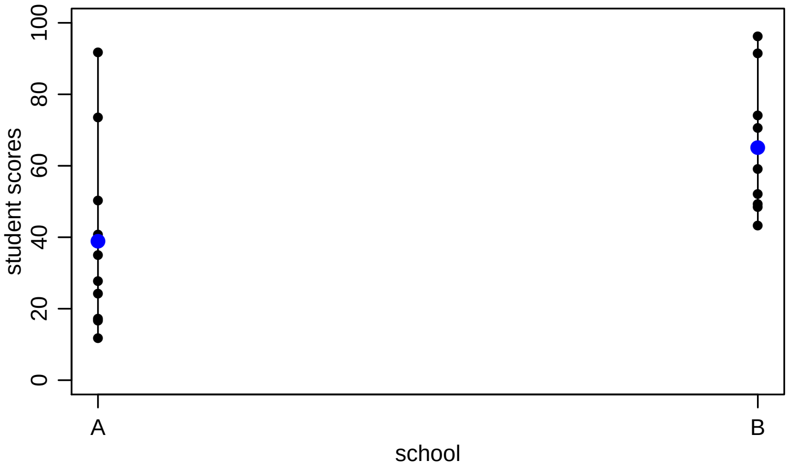
ANOVA

560 Hierarchical modeling

Peter Hoff

Statistics, University of Washington

Two group comparison



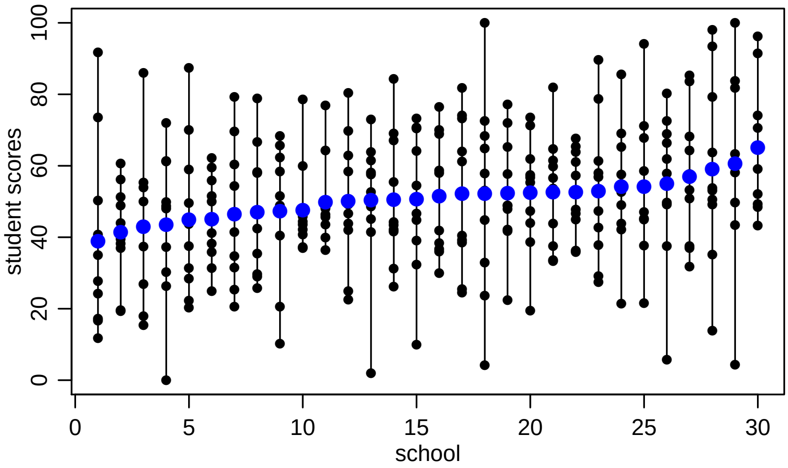
Two group comparison

```
t.test(yA,yB)

##
##  Welch Two Sample t-test
##
## data:  yA and yB
## t = -2.5922, df = 16.037, p-value = 0.01962
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -47.543740  -4.770049
## sample estimates:
## mean of x mean of y
##  38.88928  65.04617
```

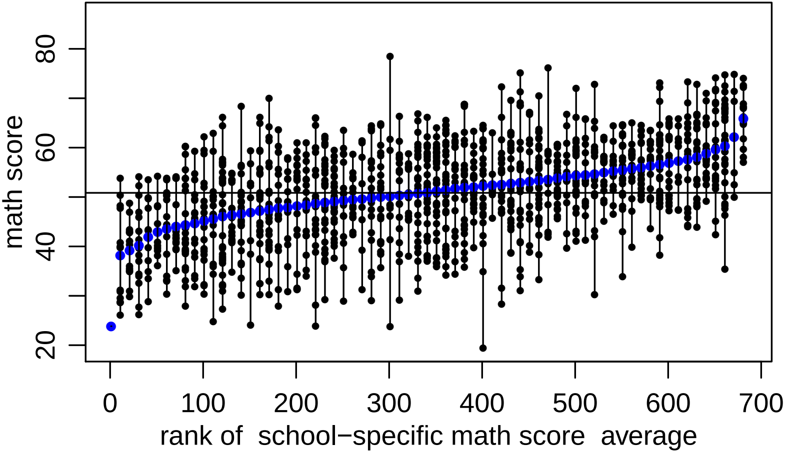
Two group comparison

```
t.test(yA,yB,var.equal=TRUE)

##
##  Two Sample t-test
##
## data:  yA and yB
## t = -2.5922, df = 18, p-value = 0.0184
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -47.356240  -4.957549
## sample estimates:
## mean of x mean of y
##  38.88928  65.04617
```

Multi group comparisons



NELS data



math score
20
40
60
80
0
100
200
300
400
500
600
700
rank of school−specific math score average

Data analysis goals

Descriptions of center:

- overall mean
- within group means

Descriptions of variability:

- across group variability
- within group variability

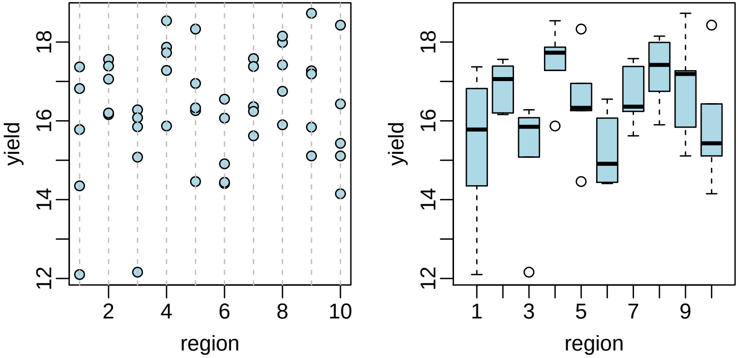
Tests and comparisons:

- Overall evaluation: Is there evidence of differences between groups?
- Specific comparisons:

  Which group has the largest mean?
  Which has the smallest?
  How confident are we in these evaluations?

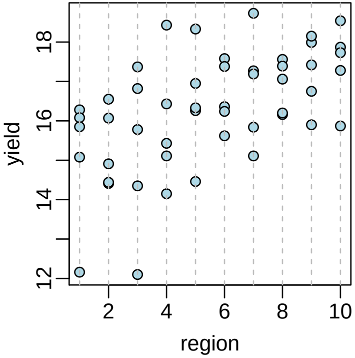
Example (wheat yield):

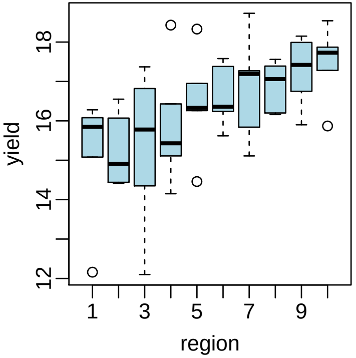
- $m = 10$ regions of land were randomly selected,
- $n = 5$ plots of land were seeded within each region.
- $y_{i,j}$=the yield of plot i in region j.



Example (wheat yield):

- $m = 10$ regions of land were randomly selected,
- $n = 5$ plots of land were seeded within each region.
- $y_{i,j}$=the yield of plot i in region j.





One-way ANOVA model

$$
\begin{aligned}
y_{i,j} =& \mu + a_j + \epsilon_{i,j} && \text{(treatment effects model) , or} \\
y_{i,j} =& \mu_j + \epsilon_{i,j} && \text{(treatment means model),}
\end{aligned}
$$

where $\mu_j = \mu + a_j$.

- μ is expected yield across all regions;
- μ_j is expected yield from region j;
- a_j is the deviation of region-specific expected yield from μ;

$$\mu_j = \mu + a_j \;\Leftrightarrow a_j = \mu_j - \mu$$

- $\epsilon_{i,j}$ deviation of an observed plot yield from its region-specific expectation.

Identifiability

The standard ANOVA model parameterizes things so that

- $\sum_j a_j = 0$ (sum-to-zero side conditions),
- $\{\epsilon_{i,j}\} \sim$ i.i.d. $p(\epsilon)$, with $\mathrm{E}[\epsilon_{i,j}] = 0$ within all groups.

In this case,

$$\begin{aligned}
\mathrm{E}[y_{i,j}|\mu, a_1, \ldots, a_m] &= \mathrm{E}[\mu + a_j + \epsilon_{i,j}|\mu, a_1, \ldots, a_m] \\
&= \mathrm{E}[\mu|\mu, a_1, \ldots, a_m] + \mathrm{E}[a_j|\mu, a_1, \ldots, a_m] + \mathrm{E}[\epsilon_{i,j}|\mu, a_1, \ldots, a_m] \\
&= \mu + a_j \\
&= \mu_j
\end{aligned}$$

If we assume $p(\epsilon)$ is the normal$(0, \sigma^2)$ distribution, then the model is

$$\begin{aligned}
y_{i,j} &\sim \text{normal}(\mu + a_j, \sigma^2) \quad \text{or equivalently,} \\
y_{i,j} &\sim \text{normal}(\mu_j, \sigma^2).
\end{aligned}$$

Parameter estimates

Parameters to estimate include

- $\{\mu_1, \dots, \mu_m, \sigma^2\}$, or equivalently
- $\{\mu, a_1, \dots, a_m, \sigma^2\}$

If $\hat{\mu}_j$ is an estimate of μ_j, we say that

- $\hat{y}_{i,j} = \hat{\mu}_j$ is the *fitted value* of $y_{i,j}$;
- $\hat{\epsilon}_{i,j} = y_{i,j} - \hat{y}_{i,j} = y_{i,j} - \hat{\mu}_j$ is the *residual* for $y_{i,j}$.

OLS estimation: The OLS estimates are the values that minimize

$$
\begin{aligned}
\mathrm{SSE}(\hat{\mu}_1, \dots, \hat{\mu}_m) &= (y_{1,1} - \hat{\mu}_1)^2 + (y_{2,1} - \hat{\mu}_1)^2 + \cdots + (y_{n-1,m} - \hat{\mu}_m)^2 + (y_{n,m} - \hat{\mu}_m)^2 \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n} (y_{i,j} - \hat{\mu}_j)^2
\end{aligned}
$$

Minimizing sums of squares

Task: Find the value $\hat{\mu}$ that minimizes

$$\sum_{i=1}^{n} (y_i - \hat{\mu})^2$$

Solution:

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \hat{\mu})^2 &= \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y} - \hat{\mu})^2 \\
&= \sum_{i=1}^{n} [(y_i - \bar{y})^2 + 2(y_i - \bar{y})(\bar{y} - \hat{\mu}) + (\bar{y} - \hat{\mu})^2] \\
&= \sum (y_i - \bar{y})^2 + n(\bar{y} - \hat{\mu})^2,
\end{aligned}$$

- The sum of squares is minimized by $\hat{\mu} = \bar{y}$;
- The minimim value is $SSE = \sum (y_i - \bar{y})^2$;
- Recall that the sample variance is given by $\frac{1}{n-1} SSE$.

OLS estimates

Not surprisingly,

$$(\hat{\mu}_1, \ldots, \hat{\mu}_m) = (\bar{y}_1, \ldots, \bar{y}_m).$$

For the "treatment effects" parametrization,

$$\hat{\mu} = \bar{y}_{\cdot\cdot}$$
$$\hat{a}_j = (\hat{\mu}_j - \hat{\mu}) = (\bar{y}_j - \bar{y}_{\cdot\cdot})$$

Exercises: Show that (in this case of equal group sample sizes),

- $\hat{\mu} = \sum \hat{\mu}_j / m$;
- $\sum \hat{a}_j = 0$.

ANOVA decomposition

The OLS estimates provide a "decomposition" of the data:

$$\begin{aligned} y_{i,j} = & \quad \bar{y}_{\cdot\cdot} & + & \quad (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) & + & \quad (y_{i,j} - \bar{y}_{\cdot j}) \\ & \quad \hat{\mu} & + & \quad \hat{a}_j & + & \quad \hat{\epsilon}_{i,j}. \end{aligned}$$

ANOVA decomposition

Total		Group		Error
$y_{11} - \bar{y}_{..}$	$=$	$(\bar{y}_{.1} - \bar{y}_{..})$	$+$	$(y_{11} - \bar{y}_{.1})$
$y_{21} - \bar{y}_{..}$	$=$	$(\bar{y}_{.1} - \bar{y}_{..})$	$+$	$(y_{21} - \bar{y}_{.1})$
.	$=$	.	$+$	.
.	$=$	.	$+$	.
.	$=$	.	$+$	.
$y_{n1} - \bar{y}_{..}$	$=$	$(\bar{y}_{.1} - \bar{y}_{..})$	$+$	$(y_{n1} - \bar{y}_{.1})$
$y_{12} - \bar{y}_{..}$	$=$	$(\bar{y}_{.2} - \bar{y}_{..})$	$+$	$(y_{12} - \bar{y}_{.2})$
.	$=$	.	$+$	.
.	$=$	.	$+$	.
.	$=$	.	$+$	.
$y_{n2} - \bar{y}_{..}$	$=$	$(\bar{y}_{.2} - \bar{y}_{..})$	$+$	$(y_{n2} - \bar{y}_{.2})$
$\vdots$		$\vdots$		$\vdots$
$y_{1m} - \bar{y}_{..}$	$=$	$(\bar{y}_{.m} - \bar{y}_{..})$	$+$	$(y_{1m} - \bar{y}_{.m})$
.	$=$	.	$+$	.
.	$=$	.	$+$	.
.	$=$	.	$+$	.
$y_{nm} - \bar{y}_{..}$	$=$	$(\bar{y}_{.m} - \bar{y}_{..})$	$+$	$(y_{nm} - \bar{y}_{.m})$
SST	$=$	SSG	$+$	SSE
$mn - 1$	$=$	$m - 1$	$+$	$m(n - 1)$

ANOVA decomposition

SST: Total sum of squares variation = variation around $\bar{y}_{\cdot\cdot}$.

SSG: Total group variation = variation of group means around grand mean.

SSE: Error or residual variation = variation of data around group means.

Sum of squares decomposition: You can show that

$$\begin{array}{ccccc} \text{SST} & = & \text{SSG} & + & \text{SSE} \\ \text{total variation} & = & \text{between group variation} & + & \text{within group variation} \end{array}$$

ANOVA for wheat yield

```
y

##  [1] 17.37 15.78 14.35 12.10 16.82 16.16 16.20 17.56 17.39 17.06 16.28
## [12] 16.08 15.08 12.16 15.85 17.87 17.73 15.87 17.28 18.54 18.33 16.26
## [23] 16.95 16.33 14.46 14.41 14.44 14.91 16.55 16.07 16.36 16.24 17.58
## [34] 17.38 15.62 15.90 17.42 17.99 16.75 18.15 17.27 15.84 17.19 15.11
## [45] 18.73 18.43 15.11 14.15 16.43 15.43

g

##  [1]  1  1  1  1  1  2  2  2  2  2  3  3  3  3  3  4  4  4  4  4  5  5  5
## [24]  5  5  6  6  6  6  6  7  7  7  7  7  8  8  8  8  8  9  9  9  9  9 10
## [47] 10 10 10 10

mu.grand<-mean(y)
mu.group<-tapply(y,g,mean)

mu.grand

## [1] 16.3064

mu.group

##      1      2      3      4      5      6      7      8      9     10
## 15.284 16.874 15.090 17.458 16.466 15.276 16.636 17.242 16.828 15.910

mean(mu.group)

## [1] 16.3064
```

ANOVA for wheat yield

```
SST<-sum( (y-mu.grand)^2 )
SST

## [1] 104.8566

mu.group[ g ]

##      1      1      1      1      1      2      2      2      2      2
## 15.284 15.284 15.284 15.284 15.284 16.874 16.874 16.874 16.874 16.874
##      3      3      3      3      3      4      4      4      4      4
## 15.090 15.090 15.090 15.090 15.090 17.458 17.458 17.458 17.458 17.458
##      5      5      5      5      5      6      6      6      6      6
## 16.466 16.466 16.466 16.466 16.466 15.276 15.276 15.276 15.276 15.276
##      7      7      7      7      7      8      8      8      8      8
## 16.636 16.636 16.636 16.636 16.636 17.242 17.242 17.242 17.242 17.242
##      9      9      9      9      9     10     10     10     10     10
## 16.828 16.828 16.828 16.828 16.828 15.910 15.910 15.910 15.910 15.910

SSG<-sum( (mu.group[ g ]-mu.grand)^2 )
SSG

## [1] 33.36831

n*sum( (mu.group-mu.grand)^2 )

## [1] 33.36831
```

ANOVA for wheat yield

```
SSE<-sum( (y-mu.group[ g ])^2 )
SSE

## [1] 71.48824

SSE+SSG

## [1] 104.8566

SST

## [1] 104.8566
```

ANOVA table

The ANOVA decomposition is usually summarized with an ANOVA table:

source	deg of freedom	SS	MS	F-ratio
groups	$m-1$	SSG	MSG=SSG$/(m-1)$	MSG/MSE
residuals	$m(n-1)$	SSE	MSE=SSE$/m(n-1)$	
total	$mn-1$	SST		

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

## Analysis of Variance Table
##
## Response: y
##              Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(g)  9 33.368  3.7076  2.0745 0.0555 .
## Residuals    40 71.488  1.7872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA table

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

## Analysis of Variance Table
##
## Response: y
##              Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(g)  9 33.368  3.7076  2.0745 0.0555 .
## Residuals    40 71.488  1.7872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

SSG

## [1] 33.36831

SSG/(m-1)

## [1] 3.70759

SSE

## [1] 71.48824

SSE/(m*(n-1))

## [1] 1.787206

(SSG/(m-1)) / (SSE/(m*(n-1)))

## [1] 2.074518
```

ANOVA decomposition as a description

The ANOVA decomposition and sums of squares provide

Descriptions of center:

- overall mean: $\bar{y}_{\cdot\cdot}$
- group means: $\bar{y}_1, \ldots, \bar{y}_m$
- group effects: $\bar{y}_1 - \bar{y}_{\cdot\cdot}, \ldots, \bar{y}_1 - \bar{y}_{\cdot\cdot}$

Descriptions of variability:

- across group variability

$$\begin{aligned} \text{SSG} &= \sum_j \sum_i (\bar{y}_j - \bar{y}_{\cdot\cdot})^2 \\ &= n \sum_j (\bar{y}_j - \bar{y}_{\cdot\cdot})^2 = n \times (m-1) \times \text{sample variance}(\bar{y}_1, \ldots, \bar{y}_m) \end{aligned}$$

- within group variability

$$\text{SSE} = \sum_j \sum_i (y_{i,j} - \bar{y}_j)^2 = \sum_j (n-1) s_j^2$$

SSG:

```
SSG

## [1] 33.36831

n*(m-1)*var(mu.group)

## [1] 33.36831
```

SSE:

```
SSE

## [1] 71.48824

tapply(y,g,var)

##       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

sum( ( n-1)* tapply(y,g,var)  )

## [1] 71.48824
```

Tests and comparisons:

- Overall evaluation: Is there evidence of differences between groups?
- Specific comparisons:

  Which group has the largest mean?
  Which has the smallest?
  How confident are we in these evaluations?

The first of these can be addressed with an *F-test*.

Testing for across-group heterogeneity

Model:

$$y_{i,j} = \mu + a_j + \epsilon_{i,j} \quad \{\epsilon_{i,j}\} \sim \text{iid } N(0, \sigma^2)$$

Hypotheses:
Consider deciding between the following hypotheses:

$$\begin{aligned} H_0 &: a_j = 0 \text{ for all } j \\ H_1 &: a_j \neq 0 \text{ for some } j \end{aligned}$$

H_0 imples all group means are the same, H_1 implies the opposite.

Statistical inference:
How can we evaluate H_1 versus H_0 based on the observed data?

SSG as a measure of heterogeneity

$$MSG = n \times \text{sample variance}(\bar{y}_1, \ldots, \bar{y}_m)$$

Proceeding heuristically,

$$\begin{aligned} \mathrm{E}[\bar{y}_j] &= \mu + a_j \\ \bar{y}_j &\approx \mu + a_j \end{aligned}$$

$$\begin{aligned} \text{sample variance}(\bar{y}_1, \ldots, \bar{y}_m) &\approx \text{sample variance}(\mu + a_1, \ldots, \mu + a_m) \\ &= \text{sample variance}(a_1, \ldots, a_m) \\ &= \frac{1}{m-1} \sum a_j^2 \end{aligned}$$

Intuitively,

$$H_0 \text{ true} \Leftrightarrow \frac{1}{m-1} \sum a_j^2 = 0 \Leftrightarrow \text{small MSG}$$

$$H_1 \text{ true} \Leftrightarrow \frac{1}{m-1} \sum a_j^2 > 0 \Leftrightarrow \text{large MSG}$$

Expected mean squares

$$
\begin{aligned}
MSG &= n \times \text{sample variance}(\bar{y}_1, \ldots, \bar{y}_m) \\
&\approx n \times \text{sample variance}(a_1, \ldots, a_m) \\
&= n \times \frac{1}{m-1} \sum a_j^2
\end{aligned}
$$

More precisely, one can show that

$$\mathrm{E}[MSG] = \sigma^2 + n \times \frac{1}{m-1} \sum a_j^2,$$

where the σ^2 comes from the fact that $\bar{y}_j$ only approximates a_j.

Letting $\tau^2 = \frac{1}{m-1} \sum a_j^2$, we have

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

where τ^2 is the *across-group variability*.

Comparison to σ^2

Idea:

$$MSG \approx \sigma^2 \;\Rightarrow\; \tau^2 \text{ is small or zero} \;\Rightarrow\; \text{accept } H_0$$

$$MSG > \sigma^2 \;\Rightarrow\; \tau^2 \text{ is not zero} \;\Rightarrow\; \text{accept } H_1$$

Problem: We don't know what σ^2 is.

Solution: Compare MSG to an estimate of σ^2.

Comparison to MSE

$$
\begin{aligned}
MSE = SSE/m(n-1) &= \frac{1}{m(n-1)} \sum_j \sum_i (y_{i,j} - \bar{y}_j)^2 \\
&= \frac{1}{m} \sum_j \frac{1}{n-1} \sum_i (y_{i,j} - \bar{y}_j)^2 \\
&= \frac{1}{m} \sum_j s_j^2 .
\end{aligned}
$$

Recall that $\mathrm{E}[s_j^2] = \sigma^2$, and so

$$\mathrm{E}[MSE] = \sigma^2$$

The F-statistic

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

Let $F = MSG/MSE$. Then

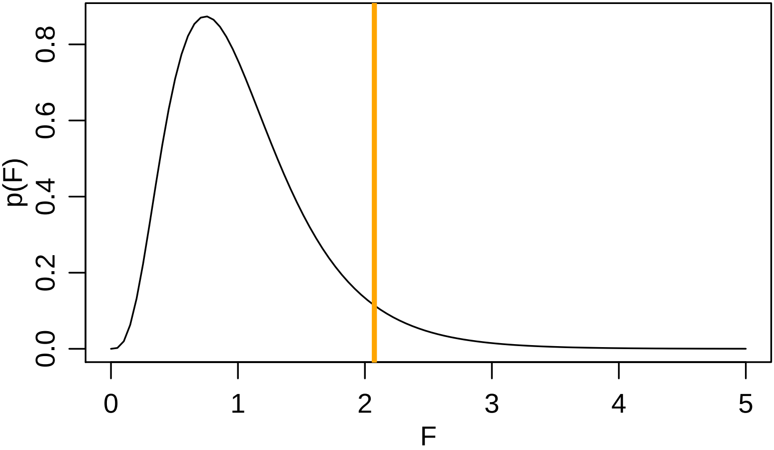
under H_0, MSG/MSE should be around 1,

under H_1, MSG/MSE should be bigger than 1.

Null distribution

Under the normal model $y_1, \ldots, y_n \sim$ iid $N(\mu, \sigma^2)$,

$$MSG/MST = F \sim F_{m-1, m(n-1)}$$



Classical testing for across-group heterogeneity

- We expect an $F_{m-1,m(n-1)}$-distribution under H_0.
- We observe $F(\boldsymbol{y}) = MSG/MSE$.
- Discrepancy between $F_{m-1,m(n-1)}$ and $F(\boldsymbol{y})$ is evidence against H_0.

$$p\text{-value} = \Pr(F_{m-1,m(n-1)} \geq F(\boldsymbol{y}))$$

```
MSG<-SSG/(m-1)
MSE<-SSE/(m*(n-1))
MSG/MSE

## [1] 2.074518

1-pf( MSG/MSE, m-1,m*(n-1))

## [1] 0.05550019
```

ANOVA table

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

## Analysis of Variance Table
##
## Response: y
##              Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(g)  9 33.368  3.7076  2.0745 0.0555 .
## Residuals    40 71.488  1.7872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Classical data analysis and estimation

The "classical" hypothesis testing and parameter estimation procedure is

If the p-value < 0.05,

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

If the p-value > 0.05,

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

Note that the esitmator 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

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

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.