Checking assumptions

560 Hierarchical modeling

Peter Hoff

Statistics, University of Washington

Assumptions of the HNM

$$
\begin{aligned}
y_{i,j} &= \mu_j + \epsilon_{i,j} \\
\{\epsilon_{i,j}\} &\sim \text{iid } N(0, \sigma^2) \qquad (1) \\
\mu_1, \ldots, \mu_m &\sim \text{iid } N(\mu, \tau^2) \qquad (2)
\end{aligned}
$$

Assumptions concerning within-group variation: Item (1) implies

- the $\epsilon_{i,j}$'s are independent;
- the $\epsilon_{i,j}$'s have the same variance in each group;
- the $\epsilon_{i,j}$'s are normally distributed.

Assumptions concerning between-group variation: Item (2) implies

- the μ_j's are independent;
- the μ_j's are normally distributed.

Hierarchy of micro-level assumptions

Some assumptions are more important than others. Statistical folklore (and theoretical results) suggest the order of importance of the assumptions is

independence: the $\epsilon_{i,j}$'s are independent;

constant variance: the $\epsilon_{i,j}$'s have the same variance in each group;

normality: the $\epsilon_{i,j}$'s are normally distributed.

Cautions: Ignoring violations can lead to invalid inference

dependence: can lead to inaccurate p-values and confidence intervals;

nonconstant variance: can affect type I error rates and estimation efficiency;

nonnormality: our procedures are somewhat robust to nonnormality (CLT).

Checking micro-level assumptions with residuals

We don't observe the $\epsilon_{i,j}$'s, so we can't check these assumptions directly. Standard practice is to evaluate the residuals:

$$y_{i,j} = \mu_j + \epsilon_{i,j}$$
$$\epsilon_{i,j} = y_{i,j} - \mu_j$$

If $\hat{\mu}_j \approx \mu_j$, then

$$\epsilon_{i,j} = y_{i,j} - \mu_j \approx y_{i,j} - \hat{\mu}_j = \hat{\epsilon}_{i,j}$$

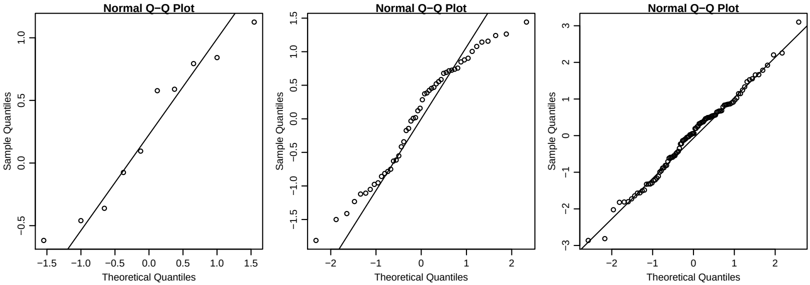
Here, $\hat{\mu}_j$ could be either $\bar{y}_j$ or the shrinkage estimator.

Standard practice is to use $\bar{y}_j$.

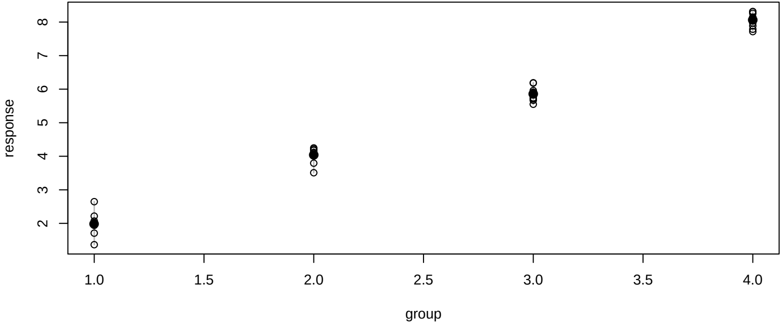
Checking normality

Q-Q plots: A useful visual tool for checking normality is the normal scores plot. This plots the sample quantiles versus those of the normal distribution.

```
y10<-rnorm(10) ; y50<-rnorm(50) ; y100<-rnorm(100)

qqnorm(y10) ; qqline(y10)
qqnorm(y50) ; qqline(y50)
qqnorm(y100) ; qqline(y100)
```

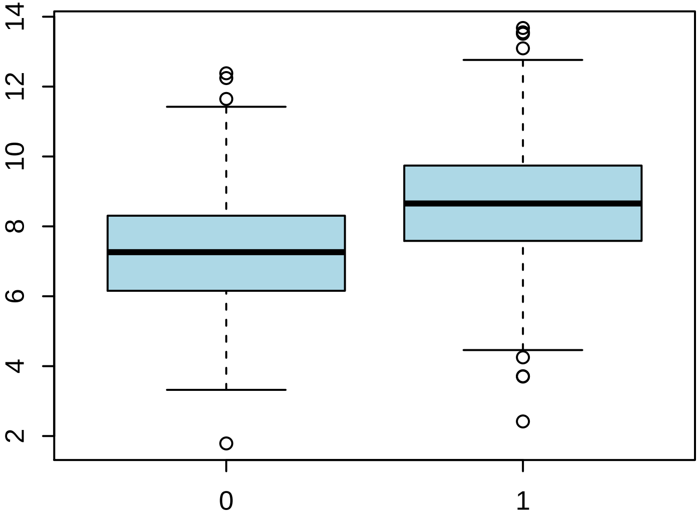


The wrong way to check normality



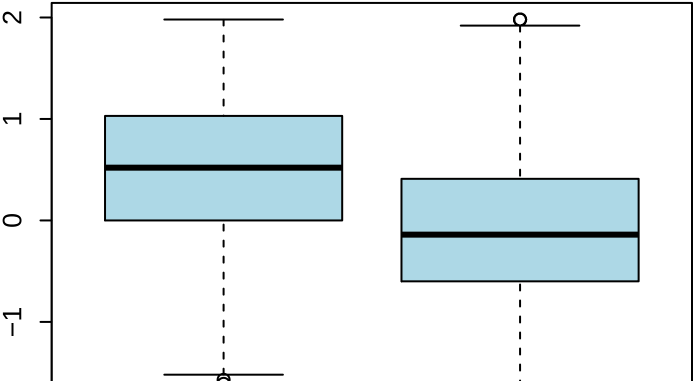
The wrong way to check normality

```
par(mfrow=c(1,2))
hist(y)
qqnorm(y) ; qqline(y)
```

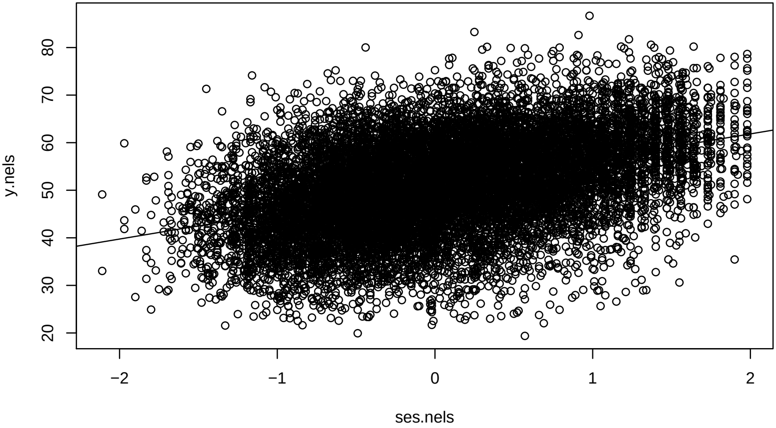


The right way to check normality

```
par(mfrow=c(1,2))

fit<-lm(y~as.factor(g))

res<-fit$res

hist(res)

qqnorm(res) ; qqline(res)
```

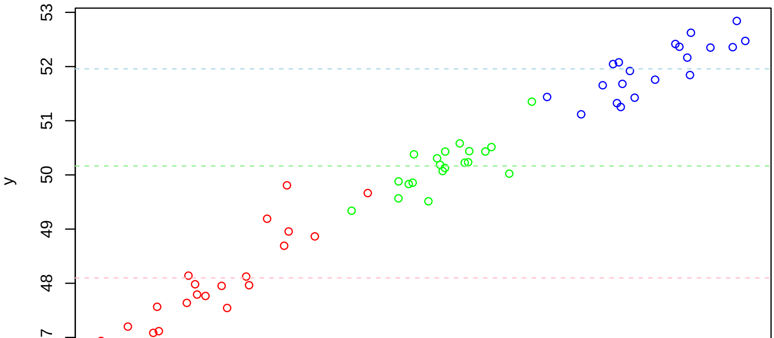


Example: Wheat yield

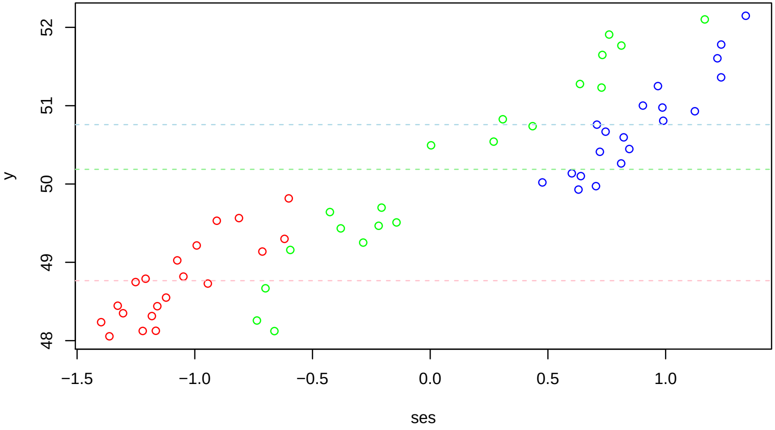


Example: Wheat yield

```
par(mfrow=c(1,2))

fit.wheat<-lm(y.wheat~as.factor(g.wheat))

res<-fit.wheat$res

hist(res)

qqnorm(res) ; qqline(res)
```

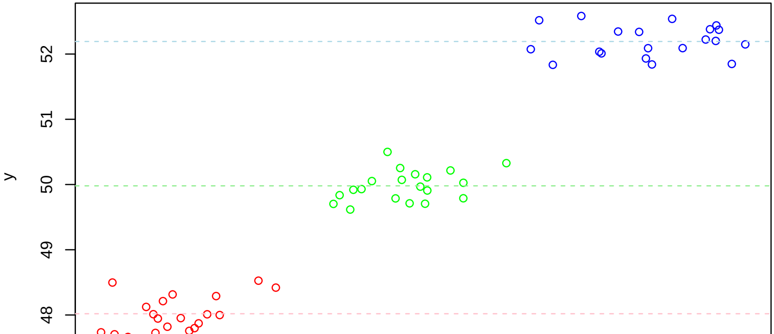


Example: Nels data



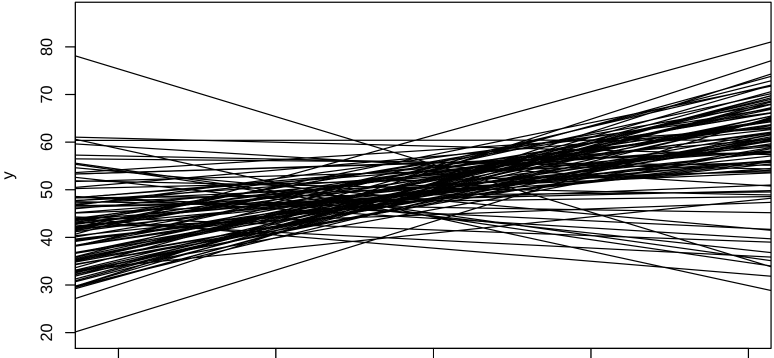
Example: NELS data

```
par(mfrow=c(1,2))

fit.nels<-lm(y.nels~as.factor(g.nels))

res<-fit.nels$res

hist(res)

qqnorm(res) ; qqline(res)
```



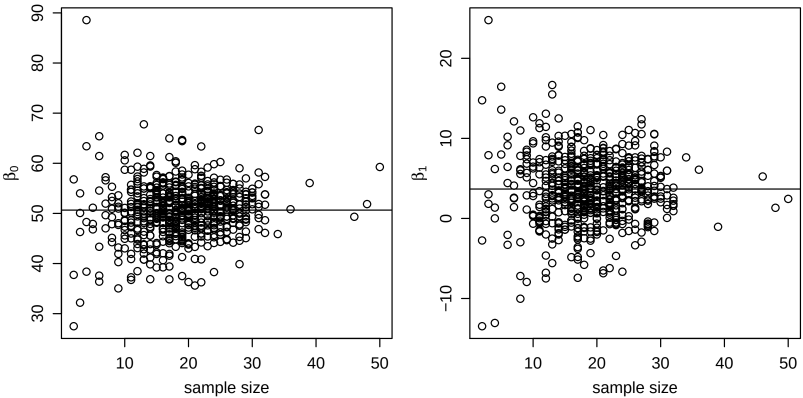
Example: Grouse ticks

```
grouseticks[1:5,]

##   INDEX TICKS BROOD HEIGHT YEAR LOCATION   cHEIGHT
## 1     1     0   501    465   95       32  2.759305
## 2     2     0   501    465   95       32  2.759305
## 3     3     0   502    472   95       36  9.759305
## 4     4     0   503    475   95       37 12.759305
## 5     5     0   503    475   95       37 12.759305
```



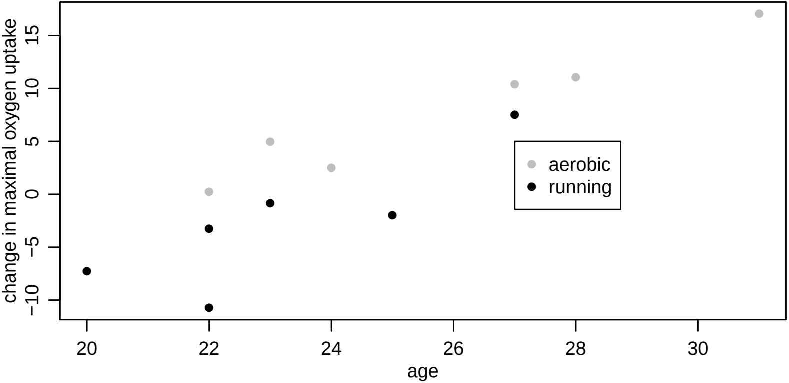
Example: Grouse ticks

```
par(mfrow=c(1,2))

fit.grouse<-lm(y.grouse~as.factor(g.grouse))

res<-fit.grouse$res

hist(res)

qqnorm(res) ; qqline(res)
```

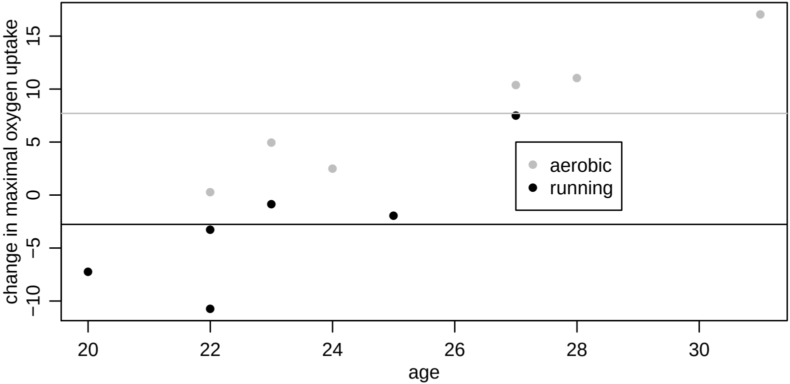


Example: Grouse ticks normality evaluation, the wrong way

```
par(mfrow=c(1,2))

hist(y.grouse)

qqnorm(y.grouse) ; qqline(y.grouse)
```



What should my residuals look like?



Within-group variance

$$\{\epsilon_{i,j}\} \sim N(0, \sigma^2)$$

This implies that not only are the errors normal, but their *variance is the same for all groups*.

How might we evaluate this assumption?

Idea: Suppose $\epsilon_{1,j}, \ldots, \epsilon_{n_j,j} \sim \text{iid } N(0, \sigma_j^2)$

- $s_j^2 \approx \sigma_j^2$
- differences between σ_j^2's can be evaluated by differences between s_j^2's.

Example: wheat yield

```
s2.wheat<-c(tapply(y.wheat,g.wheat,var))

s2.wheat

##       1       2       3       4       5       6       7       8       9
## 4.49173 0.43388 2.88970 0.99197 1.94843 0.95908 0.67748 0.86467 1.96792
##      10
## 2.64720

max(s2.wheat)/min(s2.wheat)

## [1] 10.35247
```

Is the heterogeneity large? Remember $n_j = 5$ for all groups.

Fmax test: A test of equality of variances - reject $H_0 : \sigma_j^2 = \sigma^2$ if

$$s^2_{max}/s^2_{min} > Fmax_{1-\alpha,m,n}$$

The critical value must be looked up on a table.
It is *not* the same as the usual F-distribution.

Levene's test

Idea: If σ_j^2 is large, then $|y_{i,j} - \bar{y}_j| = |\hat{\epsilon}_{i,j}|$ should be large.

- Let $z_{i,j} = |\hat{\epsilon}_{i,j}|$
- Use the ANOVA F-test for across-group differences *in the* $z_{i,j}$*'s*

```
z.wheat<-abs( fit.wheat$res )
anova(lm(z.wheat~as.factor(g.wheat)) )

## Analysis of Variance Table
##
## Response: z.wheat
##                     Df  Sum Sq Mean Sq F value Pr(>F)
## as.factor(g.wheat)  9  4.8893 0.54325  1.0389 0.4273
## Residuals          40 20.9174 0.52294
```

Example: NELS data

```
s2.nels<-c(tapply(y.nels,g.nels,var))

max(s2.nels,na.rm=TRUE)

## [1] 187.082

min(s2.nels,na.rm=TRUE)

## [1] 3.20045

n.nels<-table(g.nels)

n.nels[ which.max(s2.nels)]

## 320
##  19

n.nels[ which.min(s2.nels)]

## 643
##   2
```

Example: NELS data

```
z.nels<-abs( fit.nels$res )
anova(lm(z.nels~as.factor(g.nels)) )

## Analysis of Variance Table
##
## Response: z.nels
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(g.nels)   683  27078  39.645  1.6092 < 2.2e-16 ***
## Residuals         12290 302776  24.636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Grouse tick data

```
s2.grouse<-c(tapply(y.grouse,g.grouse,var))

max(s2.grouse,na.rm=TRUE)

## [1] 346.3

min(s2.grouse,na.rm=TRUE)

## [1] 0

n.grouse<-table(g.grouse)

n.grouse[ which.max(s2.grouse)]

## 626
##   5

n.grouse[ which.min(s2.grouse)]

## 501
##   2
```

Example: Grouse tick data

```
z.grouse<-abs( fit.grouse$res )
anova(lm(z.grouse~as.factor(g.grouse)) )

## Analysis of Variance Table
##
## Response: z.grouse
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(g.grouse) 117 3954.0  33.795  4.8627 < 2.2e-16 ***
## Residuals           285 1980.7   6.950
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remedies

NELS data:

- The evidence suggests the residual variance is not equal across schools.
- It seems plausible that some schools are more heterogeneous than others due to observable factors (SES, for example)
- Coming up, we will evaluate heteroscedasticity after including additional micro-level information. This generally reduces the across-group heteroscedasticity.

Grouse tick data:

- The data are clearly nonnormal, and have nonconstant variance.
- One approach is to transform the data, which in some cases can remedy both problems.
- Alternatively, we can fit models that explicitly allow for the count-valued nature of the data (GLMEs).

Why are data normal?

Additiive effects: Often, an outcome is the result of many *additive* effects:

$$\begin{aligned} y_{i,j} &= \mu_j + \epsilon_{i,j} \\ &= \mu_j + x_{i,j,1} + x_{i,j,2} + \cdots + x_{i,j,p} \end{aligned}$$

CLT:
In such cases, if the $x_{i,j,k}$'s vary somewhat independently across subjects,the distribution of the $y_{i,j}$'s should look normal (even if the $x_{i,j,k}$'s are not normal).

Multiplicative effects

Multiplicative effects: Some outcomes are the result of *multiplicative* effects:

$$y_{i,j} = \mu_j \times x_{i,j,1} \times x_{i,j,2} \times \cdots \times x_{i,j,p}$$

eg., the outcome when $x_{i,j,1} = 2$ is twice that when $x_{i,j,1} = 1$.

Mean-variance relationship: Let $\epsilon_{i,j} = x_{i,j,1} \times \cdots \times x_{i,j,p}$. Then

$$\begin{aligned} y_{i,j} &= \mu_j \times \epsilon_{i,j} \\ \mathrm{Var}[y_{i,j}|\mu_j] &= \mathrm{Var}[\mu_j \times \epsilon_{i,j}|\mu_j] \\ &= \mu_j^2 \times \mathrm{Var}[\epsilon_{i,j}|\mu_j] \end{aligned}$$

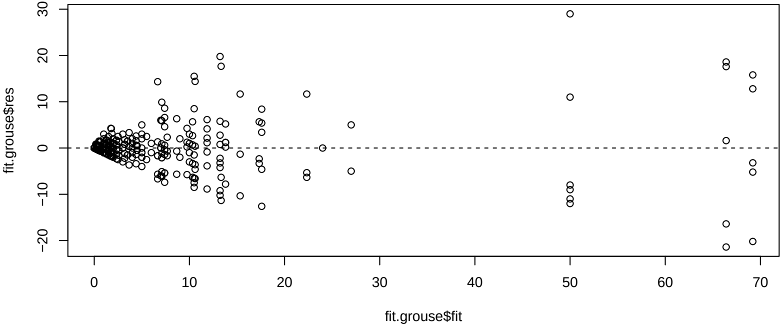
If there are multiplicative effects, we *expect* heteroscedasticity:

- groups with large means will have large variances;
- groups with small means will have small variances.

Mean-variance relationships

A mean-variance relationship can be evaluated with a *fitted versus residual* plot.

```
plot( fit.grouse$fit, fit.grouse$res)
abline(h=0,lty=2)
```



Variance stabilizing transformations

Log transformation: Suppose the multiplicative model is correct.

$$\begin{aligned}\tilde{y}_{i,j} = \log y_{i,j} &= \log(\mu_j \times x_{i,j,1} \times x_{i,j,2} \times \cdots \times x_{i,j,p}) \\ &= \log \mu_j + \log x_{i,j,1} + \log x_{i,j,2} + \cdots + \log x_{i,j,p} \\ &= \tilde{\mu}_j + \tilde{x}_{i,j,1} + \tilde{x}_{i,j,2} + \cdots + \tilde{x}_{i,j,p}\end{aligned}$$

If the variances of the $\tilde{x}_{i,j,k}$'s is constant across groups, then

- the variance of the $\tilde{y}_{i,j}$'s should be constant across groups;
- the distribution of the $\tilde{y}_{i,j}$'s should be approximately normal, within groups.

Power transformations

In many cases, the effects are neither strictly additive or multiplicative.

In such cases, we might hope that there is some value p for which

$$\tilde{y}_{i,j} = y_{i,j}^{p} = \mu_j + \epsilon_{i,j}$$

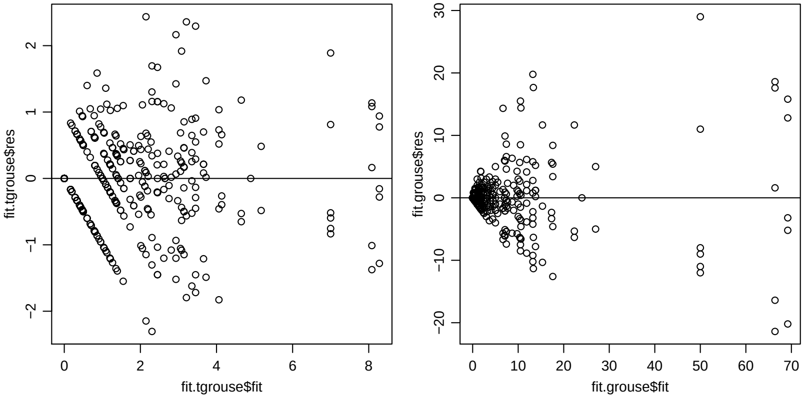
holds approximately.

Common power transformations:

p	name
1	no transformation
1/2	square-root transformation
1/4	quarter-power transformation
0	log transformation (in a limiting sense)

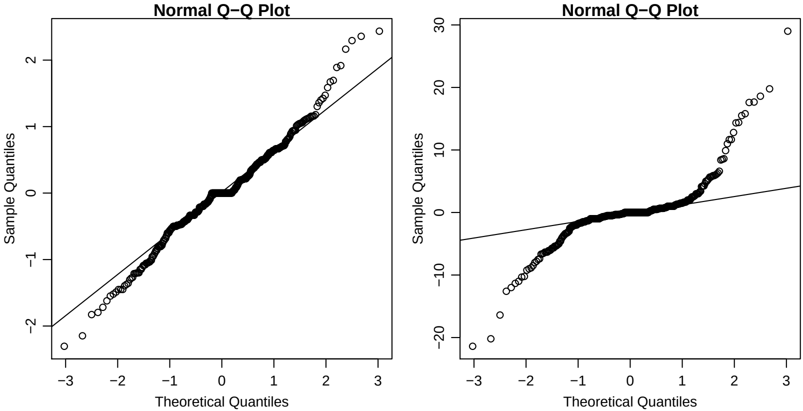
Example: Tick data

```
ty.grouse<-sqrt(y.grouse)
fit.tgrouse<-lm(ty.grouse~as.factor(g.grouse))

mpar()
par(mfrow=c(1,2))

plot(fit.tgrouse$fit, fit.tgrouse$res) ; abline(h=0)
plot(fit.grouse$fit, fit.grouse$res) ; abline(h=0)
```



What about normality?

```
mpar()
par(mfrow=c(1,2))

qqnorm(fit.tgrouse$res) ; qqline(fit.tgrouse$res)
qqnorm(fit.grouse$res) ; qqline(fit.grouse$res)
```



Recommendations

Power transformations: Pros

If your data are non-normal and exhibit a mean variance relation, a transformation can

- stabilize the variance across groups;
- make the transformed data more normally distributed (within groups).

Power transformations: Cons

A power transformation

- changes the scale on which your parameters are estimated;
- makes results possibly more difficult to interpret;
- might be less preferable than using a different model (GLME vs LME).

Macro-level assumptions

$$y_{i,j} = \mu_j + \epsilon_{i,j}$$
$$\{\epsilon_{i,j}\} \sim \text{iid } N(0, \sigma^2)$$
$$\mu_1, \ldots, \mu_m \sim \text{iid } N(\mu, \tau^2)$$

Assumptions concerning between-group variation:

- the μ_j's are independent;
- the μ_j's are normally distributed.

There is no heteroscedasticity to check.

Only normality and independence need to be considered.

Checking the macro level distribution

$$\mu_1, \ldots, \mu_m \sim \text{ iid } N(\mu, \sigma^2)$$

Evaluation via group sample means:
Assumptions about μ_j's can be assesed via the the $\bar{y}_j$'s.

$$\begin{aligned}\bar{y}_j &= \tfrac{1}{n} \sum_i (\mu_j + \epsilon_{i,j}) \\ &= \mu_j + \tfrac{1}{n_j} \sum \epsilon_{i,j} \\ &= \mu + \bar{\epsilon}_j\end{aligned}$$

Distribution of group sample means

Assume for the moment that the sample sizes are constant.

Expectation of $\bar{y}_j$: Under the assumptions,

$$\begin{aligned} \mathsf{E}[\bar{y}_j] &= \mathsf{E}[\mu_j + \bar{\epsilon}_j] \\ &= \mathsf{E}[\mu_j] + \mathsf{E}[\bar{\epsilon}_j] \\ &= \mu \end{aligned}$$

Variance of $\bar{y}_j$: Under the assumptions,

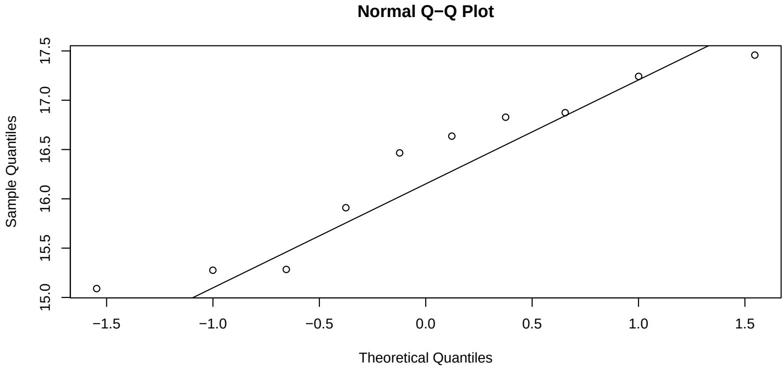
$$\begin{aligned} \mathsf{Var}[\bar{y}_j] &= \mathsf{Var}[\mu_j + \bar{\epsilon}_j] \\ &= \mathsf{Var}[\mu_j] + \mathsf{Var}[\bar{\epsilon}_j] \\ &= \tau^2 + \sigma^2/n \end{aligned}$$

Distribution of $\bar{y}_j$: If $\epsilon_{i,j}$'s are iid normal and, independently, μ_j's are iid normal, then

$$\bar{y}_1, \ldots, \bar{y}_m \sim \text{ iid } N(\mu, \tau^2 + \sigma^2/n)$$

Example: Wheat yield

```
ybar.wheat<-c(tapply(y.wheat,g.wheat,mean))

qqnorm(ybar.wheat) ; qqline(ybar.wheat)
```



No cause for alarm.

Unequal sample sizes

$$\text{Var}[\bar{y}_j] = \tau^2 + \sigma^2/n_j$$

If sample sizes are unequal, then

- $\bar{y}_1, \ldots, \bar{y}_m$'s are *not identically distributed*.
- the variance of $\bar{y}_j$ depends on its sample size.

The distribution of $\bar{y}_1, \ldots, \bar{y}_m$ will be a *scale mixture of normals*.

In practice

- If σ^2/n_j is small compared to τ^2, $\{\bar{y}_1, \ldots, \bar{y}_m\}$ should look normal.
- If σ^2/n_j is large compared to τ^2, $\{\bar{y}_1, \ldots, \bar{y}_m\}$ might not look normal, *even if the assumptions are correct*.

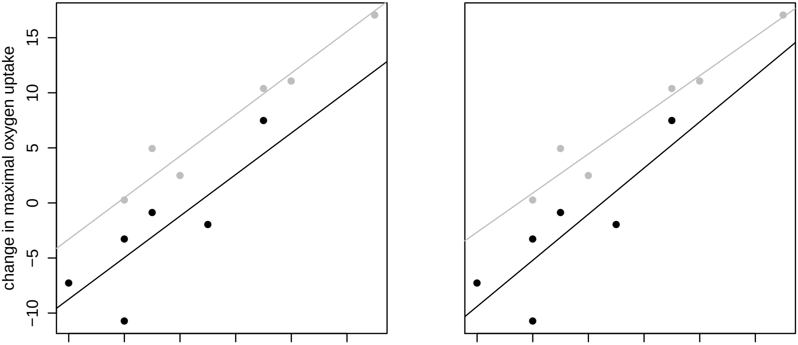
A fabricated example

```
t2<-1 ; s2<-8 ; mu<-60

m<-200

mu.group<-rnorm(m,mu,sqrt(t2))

n.sim<-y.sim<-g.sim<-NULL
for(j in 1:m)
{
  n.j<-round(1+49*rbeta(1,.1,.1))
  y.j<-rnorm(n.j,mu.group[j],sqrt(s2))

  y.sim<-c(y.sim,y.j)
  g.sim<-c(g.sim,rep(j,n.j))
  n.sim<-c(n.sim,n.j)
}
```

```
table(n.sim)

## n.sim
##  1  2  3  4  5  6  7  8  9 10 11 13 17 19 20 21 22 23 26 27 31 32 34 35 36
## 74  5  6  7  1  1  2  1  3  1  1  1  1  1  2  1  1  1  1  3  1  2  1  1  2
## 37 41 42 43 44 45 46 47 48 49 50
##  1  1  3  3  1  2  1  3  4  5 55
```

A fabricated example

```
ybar.sim<-c(tapply(y.sim,g.sim,mean))

mpar()
par(mfrow=c(1,4))
qqnorm(ybar.sim); qqline(ybar.sim)

z<-rnorm(length(ybar.sim)) ; qqnorm(z); qqline(z)

z<-rnorm(length(ybar.sim)) ; qqnorm(z); qqline(z)

z<-rnorm(length(ybar.sim)) ; qqnorm(z); qqline(z)
```



Standardized effects

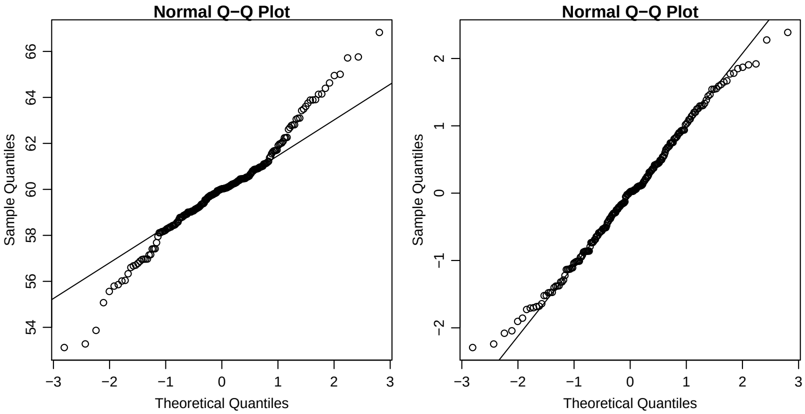
If we knew μ, σ^2, τ^2, we could standardize the $\bar{y}_j$'s appropriately:

$$\frac{\bar{y}_j - \mu}{\sqrt{\tau^2 + \sigma^2/n_j}} \sim N(0,1)$$

```
zbar.sim<- (ybar.sim -mu)/sqrt( t2+ s2/n.sim)
```

Standardized effects

```
mpar()
par(mfrow=c(1,2))
qqnorm(ybar.sim); qqline(ybar.sim)
qqnorm(zbar.sim); qqline(zbar.sim)
```



Standardized effects

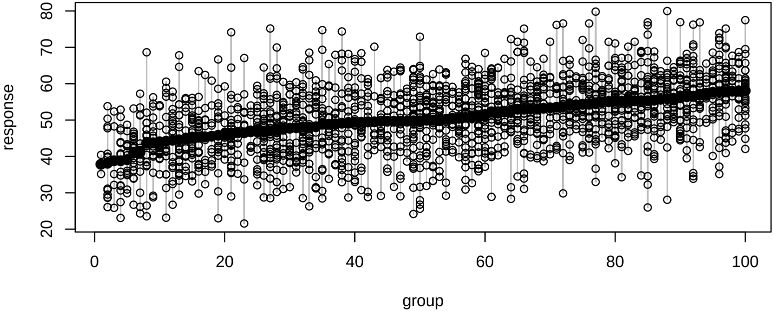
An ad-hoc approach is to replace μ, σ^2, τ^2 with their estimates:

$$\frac{\bar{y}_j - \hat{\mu}}{\sqrt{\hat{\tau}^2 + \hat{\sigma}^2/n_j}} \dot{\sim} N(0,1)$$

```
## fit mixed effects model and extract coefficients
fit.lme<-lmer(y.sim~1+(1|g.sim))
mu.mle<-fixef(fit.lme)
s2.mle<- sigma(fit.lme)^2
t2.mle <- as.numeric(VarCorr(fit.lme)$g)

## compuate standardized group means
zbar.sim<- (ybar.sim -mu.mle)/sqrt( t2.mle+ s2.mle/n.sim)
```

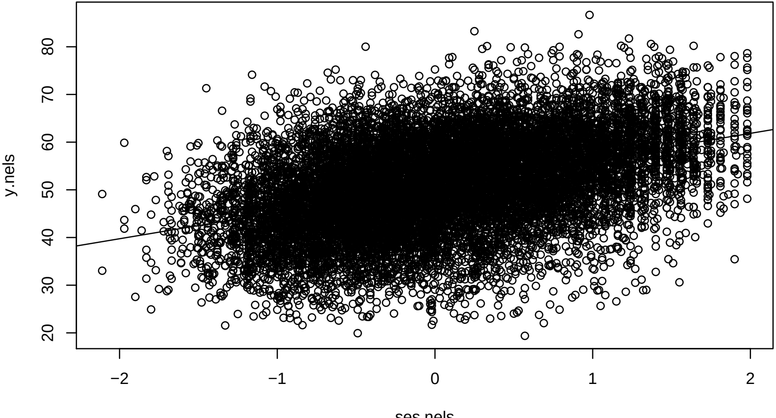
Standardized effects

```
mpar()
par(mfrow=c(1,2))
qqnorm(ybar.sim); qqline(ybar.sim)
qqnorm(zbar.sim); qqline(zbar.sim)
```



Example: NELS data

```
ybar.nels<-c(tapply(y.nels,g.nels,mean))
mpar() par(mfrow=c(1,4)) qqnorm(ybar.nels) ; qqline(ybar.nels)
z<-rnorm(length(ybar.nels)) ; qqnorm(z); qqline(z)
z<-rnorm(length(ybar.nels)) ; qqnorm(z); qqline(z)
z<-rnorm(length(ybar.nels)) ; qqnorm(z); qqline(z)
```

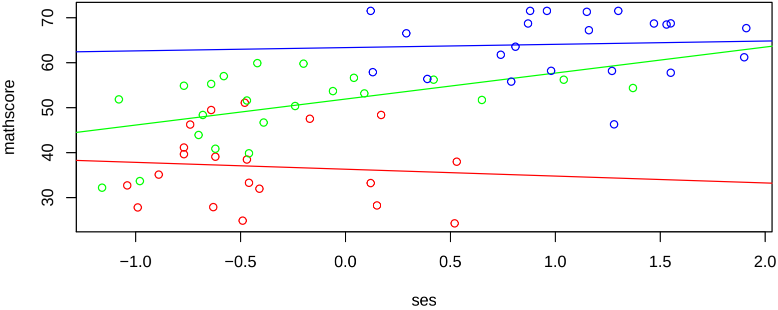


Standardized effects

```
## fit mixed effects model and extract coefficients
fit.lme<-lmer(y.nels~1+(1|g.nels))
mu.mle<-fixef(fit.lme)
s2.mle<- sigma(fit.lme)^2
t2.mle <- as.numeric(VarCorr(fit.lme)$g)

## compuate standardized group means
zbar.nels<- (ybar.nels -mu.mle)/sqrt( t2.mle+ s2.mle/n.nels)
```

Standardized effects

```
## compare qqplots
mpar()
par(mfrow=c(1,2))
qqnorm(ybar.nels); qqline(ybar.nels)
qqnorm(zbar.nels); qqline(zbar.nels)
```



Comments

QQplots of sample means should be sufficient:
It is hard to imagine erroneously rejecting normality because of a sample size difference.

Nonnormality may be due to observable group-level factors:

$$\begin{aligned} y_{i,j} &= \mu_j + \epsilon_{i,j} \\ \mu_j &= \beta_0 + \beta_1 x_j + \gamma_j \\ \gamma_1, \ldots, \gamma_m &\sim \text{ iid } N(0, \tau^2) \end{aligned}$$

We will consider such models next.