Q1) The following table shows the lives (in hrs) of four batches of electric lamps:

Batches	Life of a bulb in hrs
1	1600,1610,1650,1680,1700,1720,1800
2	1580,1640,1640,1700,1750
3	1460,1550,1600,1620,1640,1660,1740,1820
4	1510,1520,1530,1600,1680

Perform an analysis of variance.

Hypothesis

- H_0: There is no significant difference in mean life of bulbs of batches.
- H_1: Mean life of bulbs of atleast one batch differs significantly.

In [1]:
```
b1_bulb_life = c(1600,1610,1650,1680,1700,1720,1800)
b2_bulb_life = c(1580,1640,1640,1700,1750)
b3_bulb_life = c(1460,1550,1600,1620,1640,1660,1740,1820)
b4_bulb_life = c(1510,1520,1530,1570,1600,1680)
```

In [2]:
```
bulb_life_df = data.frame(bulb_life_in_hrs=c(b1_bulb_life,b2_bulb_life,b3_bulb_life,b4_bulb_life),
                          batches=factor(rep(c(1,2,3,4),times=c(length(b1_bulb_life),length(b2_bulb_life),length(b3_bulb_life),length(b4_bulb_life)))))
```

In [3]:
```
model1 = aov(bulb_life_in_hrs ~ batches,data=bulb_life_df)
```

In [4]:
```
print(anova(model1))
```

```
Analysis of Variance Table

Response: bulb_life_in_hrs
          Df Sum Sq Mean Sq F value Pr(>F)
batches    3  44361 14786.9  2.1494 0.1229
Residuals 22 151351  6879.6
```

In [5]:
```
summary(model1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
batches      3  44361   14787   2.149  0.123
Residuals   22 151351    6880
```

Conclusion

- The p-value is 0.1229, since the p-value is greater than 0.05 we cannot reject the null hypothesis. Therefore, there is no significant difference in mean life of bulbs of batches.

Q2) In 25 plots four varieties v1, v2, v3, v4 of wheat are randomly put and their yield in kg are shown below.

V1 2000	V3 2270	V2 2230	V4 2270	V4 2180
V2 2160	V1 2100	V2 2050	V3 2300	V2 2280
V1 2200	V1 2300	V4 2040	V3 2420	V1 2240
V4 2370	V1 2250	V2 2040	V2 2360	V1 2460
V3 2210	V1 2340	V2 2190	V1 2150	V3 2020

Perform the ANOVA to test whether there is any significant difference between varieties of wheat.

Hypothesis

- H_0: There is no significant difference in the effect of varieties of wheat.
- H_1: Atleast one pair of means effect of varieties of wheat is significantly different.

In [6]:
```
v1_yield = c(2000,2100,2200,2300,2240,2250,2460,2340,2150)
v2_yield = c(2230,2160,2050,2280,2040,2360,2190)
v3_yeild = c(2270,2300,2420,2210,2020)
v4_yield = c(2270,2180,2040,2370)
```

In [7]:
```
wheat_yield_df = data.frame(yield=c(v1_yield,v2_yield,v3_yeild,v4_yield),
                            varieties=factor(rep(c('v1','v2','v3','v4'),times=c(length(v1_yield),length(v2_yield),length(v3_yeild),length(v4_yield)))))
```

In [8]:
```
model2 = aov(yield~varieties,data=wheat_yield_df)
```

In [9]:
```
print(anova(model2))
```

```
Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value Pr(>F)
varieties  3  10741  3580.4  0.2011 0.8945
Residuals 21 373963 17807.8
```

In [10]:
```
summary(model2)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
varieties    3  10741    3580   0.201  0.894
Residuals   21 373963   17808
```

Conclusion

The p-value is 0.8944, since it is greater than 0.05 we cannot reject the null hypothesis. Therefore there is no difference in the effect of varieties.

Q3) A manufacturing company has purchased three new machines and wishes to determine whether one of them is faster than the others in producing a certain output. Five hourly production figures are observed at random from each maching and the results are given in the following table

	Machine A1	Machine A2	Machine A3
	25	31	24
	30	39	30
Observation	36	38	28
	38	42	25
	31	35	28

Use analysis of variance technique and determine whether the machines are significantly different in their mean speeds. Use 5% LOS.

Hypothesis

- H_0: There is no significant difference in the mean production speed of the machines.
- H_1: Mean production speed of atleast one machine differs significantly.

In [11]:
```
machine_a1 = c(25,30,36,38,31)
machine_a2 = c(31,39,38,42,35)
machine_a3 = c(24,30,28,25,28)
```

In [12]:
```
machine_prod_df = data.frame(production=c(machine_a1,machine_a2,machine_a3),
                             machine=factor(rep(c('a1','a2','a3'),times=c(length(machine_a1),length(machine_a2),length(machine_a3)))))
```

In [13]:
```
model3 = aov(production~machine,data=machine_prod_df)
```

In [14]:
```
print(anova(model3))
```

```
Analysis of Variance Table

Response: production
          Df Sum Sq Mean Sq F value   Pr(>F)
machine    2    250 125.000     7.5 0.007707 **
Residuals 12    200  16.667
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In [15]:
```
summary(model3)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
machine      2    250  125.00     7.5 0.00771 **
Residuals   12    200   16.67
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion

The p-value is 0.0077, since the p-value is less than 0.05 we can reject the null hypothesis. Therefore mean production speed of atleast one machine differs significantly.

Q4) If we have three fertilizers and we have to compare their efficacy, this could be done by a field experiment in which each fertilizer is applied to 10 plots, and then 30 plots are later harvested, with the crop field being calculated for each plot. The data were recorded in following table:

Fertilizer	Yields (in tones) from the 10 plots allocated to that fertilizer									
1	6.27	5.36	6.39	4.85	5.99	7.14	5.08	4.07	4.35	4.95
2	3.07	3.29	4.04	4.19	0.41	0.75	4.87	3.94	6.49	3.15
3	4.04	3.79	4.56	4.55	4.53	3.53	3.71	7.00	4.61	4.55

Carry out analysis of variance.

Hypothesis

- H_0: There is no significant difference in the mean effect of the fertilizers.
- H_1: Mean effect of atleast one fertilizer differs significantly.

In [16]:
```
fert1 = c(6.27,5.36,6.39,4.85,5.99,7.14,5.08,4.07,4.35,4.95)
fert2 = c(3.07,3.29,4.04,4.19,0.41,0.75,4.87,3.94,6.49,3.15)
fert3 = c(4.04,3.79,4.56,4.55,4.53,3.53,3.71,7.00,4.61,4.55)
```

In [17]:
```
fert_efficacy_df = data.frame(yields=c(fert1,fert2,fert3),
                              fertilizer=factor(rep(c(1,2,3),times=c(length(fert1),length(fert2),length(fert3)))))
```

In [18]:
```
model4 = aov(yields~fertilizer,data=fert_efficacy_df)
```

In [19]:
```
print(anova(model4))
```

```
Analysis of Variance Table

Response: yields
           Df Sum Sq Mean Sq F value   Pr(>F)
fertilizer  2 20.523 10.2615   5.984 0.007061 **
Residuals  27 46.300  1.7148
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In [20]:
```
summary(model4)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
fertilizer   2  20.52  10.261   5.984 0.00706 **
Residuals   27  46.30   1.715
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion

The p-value is 0.0070, since the p-value is less than 0.05 we can reject the null hypothesis. Therefore there is atleast one fertilizer whose mean effect differs significantly.