One-way ANOVA

MAM5120 - Statistical Concepts, Methods and Tools

Samantha Pendleton[1]

[1] Aberystwyth University
Department of Computer Science

2017

Content

Introduction to ANOVA

One-way ANOVA

- to compare three or more means
- to determine whether groups are significantly different
- useful to test differences between medicines, populations, etc.

Assumptions

- samples are independent
- response is normally distributed

Hypothesis

NULL HYPOTHESIS

H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$

ALTERNATIVE HYPOTHESIS

H_A: "at least two means are different"

Significance

if p-value is significant at 95% ($p<0.05$) then we reject the NULL Hypothesis

Table

Source	SS	df	MS	F	p(>F)
Treatment	SS_{treat}	k-1	$SS_{treat}/(k-1)$	MS_{treat}/MS_{resid}	p
Residual	SS_{resid}	N-k	$SS_{resid}/(N-k)$		
Total	$SS_{treat} + SS_{resid}$	N-1			

Annotating

- SS = Sum of Squares
- df = Degrees of Freedom
- MS = Mean of Squares
- F = f-statistic

Example

A Marine Research Centre want to observe sightings of Maui's Dolphins off the coast of New Zealand for 3 random days

Figure: Maui dolphin. Image Source: Daily Mail[1]

[1] http://www.dailymail.co.uk/sciencetech/article-2118304/Worlds-smallest-dolphin-threat-nets-species-reduced-just-55-survivors.html?ito=feeds-newsxml

Data

Sightings				
Random Day	Morning	Afternoon	Evening	Night
1	6	3	10	8
2	4	7	9	7
3	3	8	8	5
6.5	4.3333...	6	9	6.6666...

Assuming

- Sightings/Days are independent
- Counts/Numbers are normally distributed

Identifying Calculations

k = 4; N = 12; at 95%

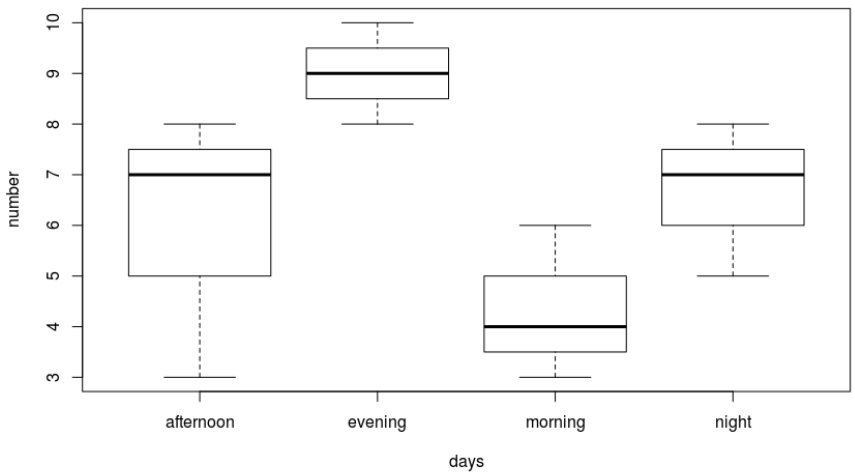
Source	SS	df	MS	F
Treatment	33.6667	3	11.2222	3.54386
Residual	25.3333	8	3.1667	
Total	59	11		

Annotating

- SS_{treat}: $3(4.3333... - 6.5)^2 + 3(6 - 6.5)^2 + ... = 33.6667$
- SS_{resid}: $((6 - 5 = 1) + (4 - 5 = -1) + (3 - 5 = -2)) = 2 + ... = 25.3333$
- $df_t = k - 1 = 4 - 1 = 3$
- $df_r = N - k = 12 - 4 = 8$
- $MS_{treat} : SS_{treat}/df_t = 33.6667/3 = 11.2222$
- $MS_{resid} : SS_{resid}/df_r = 25.3333/8 = 3.1667$
- $F : MS_{treat}/MS_{resid} = 11.2222/3.1667 = 3.54386$

Example in R

```
> summary(aov(number ~ days, data=sightings))
             Df  Sum Sq Mean Sq F value Pr(>F)
days          3   33.67  11.222   3.544 0.0677 .
Residuals     8   25.33   3.167
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Remember

if p-value is significant at 95% ($p<0.05$) then we reject the NULL Hypothesis

Results

- p-value = 0.0677
- $p > 0.05$
- not significant at 95%
- insufficient evidence to reject H_0 in favour of H_A
- observations from different times of day come from the same distribution (have the same mean)