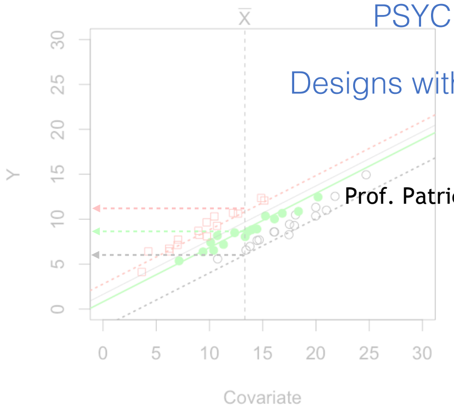
PSYCH 710

Designs with covariates

Prof. Patrick Bennett



Experimental Design: Blocking

- A study is conducted to measure effect of drug on locomotor activity in hyperactive children
- Between-subjects design:
  - 3 groups differ in drug dosage: zero, low, & high
- Dependent variable: locomotor activity
  - measured for fixed interval after drug administration
- Before study, measure baseline locomotor activity in each subject
  - baseline measure used as a **blocking variable**

Randomized Block Design

- 4 blocks of 12 Ss created using baseline locomotor activity activity measure
- subjects in each block assigned randomly to drug dose condition

		Drug Dose		
		zero	low	high
Block (Baseline Locomotor Activity)	low	4	4	4
	medium	4	4	4
	high	4	4	4
	very high	4	4	4

Do blocks differ from each other?

```
> with(theData, tapply(activity, block,mean));

      low       med      high      very
 5.153699  8.007714 10.287328 13.818771

> summary(aov(activity ~ block, data=theData) );

            Df Sum Sq Mean Sq F value Pr(>F)
block        3  483.1  161.02   106.3 <2e-16 ***
Residuals   44   66.6    1.51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect of blocking factor on $SS_{residuals}$

```
> aov.1<-aov(y~drug,data=theData)
> aov.2<-aov(y~block+drug+block:drug,data=theData)
> summary(aov.1)

            Df Sum Sq Mean Sq F value Pr(>F)
drug         2   37.1  18.539   2.173  0.126
Residuals   45  384.0   8.532
```

SS_{total} = 421

```
> summary(aov.2)

            Df Sum Sq Mean Sq F value   Pr(>F)
block        3 129.84   43.28   6.964 0.000818 ***
drug         2  37.08   18.54   2.983 0.063294 .
block:drug   6  30.37    5.06   0.815 0.565798
Residuals   36 223.75    6.22
```

SS_{total} = 421

Analysis of Covariance (ANCOVA)

- blocking allows variation in dependent variable (Y) that is associated with blocking variable to be removed from residuals
- in hyperactivity example, Y was linearly related to baseline activity
  - but blocking variable was a qualitative factor
  - did not fully take advantage of quantitative relation between Y and baseline locomotor activity
- ANCOVA quantitatively models association between dependent variable and **covariate** (baseline activity) using each subject's activity measure rather than dividing subjects into 4 factor groups

ANCOVA

```
> lm.1<-lm(y~activity+drug,data=theData)
> lm.2<-lm(y~activity,data=theData)
> anova(lm.2,lm.1)
 Analysis of Variance Table

 Model 1: y ~ activity
 Model 2: y ~ activity + drug
   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
 1     46 278.92
 2     44 239.64  2    39.286 3.6066 0.03544 *
```

Difference between models shows SS_{drug} after controlling for the linear association between Y and baseline activity

```
> anova(lm.1)

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
activity   1 142.116 142.116 26.0939 6.739e-06 ***
drug       2  39.286  19.643  3.6066   0.03544 *
Residuals 44 239.638   5.446
```

Same result obtained with sequential sums-of-squares ANOVA table for full model

Order of terms does matter

but in this case we probably should put covariate first (why?)

```
> lm.1<-lm(y~activity+drug,data=theData)
> anova(lm.1)

Analysis of Variance Table
          Df  Sum Sq Mean Sq F value    Pr(>F)
activity   1 142.116 142.116 26.0939 6.739e-06 ***
drug       2  39.286  19.643  3.6066   0.03544 *
Residuals 44 239.638   5.446
```

```
> library(car)
> Anova(lm.1,type=2)

Anova Table (Type II tests)
           Sum Sq Df F value   Pr(>F)
activity  144.324  1 26.4993 5.91e-06 ***
drug       39.286  2  3.6066  0.03544 *
Residuals 239.638 44
```

```
> lm.1b<-lm(y~drug+activity,data=theData)
> anova(lm.1b)

Analysis of Variance Table

          Df  Sum Sq Mean Sq F value   Pr(>F)
drug       2  37.078  18.539  3.4039  0.04222 *
activity   1 144.324 144.324 26.4993 5.91e-06 ***
Residuals 44 239.638   5.446
```

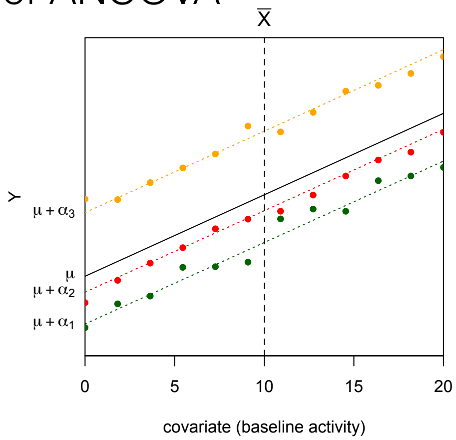
```
> Anova(lm.1b,type=2)

Anova Table (Type II tests)

           Sum Sq Df F value   Pr(>F)
drug       39.286  2  3.6066  0.03544 *
activity  144.324  1 26.4993 5.91e-06 ***
Residuals 239.638 44
```

Graphical illustration of ANCOVA

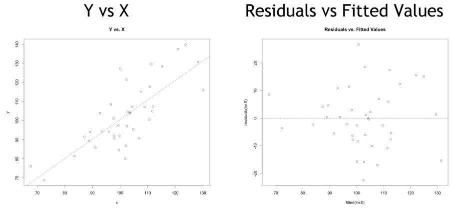
- ANCOVA computes regression lines for each group
  - equal slopes
  - variable intercepts
- group effect (α_j) corresponds to vertical shift of regression intercept



checking linear regression

Y vs X

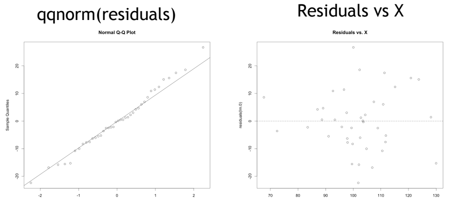
Residuals vs Fitted Values



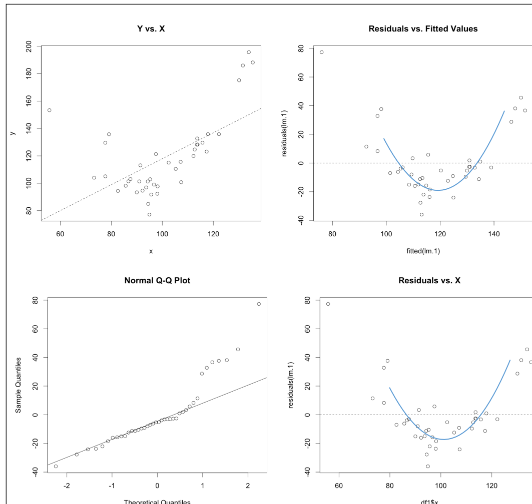
```
par(cex=1.5)
plot(residuals(lm.0)~fitted(lm.0),
    main="Residualsvs.Fitted Values")
abline(h=0,lty=2)
```

qqnorm(residuals)

Residuals vs X

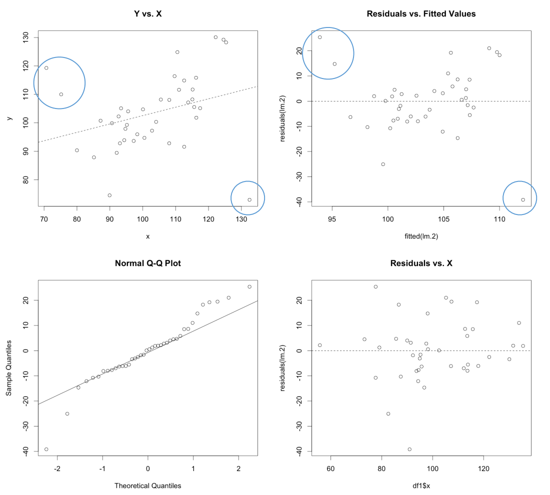


```
plot(residuals(lm.0)~df0$x,
main="Residuals vs. X")
abline(h=0,lty=2)
```




```
par(cex=1.5)
plot(residuals(lm.0)~fitted(lm.0),
+    main=“Residuals vs.Fitted Values")
abline(h=0,lty=2)
```

```
plot(residuals(lm.0)~df0$x,
+    main="Residuals vs. X")
abline(h=0,lty=2)
```

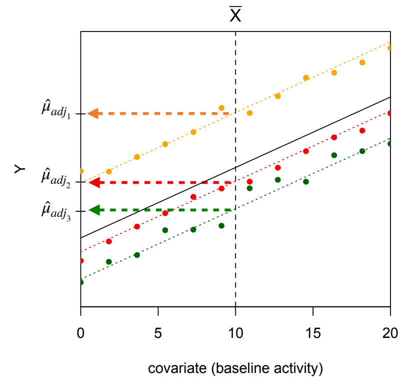



```
par(cex=1.5)
plot(residuals(lm.0)~fitted(lm.0),
+    main="Residualsvs.Fitted Values")
abline(h=0,lty=2)
```

```
plot(residuals(lm.0)~df0$x,
+    main="Residuals vs. X")
abline(h=0,lty=2)
```

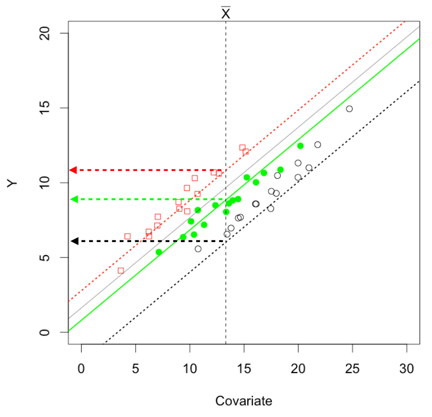
Adjusted Group Means

- point where regression line crosses grand mean of covariate
- provides estimates of group differences after removing effects of covariate



Adjusted Group Means

- point where regression line crosses grand mean of covariate
- provides estimates of group differences after removing effects of covariate
- most useful when groups differ considerably in terms of mean covariate



Computing Adjusted Means

```
> lm.1<-lm(y~activity+drug,data=theData)
> dummy.coef(lm.1);

Full coefficients are

(Intercept):       1.28519
activity:        0.5124572
drug:                  zero        low       high
             -0.8216185 -0.4386966  1.2603150
```

dummy coefficients list the parameters for the lines fit to each group

```
> library(effects)
> effect(term="drug",lm.1)

 drug effect
drug
     zero      low     high
5.238072 5.620994 7.320006
```

$$\bar{Y}_j' = \mu + \beta\bar{X} + \alpha_j$$

covariate mean = 9.3168, so
adjusted group means are:

```
> 1.28519 + 0.5124*9.3168-0.8216
[1] 5.237518
```
zero

```
> 1.28519 + 0.5124*9.3168-0.4386
[1] 5.620518
```
low

```
> 1.28519 + 0.5124*9.3168+1.2603
[1] 7.319418
```
high

```
> emmeans(lm.1,specs=~drug)
 drug emmean    SE df lower.CL upper.CL
 zero   5.24 0.583 44     4.06     6.41
 low    5.62 0.583 44     4.45     6.80
 high   7.32 0.583 44     6.14     8.50
```

linear contrasts on adjusted means

$$F = \frac{\hat{\psi}^2}{s^2_{\hat{\psi}}}$$

$$\boxed{\hat{\psi} = \sum_{j=1}^{a} c_j \bar{Y}'_j}$$

$$s^2_{\hat{\psi}} = MS_{residuals} \left[\sum_{j=1}^{a} \frac{c_j^2}{n_j} + \frac{\left(\sum_{j=1}^{a} c_j \bar{X}_j\right)^2}{\sum_{j=1}^{a}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2} \right]$$

```
> lm.1<-lm(y~activity+drug,data=theData)
> levels(theData$drug)
[1] "zero" "low"  "high"
> w1 <- c(-1,1,0)
> w2 <- c(-1,0,1)
> w3 <- c(0,-1,1)
> lm1.emm <- emmeans(lm.1,specs=~drug)
> contrast(lm1.emm,method=list(ZvsL=w1,ZvsH=w2,LvsH=w3))
 contrast estimate    SE df t.ratio p.value
 ZvsL        0.383 0.825 44 0.464   0.6449
 ZvsH        2.082 0.825 44 2.523   0.0153
 LvsH        1.699 0.825 44 2.059   0.0454
```

Association strength

(ignores covariate)

$$\omega^2 = \frac{df_{effect}(MS_{effect} - MS_{residuals})}{SS_{total} + MS_{residuals}}$$

(removes variation due to covariate)

$$\omega^2_{partial(group)} = \frac{df_{group}(F_{group} - 1)}{df_{group}(F_{group} - 1) + N}$$

```
> require(effectsize)
> mw97.aov.01 <- aov(post~pre+treatment,data=mw97)
> omega_squared(mw97.aov.01,partial=F)
Parameter | Omega2 |       90% CI
---------------------------------
pre       |   0.23 | [0.04, 0.44]
treatment |   0.12 | [0.00, 0.30]

> omega_squared(mw97.aov.01,partial=T)
Parameter | Omega2 (partial) |       90% CI
-------------------------------------------
pre       |             0.26 | [0.05, 0.47]
treatment |             0.15 | [0.00, 0.34]
```

Homogeneity of slopes assumption

- ANCOVA assumes that slope of regression line is the same in each group
  - implies that there is no covariate x group interaction
- if valid, then group differences are independent of covariate
- if not valid, then group differences vary with covariate

```
> anova(lm.4)

Analysis of Variance Table

Response: y
                Df  Sum Sq Mean Sq F value    Pr(>F)
activity.c       1 142.116 142.116 26.4881 6.604e-06
drug             2  39.286  19.643  3.6611   0.03422
activity.c:drug  2  14.298   7.149  1.3324   0.27476
Residuals       42 225.340   5.365
```

```
> lm.4 <- lm(y~activity.c + drug + activity.c:drug,data=theData)
```

Why not difference scores?

```
> theData$diff <- theData$y - theData$activity;
> diff.lm.1 <- lm(diff~drug,data=theData)
> anova(diff.lm.1);

Analysis of Variance Table

Response: diff
          Df Sum Sq Mean Sq F value  Pr(>F)
drug       2  41.46 20.7318  2.5196 0.09179 .
Residuals 45 370.27  8.2282
```

ANOVA on difference scores (activity - baseline.activity) finds drug is not significant. Why?

Why not difference scores?

$$Y_{ij} - X_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

$$Y_{ij} = \mu + \alpha_j + X_{ij} + \epsilon_{ij}$$

- ANOVA on difference scores is equivalent to ANCOVA with slope fixed at 1.0
- When slope ≠ 1, then ANCOVA provides better fit (lower MS-residuals)