Random effects ANOVA

560 Hierarchical modeling

Peter Hoff

Statistics, University of Washington

Classical data analysis and estimation

The "classical" testing and estimation procedure is as follows:

If the p-value < 0.05,

- reject H_0, and conclude there are group differences,
- estimate μ_j with $\bar{y}_{\cdot j}$.

$$\hat{\mu}_j = \bar{y}_{\cdot j}$$

If the p-value > 0.05,

- accept H_0, and conclude there is no evidence of group differences,
- estimate μ_j with $\bar{y}_{\cdot\cdot}$.

$$\hat{\mu}_j = \bar{y}_{\cdot\cdot}$$

Note that the estimator of μ_j can be written as

$$\hat{\mu}_j = w\bar{y}_j + (1 - w)\bar{y}_{\cdot\cdot}$$

Classical data analysis and estimation

Advantages of classical procedure:

- controls the type I error rate of rejecting H_0;
- is easy to implement and report.

Disadvantages:

- rejecting H_0 doesn't mean no similarities across groups
  $\Rightarrow \bar{y}_{\cdot j}$ is an inefficient estimate of μ_j
- accepting H_0 doesn't mean no differences between groups
  $\Rightarrow \bar{y}_{\cdot\cdot}$ is an inaccurate estimate of μ_j.

An alternative strategy

$$\hat{\mu}_j = w\bar{y}_j + (1 - w)\bar{y}_{\cdot\cdot}$$

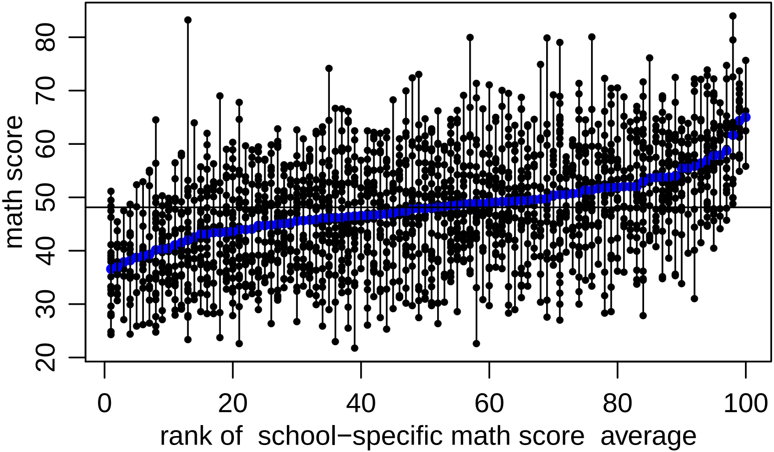
Classical approach: w is the indicator of rejecting H_0.

Multilevel approach: $w = \frac{n/\hat{\sigma}^2}{n/\hat{\sigma}^2 + 1/\hat{\tau}^2}$

The multilevel approach will allow for

- sharing of information across groups,
- the amount of sharing to be estimated from the data.

Example: Test scores



Example: Test scores

```
y.3122<-ndat$mathscore[ndat$school=="3122"]
y.2832<-ndat$mathscore[ndat$school=="2832"]

y.3122

## [1] 75.62 55.86 66.16 62.43

y.2832

##  [1] 66.26 66.12 71.22 54.90 61.98 69.42 61.22 62.99 57.99 61.33 66.85
## [12] 67.87 63.94 73.70 70.36 64.01 57.35 68.25 57.39

mean(ndat$mathscore)

## [1] 48.07446

mean(y.3122)

## [1] 65.0175

mean(y.2832)

## [1] 64.37632
```

Example: Test scores

$$\begin{array}{ccccc} 48.0744556 & < & 64.3763158 & < & 65.0175 \\ \bar{y}_{\cdot\cdot} & < & \bar{y}_{2832} & < & \bar{y}_{3122} \end{array}$$

Based on the data $\{y_{i,j}\}$, how would you estimate μ_{3122} and μ_{2832}?

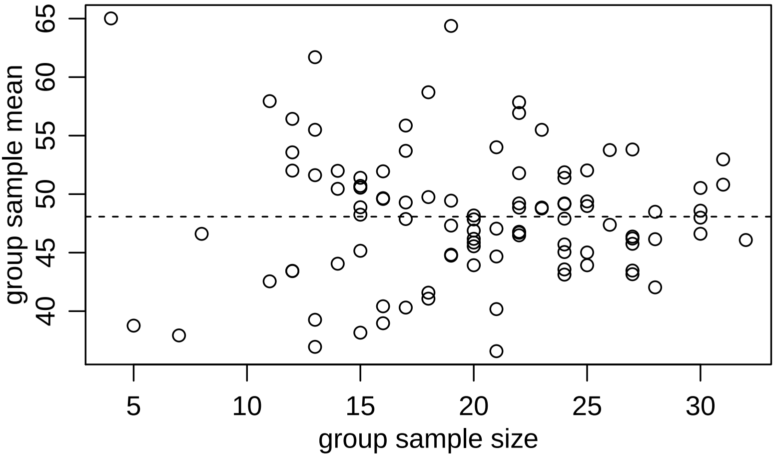
Ignoring across-group information :

- $\hat{\mu}_{2832} = \bar{y}_{2832} = 64.3763158$
- $\hat{\mu}_{3122} = \bar{y}_{3122} = 65.0175$
- $\hat{\mu}_{2832} < \hat{\mu}_{3122}$

Considering across-group information and sample size:

- $\hat{\mu}_{2832} < \bar{y}_{2832} = 64.3763158$
- $\hat{\mu}_{3122} < \bar{y}_{3122} = 65.0175$
- $\hat{\mu}_{2832} \gtrless \hat{\mu}_{3122}$?

Example: Test scores



Possible explanations for $\bar{y}_{3122}$:

- $\bar{y}_{3122}$ is large because μ_{3122} is large;
- $\bar{y}_{3122}$ is large because $\text{sd}(\bar{y}_{3122})$ is large.

Possible explanations for $\bar{y}_{2832}$:

- $\bar{y}_{2832}$ is large because μ_{2832} is large;
- $\bar{y}_{2832}$ is large because $\text{sd}(\bar{y}_{2832})$ is large.

The plausibility of the explanations will depend on

- the group specific sample sizes, $n_1, \ldots, n_m$;
- the observed across-group heterogeneity.

Example: Free throws

```
ftdat[1:20,]

##      PLAYER1     PLAYER2 TEAM  MIN FTM FTA   FT.
## 1        Sam    Jacobson  LAL   12   2   2 1.000
## 2      Steve      Henson  DET   25   2   2 1.000
## 3   Radoslav Nesterovic  MIN   30   2   2 1.000
## 4      Bryce        Drew  HOU  441   8   8 1.000
## 5    Charles    O'bannon  DET  165   8   8 1.000
## 6      Marty      Conlon  MIA   35   2   2 1.000
## 7      Mikki       Moore  DET    6   2   2 1.000
## 8       John      Crotty  POR   19   3   3 1.000
## 9     Gerald     Wilkins  ORL   28   2   2 1.000
## 10 Korleone       Young  DET   15   2   2 1.000
## 11     Brian       Evans  MIN  145   4   4 1.000
## 12      Pooh Richardson  LAC  130   4   4 1.000
## 13   Michael     Hawkins  SAC  203   3   3 1.000
## 14     Randy Livingston  PHO   22   2   2 1.000
## 15     Rusty       Larue  CHI  732  17  17 1.000
## 16      Fred     Hoiberg  IND   87   6   6 1.000
## 17      Herb    Williams  NYK   34   2   2 1.000
## 18      Ryan       Stack  CLE  199  19  20 0.950
## 19       Sam     Cassell  MIL  199  47  50 0.940
## 20    Reggie      Miller  IND 1787 226 247 0.915
```

Who does Indiana pick to shoot its technical foul free throws?

Further limitations of ANOVA

In the wheat yield example we might be interested in

(1) what the yield might be in other plots of land in these 10 regions, or

(2) what the yield might be in other regions.

For general hierarchical data, these questions translate into

(1) making inference about units within groups in our study;

(2) making inference about groups that weren't in our study.

Inference for (1) can be obtained with ANOVA.

Inference for (2) requires

- treating the m groups as a sample from a larger population;
- a statistical model for this larger population.

The hierarchical normal model

$$
\begin{aligned}
y_{i,j} &= \mu + a_j + \epsilon_{i,j} & (1) \\
\{\epsilon_{1,1}, \ldots, \epsilon_{n_1,1}\}, \ldots, \{\epsilon_{1,m}, \ldots, \epsilon_{n_m,m}\} &\sim \text{i.i.d. normal}(0, \sigma^2) & (2) \\
a_1, \ldots, a_m &\sim \text{i.i.d. normal}(0, \tau^2) & (3)
\end{aligned}
$$

The classical ANOVA model consists of (1) and (2).

The HNM assumes the sampling model (3) for the groups.

- $\{a_1, \ldots, a_m\}$ represent differences across groups
- $\{\epsilon_{i,j}\}$ represent differences within groups

The HNM represents this heterogeneity in terms of population variances:

$$
\begin{aligned}
\text{Var}[a] &= \tau^2 = \text{ across-group variance} \\
\text{Var}[\epsilon] &= \sigma^2 = \text{ within-group variance}
\end{aligned}
$$

Marginal and conditional variation

Two levels of heterogeneity require two versions of variance and covariance:

Within-group variance:

- Describes heterogeneity/variance within a particular group;
- Mathematically, is calculated *conditionally* on group-level parameters.

Population-level variance:

- Describes heterogeneity/variance across the population;
- Mathematically, is calculated *marginally* over group-level parameters.

Conditional variance and covariance

For a fixed group j,

$$\begin{aligned} \{y_{1,j}, \ldots, y_{n_j,j}\} &\sim \text{i.i.d. normal}(\mu + a_j, \sigma^2) \\ \{y_{1,j}, \ldots, y_{n_j,j}\} &\sim \text{i.i.d. normal}(\mu_j, \sigma^2) \end{aligned}$$

Variation *around the group mean μ_j* is as follows

$$\begin{aligned} \mathrm{E}[y_{i,j}|\mu, a_j] &= \mu + a_j = \mu_j \\ \mathrm{Var}[y_{i,j}|\mu, a_j] &= \sigma^2, \\ \mathrm{Cov}[y_{i_1,j}, y_{i_2,j}|\mu, a_j] &= 0. \end{aligned}$$

In words,

- sample observations *from the group* are centered around μ_j;
- the variation of the sample *around μ_j* is σ^2;
- the observations within a group are uncorrelated *around μ_j*.

Regarding correlation: Knowing how far $y_{1,j}$ is from μ_j doesn't inform you about about how far $y_{2,j}$ is from μ_j.

Within-group variance and covariance

$$
\begin{aligned}
y_{i,j} &= \mu + a_j + \epsilon_{i,j} \\
y_{i,j} &= \mu_j + \epsilon_{i,j}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}[y_{i,j}|\mu_j] &\equiv \mathrm{E}[(y_{i,j} - \mathrm{E}[y_{i,j}|\mu_j])^2|\mu_j] \\
&= \mathrm{E}[(y_{i,j} - \mu_j)^2|\mu_j] \\
&= \mathrm{E}[(\mu_j + \epsilon_{i,j} - \mu_j)^2|\mu_j] \\
&= \mathrm{E}[\epsilon_{i,j}^2|\mu_j] = \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}[y_{i_1,j}, y_{i_2,j}|\mu_j] &\equiv \mathrm{E}[(y_{i_1,j} - \mathrm{E}[y_{i_1,j}|\mu_j]) \times (y_{i_2,j} - \mathrm{E}[y_{i_2,j}|\mu_j])|\mu_j] \\
&= \mathrm{E}[(y_{i_1,j} - \mu_j) \times (y_{i_2,j} - \mu_j)|\mu_j] \\
&= \mathrm{E}[\epsilon_{i_1,j}\epsilon_{i_2,j}|\mu_j] = 0
\end{aligned}
$$

Population level variance and covariance

Across all groups,

$$\begin{aligned} a_1, \ldots, a_m &\sim \text{i.i.d. normal}(0, \tau^2) \\ \{y_{1,j}, \ldots, y_{n_j,j}\} &\sim \text{i.i.d. normal}(\mu + a_j, \sigma^2) \end{aligned}$$

For a randomly sampled observation i from a randomly sampled group j,

$$\begin{aligned} \text{E}[y_{i,j}|\mu] &= \text{E}[\mu + a_j + \epsilon_{i,j}|\mu] \\ &= \text{E}[\mu|\mu] + \text{E}[a_j|\mu] + \text{E}[\epsilon_{i,j}|\mu] \\ &= \mu + 0 + 0 = \mu \end{aligned}$$

This is the *population mean*.

Population level variance and covariance

Variation *around the population mean* μ is as follows:

$$
\begin{aligned}
\mathrm{E}[y_{i,j}|\mu] &= \mathrm{E}[\mu + a_j|\mu] = \mu + 0 = \mu, \\
\mathrm{Var}[y_{i,j}|\mu] &= \sigma^2 + \tau^2, \\
\mathrm{Cov}[y_{i_1,j}, y_{i_2,j}|\mu] &= \tau^2.
\end{aligned}
$$

In words,

- sampled observations *across groups* are centered around μ;
- the variation of the sample *around* μ is $\sigma^2 + \tau^2$;
- the observations within a group are correlated *around* μ.

Regarding correlation: Knowing how far $y_{1,j}$ is from μ *does* inform you about how far $y_{2,j}$ is from μ.

Population level variance

$$
\begin{aligned}
\text{Var}[y_{i,j}|\mu] &\equiv \text{E}[(y_{i,j} - \text{E}[y_{i,j}|\mu])^2|\mu] \\
&= \text{E}[(y_{i,j} - \mu)^2|\mu] \\
&= \text{E}[(\mu + a_j + \epsilon_{i,j} - \mu)^2|\mu] \\
&= \text{E}[(a_j + \epsilon_{i,j})^2|\mu] \\
&= \text{E}[a_j^2 + 2a_j\epsilon_{i,j} + \epsilon_{i,j}^2|\mu] \\
&= \tau^2 + 0 + \sigma^2 = \sigma^2 + \tau^2
\end{aligned}
$$

Population level covariance and correlation

$$\begin{aligned}
\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu] &\equiv \text{E}[(y_{i_1,j} - \text{E}[y_{i_1,j}|\mu]) \times (y_{i_2,j} - \text{E}[y_{i_2,j}])|\mu] \\
&= \text{E}[(y_{i_1,j} - \mu) \times (y_{i_2,j} - \mu)|\mu] \\
&= \tau^2
\end{aligned}$$

$$\begin{aligned}
\text{Cor}[y_{i_1,j}, y_{i_2,j}|\mu] &\equiv \frac{\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu]}{\sqrt{\text{Var}[y_{i_1,j}|\mu]\text{Var}[y_{i_2,j}|\mu]}} \\
&= \frac{\tau^2}{\tau^2 + \sigma^2} \equiv \rho
\end{aligned}$$

The correlation ρ is the *intraclass correlation coefficient*.

Estimation of τ^2 and ρ

The easiest way to estimate τ^2 is using the method-of-moments. Recall,

$$
\begin{aligned}
MSG &= \frac{1}{m-1}\sum_j\sum_i(\bar{y}_j - \bar{y}_{\cdot\cdot})^2 \\
&= \frac{n}{m-1}\sum(\bar{y}_j - \bar{y}_{\cdot\cdot})^2 \\
\mathrm{E}[MSG|a_1,\ldots,a_m] &= \frac{n}{m-1}\left(\frac{m-1}{n}\sigma^2 + \sum a_j^2\right) \\
&= \sigma^2 + n \times \frac{1}{m-1}\sum a_j^2.
\end{aligned}
$$

If groups are sampled, the expectation of MSG over samples is given by

$$
\begin{aligned}
\mathrm{E}[\mathrm{E}[MSG|a_1,\ldots,a_m]] &= \mathrm{E}[\sigma^2 + n \times \frac{1}{m-1}\sum a_j^2] \\
&= \sigma^2 + n \times \mathrm{E}[\frac{1}{m-1}\sum a_j^2] \\
&= \sigma^2 + n\tau^2.
\end{aligned}
$$

(In the ANOVA parameterization, $\sum a_j^2 = \sum(a_j - \bar{a})^2$ becuase $\bar{a} = 0$)

Estimation of τ^2 and ρ

The result suggests

$$\widehat{\sigma^2 + n\tau^2} = MSG.$$

How to estimate τ^2? Recall $\mathrm{E}[MSE] = \sigma^2$, so we can use

$$\hat{\sigma^2} = MSE.$$

This suggests

$$\widehat{n\tau^2} = MSG - MSE$$
$$\hat{\tau}^2 = (MSG - MSE)/n.$$

Comments:

- $MSG - MSE$ *could* be negative. If so, it is standard to set $\hat{\tau}^2 = 0$.
- If sample sizes are unequal, the formula must be modified slightly:

  $$\hat{\tau}^2 = (MSG - MSE)/\tilde{n}$$

  where there is a horrible formula for $\tilde{n}$.

Unequal sample sizes

$$\hat{\tau}^2 = (MSG - MSE)/\tilde{n}$$

$$\tilde{n} = \bar{n} - \frac{\text{sample variance}(n_1, \ldots, n_m)}{m\bar{n}}$$

where $\bar{n} = \sum_j n_j/m = \text{sample mean}(n_1, \ldots, m_m)$.

Estimation of τ^2 and ρ

It is common to use a "plug-in" estimate of ρ:

$$\hat{\rho} = \widehat{\frac{\tau^2}{\tau^2 + \sigma^2}} = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2}.$$

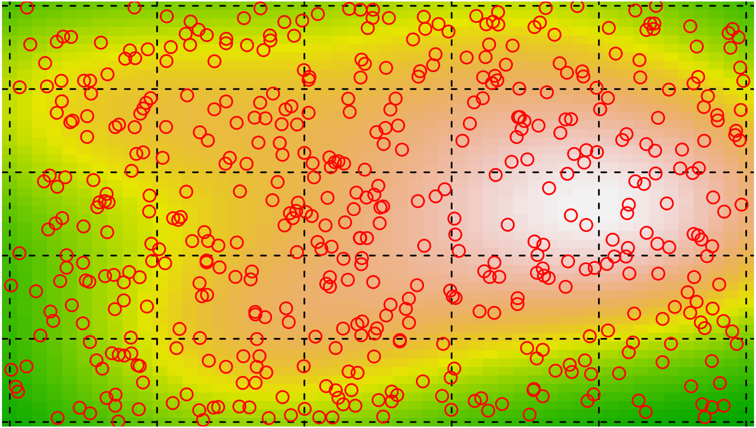
A standard error for ρ (with which we can get a CI) is

$$\text{se}(\hat{\rho}) = (1 - \hat{\rho}) \times (1 + (n-1)\hat{\rho})\sqrt{\frac{2}{n(n-1)(m-1)}}.$$

Example: Wheat

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

MSG<-fit[1,3]
MSE<-fit[2,3]

MSG

## [1] 3.70759

MSE

## [1] 1.787206

t2<-(MSG-MSE)/n

rho<-t2/(t2+MSE)

rho

##         1
## 0.1768894

se.rho<- (1-rho)*(1+(n-1)*rho)*sqrt( 2/(  n*(n-1)*(m-1)))

rho + c(-2,2)*se.rho

## [1] -0.1194179  0.4731966
```

Two-stage sampling



μ=2.1124814

Ignoring across-group heterogeneity

Task: Construct a 95% CI for the population mean.

***t*-interval for SRS:**
If $y_1, \ldots, y_n$ is an iid sample with $\text{E}[y_i] = \mu$ and $\text{Var}[y_i] = \sigma^2$,

$$\text{E}[\bar{y}] = \mu \ , \ \text{Var}[\bar{y}] = \sigma^2/n.$$

By the central limit theorem,

$$\bar{y} \dot{\sim} N(\mu, \sigma^2/n) \ , \ \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1).$$

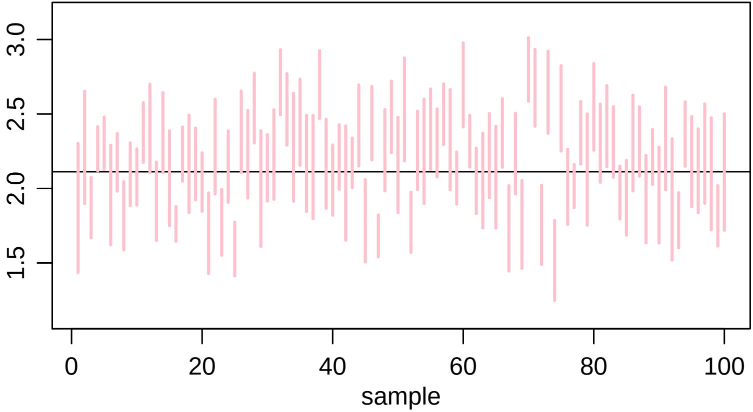
As σ^2 is generally unknown, we use

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \dot{\sim} t_{n-1}, \ \text{, where } s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2.$$

From this, we have

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n} \ \text{ is a 95\% CI for } \mu.$$

Ignoring across-group heterogeneity



Building an accurate t-interval

Recall that an approximate 95% CI for μ is given by

$$\bar{y} \pm 2 \times \text{se}(\bar{y}),$$

where $\text{se}(\bar{y})$ is an approximation to the standard deviation of $\bar{y}$.

How to find $\text{se}(\bar{y})$:

1. compute the variance v of $\bar{y}$ based on the model;
2. find an estimate $\hat{v}$ of v;
3. let $\text{se}(\bar{y}) = \sqrt{v}$.

So the first step is to find $\text{Var}[\bar{y}]$:

Variance of a group mean around population mean

$$
\begin{aligned}
\text{Var}[\bar{y}] &= \text{Var}[\frac{1}{mn}\sum_j \sum_i y_{i,j}] \\
&= \text{Var}[\frac{1}{m}\sum_j \frac{1}{n}\sum_i y_{i,j}] \\
&= \text{Var}[\frac{1}{m}\sum_j \bar{y}_j] \\
&= \frac{1}{m^2}\text{Var}[\sum_j \bar{y}_j] \\
&= \frac{1}{m^2}\sum_j \text{Var}[\bar{y}_j] \\
&= \frac{1}{m^2} m\text{Var}[\bar{y}_1] \\
&= \frac{1}{m}\text{Var}[\bar{y}_1]
\end{aligned}
$$

Variance of a group mean around population mean

What is $\text{Var}[\bar{y}_1]$? We've shown

$$\text{Var}[y_{i,1}] = \sigma^2 + \tau^2,$$

but generally,

$$\text{Var}[\bar{y}_1] \neq [\sigma^2 + \tau^2]/n.$$

Quiz: What is the smallest that $\text{Var}[\bar{y}_1]$ could be for fixed σ^2 and n? Recall

$$\text{Cor}[y_{i,1}, y_{i,2}] = \frac{\tau^2}{\tau^2 + \sigma^2}$$

Answer: When τ^2 is zero the within group samples are independent and so

$$\text{Var}[\bar{y}_1] \geq \sigma^2/n$$

Variance of a group mean around population mean

Quiz: what is the smallest that $\text{Var}[\bar{y}_1]$ could be for fixed σ^2 and τ^2?

Answer: Increasing n can reduce variation of $\bar{y}_1$ around μ_1, but across group heterogeneity remains:

$$\begin{aligned} \text{for large } n, \bar{y}_1 &\approx \mu_1 \\ \text{Var}[\mu_1] &= \tau^2 \\ \text{Var}[\bar{y}_1] &\geq \tau^2 \end{aligned}$$

Variance of a group mean around population mean

Let's compute $\mathrm{Var}[\bar{y}_1]$. For notational convenience, we'll drop the group index, and assume $\mu = 0$, so

$$\mathrm{E}[y_i] = 0 \ , \ \mathrm{E}[y_i^2] = \sigma^2 + \tau^2 \ , \ \mathrm{E}[y_i y_k] = \tau^2$$

In this case,

$$\begin{aligned}
\mathrm{Var}[\bar{y}] &= \mathrm{E}[\bar{y}^2] \\
&= \mathrm{E}[\frac{1}{n^2}(\sum y_i)^2] \\
&= \frac{1}{n^2}\mathrm{E}[\sum y_i^2 + \sum_{i \neq k} y_i y_j] \\
&= \frac{1}{n^2}(n[\sigma^2 + \tau^2] + n(n-1)\tau^2) \\
&= \frac{\sigma^2}{n} + \frac{1}{n}\tau^2 + \frac{n-1}{n}\tau^2 \\
&= \frac{\sigma^2}{n} + \tau^2
\end{aligned}$$

Exercise: Make sure the answer makes sense to you intuitively.

Variance of the sample grand mean

$$\mathrm{Var}[\bar{y}_{\cdot\cdot}] = \frac{1}{m}\mathrm{Var}[\bar{y}_j]$$
$$\mathrm{Var}[\bar{y}_j] = \frac{1}{n}\sigma^2 + \tau^2$$

$$\mathrm{Var}[\bar{y}_{\cdot\cdot}] = \frac{1}{nm}\sigma^2 + \frac{1}{m}\tau^2$$

What happens as

- $n \to \infty$ and m stays fixed?
- $m \to \infty$ and n stays fixed?

Standard error and CI

$$\widehat{\text{Var}}[\bar{y}_{\cdot\cdot}] = \frac{1}{nm}\hat{\sigma}^2 + \frac{1}{m}\tau^2$$

- $\hat{\sigma}^2 = MSE$
- $\hat{\tau}^2 = (MSG - MSE)/n$

$$\widehat{\text{Var}}[\bar{y}_{\cdot\cdot}] = \frac{1}{mn}MSG$$

This should make sense, because previously we claimed

$$\text{E}[MSG] = \sigma^2 + n \times \tau^2,$$

so

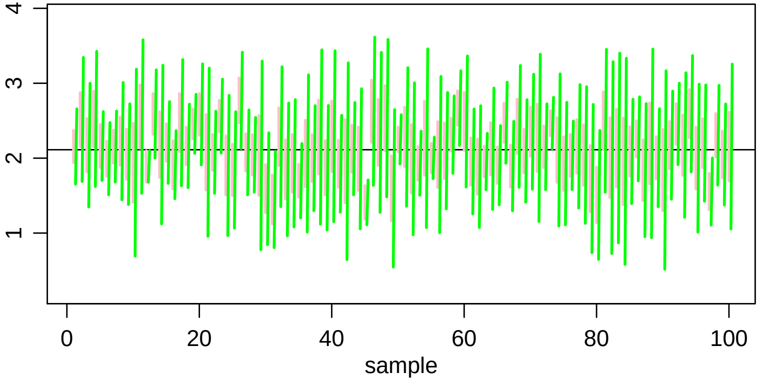
$$\text{E}[\frac{1}{mn}MSG] = \frac{1}{mn}\sigma^2 + \frac{1}{m}\tau^2 = \text{Var}[\bar{y}_{\cdot\cdot}]$$

Confidence interval

$$\bar{y}_{\cdot\cdot} \pm 2 \times \sqrt{MSG/mn}$$

```
round(y,2)

##  [1] 2.40 2.31 2.14 2.27 2.31 1.73 1.92 1.50 1.94 1.88 1.65 0.98 0.71 1.56
## [15] 1.68 3.36 3.26 3.33 3.40 3.06

g

##  [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4

anova(lm(y~as.factor(g)))

## Analysis of Variance Table
##
## Response: y
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## as.factor(g)  3 10.6178  3.5393  55.477 1.122e-08 ***
## Residuals    16  1.0207  0.0638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

MSG<-anova(lm(y~as.factor(g)))[1,3]

mean(y) + c(-2,2)*sqrt( MSG/(m*n)  )

## [1] 1.328860 3.011539

mean(y) + c(-2,2)*sqrt( var(y)/(m*n)  )

## [1] 1.820184 2.520215
```

Accounting for across-group heterogeneity




```r
mean(  CI.tss0[,1] < mu  & mu < CI.tss0[,2] )

## [1] 0.771

mean(  CI.tss1[,1] < mu  & mu < CI.tss1[,2] )

## [1] 0.931
```

Summary

$$y_{i,j} = \mu + a_j + \epsilon_{i,j}$$
$$\text{Var}[\epsilon_{i,j}] = \sigma^2$$
$$\text{Var}[a_j] = \tau^2$$

Variation around the group mean: $\mu_j = \mu + a_j$

- $\text{Var}[y_{i,j}|\mu_j] = \sigma^2$
- $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu_j] = 0$
- $\text{Var}[\bar{y}_j|\mu_j] = \sigma^2/n$

Variation around the grand mean:

- $\text{Var}[y_{i,j}|\mu] = \sigma^2 + \tau^2$
- $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu] = \tau^2$
- $\text{Var}[\bar{y}_j|\mu] = \sigma^2/n + \tau^2$
- $\text{Var}[\bar{y}_{\cdot\cdot}|\mu] = \sigma^2/(mn) + \tau^2/m$