

## Задача 1. Определение квантового выхода

Для реакции  $2COCl_2 + O_2 = 2CO_2 + Cl_2$  известно, что поглотилось  $4,4 \cdot 10^{18}$  квантов света, а изменение числа молей прореагировавшего вещества составило  $1,32 \cdot 10^{-5}$ . Определите квантовый выход  $\varphi$ .

### Описание задачи:

Известно, что:

$$\varphi = \frac{N_{\text{прореаг}}}{N_{\gamma}}, \quad (1)$$

где  $\varphi$  - квантовый выход,  $N_{\text{прореаг}}$  - число молей прореагировавшего вещества,  $N_{\gamma}$  - число квантов света.

$$N_{\text{прореаг}} = n * N_A, \quad (2)$$

где  $n$  - изменение числа молей прореагировавшего вещества,  $N_A$  - число Авогадро.

Тогда после преобразований имеем:

$$\varphi = \frac{n * N_A}{N_{\gamma}} \quad (3)$$

### Реализация в R:

#### Вариант 1.

Важно в программировании - сохранять значения переменных (variables). Присвоим переменным значения числа молей, числа Авогадро, квантов света. В R это обычно делается с помощью вот этих символов: «<-».

```
n<-1.31*10^-5
Na<-6.02*10^23
Ny<-4.4*10^18
```

Введем формулу (3) для вычисления квантового выхода и найдем ответ.

```
> x <- n*Na/Ny
> x
[1] 1.792318
```

#### Вариант 2.

Создадим функцию и найдем ответ. Для этого пишем ключевое слово **function**, в круглых скобках обозначаем переменные, с которыми

будем работать:  $n$ ,  $N_a$ ,  $N_y$ . С ними функция и будет работать. Внутри фигурных скобок пишем уравнение (3), которые будут выполняться при запуске функции. На выходе функция выдаст то, что вычисляется внутри функции `return()`. В `pow()` введем значения переменных, исходя из условия задачи.

```
> pow<-function(n,Na,Ny){
+ power<-n*Na/Ny
+ return(power)
+ }
> pow(1.31*10^-5,6.02*10^23,4.4*10^18)
[1] 1.792318
```

**Ответ:** квантовый выход  $\phi$  равен 1,79

## Задача 2. Капиллярная конденсация паров бензола

Рассчитайте и постройте интегральную кривую  $\frac{V_n}{V_{max}} = f(r)$  распределения объема пор по размерам для активного угля, пользуясь экспериментальными данными капиллярной конденсации паров бензола при 293 К:

Таблица 1

### Экспериментальные данные капиллярной конденсации паров бензола

№	1	2	3	4	5	6	7
Давление, $p \cdot 10^{-2}$ , Па	19,8	29,9	39,8	59,6	79,7	89,4	98,3
Адсорбция, $A$ , моль/кг							
Изотерма адсорбции	4,5	5,4	6,5	10,2	14,4	17,0	20,0
Изотерма десорбции	4,5	6,0	9,0	13,9	17,5	19,0	20,0

При этой температуре для бензола молярный объем  $V_M = 89 \cdot 10^{-6}$  м<sup>3</sup>/моль, поверхностное натяжение  $\sigma = 28,9$  мДж/м<sup>2</sup>, давление насыщенного пара  $p_s = 99,3 \cdot 10^2$  Па.

#### Описание задачи:

Для расчёта интегральной кривой распределения пор по размерам

используют кривую десорбции. Радиус пор рассчитывают по уравнению Кельвина:

$$r = \frac{2\sigma V_M}{RT \ln\left(\frac{p_s}{p}\right)}, \quad (4)$$

где  $V_M$  – молярный объем,  $\sigma$  - поверхностное натяжение,  $p_s$  - давление насыщенного пара,  $R$  – газовая постоянная,  $T$  – температура в Кельвинах.

Объем пор адсорбента  $V_{\Pi}$ , приходящийся на единицу массы адсорбента и заполненный при данном равновесном давлении адсорбата, находят из соотношения:

$$V_{\Pi} = A * V_M, \quad (5)$$

где  $A$  – адсорбция.

### Реализация в R:

Чтобы найти радиус пор (4), для начала найдем значения  $p_s/p$ . Создадим вектор  $x$ , это и будет отношение давления насыщенного пара к экспериментальным данным давлений (табл. 1). Чтобы создать вектор из нескольких значений, нужно воспользоваться функцией  $c()$ .

```
> x<-9930/c(1980,2992,3980,5960,7970,8940,9830)
```

```
> x
```

```
[1] 5.015152 3.318850 2.494975 1.666107 1.245922 1.110738 1.010173
```

Далее найдем натуральный логарифм этого отношения  $\log(p_s/p)$ :

```
> log(x)
```

```
[1] 1.61246363 1.19961842 0.91427866 0.51049000 0.21987599  
0.10502489
```

```
[7] 0.01012154
```

Теперь можно вычислить размер пор  $r$  по формуле (4), подставляя данные из условия задачи:

```
> r<-2*28.9*0.001*89*10^-6/8.31/293/log(x)
```

```
> r
```

```
[1] 1.310265e-09 1.761189e-09 2.310844e-09 4.138681e-09 9.608850e-09
```

```
[6] 2.011671e-08 2.087384e-07
```

Введем данные адсорбции (табл. 4) с помощью векторизации:

```
> A<-c(4.5,5.4,6.5,10.2,14.4,17,20)
```

```
> A
```

```
[1] 4.5 5.4 6.5 10.2 14.4 17.0 20.0
```

Найдем по формуле (5) объем пор абсорбента:

```
> Vn<-A*89*10^-6
```

```
> Vn
```

```
[1] 0.0004005 0.0004806 0.0005785 0.0009078 0.0012816 0.0015130
```

```
0.0017800
```

Рассчитаем отношение объема пор к максимальному значению с помощью функции `max()`.

```
> Vx<-Vn/max(Vn)*100
```

```
> Vx
```

```
[1] 22.5 27.0 32.5 51.0 72.0 85.0 100.0
```

С помощью функции `plot()` построим график распределения объема пор по размерам для активированного угля (рис. 1).

**Описание функции `plot()`:**

```
plot(x, y, xlim=range(x),ylim=range(y),type="p", main, xlab, ylab, ...)
```

x, y – координаты точек x и y.

xlim, ylim – значения для осей x и y.

Type – тип графика (« p » для точек).

Main – название графика.

Xlab, ylab – название осей.

```
> plot(r,Vx,xlab="r,нм",ylab="Vn/Vmax,%",main="Интегральная кривая распределения объема пор по размерам для активного угля",type="l")
```

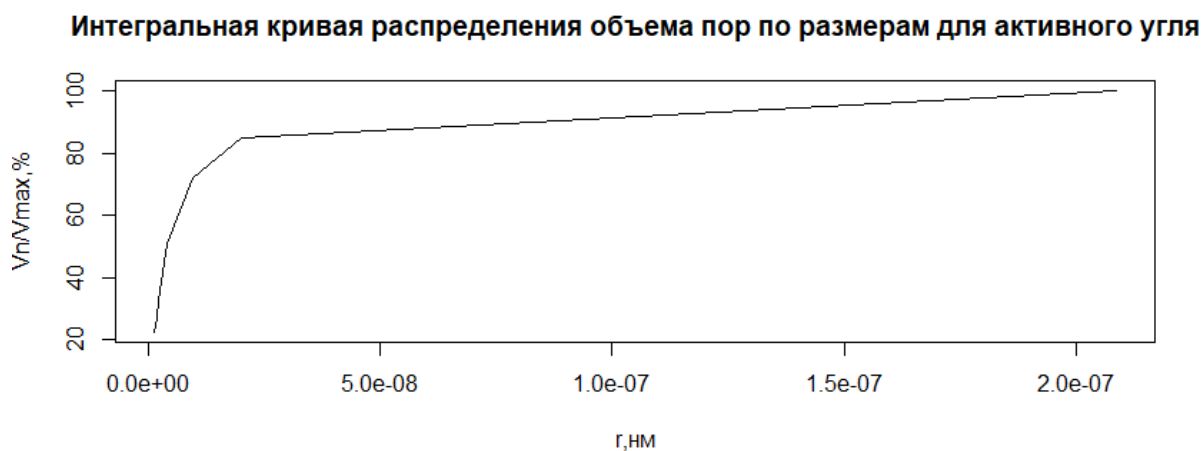


Рис. 1. Интегральная кривая распределения объема пор по размерам для активного угля

### Задача 3. Адсорбция

Ниже приведены экспериментальные данные по адсорбции азота на

TiO<sub>2</sub> при 75К.

Таблица 2

**Экспериментальные данные по адсорбции азота на TiO<sub>2</sub>**

№	1	2	3	4	5
Давление, $p \cdot 10^{-2}$ , Па	60,94	116,41	169,84	218,65	272,25
Адсорбция, А, моль/кг	0,367	0,417	0,467	0,512	0,567

Постройте график, соответствующий линейной форме уравнения Брунауэра-Эметта-Теллера (БЭТ). Найдите константы  $A_{\infty}$  и  $C$ . Рассчитайте удельную поверхность адсорбента. Давление насыщенного пара азота при указанной температуре  $p_s = 78,3 \cdot 10^3$  Па; площадь, занимаемая одной молекулой азота в адсорбционном слое,  $s_0 = 0,16$  нм<sup>2</sup>.

**Описание задачи:**

Уравнение БЭТ в линейной форме имеет вид:

$$\frac{p/p_s}{A \cdot (1 - \frac{p}{p_s})} = \frac{1}{A_{\infty} \cdot C} + \frac{C-1}{A_{\infty} \cdot C} \cdot \frac{p}{p_s}, \quad (6)$$

где  $p_s$  – давление насыщенного пара,  $A$  – адсорбция,  $A_{\infty}$  и  $C$  – константы.

Сначала рассчитывают значения  $\frac{p}{p_s}$  и  $\frac{p/p_s}{A \cdot (1 - \frac{p}{p_s})}$ .

По этим данным строят график в координатах уравнения БЭТ в линейной форме. Из графика находят отрезок, отсекаемый на оси ординат  $a$ , и тангенс угла наклона прямой:

$$\frac{1}{A_{\infty} \cdot C} = a, \quad (7)$$

$$\operatorname{tg} \alpha = \frac{C-1}{A_{\infty} \cdot C}, \quad (8)$$

Удельную поверхность адсорбента рассчитывают по уравнению:

$$S_{уд} = A_{\infty} \cdot N_A \cdot s_0, \quad (9)$$

где  $N_A$  – число Авогадро,  $s_0$  – площадь, занимаемая одной молекулой азота в адсорбционном слое.

**Реализация в R:**

Введем данные давления и адсорбции (табл. 2) с помощью векторизации:

```
> p<-c(6094,11641,16984,21865,27225)
```

```
> p
```

```
[1] 6094 11641 16984 21865 27225
```

```
> A<-c(0.367,0.417,0.467,0.512,0.567)
```

```
> A
```

```
[1] 0.367 0.417 0.467 0.512 0.567
```

Сначала рассчитаем значения  $p/ps$ :

```
> x<-p/78300
```

```
> x
```

```
[1] 0.07782886 0.14867178 0.21690932 0.27924649 0.34770115
```

Далее рассчитаем отношение  $\frac{p/ps}{A*(1-p/ps)}$ :

```
> y<-p/ps/A/(1-p/ps)
```

```
> y
```

```
[1] 0.2299657 0.4187892 0.5931292 0.7567126 0.9401052
```

Построим график (рис. 2), соответствующий линейной форме уравнения БЭТ, где введем границы для осей  $x$  и  $y$ :

```
> plot(x, y, xlim=c(0.07,0.35),ylim=c(0,0.95),type="l",  
xlab="p/ps",ylab="p/ps/A/(1-p/ps)",main="График, соответствующий  
уравнению БЭТ")
```

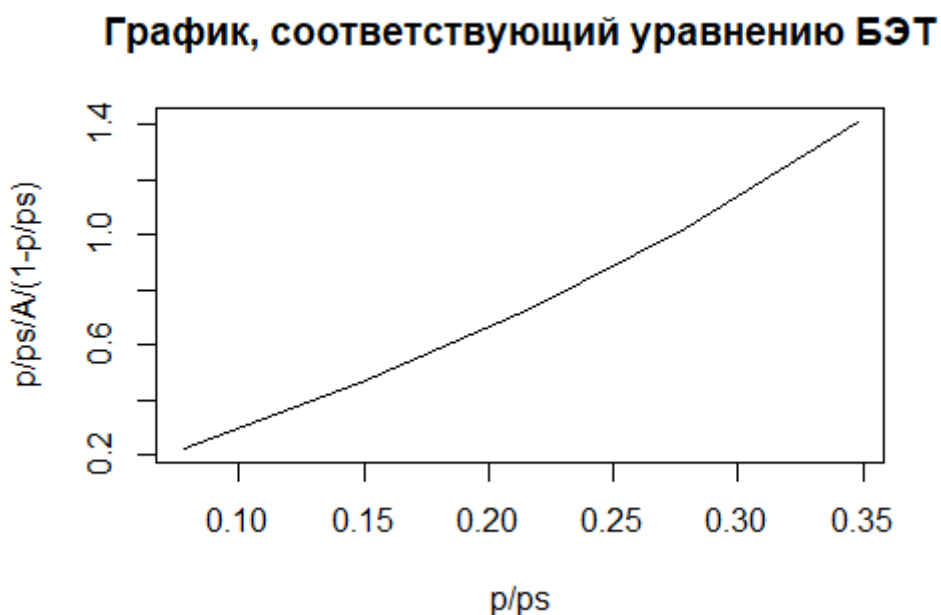


Рис. 2. График, соответствующий уравнению БЭТ

Далее, чтобы найти отрезок  $a$  и тангенс угла наклона прямой по графику, построим простую модель линейной регрессии,

используя отношение  $p/ps$  в качестве объясняющей переменной и отношение  $\frac{p/ps}{A*(1-p/ps)}$  в качестве переменной ответа, с помощью функции `summary()` (см. задачу 7):

```

> model<-lm(y~x)
> summary(model)
Call:
lm(formula = y ~ x)
Residuals:
    1      2      3      4      5
-0.0004128  0.0025911 -0.0020547 -0.0019804
 0.0018568
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02624    0.00274   9.573  0.00242
x            2.62298    0.01171 224.093 1.96e-07
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.002482 on 3 degrees of freedom
Multiple R-squared:  0.9999,          Adjusted
R-squared:  0.9999
F-statistic: 5.022e+04 on 1 and 3 DF, p-value: 1.96e-07

```

Тогда из полученных данных можно составить регрессионное уравнение:

$$y = 2.62 * x + 0.026 \quad (10)$$

Данное уравнение имеет вид:

$$y = k * x + b \quad (11)$$

где  $k$  – угловой коэффициент, равный тангенсу угла наклона этой прямой,  $b$  – пересечение прямой на оси  $y$ .

В нашем случае  $a = b = 0.026$ ,  $tga = k=2.62$

Преобразуем уравнение (7), используя (8), и найдем  $C$ :

$$C = \frac{tga+a}{a} \quad (12)$$

```
> tga=2.62
```

```
> a=0.026
```

> C<-(tga+a)/a

> C

[1] 101.7692

Найдем  $A_\infty$  уравнению (7):

> A=1/C/a

> A

[1] 0.3779289

Рассчитаем удельную поверхность адсорбента по уравнению (8):

```
> pow<-function(A,Na,s0){
```

```
+ power<-A*Na*s0
```

```
+ return(power)
```

```
+ }
```

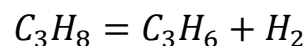
```
> pow(0.3779289,6.02*10^23,0.16*10^-18)
```

```
[1] 36402.11
```

Ответ:  $A_\infty = 0,3779$ ;  $C = 101,7692$ ;  $s = 36402,11\text{м}^2$

#### Задача 4. Расчет парциальных давлений

В каталитическом реакторе протекает реакция неокислительного дегидрирования пропана:



Задана зависимость константы равновесия реакции от температуры:

$K_p = 1,48 * 10^7 * \exp\left(\frac{-15403}{T}\right)$  и заданы условия проведения реакции

$T=723\text{К}$  и  $p=1$  атм.

Задан состав исходной смеси (%об):

$C_3H_8=80$

$H_2=20$

$C_3H_6=0$

Рассчитать равновесный состав реакционной смеси.

#### Описание задачи:

Парциальное давление – часть общего давления смеси газов, приходящаяся на отдельный газ (если бы он занимал весь объем смеси).

Определяется по формуле (закон Дальтона):



$$p_{\text{газа}} = \frac{p_{\text{смеси}} + C(\%)}{100\%} \quad (13)$$

Из реакции неокислительного дегидрирования пропана делаем вывод, что константа реакции будет иметь вид:

$$K_p = \frac{p_{\text{C}_3\text{H}_6} p_{\text{H}_2}}{p_{\text{C}_3\text{H}_8}} \quad (14)$$

Парциальное давление водорода на выходе  $p_{\text{H}_2}$  вычисляется по формуле:

$$p_{\text{H}_2} = p_{0,\text{H}_2} + p_{0,\text{C}_3\text{H}_8} - p_{\text{C}_3\text{H}_8} \quad (15)$$

Парциальное давление пропена на выходе  $p_{\text{C}_3\text{H}_6}$  вычисляется по формуле:

$$p_{\text{C}_3\text{H}_6} = p_{0,\text{C}_3\text{H}_6} + p_{0,\text{C}_3\text{H}_8} - p_{\text{C}_3\text{H}_8} \quad (16)$$

Тогда из уравнений (14) – (16) получится нелинейное алгебраическое уравнение с одним неизвестным  $p_{\text{C}_3\text{H}_8}$ :

$$K_p * p_{\text{C}_3\text{H}_8} - (p_{0,\text{C}_3\text{H}_6} + p_{0,\text{C}_3\text{H}_8} - p_{\text{C}_3\text{H}_8})(p_{0,\text{H}_2} + p_{0,\text{C}_3\text{H}_8} - p_{\text{C}_3\text{H}_8}) = 0 \quad (17)$$

где  $K_p$  – константа равновесия реакции,  $p_{0,\text{C}_3\text{H}_6}$ ,  $p_{0,\text{C}_3\text{H}_8}$ ,  $p_{0,\text{H}_2}$  – парциальное давление пропана, пропена и водорода на входе соответственно, атм,  $p_{\text{C}_3\text{H}_8}$  – парциальное давление на выходе, атм.

### Реализация в R:

Введем исходные данные:

- > T=723
- > P=1
- > C3H8=80
- > C3H6=0
- > H2=20

Зададим зависимость константы равновесия реакции от температуры:

$$> Kp=1.48*10^7*\exp(-15403/T)$$

Вычисляем парциальные давления пропана, пропена (изначально состав исходной смеси равен нулю, следовательно, парциальное давление на входе равно нулю), водорода соответственно по формуле (13):

- > p0C3H8=(P+C3H8)/100
- > p0C3H6=0
- > p0H2=(P+H2)/100

Найдем  $p_{C3H8}$  по уравнению (17). Используем функцию `function(x)`, в фигурных скобках укажем уравнение, по которому будем находить  $p_{C3H8}$  ( $x = p_{C3H8}$ ). С помощью функции `uniroot()` найдем решение.

#### Описание функции `uniroot()`:

`uniroot(f, interval, ...)`

$f$  – функция, нуль которой (т.е. корень) вычисляется;

`interval` – числовой вектор – интервал, на котором ищется корень.

Результатом функции `uniroot` является список из четырёх компонент: искомое решение  $x^*$  – `root`, значение функции в найденной точке  $f(x^*)$  – `f.root`, число итераций `iter` и точность решения `estim.prec`. Если  $x^*$  совпадает с одним из концов заданного интервала поиска, то тогда значение `estim.prec` – NA

```
> f <- function(pC3H8){f=Kp*pC3H8-(p0C3H6+p0C3H8-
pC3H8)*(p0H2+p0C3H8-pC3H8)}
```

```
> uniroot(f,c(0,1))
```

```
$root
```

```
[1] 0.7826995
```

```
$f.root
```

```
[1] 7.944708e-07
```

```
$iter
```

```
[1] 9
```

```
$estim.prec
```

```
[1] 6.103516e-05
```

Введем найденное значение и вычислим  $p_{H2}$ ,  $p_{C3H6}$  по формулам (15) и (16):

```
> pC3H8 <- 0.7719754
```

```
> pC3H6 <- p0C3H6+p0C3H8-pC3H8
```

```
> pC3H6
```

```
[1] 0.0380246
```

```
> pH2 <- p0H2+p0C3H8-pC3H8
```

```
> pH2
```

```
[1] 0.2480246
```

Ответ:  $p_{C3H8}=0,77$  атм;  $p_{H2}=0,24$  атм;  $p_{C3H6}=0,03$  атм.

### Задача 5. Сравнение нескольких дисперсий

При получении фосфора возгонкой из фосфатов измерялась степень восстановления фосфата при четырех различных температурах. В таблице

приведены результаты статистического анализа однородности дисперсий воспроизводимости результатов при разных температурах.

Таблица 3

**Результаты статистического анализа однородности дисперсий воспроизводимости**

Температура	$s_i^2$	$f_i$	$f_i s_i^2$	$lgs_i^2$	$f_i lgs_i^2$	$\frac{1}{f_i}$
$T_1$	1,72	5	8,60	0,2355	1,177	0,200
$T_2$	1,60	4	6,40	0,2041	0,816	0,250
$T_3$	1,97	6	11,82	0,2945	1,767	0,167
$T_4$	2,37	8	18,96	0,3747	2,995	0,125
$\Sigma$		23	45,78		6,755	0,742

Определить, не меняется ли точность анализа с температурой.

**Описание задачи:**

При определении оценки дисперсии по текущим измерениям по формуле:

$$s_y^2 = \frac{f_1 * s_1^2 + f_2 * s_2^2 + \dots + f_n * s_n^2}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^k f_i * s_i^2}{f}, \quad (18)$$

где  $f_i = n_i - 1$  – число степеней свободы для  $i$  – й выборки объема  $n_i$ ,  $S_i^2$  – выборочная дисперсия  $i$ -й выборки,  $f = \sum_i f_i$  – общее число степеней свободы, и  $k$  – число выборок.

Была принята нулевая гипотеза равенства соответствующих генеральных дисперсий. Проверить эту гипотезу для выборок разного объема можно по критерию Бартлета. Бартлет показал, что в условиях нулевой гипотезы отношение  $B/C$ , где:

$$B = 2,303(f lgs_y^2 - \sum_{i=1}^n f_i * lgs_i^2) \quad (19)$$

$$C = 1 + \frac{1}{3*(n-1)} (\sum_{i=1}^n \frac{1}{f_i} - \frac{1}{f}) \quad (20)$$

Гипотеза равенства генеральных дисперсий принимается, если:

$$\frac{B}{C} \leq \chi_{1-p}^2 \quad (21)$$

при выбранном уровне значимости  $p$ . Различие между выборочными дисперсиями можно считать незначимым, а сами выборочные дисперсии – однородными. Так как всегда  $C > 1$ , если окажется  $B \leq \chi_{1-p}^2$ , нулевую гипотезу следует принять; если  $B > \chi_{1-p}^2$ , критерий Бартлета вычисляют полностью.

### Реализация в R:

Сначала задаем дисперсии воспроизводимости и степени свободы (табл. 3) в виде векторов:

```
> sj<- c(1.72,1.6,1.97,2.37)
> fi <- c(5,4,6,8)
```

По этим данным найдем суммарную дисперсию воспроизводимости по формуле (18):

```
> Sy <- sum(fi*sj)/sum(fi)
> Sy
[1] 1.990435
```

Чтобы найти критерий Бартлета (19), (20), для начала найдем величину  $B$  (19):

```
> B <- 2.303*(sum(fi)*log10(Sy)-sum(fi*log10(si)))
> B
[1] 0.2692129
```

Найдем табличное значение критерия Пирсона при трех степенях свободы и уровню значимости 0,05 с помощью функции **qchisq(p,df)**, где  $p$  – доверительная вероятность,  $df$  - число степеней свободы:

```
> df<-3
> p<-0.95
> qchisq(p,df)
[1] 7.814728
```

Сравним величину  $B$  и критерий Пирсона ( $\chi_{1-p}^2$ ):

```
> B<=qchisq(p,df)
[1] TRUE
```

Так как  $B \leq \chi_{1-p}^2$ , следовательно, нулевую гипотезу можно принять, не вычисляя величину  $C$ .

Ответ: критерий Бартлета позволяет считать, что точность анализа не

зависит от температуры. Выборочные дисперсии однородны, поэтому в качестве оценки для дисперсии воспроизводимости можно взять средневзвешенную дисперсию, равную 1,915, с числом степеней свободы, равным 23.

### **Задача 6. Критерии Бартлета и Кохрена (Сравнение нескольких дисперсий нормальных генеральных совокупностей по выборкам)**

Определение содержания нефтепродуктов в питьевой воде методом ИК-спектрометрии проводилось четырьмя лабораториями с получением следующих результатов:

- 1) лаборатория 1: 0.06; 0.07; 0.06; 0.05; 0.07;
- 2) лаборатория 2: 0.06; 0.08; 0.04; 0.09; 0.05; 0.06;
- 3) лаборатория 3: 0.05; 0.07; 0.05; 0.06;
- 4) лаборатория 4: 0.06; 0.07; 0.06; 0.10; 0.08.

Необходимо с доверительной вероятностью  $p=0,95$  установить, можно ли считать равными оценки случайных погрешностей результатов анализа разных лабораторий.

#### **Описание задачи:**

Исходя из задачи, необходимо проверить гипотезу об однородности четырех выборочных дисперсий по критерию Бартлета (19) – (21). В R функция `bartlett.test()` осуществляет проверку данного критерия.

Так же если объем выборок примерно одинаковый, то можем использовать тест экстремальных значений Кохрена (Cochran) `cochran.test()` из пакета `outliers`.

Кохрен исследовал распределение максимальной выборочной дисперсии к сумме всех дисперсий:

$$G = \frac{s_{max}^2}{\sum_{i=1}^n s_i^2} \quad (22)$$

Если найденное значение критерия Кохрена окажется меньше табличного, то расхождение между дисперсиями нужно считать случайным

при выбранном уровне значимости  $p$ :

$$G < G_{1-p}(n, f), \quad (23)$$

где  $n$  – объем выборок,  $f$  – степени свободы.

### Описание функции `bartlett.test()`:

`bartlett.test(x, g...)`:

Аргументы:

$x$  – числовой вектор значений.

$g$  – вектор или фактор, дающий группу для соответствующих элементов  $x$ . Пренебрегаем, если  $x$  – список и можно использовать `bartlett.test(x)`. Если выборки еще не содержатся в списке, используют `bartlett.test(list(x...))`.

### Описание функции `cochran.test()`:

`cochran.test(x,y)`:

$x$  – числовой вектор, содержащий значения дисперсий для каждой выборки  $s_i^2$ .

$y$  – числовой вектор, содержащий объем каждой выборки.

### Реализация в R:

Запишем результаты ИК – спектрометрии:

```
> x1<-c(0.06,0.07,0.06,0.05,0.07)
> x2<-c(0.06,0.08,0.04,0.09,0.05,0.07)
> x3<-c(0.05,0.07,0.05,0.06)
> x4<-c(0.06,0.07,0.06,0.1,0.08)
```

Проверим гипотезу по критерию Бартлета:

```
> bartlett.test(list(x1,x2,x3,x4))
```

**Bartlett test of homogeneity of variances**

**data: list(x1, x2, x3, x4)**

**Bartlett's K-squared = 3.1439, df = 3, p-value = 0.37**

Bartlett's K-squared=3,1439 (значение критериальной статистики теста Бартлета), число степеней свободы 3, p-value=0.37, т.е. отвергнуть гипотезу  $H_0$  можно только при допустимой ошибке в 37%. Следовательно, гипотеза об однородности дисперсий принимается на 5% уровне значимости.

Для проверки (21) можем применить функцию `qchisq(p,df)`, где  $p$  –

доверительная вероятность, df - число степеней свободы:

```
> p<-0.95
> df<-3
> qchisq(p,df)
[1] 7.814728
> B<-3.1439
> B<=qchisq(p,df)
[1] TRUE
```

Поэтому выборочные дисперсии могут быть признаны однородными, а оценки случайных погрешностей результатов ИК-спектрометрии, проведенной четырьмя лабораториями, - равными.

Проверим гипотезу Кохрена. Для этого установим пакет outliers:

```
> install.packages("outliers")
package 'outliers' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
  C:\Users\Ангелина\AppData\Local\Temp\RtmpMD82mw\downloaded_packages
> library("outliers")
```

Далее проверим гипотезу Кохрена, с указанием объема выборок (5, 6, 4, 5) элементов, var() содержит значения дисперсий для каждой выборки:

```
> cochrans.test(object=c(var(x1),var(x2),var(x3),var(x4)),data=c(5,6,4,5))
Cochran test for outlying variance
data: c(var(x1), var(x2), var(x3), var(x4))
C = 0.44211, df = 5, k = 4, p-value = 0.4405
alternative hypothesis: Group 2 has outlying variance
sample estimates:
      1          2          3          4
7.000000e-05 3.500000e-04 9.166667e-05 2.800000e-04
```

Cochran C = 0,44211 (значение критериальной статистики теста Кохрена), число степеней свободы (средний объем выборки) 5, число групп 4, p-value 0,4405. Альтернативная гипотеза – дисперсия второй выборки значительно больше остальных (является «выбросом»). Поскольку p-value = 0.4405, то отвергнуть гипотезу  $H_0$  можно только при допустимой ошибке в 44,05%. Следовательно, гипотеза об однородности дисперсий принимается на 5% уровне значимости.

Так же можем проверить (23), используя функцию **qcochran(p, n, k)** из

того же пакета, где  $p$  – доверительная вероятность,  $n$  – объем одной выборки (если объемы различаются, то берется среднее значение),  $k$  – число выборок:

```
> n<-5  
> p<-0.95  
> k<-4  
> qcochran(p,n,k)  
[1] 0.6287245
```

Значение критериальной статистики теста Кохрена:

```
> C<-0.44211
```

Далее проверим (23):

```
> C<=qcochran(p,n,k)  
[1] TRUE
```

Ответ: исходя из данных ИК-спектрометрии, с использованием критериев Кохрена и Бартлета с доверительной вероятностью  $p=0,95$  установили, что оценки случайных погрешностей результатов анализа разных лабораторий можно считать равными, так как (21), (23) выполняются.

### Задача 7. Линейная регрессия

Требуется определить зависимость растворимости хлорида бария в воде ( $y$ ) при  $70^\circ\text{C}$ . Объем выборки  $n=6$ . Экспериментальные данные приведены ниже:

Таблица 4

#### Экспериментальные данные хлорида кальция и растворимости хлорида бария в воде

№	1	2	3	4	5	6
$x_{CaCl_2}, \%$	0	5	8	10	15	20
$y$ (растворимость $BaCl_2, \%$ )	32	25	20	17	11	5

#### Описание задачи:

Требуется определить по методу наименьших квадратов коэффициенты линейного уравнения регрессии по выборке объема  $n$ :

$$\hat{y} = b_0 + b_1 * x \quad (24)$$

где



$$b_0 = \frac{\sum_{i=1}^n y_i * \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i * \sum_{i=1}^n x_i * y_i}{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (25)$$

$$b_0 = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (26)$$

$b_0$  проще найти по известному  $b_1$  из уравнения (24):

$$b_0 = \bar{y} - b_1 * \bar{x} \quad (27)$$

Чтобы построить линейную модель регрессии используем функцию **lm(formula=f)**.

**Описание функции:**

**lm(formula= y~x)**

Функция содержит только формулу от переменных (векторов, в случае парной выборки)

«y~x» - строится модель зависимости y от x

**Реализация в R:**

Запишем экспериментальные данные (табл. 4) как векторы:

```
> x<-c(0,5,8,10,15,20)
> y<-c(32,25,20,17,11,5)
```

Построим линейную модель регрессии при помощи функции **lm(formula=f)**:

```
> p.lm<-lm(formula=y~x)
```

Найдем полную информацию о построенной модели с помощью команды **summary()**:

```
> summary(p.lm)
```

**Call:**

**lm(formula = y ~ x)**

**Residuals:**

	1	2	3	4	5	6
	0.5658	0.3421	-0.5921	-0.8816	-0.1053	0.6711

**Coefficients:**

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	31.43421	0.52085	60.35	4.51e-07 ***
x	-1.35526	0.04472	-30.31	7.06e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7117 on 4 degrees of freedom

**Multiple R-squared: 0.9957, Adjusted R-squared: 0.9946**  
**F-statistic: 918.5 on 1 and 4 DF, p-value: 7.06e-06**

Команда `summary()` выдает полную информацию о построенной модели:

Значения остатков (`residuals` – разность модельных и истинных значений переменной  $y$ ).

Коэффициенты модели и оценку их значимости по критерию Стьюдента (в этом случае все коэффициенты значимы, поскольку вероятности равны  $4,51 \cdot 10^{-7}$  и  $7,06 \cdot 10^{-6}$ , они меньше 0.05, следовательно, считаем, что существует линейная зависимость между  $x$  и  $y$ ).

По результатам регрессионного анализа мы вычислили коэффициенты регрессионного уравнения:

$$b_0 = 31,43$$

$$b_1 = -1,355$$

Тогда регрессионное уравнение (24) будет выглядеть следующим образом:

$$\hat{y} = 31,43 - 1,355x$$

Оценка значимости зависимости по критерию Фишера:  $F=918.5$   $p\text{-value}=7.06 \cdot 10^{-6}$ , то есть критерий Фишера значим, так как значение  $p\text{-value}$  меньше 0,05.

Квадрат коэффициента корреляции  $R\text{-squared}= 0.99$ , который показывает долю дисперсии  $y$ , объясненной с использованием модели.

Для визуализации построенной модели (рис. 3) используем функции `plot()` и `abline()`.

**Описание функции `abline()`:**

`abline(x, y, untf = FALSE, ...)`

`abline(h=, untf = FALSE, ...)`

`abline(v=, untf = FALSE, ...)`

$x, y$  – параметры в линейном уравнении.

`Untf` – если TRUE, то рисует линию в преобразованных координатах.

$h, v$  –  $y$  и  $x$  значения для горизонтальной и вертикальной линии соответственно.

**abline():** строит прямую по найденным  $b_0$  и  $b_1$ ,

**plot():** строит экспериментальные точки).

```
> plot(x,y, main="Зависимость растворимости хлорида бария в воде")
```

```
> abline(lm(y~x))
```

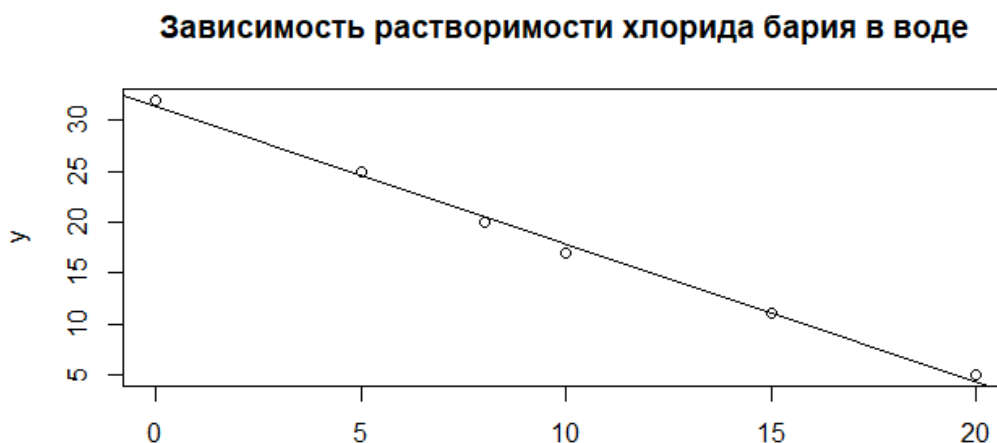


Рис. 3. Зависимость растворимости хлорида бария в воде

### Задача 8. Дисперсионный анализ

Рассмотрим применение однофакторного дисперсионного анализа для выяснения влияния вида галоидного алкила (фактор А) на процесс радикальной полимеризации. Изучалось влияние на выход полимера ( $y\%$ ) пяти различных галоген-алкилов:  $CH_3J_{(a1)}$ ;  $C_3H_7J_{(a2)}$ ;  $C_4H_9J_{(a3)}$ ;  $C_2H_5Br_{(a4)}$ ;  $C_3H_7Br_{(a5)}$ . Результаты эксперимента с различными галоидными алкилами (фиксированные уровни фактора А) приведены в таблице 5:

Таблица 5

#### Результаты эксперимента с галоидными алкилами

Номер наблюдения	Уровни фактора А				
	a1	a2	a3	a4	a5
1	79,80	87,30	42,45	76,00	70,70
2	86,30	69,60	64,30	83,50	64,65
3	86,50	81,75	78,90	72,80	38,50
4	92,30	77,95	61,00	89,00	77,00
5	76,50	83,65	31,30	76,50	91,30

Номер наблюдения	Уровни фактора А				
	a1	a2	a3	a4	a5
6	87,05	64,80	72,85	87,45	68,00
7	82,50	67,30	58,65	74,50	38,05
8	90,00	75,45	52,50	43,15	79,95
Итоги	A1=980,95	A2=607,8	A3=461,95	A4=652,9	A5=528,63

### Описание задачи:

Данный метод основан на разложении общей дисперсии численного признака на составляющие ее компоненты, сравнивая которые с друг другом посредством F-критерия Фишера можно определить, какую долю (по отношению к совокупности случайных причин) общей вариации признака обуславливает действие на него известных величин (факторов).

Метод основан на сравнении межгрупповой и внутригрупповой изменчивости признака. Каждую группу образуют значения признака при фиксированных значениях (уровнях) известных факторов, поэтому единственным источником дисперсии (изменчивости) внутри каждой группы является суммарное воздействие совокупности случайных причин.

В простейшем случае, если имеется всего один фактор, такое разложение представляется в виде таблицы дисперсионного анализа:

Таблица 6

### Однофакторный дисперсионный анализ

Источник дисперсии	SS сумма квадратов	Степени свободы	Средний квадрат	F статистика
Фактор (межгрупповая)	$SS_{\text{факт}} = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y}_{..})^2$	$k - 1$	$SS_{\text{факт}}^2 = \frac{SS}{k - 1}$	$F = \frac{SS_{\text{факт}}^2}{SS_{\text{сл}}^2}$
Случайная составляющая	$SS_{\text{сл}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$	$\sum_{i=1}^k (n_i - 1)$	$SS_{\text{сл}}^2 = \frac{SS}{\sum_{i=1}^k (n_i - 1)}$	
Общая	$SS_{\text{общ}} = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$	$N - 1$	$SS_{\text{общ}}^2 = \frac{SS}{N - 1}$	

где  $k$  – число групп,  $n_i$  – число наблюдений в  $i$ -ой группе,  $N = \sum_i n_i$  –

общее число наблюдений.

Для проведения однофакторного дисперсионного анализа в R используется линейная модель, в которой единственной независимой переменной выступает этот фактор.

### Описание функции `anova(object)`:

`object` – объект класса `lm`, `glm`.

В случае однофакторной модели (табл. 6) дисперсионного анализа совпадает с результатом дисперсионного анализа модели – сравнением остаточной и модельной дисперсий (последняя является суммарным вкладом всех факторов).

### Реализация в R:

Введем экспериментальные данные  $CH_3J_{(a1)}$ ;  $C_3H_7J_{(a2)}$ ;  $C_4H_9J_{(a3)}$ ;  $C_2H_5Br_{(a4)}$ ;  $C_3H_7Br_{(a5)}$  по данным таблицы 5:

```
> a1<-c(79.8,86.3,86.5,92.3,76.5,87.05,82.5,90)
> a2<-c(87.3,69.6,81.75,77.95,83.65,64.8,67.3,75.45)
> a3<-c(42.45,64.3,78.9,61,31.3,72.85,58.65,52.5)
> a4<-c(76,83.5,72.8,89,76.5,87.45,74.5,93.15)
> a5<-c(70.7,64.65,38.5,77,91.50,68,38.05,79.95)
```

Формируем набор данных, включающий 40 значений признака (вектор `weight`, по 8 значений для каждого из пяти уровней фактора и вектор значений фактора `group`).

```
> group<-gl(5,8,40,label=c("a1","a2","a3","a4","a5"))
> weight<-c(a1,a2,a3,a4,a5)
```

Строим модель зависимости признака от фактора (рис. 4) с помощью функции `boxplot()`, и выполняется дисперсионный анализ.

```
> boxplot(weight~group)
> lm.D<-lm(weight~group)
```

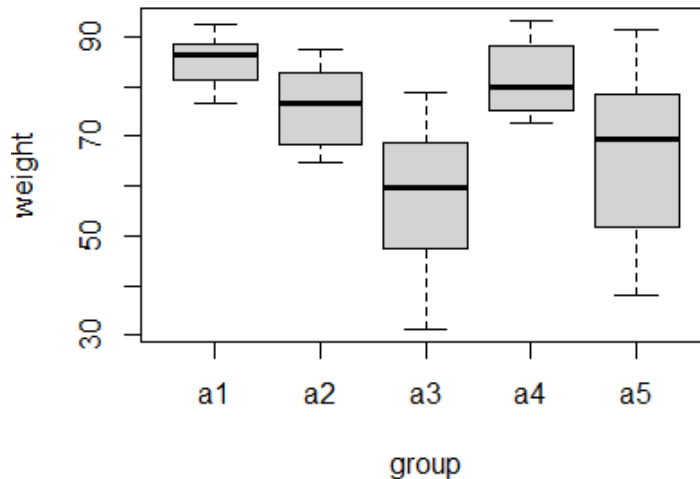


Рис. 4. Модель зависимости признака от фактора

Выполняем дисперсионный анализ с помощью команды `summary()` и `anova()`:

```
> summary(lm.D)
```

Call:

```
lm(formula = weight ~ group)
```

Residuals:

```
  Min   1Q   Median   3Q   Max
-27.994 -5.803  1.634  6.713 25.456
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	85.119	4.347	19.582	< 2e-16 ***
groupa2	-9.144	6.147	-1.487	0.14586
groupa3	-27.375	6.147	-4.453	8.26e-05 ***
groupa4	-3.506	6.147	-0.570	0.57207
groupa5	-19.075	6.147	-3.103	0.00378 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.29 on 35 degrees of freedom

Multiple R-squared: 0.4357, Adjusted R-squared: 0.3712

**F-statistic: 6.756 on 4 and 35 DF, p-value: 0.0003867**

```
> anova(lm.D)
```

Analysis of Variance Table

Response: weight

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	4	4084.7	1021.17	6.7556	0.0003867 ***
Residuals	35	5290.6	151.16		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Как информация по модели `summary()`, так и вызов `anova()` выдают для однофакторной модели те же результаты: значение F-статистики 6,7556 при 4 и 35 степенях свободы, нулевая гипотеза  $H_0$  гласит, что фактор `group` влияет на признак `weight`, уровень значимости (`p-value`) – 0,0003867, что означает, что гипотеза может быть отвергнута только если допустить 0,03867% ошибки. Таким образом, гипотеза о влиянии фактора принимается на 5% уровне значимости.

**Ответ:** влияние вида галоидного алкила (фактор A) на процесс радикальной полимеризации принимаем значимым.