

BIO5312: R Session 7

Multisample Hypothesis testing

Yujin Chung

October 11th, 2016

Fall, 2016

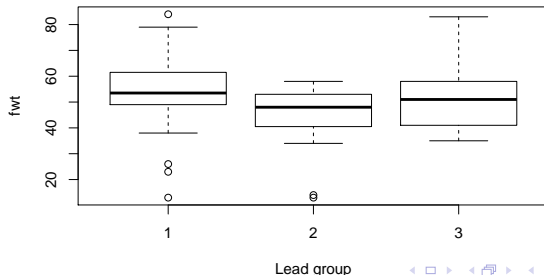
Today's R session

- ANOVA
 - ▶ `lm()`
 - ▶ `anova()`
 - ▶ `LevenTest()`
- `tapply()`

Boxplot

```
> lead = read.table("LEAD.DAT.txt",header=T)
> # remove individuals with missings 99
> ids=lead$maxfwt!=99
> fwt = lead$maxfwt[ids]
> grp = factor(lead$lead_grp[ids])

> boxplot(fwt~grp,xlab="Lead group",ylab="fwt")
```



One-way ANOVA

- `lm()`: fitting a linear model $Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$ (will revisit this function in week 10)
- `anova()`: performs a ANOVA, its argument is the output of `lm()`

```
> fit1 = lm(fwt~grp)
```

```
> anova(fit1)
```

Analysis of Variance Table

Response: fwt

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
grp	2	1600.1	800.04	5.2773	0.006692	**
Residuals	96	14553.8	151.60			

LSD: pairwise.t.test()

```
> pairwise.t.test(fwt,grp,p.adjust.method = "none")
```

Pairwise comparisons using t tests with pooled SD

data: fwt and grp

	1	2
2	0.0016	-
3	0.3955	0.0758

P value adjustment method: none

LSD

```
> xbar <- tapply(fwt, grp, mean, na.rm = TRUE) # group mean
> s <- tapply(fwt, grp, sd, na.rm = TRUE) # group s.d
> n <- tapply(!is.na(fwt), grp, sum) # group sample size
> degf <- n - 1 # d.f. of groups
> total.degf <- sum(degf) # total d.f.
> ## the pooled variance
> pooled.sd <- sqrt(sum(s^2 * degf)/total.degf)
> # for pair i and j
> i=1; j=2
> dif <- xbar[i] - xbar[j]
> se.dif <- pooled.sd * sqrt(1/n[i] + 1/n[j])
> t.val <- dif/se.dif # test statistic
> t.val
3.244684
> 2 * pt(abs(t.val), total.degf, lower.tail=F)
0.001618783
```

Linear contrasts: contrastmeans()

```
> install.packages("predictmeans")
> library(predictmeans)
> ctr = matrix(c(.5,.5,-1),nrow=1,ncol=3) # contrast matrix
> ctr
      [,1] [,2] [,3]
[1,]  0.5  0.5  -1
> contrastmeans(fit1,"grp",ctrmatrix = ctr)
$'The t tests of the specified contrasts'
      Estimate Std. Error t value df Pr(>|t|)
[1,]  -2.2813      3.473 -0.6568 96  0.5129
attr(,"Note")
[1] "The p-value is adjusted by none method, if p-value = 0 me

$K
      1  2  3
[1,] 0.5 0.5 -1
```

Levene Test: leveneTest()

```
> leveneTest(fwt,grp,center="mean")
```

```
Levene's Test for Homogeneity of Variance (center = "mean")
```

```
      Df F value Pr(>F)
```

```
group  2  0.1169 0.8898
```

```
      96
```

```
> m.fwt=fwt
```

```
> m.fwt[grp=="1"]=xbar[1]; m.fwt[grp=="2"]=xbar[2]
```

```
> m.fwt[grp=="3"]=xbar[3]
```

```
> z = abs(fwt-m.fwt)
```

```
> anova(lm(z~grp))
```

```
Analysis of Variance Table
```

```
Response: z
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
```

```
grp      2    15.9    7.938  0.1169 0.8898
```

```
Residuals 96 6521.1  67.928
```


Two-way ANOVA

```
> fit2 = lm(fwt~grp+sex)
```

```
> anova(fit2)
```

Analysis of Variance Table

Response: fwt

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
grp	2	1600.1	800.04	5.2348	0.006971	**
sex	1	34.8	34.79	0.2277	0.634354	
Residuals	95	14519.0	152.83			

Interaction effect

```
> fit3 = lm(fwt~grp*sex)
```

```
> anova(fit3)
```

Analysis of Variance Table

Response: fwt

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
grp	2	1600.1	800.04	5.3545	0.006295	**
sex	1	34.8	34.79	0.2329	0.630535	
grp:sex	2	623.3	311.67	2.0860	0.129960	
Residuals	93	13895.6	149.42			