Chapter 13: Analysis of Variance

DSCC 462

Computational Introduction to Statistics

Anson Kahng

Fall 2022

Some Midterm Comments

- Trends:
  - Some people did **well**
  - Some people did **poorly**
  - Some people **cheated…**
- Actions:
  - There will almost surely be a **curve** (yet to be determined)
  - If you cheated, prepare to hear from me next week

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17
  - HW 5 due: Thursday, December 6 (2 weeks to complete it)
- Project:
  - Think about project groups (3-4 people per group)
  - Project description will be released next Tuesday, November 15
  - Groups by next Thursday, November 18

Plan for Today

- How do we compare sample means for more than two groups?
  - One-way ANOVA
- Motivating illustration:

Analysis of Variance: Motivation

- We previously discussed the use of t-tests for comparing sample means of two groups
- What if we want to compare sample means for more than two groups?
  - Use analysis of variance (ANOVA)

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old
  - Group 3: Over 50 years old
- These three groups have means $\mu_1, \mu_2,$ and μ_3, respectively
- We want to test the null hypothesis that the population means are identical
  - $H_0 : \mu_1 = \mu_2 = \mu_3$
  - H_1 : At least one of the population means differs from one of the others

One-Way ANOVA

- In general, we are interested in comparing k different populations
- Assume the k populations are independent and normally distributed
- Draw a sample of size n_i from group i that has population mean μ_i and population variance σ_i^2
- For this sample, we get sample mean $\bar{x}_i$ and sample variance s_i^2
- Note: The number of observations in each sample does not need to be the same

One-Way ANOVA

- Consider our weight example over the three age groups, and let's say we have the following data:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$
  - Group 2: $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$
  - Group 3: $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$
- Question: Are the average weights for the three age groups the same?

One-Way ANOVA

- We could answer this question by doing $\binom{3}{2} = 3$ different paired t-tests
  - Group 1 vs. Group 2, Group 1 vs. Group 3, and Group 2 vs. Group 3
- This may seem feasible for three groups, but as the number of groups increases, it quickly becomes infeasible
  - For instance, if we have $k = 10$ groups, we need to do $\binom{10}{2} = 45$ paired t-tests

One-Way ANOVA

- There is a more important problem, though: if we perform all possible two-sample t-tests, we are likely to reach an incorrect conclusion!
- Suppose we do the three hypothesis tests for $k = 3$ groups, each at a significance level of $\alpha = 0.05$
- What's the probability of a type I error (rejecting the null hypothesis given that the null hypothesis is true)?

$$\begin{aligned}\Pr(\text{Reject } H_0 \text{ given } H_0 \text{ is true}) &= 1 - \Pr(\text{Fail to reject in all three tests}) \\ &= 1 - (1 - 0.05)^3 \\ &= 1 - 0.857 = 0.143\end{aligned}$$

One-Way ANOVA

- The probability of rejecting the null hypothesis in at least one of these tests is higher than the α we use for each pairwise test
- In other words, if we know that the null hypothesis is true, then this type I error rate (0.143) is much higher than the desired standard (0.05)
- With one-way ANOVA, we are able to keep the desired significance level α, unlike if we perform multiple t-tests

One-Way ANOVA

- Key idea: One-way ANOVA is dependent on estimates of spread or dispersion
- “One-way” indicates that there is a single factor or characteristic that distinguishes the populations from each other
    - In our example, age is the distinguishing factor

One-Way ANOVA: Assumptions

- Three assumptions must hold:
  - Normality: Each group follows a normal distribution
  - Equal variances: Population variances for each group are equal
  - Independence: Observations are not correlated

Sources of Variation

- We assume that there is a common variance σ^2 for the populations
- There are two sources of variation in this ANOVA setup:
    - **Within-group variability** (s_w^2): Variation of the individual values around their population means
        - All groups are assumed to have the same variability
    - **Between-group variability** (s_b^2): Variation of the population means around the grand (overall) mean
- When the variability *within* the k populations is small relative to the variability among their respective means, this suggests that the population means are indeed different
    - Intuition: tightly clustered and separated from each other

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}$$
$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k}$$

- **Between groups**: Variability of the population means around the grand (overall) mean

$$s_b^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

- The *grand mean*, $\bar{x}$, is defined as

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n}.$$

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Our proxy question: Do sample means vary around the grand mean *more* than the individual observations around the sample means?
    - I.e., is $s_b^2 > s_w^2$? (Recall that s_w^2 measures within-group variability, s_b^2 measures between-group variability)
- If the answer to this question is "yes", this provides evidence that the population means are different
- How can we compare s_w^2 and s_b^2?
    - Thinking back to variances, s_b^2 / s_w^2 follows an F distribution

ANOVA Testing Procedures

- To test the null hypothesis with a certain significance level α, se use the following test statistic: $F = s_b^2/s_w^2$
- Under the null hypothesis H_0, both s_b^2 and s_w^2 estimate the true σ^2
- Thus, we expect F to be close to 1
- If a difference exists between population means, $s_b^2 > s_w^2$ and thus F will be larger than 1
    - If $s_b^2 \leq s_w^2$, then there is no difference between population means

ANOVA Testing Procedures

- Under H_0, $F = s_b^2/s_w^2$ has an F distribution with $k-1$ degrees of freedom in the numerator and $n-k$ degrees of freedom in the denominator
- If $k=2$, this F-test reduces to a two-sample t-test

ANOVA Testing Procedures

- Once F is calculated, we can calculate a p-value, p, based on the F distribution with degrees of freedom df1 and df2
  - Reject H_0 if $p \leq \alpha$
- To get p-values in R (recall that we are only interested in the upper tail probability): `1-pf(F,df1,df2)`
- Or, we can compare our test statistic to the critical value that cuts of the upper $\alpha \cdot 100\,\%$ of the F distribution with degrees of freedom df1 and df2

ANOVA Example

- Consider our weight example over the three age groups
- We are interested in comparing the mean weight for three age groups: 18-30 years old, 31-50 years old, and 51+ years old
- At the $\alpha = 0.05$ significance level, we want to test $H_0 : \mu_1 = \mu_2 = \mu_3$ against H_1 : at least one of the age groups has an average weight that is different from at least one of the other age groups

ANOVA Example

- Recall our sample summary statistics:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$
  - Group 2: $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$
  - Group 3: $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$
- Using this information, we can calculate s_w^2, s_b^2, and our F-statistic

ANOVA Example

- Given $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$, $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$, and $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$

- $s_w^2 =$

- $\bar{x} =$

- $s_b^2 =$

ANOVA Example

- Given $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$, $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$, and$n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{n_1 + n_2 + n_3 - 3}$$

- $$= \frac{(26 - 1)8.9^2 + (31 - 1)11.4^2 + (44 - 1)9.9^2}{26 + 31 + 44 - 3}$$

$= 102.995$ pounds^2 .

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$$

- $$= \frac{26(151) + 31(174) + 44(162)}{26 + 31 + 44}$$

$= 162.85$ pounds .

$$s_b^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2}{3 - 1}$$

- $$= \frac{26(151 - 162.85)^2 + 31(174 - 162.85)^2 + 44(162 - 162.85)^2}{2}$$

$= 7520.877$ pounds^2 .

ANOVA Example

- Therefore, our test statistic is $F = s_b^2 / s_w^2 =$
- For an F distribution with $k - 1 =$ and $n - k =$ degrees of freedom, we get a p-value of $p =$
- Conclusion:

ANOVA Example

- Therefore, our test statistic is $F = s_b^2/s_w^2 = 73.02$
- For an F distribution with $k - 1 = 3 - 1 = 2$ and $n - k = 26 + 31 + 44 - 3 = 98$ degrees of freedom, we get a p-value of $p = 1 - \texttt{pf(73.02, 2, 98)} = 3.8 \times 10^{-30}$
- Conclusion: Since $p < \alpha$, we reject the null hypothesis and conclude that at least one of the age groups differs from one of the others in height

ANOVA Table

Source of Variation	Sum of Squares (SS)	df	Mean Squares (MS)	F	P-value
Between (treatment)	$SSB = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$	$k - 1$	$s_b^2 = \frac{SSB}{k-1}$	$\frac{s_b^2}{s_w^2}$	p
Within (error)	$SSE = \sum_{i=1}^{k} (n_i - 1)s_i^2$	$n - k$	$s_w^2 = \frac{SSE}{n-k}$		
Total	$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$n - 1$			

ANOVA Table for Example

Source of Variation	Sum of Squares (SS)	df	Mean Squares (MS)	F	P-value
Between (treatment)	15041.754				
Within (error)	10093.51				
Total					

ANOVA Table for Example

Source of Variation	**Sum of Squares (SS)**	**df**	**Mean Squares (MS)**	**F**	**P-value**
Between (treatment)	15041.754	2	7520.877	73.02	0
Within (error)	10093.51	98	102.995		
Total	25135.264	100			

ANOVA

- Our null hypothesis is $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Once H_0 is rejected, we can conclude that the population means are not all equal
- However, we don't know exactly which means differ!
- We need to conduct additional tests to find where the differences are
- In this case, we are performing *multiple comparisons*

Multiple Comparisons

- Typically, we will be interested in comparing each pair of means individually
- Recall that in performing the $\binom{k}{2}$ possible two-sample t-tests, we increase our overall probability of committing a type I error
- We correct for this by being more conservative in the individual comparisons
- Make it more difficult to reject each individual comparison so that the overall significance level remains at α
    - We call this the *familywise type I error*, or α_{FWE}

Multiple Comparisons

- Intuition: If we are performing $\binom{k}{2}$ tests, then we can separate α evenly between these tests
  - $\alpha^* = \dfrac{\alpha}{\binom{k}{2}}$ is the significance level for an individual comparison
- This is called the *Bonferroni correction*
- For instance, if we want $\alpha = 0.05$ significance for $k = 5$ populations, then each pairwise test should have significance level $\alpha^* = 0.05/10 = 0.005$

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha* = \dfrac{\alpha}{\binom{k}{2}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$
  - This is the test statistic for a two-sample t-test
  - But we estimate σ^2 based on all populations, not just populations i and j
  - Under null hypothesis H_0, t_{ij} has a t distribution with df = $n - k$
- Calculate the p-value, p, based on a t distribution with $n - k$ degrees of freedom
- Reject H_0 if $p \leq \alpha*$

Multiple Comparisons: Example

- Return to the weight by age group example
- We found that the population means were not all identical
- Now, we must compare each pair of age groups to see where the differences are
- Total of $\binom{k}{2} = \binom{3}{2} = 3$ comparisons
- Overall desired significance $\alpha = 0.05$

Multiple Comparisons: Example

- $t_{12} =$

- $t_{13} =$

- $t_{23} =$

Multiple Comparisons: Example

- $t_{12} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_w^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{151 - 174}{\sqrt{102.995(1/26 + 1/31)}} = -8.52$
- $t_{13} = \dfrac{\bar{x}_1 - \bar{x}_3}{\sqrt{s_w^2\left(\frac{1}{n_1} + \frac{1}{n_3}\right)}} = \dfrac{151 - 162}{\sqrt{102.995(1/26 + 1/44)}} = -4.38$
- $t_{23} = \dfrac{\bar{x}_2 - \bar{x}_3}{\sqrt{s_w^2\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}} = \dfrac{174 - 162}{\sqrt{102.995(1/31 + 1/44)}} = 5.04$

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$, $t_{13} =$, and $t_{23} =$
- Calculating other parameters: $\alpha^* =$, df =
- Calculating p-values
  - $p_{12} =$
  - $p_{13} =$
  - $p_{23} =$

Multiple Comparisons: Example

- From previous slide, we have $t_{12} = -8.52$, $t_{13} = -4.38$, and $t_{23} = 5.04$
- Calculating other parameters: $\alpha^* = \dfrac{\alpha}{\binom{k}{2}} = 0.0167$, df $= n - k = 98$
- Calculating p-values
  - $p_{12} =$ `2*pt(-8.52,df=98) = 1.95 x 10`$^{-13}$
  - $p_{13} =$ `2*pt(-4.38,df=98) = 2.98 x 10`$^{-5}$
  - $p_{23} =$ `2*(1-pt(5.04,df=98) = 2.13 x 10`$^{-6}$

Multiple Comparisons: Example

- Given these p-values, what conclusions can we draw?
    - Group 1 vs. Group 2:
    - Group 1 vs. Group 3:
    - Group 2 vs. Group 3:

Multiple Comparisons: Example

- Given these p-values, what conclusions can we draw?
  - Group 1 vs. Group 2: Reject H_0 and conclude that there is a difference in the mean weights of 18-30 year olds and 31-50 year olds
  - Group 1 vs. Group 3: Reject H_0 and conclude that there is a difference in the mean weights of 18-30 year olds and 51+ year olds
  - Group 2 vs. Group 3: Reject H_0 and conclude that there is a difference in the mean weights of 31-50 year olds and 51+ year olds

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant
- This does not always have to be the case
- Some populations may be the same whereas others are different
- Conclusions from ANOVA and multiple comparisons may contradict each other!
    - Significant ANOVA and no significant pairwise comparisons: Overly conservative pairwise comparisons test
    - Non-significant ANOVA but significant pairwise comparisons: Generally consider pairwise comparisons result valid

Multiple Comparisons: Other Methods

- Other testing procedures than the Bonferroni procedure exist
    - Often called *post-hoc* analysis methods
- Bonferroni is traditionally one of the most conservative measures and can suffer from lack of power
- Some others: Tukey, Newman-Keuls, Scheffee, Dunnett, etc.

ANOVA in R

- First, create an ANOVA object using the `aov()` function
  - Let Y be the continuous variable (e.g., weight) and X be the grouping variable (e.g., age ranges)
  - `model1=aov(Y~X)`
- Summarize using the anova() function
  - `anova(model1)`