mathcal{H}_0$, F is F-distributed with $k_2 - k_1, k_1$ degrees of freedom
- The p-values should be the same as from a 2 sample t-test, but we wont know which group's mean is higher

Full and Reduced models

- Full Model: $y_i = \mu + \mu_{group} + \epsilon_i$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \hline \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \hline \vdots \\ \mu \\ \mu \end{pmatrix} + \begin{pmatrix} \mu_a - \mu \\ \mu_a - \mu \\ \vdots \\ \hline \vdots \\ \mu_b - \mu \\ \mu_b - \mu \end{pmatrix} + \begin{pmatrix} y_1 - \mu_a \\ y_2 - \mu_a \\ \vdots \\ \hline \vdots \\ y_{n-1} - \mu_b \\ y_n - \mu_b \end{pmatrix}$$

- Reduced Model: $y_i = \mu + \epsilon_i$
- Under $\mathcal{H}_0$, we are just splitting the degree of freedom of ϵ into 2 parts $\Rightarrow$ F-test using sum of squares $\longrightarrow \mathbf{F(1, n-2)}$
- If the Full model is significantly better than the reduced model, it is said that the **main effect** of group is significant

F distribution - A reminder

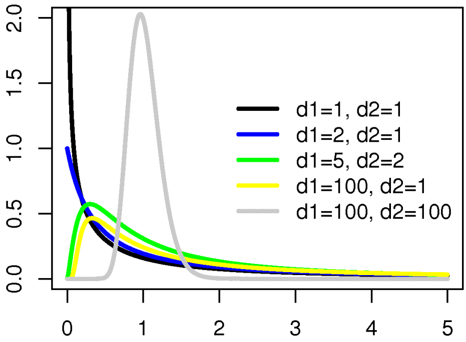


Figure: Distribution of sum of squared mean-zero normal variables

One way 2 level ANOVA example

- 2 Group IQs: Group2s = 87, 110, 93, 99, 75, 102, 90, 83, 100, 70
  Group1s = 77, 81, 64,100, 84, 72, 69, 90, 68, 70
- $t = 2.5138$, $df = 18$, $p = 0.0217 < 0.05$
- $F = 6.3192$, $df = 1, 18$, $p = 0.0217 < 0.05$
- Note that for a 2 group analysis, $F_{anova} = t^2$

One way ANOVA with 3 levels

- With GLMs, we need not restrict to 2 levels:
  Group1 = 87, 110, 93, 99, 75, 102, 90, 83, 100, 70
  Group2 = 77, 81, 64,100, 84, 72, 69, 90, 68, 70
  Group3 = 88, 87, 98, 120, 111, 97, 89, 132, 114, 126
- Same question asked: Is there any dependence on group (when #groups > 2)?
- Here IQ (y) is a modeled as a linear function of 2 group differences
- $F(2, 17) = 11.05$, $p = 0.0003$
- Degrees of freedom is the rank of the linear subspace of $\mathcal{R}^n$ spanned by each explanatory variate: The number of independent variables in the set

One way ANOVA with continuous explanatory variable: Correlation

- Does IQ depend on age?
- Same as asking 'Is IQ correlated with AGE'?
- 10 subjects:
- IQ (y)= 87, 110, 93, 99, 75, 102, 90, 83, 100, 70
  AGE (x)= 9,15,9,10,10,12,8,10,11,7
- $y = \mu + \beta x + \epsilon$ versus $y = \mu + \epsilon$: Is one significantly better than the other
- F - test would give us the answer, $p = 0.0095$
- Both age and group as factors explaining IQ $\Rightarrow$ **2 way ANOVA**

Where we are...

1 way 2 group anova revisited

Let the groups have $\frac{n}{2}$ subjects each:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \hline \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \hline \vdots \\ \mu \\ \mu \end{pmatrix} + \begin{pmatrix} \mu_a - \mu \\ \mu_a - \mu \\ \vdots \\ \hline \vdots \\ \mu_b - \mu \\ \mu_b - \mu \end{pmatrix} + \begin{pmatrix} y_1 - \mu_a \\ y_2 - \mu_a \\ \vdots \\ \hline \vdots \\ y_{n-1} - \mu_b \\ y_n - \mu_b \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \hline \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \left(\begin{array}{c|c} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ \hline \vdots & \vdots \\ 1 & -1 \\ 1 & -1 \end{array}\right) \begin{pmatrix} \mu \\ \\ \frac{1}{2}(\mu_a - \mu_b) \end{pmatrix} + \begin{pmatrix} y_1 - \mu_a \\ y_2 - \mu_a \\ \vdots \\ \hline \vdots \\ y_{n-1} - \mu_b \\ y_n - \mu_b \end{pmatrix}$$

Thus it is an instance of the GLM: $\mathbf{Y} = \mathbf{X}\beta + \epsilon$

Design matrix for 1 way 3 level ANOVA with 30 subjects

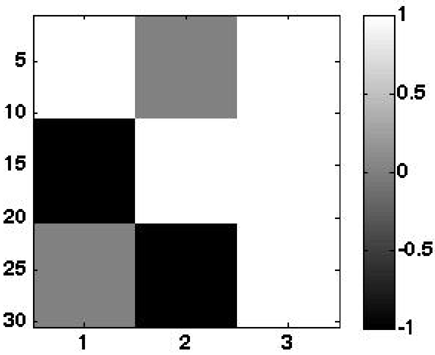


Figure: One way 3 level ANOVA has 3 experimental effects: 2 Group Differences and 1 Overall Mean

Design matrix - cell mode

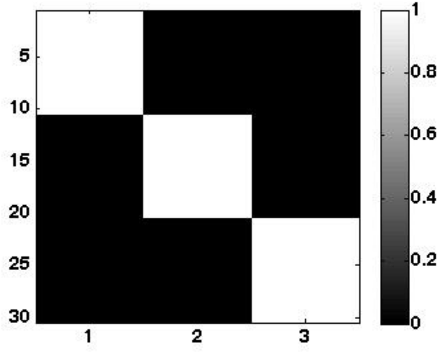


Figure: Equivalent design matrix as 2 group differences and 1 overall mean: 3 different group means

Design matrix - 2 way ANOVA: 1 categorical and 1 continuous factor

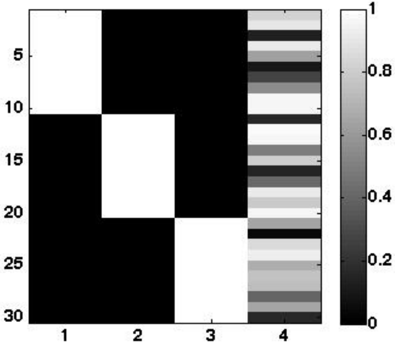


Figure: Design matrix for 30 subjects divided into 3 groups of 10 with AGE as a covariate/factor. **What is the NULL model?** - A subset of the full design matrix

Design matrix - 1 way repeated measures ANOVA with 2 conditions

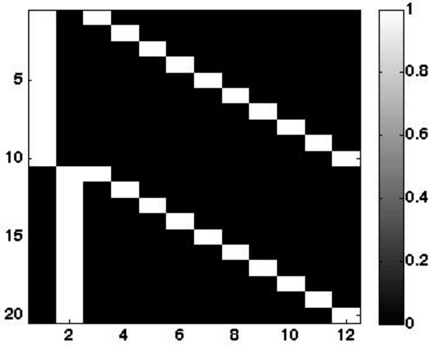


Figure: 1 way ANOVA with **block (subject)** effects, 20 subjects each measured in 2 conditions: First 2 columns are the cells corresponding to the conditions and then other 10 model effects. **What is the NULL model?**

Where we are...

In general

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon = \left(\begin{array}{cc|cc} \mathbf{G_1} & \mathbf{H_1} & \mathbf{G_0} & \mathbf{H_0} \end{array} \right) \begin{pmatrix} \gamma_1 \\ \kappa_1 \\ \hline \gamma_0 \\ \kappa_0 \end{pmatrix} + \epsilon$$

- All design using linear models and assuming a normal distribution with common error covariances are an instance of the above
- $\mathbf{G_1}$ and $\mathbf{H_1}$ are interesting categorical and continuous factors respectively
- $\mathbf{G_0}$ and $\mathbf{H_0}$ are uninteresting categorical and continuous factors
- The null model contains only the partition with $\mathbf{G_0}$ and $\mathbf{H_0}$

Where we are...

2 Groups, 2 Conditions: 2 way ANOVA with interactions

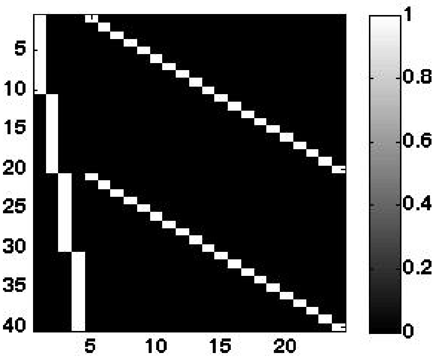
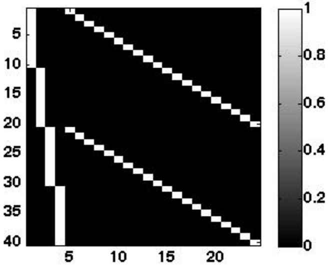


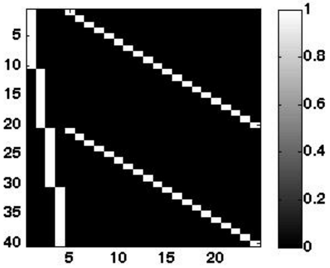
Figure: 2 way ANOVA with **block (subject)** effects, 20 subjects each measured in 2 conditions, divided into 2 groups: First 4 columns are the cells corresponding to every condition-group pair: (cond1, grp1), (cond1, grp2), (cond2,grp1) and (cond2,group2). **How do we assess the main effect of group?**

Contrast for main effect of group



- The cells for the first 4 colums are (cond1, grp1), (cond1, grp2), (cond2,grp1) and (cond2,group2)
- Let $\mathbf{c} = (1, -1, 1, -1, 0, \ldots, 0)^T$: The data in the subspace of $\mathbf{Xc}$ represent the variance because of group differences **averaging over conditions**
- Contrast matrix for **main effect of group**

Contrast for main effect of condition

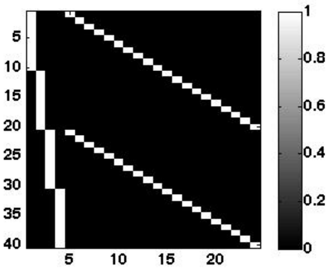


- The cells for the first 4 colums are (cond1, grp1), (cond1, grp2), (cond2,grp1) and (cond2,group2)
- Let $\mathbf{c} = (1, 1, \text{-}1, \text{-}1, 0, \ldots, 0)^T$: The data in the subspace of $\mathbf{Xc}$ represent the variance because of condition differences **averaging over groups**
- Contrast matrix for **main effect of conditions**

Interactions

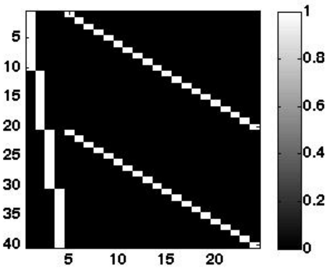
- The **effect** of one factor may depend on the **level** of another factor
- Example: Sleep hours modeled as a function of amount of exercise and weight of a person: Regular exercise increases the amount of sleep more for heavier people than for lighter people
- For our 2 group - 2 condition example: There may be group differences that are condition independent (main effects) but there might be group differences that occur only in one condition (interaction)
- If **x** and **y** are the factors, an interaction is a dependance on **xy**

Contrast for interaction between group and condition



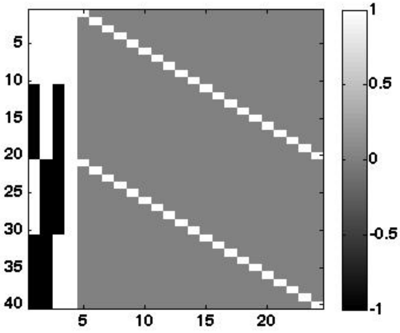
- The cells for the first 4 colums are (cond1, grp1), (cond1, grp2), (cond2,grp1) and (cond2,group2)
- Let $\mathbf{c} = (1, \text{-}1, \text{-}1, 1, 0, \ldots, 0)^T$: The data in the subspace of $\mathbf{Xc}$ represents the interaction **Difference of differences**
- Contrast matrix for **Interaction**

Contrast for overall mean



- The cells for the first 4 colums are (cond1, grp1), (cond1, grp2), (cond2,grp1) and (cond2,group2)
- Let $\mathbf{c} = (1, 1, 1, 1, 0, \ldots, 0)^T$: The data in the subspace of $\mathbf{Xc}$ represents the effects common to each of the cells
- Contrast matrix for **overall mean**

With experimental effects along the colums of the design matrix



- Column 1 is group difference, column 2 is condition
- Column 3: interaction (Note that this is column1 (dot) column2
- Column 4: Overall mean

Some remarks

- Number of experiemental effects is the number of different means. Example: A 2 way anova with 2 groups and 3 conditions has 6 experimental effects: 1 group difference, 2 condition differences, 2 condition $\times$ group interactions and 1 overall mean

Some remarks

- Number of experiemental effects is the number of different means. Example: A 2 way anova with 2 groups and 3 conditions has 6 experimental effects: 1 group difference, 2 condition differences, 2 condition $\times$ group interactions and 1 overall mean
- We can group some of the experimental effects to make omnibus inferences. Example: In the above, the 2 condition differences can be taken together to assess the main effect of condition

Some remarks

- Number of experiemental effects is the number of different means. Example: A 2 way anova with 2 groups and 3 conditions has 6 experimental effects: 1 group difference, 2 condition differences, 2 condition $\times$ group interactions and 1 overall mean
- We can group some of the experimental effects to make omnibus inferences. Example: In the above, the 2 condition differences can be taken together to assess the main effect of condition
- The null model is basically whatever is in the full model design excluding the effects of interest

Some remarks

- Number of experiemental effects is the number of different means. Example: A 2 way anova with 2 groups and 3 conditions has 6 experimental effects: 1 group difference, 2 condition differences, 2 condition $\times$ group interactions and 1 overall mean
- We can group some of the experimental effects to make omnibus inferences. Example: In the above, the 2 condition differences can be taken together to assess the main effect of condition
- The null model is basically whatever is in the full model design excluding the effects of interest
- In practice the 2 models need not be fitted separately, concept of projections and orthogonal subspaces can be used from linear algebra

Where we are...

Permutation tests: Example 1

IQ of 2 groups of 10 subjects each: Use our **recipe**

- Let T = mean IQ of group 1 - mean IQ of group 2
- Under $\mathcal{H}_0$, we want to know what the distribution of T is
- Non-parametric approach: Under $\mathcal{H}_0$, group does not have any effect on data $\Rightarrow$ We can assign group to subjects randomly
- Thus we can get many groupings, here we can have up to $\binom{20}{10} > 180,000$ permutations where the subjects from the 2 groups are mixed
- For each of these permutations we can get a T_{perm} $\Rightarrow$ We have a distribution for T under $\mathcal{H}_0$
- Is this generalizable to the population or applicable only to the cohort?

Permutation tests: Example 2

10 subjects who are politicians or have an IQ score less than 80 or both

$\mathcal{H}_1$: Politicians are more likely to have IQ < 80

$\mathcal{H}_0$: They are unrelated attributes

- The data is not normally distributed, its categorical
- Let $x_1 = (0\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1)$ and $x_2 = (1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1)$ denote politician or not and IQ less than 80 or not respectively for the 10 subjects
- $\mathbf{d} = \text{norm}(x_1 - x_2)$ is a good measure of the conjunction between the 2 attributes, $d = \sqrt{2}$ here
- Permute x_1 or x_2 values randomly and get a distribution for $\mathbf{d}$ and find $p(d \leq \sqrt{2})$

Where we are...

MCP

$\mathcal{H}_1$: Coin is biased
$\mathcal{H}_0$: Coin is unbiased

- Test: Toss coin 10 times, if Head or Tail shows up 9 or more times, reject $\mathcal{H}_0$ (p(9 or more heads/$\mathcal{H}_0$) ≈ 0.02)
- This means if we repeat the test 100 times we'll get 9 or more heads only 2 times on an average
- What if we have a million coins and test each of them with this test?
- On an average 20,000 coins will turn head more than 9 times even when non of them are biased $\Rightarrow$ We have a family wise error which we **must** correct for

Why should we worry about MCP

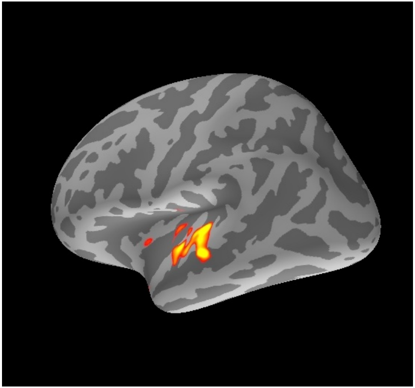


Figure: When we do any voxelwise or a vertexwise statistical test (ANOVA or t-test or anything), we encounter typically an MCP of the order of 10,000. Worse, these tests are not independent of each other (unlike the coins problem)

Multiple testing corrections

- For discrete tests (example: multiple end point drug trials): Non-parametric family-wise testing or False Discovery Rate (**FDR**) approaches
- For data sets with an inherent topology (example: Time courses, whole brain signals, time-frequency maps): Random field theory or Non-parametric topological inference tests

References

1. B.J. Winer, D.R. Brown, K.M. Michels (1971): Statistical principles in experimental design. McGraw-Hill.
2. K.J. Friston, A.P. Holmes, K.J. Worsley, J.-P. Poline, C.D. Frith, R.S.J. Frackowiak (1995): Statistical Parametric Maps in Functional Imaging: A General Linear Approach. Human Brain Mapping 2:189-210.
3. Y. Benjamini and Y. Hochberg (1995): Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B. 57: 289 - 300.