

Exam: 2021/06/18

Marco Scarpelli

05 giugno, 2024

Dataset exploration

```
## price location type
## 1 1104 Chiavari hotel
## 2 677 Rapallo bb
## 3 834 Rapallo hotel
## 4 960 Rapallo bb
## 5 1117 Chiavari bb
## 6 965 Rapallo apt
```

Point A

Assumptions: all data belonging to each combination of factors is univariate Gaussian.

Here, we will choose a p-value that is 5% divided by the amount of combined factors:

```
## [1] 0.008333333
```

p-values for Gaussianity:

```
## Ps
## 1 0.2579667
## 2 0.6449031
## 3 0.9418255
## 4 0.6741621
## 5 0.2909484
## 6 0.3074365
```

We accept Gaussianity at every level with large confidence.

We will now check that they have the same covariance structure. We can use Bartlett's test, keeping in mind that it will fail easily if the data is not Gaussian enough:

```
##
## Bartlett test of homogeneity of variances
##
## data: predicted_v and combined_factors
## Bartlett's K-squared = 6.6585, df = 5, p-value = 0.2473
```

The test succeeds, so we may proceed.

Complete ANOVA model

```
## Df Sum Sq Mean Sq F value Pr(>F)
## factor_1 1 2623 2623 0.130 0.719
```

```
## factor_2          2  902414  451207  22.424 6.13e-09 ***
## factor_1:factor_2  2   38296   19148   0.952   0.389
## Residuals        114 2293883   20122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Point B

From the last summary we can see that the interaction has low significance; we will now use an additive model and figure out whether the first factor (location) still has low significance.

```
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## factor_1      1    2623     2623    0.13    0.719
## factor_2      2  902414  451207   22.44 5.76e-09 ***
## Residuals    116 2332179   20105
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the first factor has a high p-value, we will remove it too:

```
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## factor_2      2  902414  451207   22.61 4.98e-09 ***
## Residuals    117 2334802   19956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let us verify this model's assumptions; again, Gaussianity and same covariance structure.

p-values for Gaussianity, keeping in mind that now alpha is divided by the 3 levels of factor 2:

```
##              Ps
## 1 0.1935144
## 2 0.8060580
## 3 0.5141077
```

Covariance structure:

```
##
## Bartlett test of homogeneity of variances
##
## data:  predicted_v and factor_v
## Bartlett's K-squared = 4.1785, df = 2, p-value = 0.1238
```

Both tests are OK.

Point C

```
##      Mean  beta.appt  beta.bb beta.hotel
## 872.55833 -80.13333 -40.33333 120.46667
```

Point D

```
##              Inf      Sup
## apt-bb    -116.5228  36.92284
## apt-hotel -277.3228 -123.87716
## bb-hotel  -237.5228 -84.07716
```

At a global level of 95%, it seems that there is always a difference between Bed&Breakfast and Hotels; the other confidence intervals include 0, so we cannot rule out that their differences will be 0.