Question 1: ANOVA

⋆ Given

Four different, though supposedly equivalent, methods of applying torque to a bolt on to the chassis of a machine were performed on each of 5 bolts (in a random order), and the following are the measured torque applied in $ft \cdot lbf$:

Bolt	Method A	Method B	Method C	Method D
1	75	83	86	73
2	73	72	61	67
3	59	56	53	62
4	69	70	72	79
5	84	92	88	95
$\bar{x}_{method}$	72.0	74.6	72.0	75.2
s_{method}	9.11	13.67	15.28	12.77

⋆ Find

Perform an ANOVA [you may use R] to test whether it is reasonable to treat the 4 methods as equivalent at a significance level of $\alpha = 0.05$.

⋆ Solution

- Enter data into R as a CSV file with 20 rows of data with columns of 'Method and 'Torque'
- R Code to read data and perform ANOVA

```
    myData <- read.csv("ME_488_HWK_2_1_DATA.CSV")
    myData # Confirm Data is loaded
    myData$Method <- factor(myData$Method)
    m <- aov(Torque~Method, data=myData)
    anova(m)
```

- ANOVA Output from R

```
Analysis of Variance Table

Response: Torque
          Df  Sum Sq Mean Sq F value Pr(>F)
Method     3   42.95  14.317  0.0859 0.9667
Residuals 16 2666.00 166.625
```

- There is a 96.67% chance of making a Type I error if we reject the H_0, which means at an $\alpha = 0.05$ we should definitely accept H_0 and claim **there is no evidence that the method of torquing bolts matters**.

Question 2: Completing an ANOVA Table

⋆ Given

An experiment is designed to determine which of six different oils provides the best lubrication for a complex mechanism. Each oil is run in the mechanism eight times. The run order is completely random.

Source	df	SS	MS	F	p
Oil	?	4525	?	?	?
Error	?	14742	?		
Total	?	?			

⋆ Find

Use this information to complete the following ANOVA table. Is there evidence that one or more of the oils is different from the others?

⋆ Assumptions

$\alpha = 0.05$

⋆ Solution

- $k = 6$: Different kinds of oils
- $n = 8$: Number of times each oil is run
- $df_{treatment} = df_{Oil} = k - 1 = 6 - 1 = 5$
- $df_{error} = k(n-1) = 6(8-1) = 42$
- Now fill in the blanks

Source	df	SS	MS	F	p
Oil	5	4525	905	2.58	?
Error	42	14742	351		
Total	47	19267			

- p value for the F you can get with the **pf** function in R

```
1-pf(2.58,5,42)
[1] 0.04011714
```

- Remember, pf function in R gives the area to the LEFT of the cutoff , so that's why we subtract from 1
- $p_{value} = 0.04011714 = 0.04 < \alpha$ so we reject the H_0, and **claim that there is a significant difference between at least one pair of oils**

Question 3: More Complex ANOVA

⋆ Given

A platinum thermal deflection sensor was mounted on a stationary probe craft that landed on the surface of Mars. For a specific period of 5 Martian days each Earth year, the thermally induced deflection of a strain gage is measured as a relative deflection from the previous day and transmitted back to mission control in Huston. The data for the last 6 years [2001-2007] is included in the CSV file (downloadable from the course website). It was discovered, after a preliminary data analysis, that the highest average deflection tends to occur on Day 3.

⋆ Find

- Use ANOVA to confirm that the third day difference in relative deflection is statistically meaningful.
- Next, normalize the deflection per day by the total deflection during the 5 day period for each year. For example if the deflections were [10,20,50,80,40] (total = 200) then the normalized (divide each one by the overall total) ones would be [0.05, 0.1, 0.25, 0.4, 0.2] (total = 1). Use ANOVA to determine if the normalized relative deflection rate difference is real or due purely to chance occurrence.

⋆ Assumptions

$\alpha = 0.05$

⋆ Solution

Solution Part I

- R Code to read data and perform ANOVA

```
    myData <- read.csv("ME488_HWK_2_P_3_DATA.csv")
    myData # Confirm Data is loaded
    myData$Day <- factor(myData$Day)
    m <- aov(Deflection~Day, data=myData)
    anova(m)
```

- ANOVA Output from R

```
Analysis of Variance Table

Response: Deflection
          Df  Sum Sq Mean Sq F value Pr(>F)
Day        4  266.46  66.614  1.0412 0.4025
Residuals 30 1919.43  63.981
```

- There is a 40.25% chance of making a Type I error if we reject the H_0, which means at an $\alpha = 0.05$ we should definitely accept H_0 and claim **there is no evidence that the day matters to the deflection values**.

Solution Part II

- First open the data set up and normalize it by converting to percentages of the annual deflection
- R Code to read data and perform ANOVA

```
myData <- read.csv("ME488_HWK_2_P_3_DATA_CONVERTED.csv")
myData # Confirm Data is loaded
myData$Day <- factor(myData$Day)
m <- aov(Relative_Deflection~Day, data=myData)
anova(m)
```

- Listing of data converted [example, yours may look different]

```
     Year Day Deflection Relative_Deflection
1  2001   1         13          0.16250000
2  2001   2         20          0.25000000
3  2001   3         16          0.20000000
4  2001   4         18          0.22500000
5  2001   5         13          0.16250000
6  2002   1         24          0.29629630
7  2002   2         22          0.27160494
8  2002   3         12          0.14814815
9  2002   4          8          0.09876543
10 2002   5         15          0.18518519
11 2003   1         22          0.19130435
12 2003   2         37          0.32173913
13 2003   3         33          0.28695652
14 2003   4         17          0.14782609
15 2003   5          6          0.05217391
16 2004   1         15          0.21126761
17 2004   2         10          0.14084507
18 2004   3         22          0.30985915
19 2004   4         10          0.14084507
20 2004   5         14          0.19718310
21 2005   1          2          0.06250000
22 2005   2          9          0.28125000
23 2005   3         10          0.31250000
24 2005   4          7          0.21875000
25 2005   5          4          0.12500000
26 2006   1          4          0.12903226
27 2006   2          6          0.19354839
28 2006   3          8          0.25806452
29 2006   4          5          0.16129032
30 2006   5          8          0.25806452
31 2007   1          5          0.10638298
32 2007   2          7          0.14893617
33 2007   3         16          0.34042553
34 2007   4         11          0.23404255
35 2007   5          8          0.17021277
```

- ANOVA Output from R

```
Analysis of Variance Table

Response: Relative_Deflection
          Df   Sum Sq   Mean Sq F value  Pr(>F)
Day        4 0.057358 0.0143395  3.2353 0.02545 *
Residuals 30 0.132968 0.0044323
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- There is a 2.545% chance of making a Type I error if we reject the H_0, which means at an $\alpha = 0.05$ we should reject H_0 and claim **there is evidence that the day matters to the RELATIVE deflection values**.

Bonus Question: Systemic Experimental Thinking

⋆ Given

You have ten shipments of diesel engines, each shipment contains 500 or more engines. The manufacturer just sent you a bulletin that one of the shipments is made of engines that are missing a hard to locate internal component, and would like you to ship it back, at their expense. Your large scale is a little finicky and can only be trusted for the first measurement of the day [so you only get one shot at it]

The missing component has a mass of 5kg, and the normal diesel engine has a mass of 1000kg. The scale has a capacity of 75,000kg and an abilty to detect a difference of 2.5kg.

⋆ Find

Describe a plan to discover the shipment with incomplete engines, with only one scale measurement allowed.

⋆ Solution

- Mark each shipments 1 through 10
- Take $i - 1$ engines from the i^{th} shipment
- Take the whole batch and weigh it once
- Take the Scale reading and subtract it from 10000kg
- Integer Divide the result by 5kg
- Result+1 is which shipment is defective

END OF ASSIGNMENT