Варіант 4

В трьох регіонах протягом 8 місяців вимірювалася кількість опадів в мм. Перевірити на рівні значущості $\alpha = 0.1$ гіпотезу про однакову середню кількість опадів а регіонах.

вибірки вважаються витягнутими з нормально розподілених генеральних сукупностей з однаковими дисперсіями

Місяць спостереження	Регіон (фактор F)		
	F_1	F_2	F_3
1	33	16	20
2	28	17	21
3	21	20	24
4	25	19	19
5	21	20	19
6	41	18	21
7	23	19	27
8	19	27	36

In [1]:

```
from sys import path
path.append('..')
from mathstat import *
from IPython.display import HTML, Latex, display
alpha = 0.1
n = 8 # number of measurements (rows)
m = 3 # factor level (cols)
data = np.array([33, 16, 20, 28, 17, 21, 21, 20, 24, 25, 19, 19, 21, 20, 19, 41, 18,
21, 23, 19, 27, 19, 27, 36]).reshape(n,m)
```

Для перевірки гіпотези про рівність групових середніх обчислюємо за вибірковими даними $S_{\text{заг.}}$ та $S_{\text{факт.}}$ проте задля спрощення обчислень можемо знайти середнє значення спостережень $\bar{x}_{**} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}$ віднімемо $\mathrm{C} = \bar{x}_{**}$ з кожного спостережуваного значення $y_{ij} = x_{ij} - C$, обчислимо допоміжні значення. Складемо розрахункову табл. 2

In [2]:

```
C = int(round(data.sum()/(n*m)))
print(f"{C = }")
Y = data-C
Y2 = Y**2
Q = Y2.sum(axis=0)
T = Y.sum(axis=0)
T2 = T**2
Q_sum = Q.sum()
T_sum = T.sum()
T2_sum = T2.sum()
style = """<style type="text/css">
.tg  {border-collapse:collapse;border-spacing:0;}
.tg td{border-color:black;border-style:solid;border-width:1px;font-family:Arial, san
s-serif;font-size:14px;
  overflow:hidden;padding:10px 15px;word-break:normal;}
.tg th{border-color:black;border-style:solid;border-width:1px;font-family:Arial, san
s-serif;font-size:14px;
  font-weight:normal;overflow:hidden;padding:10px 5px;word-break:normal;}
.tg .tg-9wq8{border-color:inherit;text-align:center;vertical-align:middle}
.tg .tg-c3ow{border-color:inherit;text-align:center;vertical-align:top}
.tg .tg-0pky{border-color:inherit;text-align:center;vertical-align:middle}
.tg .tg-baqh{text-align:center;vertical-align:middle}
</style>
"""
new_line = '\n'
F_i_template = lambda i: f'<th class="tg-c3ow" colspan="2">$F_{i}$</th>'
y_i_template = lambda i: '<td class="tg-c3ow">$y_{'+str(i)+'j}$</td><td class="tg-c3
ow">$y_{'+str(i)+'j}^2$</td>'
thead = f"""<table class="tg">
<thead>
<tr>
    <th class="tg-9wq8" rowspan="2">j</th>
    {new_line.join(F_i_template(i) for i in range(1,m+1))}
    <th class="tg-0pky" rowspan="2">$\sum$</th>
</tr>
<tr>
    {new_line.join(y_i_template(i) for i in range(1,m+1))}
</tr>
</thead>"""
row_template = lambda cols: '<tr>'+cols+'</tr>'
col_template = lambda cell: '<td class="tg-baqh">'+str(cell)+'</td>'
from itertools import chain
# prepare list of data
list_in_row = lambda j: [j]+(list(chain(*zip(Y[j-1],Y2[j-1]))))+['']
cols_in_row = lambda j: new_line.join(map(col_template, list_in_row(j)))

Q_row = col_template(r"$$Q_i=\sum_{j=1}^{"+str(n)+"}y_{ij}^2$$")+\
    new_line.join(map(lambda c: col_template('')+c,map(col_template, Q)))+\
    col_template('$$\sum_{i=1}^{'+str(m)+'}Q_i='+str(Q_sum)+'$$')
T_row = col_template(r"$$T_i=\sum_{j=1}^{"+str(n)+"}y_{ij}$$")+\
    new_line.join(map(lambda c: c+col_template(''),map(col_template, T)))+\
    col_template('$$\sum_{i=1}^{'+str(m)+'}T_i='+str(T_sum)+'$$')
T2_row = col_template(r"$T_i^2$")+\
    new_line.join(map(lambda c: c+col_template(''),map(col_template, T2)))+\
    col_template('$$\sum_{i=1}^{'+str(m)+'}T_i^2='+str(T2_sum)+'$$')

display(HTML(f"""
{style}
{thead}
```

```
{new_line.join(row_template(cols_in_row(j)) for j in range(1,n+1))}
{row_template(Q_row)}
{row_template(T_row)}
{row_template(T2_row)}
</table>"""
))
```

C = 23

j	F_1		F_2		F_3		$\sum$
	y_{1j}	y_{1j}^2	y_{2j}	y_{2j}^2	y_{3j}	y_{3j}^2	
1	10	100	-7	49	-3	9	
2	5	25	-6	36	-2	4	
3	-2	4	-3	9	1	1	
4	2	4	-4	16	-4	16	
5	-2	4	-3	9	-4	16	
6	18	324	-5	25	-2	4	
7	0	0	-4	16	4	16	
8	-4	16	4	16	13	169	
$Q_i = \sum_{j=1}^{8} y_{ij}^2$		477		176		235	$\sum_{i=1}^{3} Q_i = 888$
$T_i = \sum_{j=1}^{8} y_{ij}$	27		-28		3		$\sum_{i=1}^{3} T_i = 2$
T_i^2	729		784		9		$\sum_{i=1}^{3} T_i^2 = 1522$

In [3]:

```
print(f"Кількість рівнів фактора {m = }\nКількість випробувань на кожному рівні {n = }")
```

Кількість рівнів фактора m = 3
Кількість випробувань на кожному рівні n = 8

Знайдемо загальну і факторну суми квадратів відхилень:

$$S_{\text{заг.}} = \sum_{i=1}^{m} Q_i - \frac{1}{mn}(\sum_{i=1}^{m} T_i)^2$$

$$S_{\text{факт.}} = \frac{1}{n}\sum_{i=1}^{m} (T_i)^2 - \frac{1}{mn}(\sum_{i=1}^{m} T_i)^2$$

Знайдемо залишкову суму квадратів відхилень: $S_{\text{зал.}} = S_{\text{заг.}} - S_{\text{факт.}}$

In [4]:

```
S_gen = Q_sum-T_sum**2/(m*n)
S_fac = T2_sum/n-T_sum**2/(m*n)
S_rem = S_gen-S_fac
Latex("$$S_{заг.} ="+str(S_gen)+"$$$$S_{факт.} ="+str(S_fac)+"$$$$S_{зал.} ="+str(S_re
m)+"$$")
```

Out[4]:

$$S_{\text{заг.}} = 887.8333333333334$$
$$S_{\text{факт.}} = 190.08333333333334$$
$$S_{\text{зал.}} = 697.75$$

Знайдемо факторну і залишкову дисперсії:

$$s^2_{\text{факт.}} = \frac{S_{\text{факт.}}}{m-1}; s^2_{\text{зал.}} = \frac{S_{\text{зал.}}}{m(n-1)}$$

In [5]:

```
s2_fac = S_fac/(m-1)
s2_rem = S_rem/(m*(n-1))
Latex("$$s^2_{факт.} ="+str(s2_fac)+"$$$$s^2_{зал.} ="+str(s2_rem)+"$$")
```

Out[5]:

$$s^2_{\text{факт.}} = 95.04166666666667$$
$$s^2_{\text{зал.}} = 33.226190476190474$$

Порівняємо факторну і залишкову дисперсії за критерієм Фішера, для чого знайдемо спостережуване значення критерію:

$$F_{\text{спост.}} = \frac{s^2_{\text{факт.}}}{s^2_{\text{зал.}}}$$

In [6]:

```
F_obs = round(s2_fac/s2_rem,4)
Latex()
k1 = m-1
k2 = m*(n-1)
Latex(r"$$F_{спост.}="+str(F_obs)+"$$"+r"Оскільки кількість ступенів вільності чисельника $k_1 = "+str(k1)+"$, а знаменника $k_2 = "+str(k2)+r"$ і рівень значущості $\alpha = "+str(alpha)+"$, за таблицею для критерію Фішера знаходимо критичну точку (в нашому випадку функцією з `scipy.stats`)")
```

Out[6]:

$F_{спост.} = 2.8604$

Оскільки кількість ступенів вільності чисельника $k_1 = 2$, а знаменника $k_2 = 21$ і рівень значущості $\alpha = 0.1$, за таблицею для критерію Фішера знаходимо критичну точку (в нашому випадку функцією з `scipy.stats`)

In [8]:

```
from scipy.stats import f
Latex(r"$$F_{крит.}="+str(round(f.ppf(1-alpha, k1, k2),4))+"$$")
```

Out[8]:

$F_{крит.} = 2.5746$

Оскільки $F_{спост.} > F_{кр.}$, гіпотезу про рівність групових середніх відкидаємо