STAT509: One-way ANOVA

Peijie Hou

University of South Carolina

November 19, 2014

Motivation of Inference on Two Samples

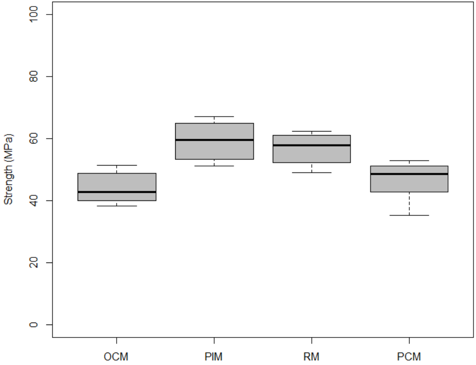
- When we want to compare the mean of two normal distributions, we can use a t-test to detect the difference.
- What if we want to compare the means of several normal populations?
- A t-test does not suffice to detect the difference.
- We need to use **one-way analysis of variance** technique (frequently abbreviated ANOVA).

An Example

Four types of mortars: ordinary cement mortar (OCM), polymer impregnated mortar (PIM), resin mortar (RM), and polymer cement mortar (PCM), were subjected to a compression test to measure strength (MPA). Here are the strength measurements taken on different mortar specimens (36 in all).

OCM:	51.45	42.96	41.11	48.06	38.27	38.88	42.74	49.62		
PIM:	64.97	64.21	57.39	52.79	64.87	53.27	51.24	55.87	61.76	67.15
RM:	48.95	62.41	52.11	60.45	58.07	52.16	61.71	61.06	57.63	56.80
PCM:	35.28	38.59	48.64	50.99	51.52	52.85	46.75	48.31		

Mortar example



- "Treatment" = mortar type (OCM, PIM, RM, and PCM). There are $t = 4$ treatment groups.
- This is an example of an **observational study**; not an **experiment**. That is, we do not physically apply a treatment

Data Display of Kenton Food Example

QUERY: An initial question that we might have is the following:

"Are the treatment (mortar type) population means equal? Or, are the treatment population means different?"

One-Way ANOVA Design

- In a one-way ANOVA design, the only way in which individuals are "classified" is by the treatment group assignment. Hence, such an arrangement is called **a one-way classification**

- **Assumptions**: In general, we have independent random samples from $p \geq 2$ normal distributions with same variance:

$$\begin{aligned}
\text{Sample } 1 &: Y_{11}, Y_{12}, \ldots, Y_{1n_1} \sim \mathcal{N}(\mu_1, \sigma^2) \\
\text{Sample } 2 &: Y_{21}, Y_{22}, \ldots, Y_{2n_2} \sim \mathcal{N}(\mu_2, \sigma^2) \\
&\vdots \qquad\qquad\qquad \vdots \\
\text{Sample } p &: Y_{p1}, Y_{p2}, \ldots, Y_{pn_p} \sim \mathcal{N}(\mu_p, \sigma^2).
\end{aligned}$$

- We want to test

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$$
$$H_a : \text{the population means } \mu_i \text{ are not all equal.}$$

- The **null hypothesis** H_0 says that there is "no treatment difference", that is, all treatment population means are the same.
- The **alternative hypothesis** H_a says that a difference among the t population means exists somewhere (but does not specify how the means are different).

The procedures we develop is formulated by deriving two estimators for σ^2. These two estimators are formed by

- looking at the variance of the observations **within** samples
  - Unbiased for σ^2: $MS_{res} = \frac{SS_{res}}{N-t}$
- looking at the variance of the sample means **across** the t samples
  - Unbiased for σ^2 **under** H_0: $MS_{trt} = \frac{SS_{trt}}{t-1}$

where

$$SS_{total} = SS_{trt} + SS_{res}$$

- Sampling distribution: when H_0 is true:

$$F = \frac{MS_{trt}}{MS_{res}} \sim F(t-1, N-t)$$

Such decomposition is usually presented in a so-called ANOVA table of the following form:

Source of Variation	SS	df	MS	F
Treatments	SS_{trt}	$t-1$	$MS_{trt} = \frac{SS_{trt}}{t-1}$	$F = \frac{MS_{trt}}{MS_{res}}$
Error	SS_{res}	$N-t$	$MS_{res} = \frac{SS_{res}}{N-t}$	
Total	SS_{total}	$N-1$		

- SS_{total} measures the total variation in all the data
- SS_{trt} measures how much of the total variation is due to the treatments
- SS_{res} measures what is "left over"

Test Procedure

1. State the null and alternative hypotheses

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$$
$$H_a : \text{the population means } \mu_i \text{ are not all equal.}$$

2. The test statistic is $F_0 = \frac{MS_{trt}}{MS_{res}}$, which can be found from ANOVA table.
3. Draw conclusion based on the p-value of the test.
   - p-value$< \alpha$: We reject the H_0, and conclude not all treatments have the same effect
   - p-value$\geq \alpha$: We do not reject the H_0, and conclude that there is not difference among the treatments

Mortar example

```
> OCM = c(51.45,42.96,41.11,48.06,38.27,38.88,42.74,49.62)
> PIM = c(64.97,64.21,57.39,52.79,64.87,53.27,51.24,55.87,61.76,67.15)
> RM = c(48.95,62.41,52.11,60.45,58.07,52.16,61.71,61.06,57.63,56.80)
> PCM = c(35.28,38.59,48.64,50.99,51.52,52.85,46.75,48.31)

> strength = c(OCM,PIM,RM,PCM)
> # Create a treatment indicator variable
> mortar.type = c(
+     rep("OCM",length(OCM)),
+     rep("PIM",length(PIM)),
+     rep("RM",length(RM)),
+     rep("PCM",length(PCM))
+ )
> mortar.type = factor(mortar.type)
> # Create data frame
> # R wants the data in this format to do the ANOVA
> data = data.frame(strength,mortar.type)
```

- Now, the dataset looks like

```
      strength mortar.type
 [1,]    51.45           1
 [2,]    42.96           1
 [3,]    41.11           1
............................
[33,]    51.52           2
[34,]    52.85           2
[35,]    46.75           2
[36,]    48.31           2
```

- The R output is given by

```
> anova(lm(strength ~ mortar.type))
Analysis of Variance Table

Response: strength
            Df  Sum Sq Mean Sq F value    Pr(>F)
mortar.type  3 1520.88  506.96  16.848 9.576e-07 ***
Residuals   32  962.86   30.09
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Source of Variation	SS	df	MS	F
Treatments	1520.88	3	506.96	16.848
Error	962.86	32	30.09	
Total	2483.74	35		

The p-value of the test is $9.576 * 10^{-7}$, so we reject the H_0 at any reasonable level of significance, and conclude that we have sufficient evidence that the strength for four mortars are not equal.

Multiple Comparisons Following the ANOVA

- When the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$ is rejected in the ANOVA, we know that some of the treatment or factor level means are different.
- However, the ANOVA doesn't identify which means are different.
- Methods for investigating this issue are called *multiple comparisons methods*.
- To do this, we will construct Tukey pairwise confidence intervals for all population treatment mean differences $\mu_i - \mu_{i'}$, $1 \leq i < i' \leq p$.
- In the Kenton food Example, we have 4 treatments, we have total $\binom{4}{2} = 6$ Tukey intervals to construct. Namely,

$$\mu_1 - \mu_2, \mu_1 - \mu_3, \mu_1 - \mu_4, \mu_2 - \mu_3, \mu_3 - \mu_4.$$

Multiple Comparisons Following the ANOVA

- If we construct multiple confidence intervals (here, 6 of them), and if we construct each one at the $100(1-\alpha)$ percent confidence level, then the overall confidence level for all 6 intervals will be less than $100(1-\alpha)$ percent.
- *GOAL*: Construct confidence intervals for all pairwise intervals $\mu_i - \mu_{i'}$, $1 \leq i < i' \leq p$, and have our overall confidence level still be at $100(1-\alpha)$ percent.
- *SOLUTION*: Simply increase the confidence level associated with each individual interval! Tukeys method is designed to do this. The interval are of the form:

$$(\overline{Y}_i - \overline{Y}_{i'}) \pm q_{\alpha,p,N-p}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)},$$

where $q_{\alpha,p,N-p}$ is the Tukey upper percentage point. We will use R to construct such intervals.

Multiple Comparisons: Mortar

```
> TukeyHSD(aov(lm(strength ~ mortar.type)),conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = lm(strength ~ mortar.type))

$$$mortar.type
            diff        lwr       upr     p adj
PCM-OCM  2.48000 -4.950955  9.910955 0.8026758
PIM-OCM 15.21575  8.166127 22.265373 0.0000097
RM-OCM  12.99875  5.949127 20.048373 0.0001138
PIM-PCM 12.73575  5.686127 19.785373 0.0001522
RM-PCM  10.51875  3.469127 17.568373 0.0016850
RM-PIM  -2.21700 -8.863448  4.429448 0.8029266
```

Multiple Comparisons: Read Output

- The columns labeled `lwr` and `upr` give the lower and upper confidence limits of the pairwise confidence interval.
- We are (at least) 95 percent confident that the difference between the population mean strengths for the PCM and OCM mortars is between -4.95 and 9.91 MPa.
  - Note that this confidence interval includes "0", which suggests that these two population means are not different.
  - An equivalent finding is that the **adjusted p-value**, given in the p adj column, is large.
- We are (at least) 95 percent confident that the difference between the population mean strengths for the PIM and OCM mortars is between 8.17 and 22.27 MPa.
  - Note that this confidence interval does not include "0", which suggests that these two population means are different.
  - An equivalent finding is that the **adjusted p-value**, given in the p adj column, is small.

Multiple Comparisons: Summary

- Therefore, for the strength/mortar type data, the following pairs of population means are declared to be different:

  PIM-OCM, RM-OCM, PIM-PCM, RM-PCM

- The following pairs of population means are declared to be not different:

  PCM-OCM, RM-PIM