Experiments and ANOVA

- Experiments
  - Experiment is a test done in order to discover something under different conditions.
  - Generally, it aims to observe the effect of a certain factor on the outcome.
  - Design of experiment is a plan to make the outcome reach a desired goal by the results of experiments.

한양대학교 HANYANG UNIVERSITY

Experiments and ANOVA

- Some concepts
  - Factor : independent variable.
  - Levels: a set of different values in a factor.
  - Treatments: a certain experiment conditions
  - Response value : outputs by the treatment

한양대학교 HANYANG UNIVERSITY

Experiments and ANOVA

- Why?
  - To compare the results of response values by the different factor levels..
  - To find the factor levels to affect response values.

Experiments and ANOVA

- Principles
  - Randomization
    : random assignment of subjects to the treatment in an experiment.
  - Replication
    : Repeated experiment in the same treatment to control errors.
  - Blocking
    : Similar characterized subjects are treated as a block to increase the precision of the experiment.

Experiments and ANOVA

- Procedures
  - Goal setting
  - Understanding constraints
  - Modeling
  - Determine the number of repetition
  - Randomization and design
  - Prior testing
  - Data collection
  - Analysis

Experiments and ANOVA

- One-way layout
  - To compare the response values by the different k levels.

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_k$$

Level 1	Level 2	…	Level k
y_{11}	y_{12}	…	y_{1k}
y_{21}	y_{22}		y_{2k}
…	…		…
y_{n1}	y_{n2}	…	y_{nk}

Experiments and ANOVA

- One-way ANOVA

  - One-way ANOVA model

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

  - μ is the overall mean and τ_j is the j-th treatment effect. ε_{ij} is error.
  - $\varepsilon_{ij} \sim N(0, \sigma^2)$

한양대학교 HANYANG UNIVERSITY

Experiments and ANOVA

- By the model

$$L = \sum_{i=1}^{n}\sum_{j=1}^{k}\varepsilon_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}(y_{ij} - \mu - \tau_j)^2$$

  - As the same way,

$$\frac{\partial L}{\partial \mu} = -2\sum_{i=1}^{n}\sum_{j=1}^{k}(y_{ij} - \mu - \tau_i) = 0$$

$$\frac{\partial L}{\partial \tau_i} = -2\sum_{i=1}^{n}(y_{ij} - \mu - \tau_j) = 0, \qquad j = 1, \dots, k$$

한양대학교 HANYANG UNIVERSITY

Experiments and ANOVA

- We will get k+1 equations…

$$
\begin{aligned}
&N\mu + n\tau_1 + n\tau_2 + \cdots + n\tau_k = y_{..} \\
&n\mu + n\tau_1 = y_{.1} \\
&n\mu + n\tau_2 = y_{.2} \\
&\cdots \\
&n\mu + n\tau_k = y_{.k}
\end{aligned}
$$

  - Can we solve this??

Experiments and ANOVA

- Finally, we get

$$
\hat{\mu} = \overline{y_{..}}
$$

$$
\hat{\tau_j} = \overline{y_{.j}} - \overline{y_{..}}
$$

Experiments and ANOVA

- One-way ANOVA

$$SST = SSt + SSE$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij} - \overline{y}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n} {y_{ij}}^2 - \frac{{y_{..}}^2}{k \cdot n}$$

$$SSt = \sum_{j=1}^{k}\sum_{i=1}^{n}(\overline{y_{.j}} - \overline{y}_{..})^2 = n\sum_{j=1}^{k}(\overline{y_{.j}} - \overline{y}_{..})^2$$

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij} - \overline{y_{.j}})^2$$

한양대학교 HANYANG UNIVERSITY

Experiments and ANOVA

- One-way ANOVA Table

	SS	df	MS	F
Treatment	SSt	$k-1$	$MSt = \frac{SSt}{k-1}$	$\frac{MSt}{MSE}$
Error	SSE	$k\cdot(n-1)$	$MSE = \frac{SSE}{k\cdot(n-1)}$	
Total	SST	$nk-1$		

한양대학교 HANYANG UNIVERSITY

Experiments and ANOVA

- One-way ANOVA Table
  - $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k$

$$T = \frac{MSt}{MSE} \sim F_{(k-1, k\cdot(n-1))}$$

  - Reject if $\mathrm{T} > F_{(k-1, k\cdot(n-1);)}$

Experiments and ANOVA

pd.groupby(column)

-. *Group dataframe using a mapper or by a series of columns.*

pd.agg(func, axis)
-. *Aggregate using one or more operations over the specified axis.*
func : function to use for aggregating data
axis = 0 or 1 : row or column

Experiments and ANOVA

statsmodels.formula.api
-. *A interface for specifying models*

ols(formula = ‘y~x’, data).fit()
-. *Fit data using formula*
data : dataframe
y : response variable
x : independent variable. If x is a string, use C(x, sum) instead

Experiments and ANOVA

statsmodels.api
-. *A general models and methods*

statsmodels.stats.anova_lm(*args)
-. *Return ANOVA table for one or more fitted lines*
**args : fitted linear model results instance*

Experiments and ANOVA

- Practice
  - We collected data about the crop yields from 16 different areas, and 4 different fertilizers were used. Test if crop yields are affected by the fertilizer at α=0.05.

Experiments and ANOVA

- Practice

In [33]:
```
import pandas as pd
import statsmodels.formula.api as smf
import statsmodels.api as sm
```

In [34]:
```
data1 =pd.read_csv("harvest_8.csv")
data1.head()
```

Out[34]:

Response variable → Yield; Factor → Fertil

	Yield	Fertil
0	148	F1
1	76	F1
2	134	F1
3	98	F1
4	166	F2

Experiments and ANOVA

- Practice
  - Observe the response values by the group

In [37]:
```
data1.groupby("Fertil").agg({'Yield':['mean','std','min','max']})
```

Out[37]:

	Yield			
	mean	std	min	max
Fertil				
F1	114.0	32.944398	76	148
F2	209.5	58.094750	153	264
F3	295.0	55.575774	214	335
F4	426.0	35.336478	380	465

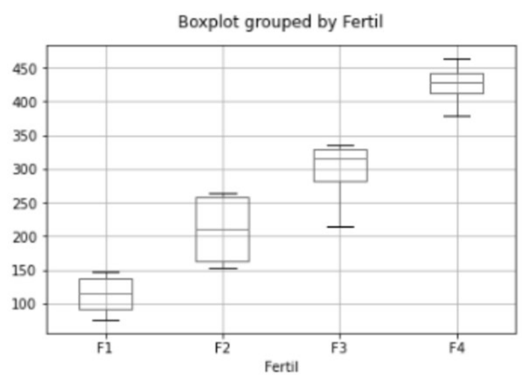
Experiments and ANOVA

- Practice

In [36]:
```
import matplotlib.pyplot as plt

data1.boxplot("Yield", by="Fertil")
plt.title("")
```

Out[36]: Text(0.5, 1.0, '')



Experiments and ANOVA

- Practice

```
In [43]: harvest_fit = smf.ols("Yield ~ C(Fertil, Sum)",data=data1).fit()
         table = sm.stats.anova_lm(harvest_fit)
         print(table)

                  df     sum_sq       mean_sq          F    PR(>F)
C(Fertil, Sum)   3.0  210568.75  70189.583333  31.912818  0.000005
Residual        12.0   26393.00   2199.416667        NaN       NaN
```

  - What is your conclusion ?

Experiments and ANOVA

- Multiple comparison
  - Bonferroni's method

  : Force to adjust the sum of confidence levels to $1 - \alpha$!

$$P\left(\bigcap_{i=1}^{m} E_i\right) = 1 - P\left(\bigcup_{i=1}^{m} E_i^c\right) \geq 1 - \sum_{i=1}^{m} P(E_i^c)$$

Experiments and ANOVA

stats.multicomp.MultiComparison(data, groups)
-. *Test for multiple comparison*
   data : independent data array
   groups : group labels

.allpairtest(testfunc[method])
-. *Run a pairwise test on all pairs with multiple test correction*
   testfunc : test functions
   method : method of multiple tests

Experiments and ANOVA

- Practice

In [49]:
```
import statsmodels.stats.multicomp as mc
from scipy import stats

comp = mc.MultiComparison(data1['Yield'], data1['Fertil'])
comptable, _, _ = comp.allpairtest(stats.ttest_ind, method="bonf")
comptable
```

Out [49]:

Test Multiple Comparison ttest_ind FWER=0.05
method=bonf alphacSidak=0.01, alphacBonf=0.008

group1	group2	stat	pval	pval_corr	reject
F1	F2	-2.8599	0.0288	0.1728	False
F1	F3	-5.6032	0.0014	0.0083	True
F1	F4	-12.9162	0.0	0.0001	True
F2	F3	-2.1269	0.0775	0.4652	False
F2	F4	-6.3679	0.0007	0.0042	True
F3	F4	-3.9782	0.0073	0.0438	True