Zápočtová úloha z navrhovania experimentov

Vypracoval: **Martin Kubiš**

1. Zadanie

V tabuľke sú uvedené počty predmetov, ktoré boli za jednotku času umyté v roztoku skúmaných prostriedkov. Testujte tiež významnosť elementárnych kontrastov. Nie úplne celkom vyvážené bloky.

mycí prostr.	A	B	C	D	E
blok					
1	27	26	30		
2	28	26		30	
3	30	29			26
4	31		34	33	
5	29		32		24
6				34	25
7		30	34	31	
8		21	31		23
9		26		33	24
10			33	31	26

2. Model s pevnými efektmi

Podľa zadania ide o model analýzy rozptylu, kde položíme ako základný model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

kde μ je pevný efekt (všeobecný priemer), α_i je pevný efekt pre i-tý saponát, β_j sú efekty pre factor(blok) a y_{ij} sú napozorované hodnoty umytých predmetov, ktorých atribúty su nasledovné

```
> summary(data)
 saponat    hodnota
 A:6     Min.   :21.0
 B:6     1st Qu.:26.0
 C:6     Median :30.0
 D:6     Mean   :28.9
 E:6     3rd Qu.:31.0
         Max.   :34.0
```

Pre ilustraciu prikladam boxplot.

Model v programe R popíšeme nasledovne:

```
>lm1<-lm(hodnota~factor(blok)+factor(saponat))
```

Z modelu nemôžeme vynechať ani faktor saponatu ani bloku, ako nám ukáže test podmodelu:

```
> lm0a<-lm(hodnota~saponat)
> anova(lm0a,lm1)
Analysis of Variance Table

Model 1: hodnota ~ factor(saponat)
Model 2: hodnota ~ factor(blok) + factor(saponat)
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     25 92.833
2     16 28.978  9    63.856 3.9175 0.008503 **

> lm0b<-lm(hodnota~factor(blok))
> anova(lm0b,lm1)
Analysis of Variance Table

Model 1: hodnota ~ factor(blok)
Model 2: hodnota ~ factor(blok) + factor(saponat)
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
```

```
1     20 238.667
2     16  28.978  4   209.689 28.945 3.792e-07 ***
> anova(lm1)
Analysis of Variance Table

Response: hodnota
                Df  Sum Sq Mean Sq F value    Pr(>F)
factor(blok)     9 130.033  14.448  7.9775 0.0001916 ***
factor(saponat)  4 209.689  52.422 28.9448 3.792e-07 ***
Residuals       16  28.978   1.811
```

V oboch prípadoch sme zamietli hypotézu, že vplyv saponátu, či bloku je nepodstatný. Ďalší argument je koeficient R^2, čiže miera vysvetlenia jednotlivých modelov:

```
> summary(lm1)
Adjusted R-squared: 0.8575
> summary(lm0a)
Adjusted R-squared: 0.7079
> summary(lm0b)
Adjusted R-squared: 0.06139
```

Použijeme teda plný model.

K overeniu predpokladov použijeme *qqplot* a rezidua a aspoň od pohľadu môžeme povedať, že pozorovania sú približne normálne a platí homoskedasticita.

Uvedený model $lm1$ nám dáva v progreme R následovné výstupy:

```
> anova(lm1)
              Df  Sum Sq Mean Sq F value    Pr(>F)
saponat        4 275.867  68.967 38.0798 5.493e-08 ***
factor(blok)   9  63.856   7.095  3.9175  0.008503 **
Residuals     16  28.978   1.811
```

Pre overenie vidíme, že model má 4 stupne voľnosti pro *saponat* čo odpovedá piatim druhom saponátu a 9 stupňov voĺnosti pre *blok*, čo odpovedá desiatim blokom. Z uvedeného výstupu je tiež patrné že zamietam nulovosť nielen efekty saponátov $\alpha_i, i = 1..5$ ale aj efektu bloku.

Obrázek 1: qqplot a homoskedasticita

```
> coefficients(lm1)
   (Intercept)       saponatB       saponatC       saponatD       saponatE
    27.1555556     -2.1333333      3.6666667      2.4000000     -4.2666667
 factor(blok)2  factor(blok)3  factor(blok)4  factor(blok)5  factor(blok)6
     0.7555556      3.3111111      3.4888889      1.3777778      3.1333333
 factor(blok)7  factor(blok)8  factor(blok)9 factor(blok)10
     3.2000000     -1.2444444      1.8444444      2.2444444
```

Ďalej možeme vidieť že všetky efekty saponátu sú preukázateľne nenulové, čo sa ale nedá povedať o všetkých efektoch bloku.

```
> summary(lm1)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    27.1556     0.9193  29.538 2.19e-15 ***
saponatB       -2.1333     0.8511  -2.506 0.023370 *
saponatC        3.6667     0.8511   4.308 0.000542 ***
saponatD        2.4000     0.8511   2.820 0.012329 *
saponatE       -4.2667     0.8511  -5.013 0.000128 ***
factor(blok)2   0.7556     1.1349   0.666 0.515042
factor(blok)3   3.3111     1.1349   2.918 0.010065 *
factor(blok)4   3.4889     1.1349   3.074 0.007260 **
factor(blok)5   1.3778     1.1349   1.214 0.242344
factor(blok)6   3.1333     1.1698   2.679 0.016480 *
```

```
factor(blok)7    3.2000     1.1349   2.820 0.012329 *
factor(blok)8   -1.2444     1.1349  -1.097 0.289062
factor(blok)9    1.8444     1.1698   1.577 0.134420
factor(blok)10   2.2444     1.1698   1.919 0.073045 .
```

Ako sa však presne líšia medzi sebou jednotlivé saponáty *Tukeyho metódou*:

```
> a.1<-aov(hodnota~factor(saponat)+factor(blok))
> TukeyHSD(a.1)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = hodnota ~ saponat + factor(blok), data = data)

          diff         lwr        upr
B-A -2.8333333  -5.2137573 -0.4529093
C-A  3.1666667   0.7862427  5.5470907
D-A  2.8333333   0.4529093  5.2137573
E-A -4.5000000  -6.8804240 -2.1195760
C-B  6.0000000   3.6195760  8.3804240
D-B  5.6666667   3.2862427  8.0470907
E-B -1.6666667  -4.0470907  0.7137573
D-C -0.3333333  -2.7137573  2.0470907
E-C -7.6666667 -10.0470907 -5.2862427
E-D -7.3333333  -9.7137573 -4.9529093
```

3. Náhodné efekty

Zaujíma nás či je výhodné prejsť k modelu s náhodnými efektmi pre blok, teda k modelu

$$y_{ij} = \mu + \alpha_i + U_j + \epsilon_{ij}$$

kde ϵ_{ij} je z rozdelenia $N(0, \sigma_\epsilon^2)$ a náhodná veličina U_j má rozdelenie $N(0, \sigma_U^2)$. Tento model v programe R zadáme po načítaní knižnice *nlme*:

```
lme2<-lme(hodnota~saponat,random=~1|factor(blok) )
```

Pre saponát je použitý pevný efekt, pre blok náhodný.

```
> anova.lme(lm1,lme2)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
lm1      1 15 100.7235 112.3124 -35.36176
lme2     2  7 120.9691 129.5013 -53.48456 1 vs 2 36.24561  <.0001
```

Z predchádzajúceho testu je vidno, že nemôžem zamietnuť vplyv náhodných efektov. Model $lme2$ má väčšiu hodnotu Akaikovho kritéria aj BIC, popisuje teda lepšie danú situáciu.

```
> summary(lme2)
Random effects:
 Formula: ~1 | factor(blok)
        (Intercept) Residual
StdDev:    1.375883 1.344978

Fixed effects: hodnota ~ saponat
                Value Std.Error DF  t-value p-value
(Intercept) 29.021973  0.725051 16 40.02749  <.0001
saponatB    -2.326906  0.830805 16 -2.80078  0.0128
saponatC     3.528400  0.830805 16  4.24697  0.0006
saponatD     2.519831  0.830805 16  3.03300  0.0079
saponatE    -4.331191  0.830805 16 -5.21325  0.0001
```

Vidíme, že $\sigma_\epsilon = 1.344978$ a $\sigma_U = 1.375883$. U všetkých druhov saponátu, tak ako v predchádzajúcej časti, sme zamietli hypotézu, že ich pevný efekt je nulový (p-value).

4. Výsledky, záver

Môžeme to zhrnúť, tak sa navzájom signifikantne na hladine 95% líšia všetky saponáty okrem E od B a D od C, tie sa navzájom štatisticky nelíšia.

Najlepšie popisuje výdrž saponátov model s náhodnými efektmi pre blok, kde výsledné počty umytých predmetov sú v tabuľke (Intercept):

Saponát	A	B	C	D	E
Efekty	$29,021973$	$26,695067$	$32,550373$	$31,541804$	$24,690782$

Štandardné odchylky sú

$$\sigma_\epsilon = 1.344978,\ \sigma_U = 1.375883$$

Chcel by som ešte podotknúť, že *Tukeyho metóda* porovnala aj kontrasty blokov a zaujímavé je, že hodnota bloku 8 na hladine 95% sa líši od blokov $3, 4$ a 7. Zrejme neboli všetky podmienky experimentu pri meraní bloku 8 rovnaké ako pri iných blokoch.