Analysis of Variance an Experimental Design
final proposal

Factors that affect Taiwanese to buy new cars

Group 3
110304039 邱士芳
110304036 石瀞捷

We are interested in the factors which affect Taiwanese people's intention to buy a new car. To explore this topic, we designed an experiment of variance analysis. The factors we chose are seasons, years and brands. The following y, which is the monthly amount of new cars being sold.

Y: the monthly amount of new cars being sold in Taiwan.
Factors:

1. Years: 102, 103, 104, 105, 106, 107, 108, 109, 110, 111
2. Seasons: Spring, Summer, Fall, Winter
3. Brands: Lexus, Mitsubishi, Subaru, Suzuki, Nissan, Honda, Mazda, Toyota, Volkswagen, Mercedes Benz, Volvo, Porsche, Skoda, Audi, BMW
   *selection criterion: the brand with high market share in Taiwan

Hypothesis:

1. H10：There is a significant difference between seasons.
   H11：There isn't a significant difference between seasons.

2. H20：There is a significant difference between the brands.
   H21：There is a significant difference between the brands.

3. H30：There is a significant difference between the years.
   H31：There is a significant difference between the years.

4. H40：There exist interaction effects
   H41：There are no interaction effects

Data schema:

	Season	Year	Brand	y
1	fall	102	LEXUS	2630
2	fall	102	MITSUBISHI	88
3	fall	102	SUBARU	1097
4	fall	102	SUZUKI	1010
5	fall	102	TOYOTA	4573
6	fall	102	MAZDA	2998

ANOVA:

```
> model3 = aov(y ~ Year * Season * Brand, data = df3)
> summary(model3)
                    Df    Sum Sq   Mean Sq
Year                 9 1.073e+08  11923995
Season               3 1.650e+07   5501485
Brand               14 3.734e+09 266698738
Year:Season         27 1.927e+07    713837
Year:Brand         126 3.368e+08   2673047
Season:Brand        42 2.789e+07    664056
Year:Season:Brand  378 1.107e+08    292908
```

We found out that there is a problem of overfitting. To deal with this problem, we transformed the ‘year’ factor from 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, which have 9 degrees of freedom, to 102-106 and 107-111, which have 1 degree of freedom. The following table are the new data schema and ANOVA:

new data schema:

```
> head(df7)
  Season Year      Brand    y
1   fall    1      LEXUS 2630
2   fall    1 MITSUBISHI   88
3   fall    1     SUBARU 1097
4   fall    1     SUZUKI 1010
5   fall    1     TOYOTA 4573
6   fall    1      MAZDA 2998
```

new ANOVA

```
> summary(model7)
                   Df    Sum Sq   Mean Sq F value   Pr(>F)
Year                1 5.467e+07  54665639  73.122  < 2e-16 ***
Season              3 1.650e+07   5501485   7.359 7.88e-05 ***
Brand              14 3.734e+09 266698738 356.743  < 2e-16 ***
Year:Season         3 9.142e+05    304733   0.408    0.748
Year:Brand         14 1.501e+08  10721558  14.341  < 2e-16 ***
Season:Brand       42 2.789e+07    664056   0.888    0.673
Year:Season:Brand  42 9.586e+06    228236   0.305    1.000
Residuals         480 3.588e+08    747594
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

from the new ANOVA, we can conclude that:

- **Season:** p = 7.88e - 05 < 0.05
  - → There is significant difference between seasons
- **Year :** p < 2e - 16 < 0.05
  - → There is significant difference between the years
- **Brand :** p < 2e - 16 < 0.05
  - → There is significant difference between the brands
- **Year:Season :** p = 0.748 > 0.05
  - → There is no interaction effect between the year and season
- **Year:Brand :** p < 2e - 16 < 0.05
  - → There is interaction between between the year and the brands
- **Season:Brand :** p = 0.673 > 0.05
  - → There is no interaction effect between seasons and the brands
- **Season:Year:Brand :** p = 1.000 > 0.05
  - →There is no interaction effect between season, the year and the brands.

emeans analysis:

1. Seasons

```
> summary(posthoc_season)
$emmeans
 Season emmean   SE  df lower.CL upper.CL
 fall     2970 70.6 480     2831     3109
 spring   2953 70.6 480     2814     3092
 summer   2983 70.6 480     2844     3121
 winter   3351 70.6 480     3212     3490

Results are averaged over the levels of: Year, Brand
Confidence level used: 0.95

$contrasts
 contrast        estimate   SE  df t.ratio p.value
 fall - spring       16.9 99.8 480   0.169  0.9983
 fall - summer      -12.6 99.8 480  -0.126  0.9993
 fall - winter     -380.8 99.8 480  -3.815  0.0009
 spring - summer    -29.4 99.8 480  -0.295  0.9911
 spring - winter   -397.7 99.8 480  -3.983  0.0005
 summer - winter   -368.3 99.8 480  -3.688  0.0014

Results are averaged over the levels of: Year, Brand
P value adjustment: tukey method for comparing a family of 4 estimates
```

- **fall - spring:** The difference in average values between these two seasons is 16.9, the t-ratio is 0.169, and the p-value is 0.9983, indicating that the difference is not significant.
- **fall - summer:** The difference in average values between these two seasons is -12.6, the t-ratio is -0.126, and the p-value is 0.9993, suggesting that the difference is not significant.
- **fall - winter:** The difference in average values between these two seasons is -380.8, the t-ratio is -3.815, and the p-value is 0.0009, indicating a significant difference.
- **spring - summer:** The difference in average values between these two seasons is -29.4, the t-ratio is -0.295, and the p-value is 0.9911, suggesting that the difference is not significant.
- **spring - winter:** The difference in average values between these two seasons is -397.7, the t-ratio is -3.983, and the p-value is 0.0005, indicating a significant difference.
- **summer - winter:** The difference in average values between these two seasons is -368.3, the t-ratio is -3.688, and the p-value is 0.0014, indicating a significant difference.

2. Years:

```
> summary(posthoc_year)
$emmeans
 Year emmean   SE  df lower.CL upper.CL
 1      2762 49.9 480     2664     2860
 2      3366 49.9 480     3268     3464

Results are averaged over the levels of: Season, Brand
Confidence level used: 0.95

$contrasts
 contrast      estimate   SE  df t.ratio p.value
 Year1 - Year2     -604 70.6 480  -8.551  <.0001

Results are averaged over the levels of: Season, Brand
```

Year 1 - Year 2: The estimated contrast (difference) in means between Year 1 and Year 2 is -604. The standard error (SE) is 70.6, and the t-ratio is -8.551, with a p-value less than 0.0001, indicating a highly significant difference.

3. Brands:

```
> summary(posthoc_brand)
$emmeans
 Brand         emmean  SE  df lower.CL upper.CL
 AUDI             973 137 480    704.8     1242
 BMW             4366 137 480   4097.5     4635
 HONDA           7322 137 480   7053.3     7591
 LEXUS           4193 137 480   3924.5     4462
 MAZDA           4545 137 480   4276.8     4814
 MERCEDES-BENZ   6236 137 480   5967.0     6504
 MITSUBISHI       212 137 480    -56.6      481
 NISSAN           442 137 480    173.2      710
 PORSCHE          941 137 480    672.3     1210
 SKODA           1238 137 480    969.3     1507
 SUBARU          1465 137 480   1196.2     1733
 SUZUKI          1427 137 480   1158.0     1695
 TOYOTA          8074 137 480   7805.3     8343
 VOLKSWAGEN      3023 137 480   2754.1     3291
 VOLVO           1506 137 480   1237.2     1774

Results are averaged over the levels of: Year, Season
Confidence level used: 0.95
```

```
> summary(posthoc_brand)
$emmeans
 Brand          emmean  SE  df lower.CL upper.CL
 AUDI              973 137 480    704.8     1242
 BMW              4366 137 480   4097.5     4635
 HONDA            7322 137 480   7053.3     7591
 LEXUS            4193 137 480   3924.5     4462
 MAZDA            4545 137 480   4276.8     4814
 MERCEDES-BENZ    6236 137 480   5967.0     6504
 MITSUBISHI        212 137 480    -56.6      481
 NISSAN            442 137 480    173.2      710
 PORSCHE           941 137 480    672.3     1210
 SKODA            1238 137 480    969.3     1507
 SUBARU           1465 137 480   1196.2     1733
 SUZUKI           1427 137 480   1158.0     1695
 TOYOTA           8074 137 480   7805.3     8343
 VOLKSWAGEN       3023 137 480   2754.1     3291
 VOLVO            1506 137 480   1237.2     1774

Results are averaged over the levels of: Year, Season
Confidence level used: 0.95
```

```
$contrasts
 contrast                  estimate  SE  df t.ratio p.value
 AUDI - BMW                -3392.7 193 480 -17.548  <.0001
 AUDI - HONDA              -6348.5 193 480 -32.836  <.0001
 AUDI - LEXUS              -3219.7 193 480 -16.653  <.0001
 AUDI - MAZDA              -3572.0 193 480 -18.476  <.0001
 AUDI - (MERCEDES-BENZ)    -5262.2 193 480 -27.218  <.0001
 AUDI - MITSUBISHI           761.4 193 480   3.938  0.0081
 AUDI - NISSAN               531.6 193 480   2.750  0.2784
 AUDI - PORSCHE               32.5 193 480   0.168  1.0000
 AUDI - SKODA               -264.5 193 480  -1.368  0.9896
 AUDI - SUBARU              -491.4 193 480  -2.542  0.4127
 AUDI - SUZUKI              -453.1 193 480  -2.344  0.5582
 AUDI - TOYOTA             -7100.5 193 480 -36.726  <.0001
 AUDI - VOLKSWAGEN         -2049.3 193 480 -10.600  <.0001
 AUDI - VOLVO               -532.4 193 480  -2.753  0.2763
```

BMW - HONDA	-2955.8	193	480	-15.288	<.0001
BMW - LEXUS	173.0	193	480	0.895	0.9999
BMW - MAZDA	-179.3	193	480	-0.928	0.9998
BMW - (MERCEDES-BENZ)	-1869.5	193	480	-9.670	<.0001
BMW - MITSUBISHI	4154.1	193	480	21.486	<.0001
BMW - NISSAN	3924.3	193	480	20.298	<.0001
BMW - PORSCHE	3425.2	193	480	17.716	<.0001
BMW - SKODA	3128.2	193	480	16.180	<.0001
BMW - SUBARU	2901.3	193	480	15.006	<.0001
BMW - SUZUKI	2939.6	193	480	15.204	<.0001
BMW - TOYOTA	-3707.8	193	480	-19.178	<.0001
BMW - VOLKSWAGEN	1343.4	193	480	6.948	<.0001
BMW - VOLVO	2860.3	193	480	14.795	<.0001
HONDA - LEXUS	3128.8	193	480	16.183	<.0001
HONDA - MAZDA	2776.5	193	480	14.361	<.0001
HONDA - (MERCEDES-BENZ)	1086.3	193	480	5.619	<.0001
HONDA - MITSUBISHI	7109.9	193	480	36.774	<.0001
HONDA - NISSAN	6880.1	193	480	35.586	<.0001
HONDA - PORSCHE	6381.0	193	480	33.004	<.0001
HONDA - SKODA	6084.0	193	480	31.468	<.0001
HONDA - SUBARU	5857.1	193	480	30.295	<.0001
HONDA - SUZUKI	5895.4	193	480	30.492	<.0001
HONDA - TOYOTA	-752.0	193	480	-3.890	0.0097
HONDA - VOLKSWAGEN	4299.2	193	480	22.237	<.0001
HONDA - VOLVO	5816.1	193	480	30.083	<.0001
LEXUS - MAZDA	-352.4	193	480	-1.822	0.8881
LEXUS - (MERCEDES-BENZ)	-2042.5	193	480	-10.565	<.0001
LEXUS - MITSUBISHI	3981.1	193	480	20.591	<.0001
LEXUS - NISSAN	3751.3	193	480	19.403	<.0001
LEXUS - PORSCHE	3252.2	193	480	16.821	<.0001
LEXUS - SKODA	2955.2	193	480	15.285	<.0001
LEXUS - SUBARU	2728.3	193	480	14.111	<.0001
LEXUS - SUZUKI	2766.5	193	480	14.309	<.0001
LEXUS - TOYOTA	-3880.8	193	480	-20.073	<.0001
LEXUS - VOLKSWAGEN	1170.4	193	480	6.054	<.0001
LEXUS - VOLVO	2687.3	193	480	13.900	<.0001

MAZDA - (MERCEDES-BENZ)	-1690.2	193	480	-8.742	<.0001
MAZDA - MITSUBISHI	4333.4	193	480	22.414	<.0001
MAZDA - NISSAN	4103.7	193	480	21.225	<.0001
MAZDA - PORSCHE	3604.6	193	480	18.644	<.0001
MAZDA - SKODA	3307.5	193	480	17.107	<.0001
MAZDA - SUBARU	3080.6	193	480	15.934	<.0001
MAZDA - SUZUKI	3118.9	193	480	16.132	<.0001
MAZDA - TOYOTA	-3528.5	193	480	-18.250	<.0001
MAZDA - VOLKSWAGEN	1522.7	193	480	7.876	<.0001
MAZDA - VOLVO	3039.7	193	480	15.722	<.0001

(MERCEDES-BENZ) - MITSUBISHI	6023.6	193	480	31.156	<.0001
(MERCEDES-BENZ) - NISSAN	5793.9	193	480	29.968	<.0001
(MERCEDES-BENZ) - PORSCHE	5294.8	193	480	27.386	<.0001
(MERCEDES-BENZ) - SKODA	4997.7	193	480	25.850	<.0001
(MERCEDES-BENZ) - SUBARU	4770.8	193	480	24.676	<.0001
(MERCEDES-BENZ) - SUZUKI	4809.1	193	480	24.874	<.0001
(MERCEDES-BENZ) - TOYOTA	-1838.3	193	480	-9.508	<.0001
(MERCEDES-BENZ) - VOLKSWAGEN	3212.9	193	480	16.618	<.0001
(MERCEDES-BENZ) - VOLVO	4729.9	193	480	24.464	<.0001

data file:
https://drive.google.com/drive/folders/1mz-aHphb4PrHY99LLbVRSvrcw1JexmWs

data resource:
https://stat.thb.gov.tw/hb01/webMain.aspx?sys=100&funid=11200