Outline

1. The traditional approach
2. The Mean Squares approach for the Completely randomized design (CRD)
   - CRD and one-way ANOVA
   - Variance components and the F test
   - Inference about the intercept
   - Sample vs. subsample size determination
3. The Mean Squares approach for CRD with subsampling
   - Example and Model
   - ANOVA table
   - F tests
   - Pairwise treatment comparisons with LSD
   - Sample and subsample size determination

Traditional methods for experimental studies

- An experimental design is chosen from a finite set (catalog) of experimental designs.
- Designs are balanced, at least in some way.
- Designs are named.
- Each design is associated to a particular model.
- The traditional analysis of these models uses F-tests, as in ANOVA.

Because: there are formulas for these F-tests on nicely balanced data sets. (Remember?)

At the time, formulas were necessary for the calculations to be computationally tractable.

Traditional methods

You already know:

- Many of these designs (we just need to name them!)
- All their associated models (just need to recognize them!)
- The model for a particular ‘traditional’ design is either a linear model or a random-effect or a mixed-effect model, with a particular set of predictors and a particular set of interactions between them.
- The models we have already covered are not restricted to catalog.

What we will cover:

- the names of designs in the catalog, and the predictors in associated models
- The F-tests associated with these nicely balanced designs

Traditional methods

Benefits:

- Communication –names do help
- Organizing one's thoughts –a catalog may help
- The F-test and the likelihood ratio test are often similarly powerful, but the F-test can be more powerful

Overall comparison of the F-test and LRT

- F-test restricted in use, LRT always available (ex: logistic)
- On fixed-effect models: the F-test is exact, the χ^2_k-based LRT is approximate.
- On mixed-effect models: both tests are approximate.
  - The F-test assumes that all variance components are known for sure,
  - The χ^2_k-based LRT might be so conservative or so anti-conservative that we need a parametric-bootstrap based p-value.
- With unbalanced data, it is unclear what F-distribution the F-value should be compared with. Use approximate denominator df.
- The F-test can be more powerful.

Outline

1. The traditional approach
2. The Mean Squares approach for the Completely randomized design (CRD)
   - CRD and one-way ANOVA
   - Variance components and the F test
   - Inference about the intercept
   - Sample vs. subsample size determination
3. The Mean Squares approach for CRD with subsampling
   - Example and Model
   - ANOVA table
   - F tests
   - Pairwise treatment comparisons with LSD
   - Sample and subsample size determination

Completely Randomized Design and one-way ANOVA

$$Y_i = \mu + \alpha_{j[i]} + e_i \quad \text{with } e_i \sim \mathcal{N}(0, \sigma_e^2) \text{ independent}$$

- **Balanced** design: each of the k treatments is sampled n times: $n_j = n$.
  Example: loblolly pine needles and stomata density.
  $k = 10$ needles, $n = 4$ rows from each needle.
  Example: corn yield, $k = 8$ sites, $n = 8$ plots from each site.
- **Fixed effects**: we will constrain the k deviations from the overall mean, α_j, to sum up to 0.
  In R, we will use the "sum" contrast. The default "treatment" contrast uses one reference level and constrains its $\alpha_{\text{ref } i}$ to be 0.
- **Random effects**: assume $\alpha_j \sim \mathcal{N}(0, \sigma_\alpha^2)$ and independent.

Analysis in R, default “treatment” contrast option

```
> corn = read.table("corn.txt", header=T)
> getOption("contrasts")
       unordered           ordered
"contr.treatment"      "contr.poly"

> lm(ears ~ site, data=corn)
(Intercept)     siteLFAN     siteNSAN     siteORAN     siteOVAN
      44.38        -1.13       -17.88        -1.13        -4.88
   siteTEAN     siteWEAN     siteWLAN
      -6.63         0.25         6.12

> with(corn, mean(ears[site=="DBAN"]))
[1] 44.38
>  44.38 + 6.12 # last site, WLAN
[1] 50.5
> with(corn, mean(ears[site=="WLAN"]))
[1] 50.5
```

Analysis in R, using the “sum” contrast option

```
> oldoptions = options(contrasts=c("contr.sum","contr.poly"))
> getOption("contrasts")
[1] "contr.sum"  "contr.poly"
> oldoptions$contrasts       # saved, in case we want to go back
       unordered          ordered
"contr.treatment"      "contr.poly"

> fit.lm = lm(ears ~ site, data=corn)   # re-fit same model
> coef(fit.lm)
(Intercept)        site1        site2        site3        site4
      41.22         3.16         2.03       -14.72         2.03
      site5        site6        site7
      -1.72        -3.47         3.41
> with(corn, mean(ears))  # overall mean
[1] 41.22

how much does the last site deviate from the overall mean? = alpha_8
> -sum( coef(fit.lm)[2:8])
[1] 9.28
> 41.22 + 9.28
[1] 50.5        # indeed, this is the mean of last site, WLAN
```

The F test

$$Y_i = \mu + \alpha_{j[i]} + e_i \quad \text{with } e_i \sim \mathcal{N}(0, \sigma_e^2) \text{ independent}$$

	Fixed Effects	Random Effects
Model	$\sum_{i=j}^{k} \alpha_i = 0$	$\alpha_j \sim$ iid $\mathcal{N}(0, \sigma_\alpha^2)$
F-Test	$H_0 : \alpha_j = 0$ for all j	$H_0 : \sigma_\alpha^2 = 0$

ANOVA table, with an extra column: expected Mean Squares.

Source	df	SS	MS	$\mathbb{E}$(MS) fixed	random
Trt	$k-1$	SSTrt	MSTrt	$\sigma_e^2 + \frac{n\sum_{i=1}^{k}\alpha_i^2}{k-1}$	$\sigma_e^2 + n\sigma_\alpha^2$
Error	$k(n-1)$	SSErr	MSErr	σ_e^2	σ_e^2
Total	$kn-1$	SSTot			

F test: uses the fact that $F = \frac{\text{MSTrt}}{\text{MSErr}} \sim F_{k-1,k(n-1)}$ if H_0 is true (no needle effect), in both models.

Variance component estimation with Mean Squares

Random effect model:

Source	df	SS	MS	$\mathbb{E}$(MS)
Trt	$k-1$	SSTrt	MSTrt	$\sigma_e^2 + n\sigma_\alpha^2$
Error	$k(n-1)$	SSErr	MSErr	σ_e^2
Total	$kn-1$	SSTot		

We can use E(MS) to estimate:

$$\hat{\sigma}_e^2 = \text{MSErr} \qquad \text{and } \hat{\sigma}_\alpha^2 = \frac{\text{MSTrt} - \text{MSErr}}{n}$$

- These estimates may differ from the ML or REML estimates.
- This $\hat{\sigma}_\alpha^2$ calculation may give a negative value...
  In this case, take $\hat{\sigma}_\alpha^2 = 0$.

Variance component estimation with Mean Squares

With a balanced design, the REML estimates of σ^2_e and σ^2_α and the MS-based estimates are equal.

```
> fit.lm = lm(ears ~ site, data=corn)
> anova(fit.lm)
          Df Sum Sq Mean Sq F value    Pr(>F)
site       7 2780.7  397.24  18.230 2.205e-12 ***
Residuals 56 1220.2   21.79

> ( 397.24 - 21.79 ) / 8  # n=8 plots per site
[1] 46.93125              # that is s^2_site, Mean Square estimation

# compare with REML:
> fit.lmer = lmer(ears ~ (1|site), data=corn)
> fit.lmer
...
Random effects:
 Groups   Name        Variance Std.Dev.
 site     (Intercept) 46.931   6.8506
 Residual             21.790   4.6680
Number of obs: 64, groups: site, 8
```

Inference about μ with Mean Squares

Assuming we have a balanced design $n_j = n$, the intercept estimate is the grand mean:

$$\hat{\mu} = \bar{y}_{\cdot\cdot} = \frac{\sum_i y_i}{nk} = \mu + \frac{\sum_j \alpha_j}{k} + \frac{\sum_i e_i}{nk}$$

so its variance is

$$\text{var}(\hat{\mu}) = \frac{\sigma_\alpha^2}{k} + \frac{\sigma_e^2}{nk} = \frac{n\sigma_\alpha^2 + \sigma_e^2}{nk}$$

Since $\mathbb{E}(\text{MSTrt}) = n\sigma_\alpha^2 + \sigma_e^2$, we estimate $\text{var}(\hat{\mu})$ by $\dfrac{\text{MSTrt}}{nk}$ and

$$\text{SE of } \hat{\mu} = \sqrt{\frac{\text{MSTrt}}{nk}} \text{ on } \text{df}_{\text{MSTrt}} = k - 1\,.$$

Confidence intervals obtained using the t-distribution, df$= k - 1$.

Inference about μ with Mean Squares

```
> anova(fit.lm)
          Df Sum Sq Mean Sq F value    Pr(>F)
site       7 2780.7  397.24  18.230 2.205e-12 ***
Residuals 56 1220.2   21.79

> sqrt(397.24 / (8*8))  # SE for mu: sqrt of MSTrt/nk
[1] 2.49136

> summary(fit.lmer)
Fixed effects:
            Estimate Std. Error t value
(Intercept)   41.219      2.491   16.55

# T-test for mu=0, in the random effect model:
> 2*pt(16.55, df=8-1, lower.tail=F)  # df = df_sites = #sites -1
[1] 7.17933e-07

# 95% confidence interval for mu:
> qt(.025, df=8-1, lower.tail=F) # df = df_sites = #sites -1
[1] 2.364624
> 41.219 + 2.365 * 2.491
[1] 47.11022                # upper bound
> 41.219 - 2.365 * 2.491
[1] 35.32778                # lower bound. CI: (35.33, 47.11) ears
```

Sample and subsample size

Goal: confidence interval for mean # ears
How to choose # of sites k and # plots k per site?

$$\text{Minimize SE of } \hat{\mu}, \text{ or } \text{var}(\hat{\mu}) = \frac{n\sigma_\alpha^2 + \sigma_e^2}{nk}$$

We will also want $\text{df}_{\text{site}} = k - 1$ not too small.

- In general, $\text{var}(\hat{\mu})$ goes down whenever k or n increases.
- If there is funding for a total of 16 measurements, what k and n should we choose? for a total of 32 measurements?

Sample and subsample size

Suppose $\sigma^2_\alpha = 47 > \sigma^2_e = 22$.

```
> se = function(n, k, sa2=47, se2=22){ (n*sa2+se2)/(n*k) }
> se(n=4, k=8)
[1] 6.6
> se(n=c(1,2,4,8,100), k=8)
[1] 8.6 7.2 6.6 6.2 5.9
```

	$k = 4$	$k = 8$	$k = 16$
n	$\text{var}(\hat{\mu}) = (n \times 4 + 1)/(nk)$		
1	17.2	8.6	4.3
2	14.5	7.2	3.6
4	13.1	6.6	3.3
8	12.4	6.2	3.1
100	11.8	5.9	3.0

When $\sigma^2_\alpha > \sigma^2_e$:

- k has a larger effect than subsample size n
- Better to have many sites, than many plots within each site.

Sample and subsample size

Now suppose $\sigma_\alpha^2 = 10$ (sites) $< \sigma_e^2 = 50$ (plots)

	$k = 5$	$k = 10$	$k = 20$
n	$\text{var}(\hat{\mu}) = (22n + 53)/(nk)$		
1	15.0	7.5	3.8
2	8.8	4.4	2.2
4	5.6	2.8	1.4
8	4.1	2.0	1.0
100	2.62	1.31	0.66

```
> f = function(n,k){ (n*10+50)/(n*k) }
> f(4,8)
[1] 2.8
> f( c(1,2,4,8,100), 8)
[1] 7.5 4.4 2.8 2.0 1.3
```

When $\sigma_\alpha^2 < \sigma_e^2$:

- Subsampling helps much more.
- Extremely large subsample size is not worthwhile.

Sample/subsample size determination for estimating μ:

- Relies on some preliminary estimates of σ_α^2 and σ_e^2.
- Consider costs in deciding sample and subsample sizes.

Outline

1. The traditional approach
2. The Mean Squares approach for the Completely randomized design (CRD)
   - CRD and one-way ANOVA
   - Variance components and the F test
   - Inference about the intercept
   - Sample vs. subsample size determination
3. The Mean Squares approach for CRD with subsampling
   - Example and Model
   - ANOVA table
   - F tests
   - Pairwise treatment comparisons with LSD
   - Sample and subsample size determination

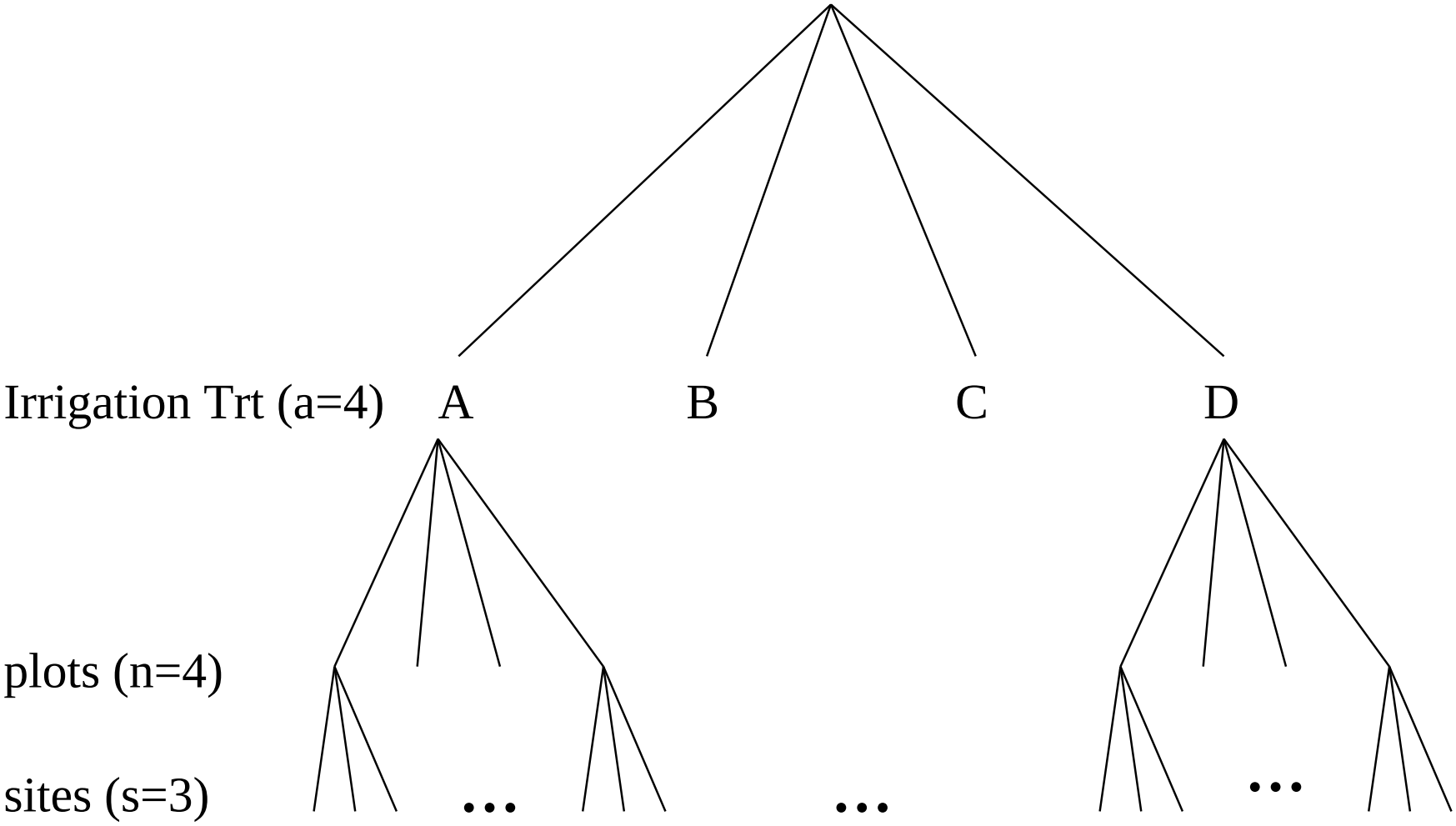
CRD with subsampling: Irrigation example

Compare the effect of $a = $ **4 irrigation methods** on soil moisture. CRD with $n = $ **4 plots** for each irrigation methods. Layout of plots and treatment allocation:

A	B	A	C
B	D	D	C
A	B	D	A
C	D	B	C

It is too costly to measure soil moisture for an entire plot. Select $s = $ **3 sites** at random within each plot and measure the soil moisture. The 3 measurements are used to estimate the soil moisture for the entire plot.

CRD with subsampling: Design



ANOVA table | aov output

CRD with subsampling: Model

response $\sim$ treatment + plot + error (=sites or subsamples)

$$Y_i = \mu + \alpha_{j[k[i]]} + \beta_{k[i]} + e_i$$

- α_j is a fixed treatment effect. Constraint: $\sum_{j=1}^{a} \alpha_j = 0$.
- $\beta_k \sim$ iid $\mathcal{N}(0, \sigma_\beta^2)$ error term for variation among samples, nested in treatment. β's bave to be treated as random.
- $e_i \sim$ iid $\mathcal{N}(0, \sigma_e^2)$ is an error term for variation among subsamples within a sample.

Soil moisture: $a = 4$ irrigation treatments, $n = 4$ plots per treatment, $s = 3$ sites per plot. Total: $ans = 4 * 4 * 3 = 48$ observations.

CRD with subsampling: ANOVA table

$$Y_i = \mu + \alpha_{j[k[i]]} + \beta_{k[i]} + e_i \ , \qquad \beta_k \sim \text{ iid } \mathcal{N}(0, \sigma^2_\beta), \ e_i \sim \text{ iid } \mathcal{N}(0, \sigma^2_e)$$

ANOVA table, with a balanced design:

Source	df	SS	MS	$\mathbb{E}$(MS)
Trt	$a-1$	SSTrt	MSTrt	$\sigma^2_e + s\sigma^2_\beta + \frac{ns\sum_{i=1}^{a}\alpha_i^2}{a-1}$
Plot Err	$a(n-1)$	SSPE	MSPE	$\sigma^2_e + s\sigma^2_\beta$
Subsamp. Err	$an(s-1)$	SSSSE	MSSSE	σ^2_e
Total	$ans-1$	SSTot		

- df's for trt and plot error are the same as those in CRD without subsampling.
- df for subsampling error is $an(s-1)$, because there are an plots and $(s-1)$ df per plot.

▸ design cartoon

▸ aov output

CRD with subsampling: ANOVA table

$$
\begin{aligned}
\text{SSTot} &= \sum_{i=1}^{ans} (y_i - \bar{y}_{...})^2 = \text{SSTrt} + \text{SSPE} + \text{SSSSE} \\
\text{SSSSE} &= \sum_{i=1}^{ans} (y_i - \bar{y}_{k[i]\cdot})^2 \\
\text{SSPE} &= \sum_{i=1}^{ans} (\bar{y}_{k[i]\cdot} - \bar{y}_{j[k[i]]\cdot\cdot})^2 = s \sum_{j=1}^{a} \sum_{k=1}^{n} (\bar{y}_{k\cdot} - \bar{y}_{j[k]\cdot\cdot})^2 \\
\text{SSTrt} &= \sum_{i=1}^{ans} (\bar{y}_{j[k[i]]\cdot\cdot} - \bar{y}_{...})^2 = ns \sum_{j=1}^{a} (\bar{y}_{j\cdot\cdot} - \bar{y}_{...})^2
\end{aligned}
$$

- SSTrt and SSPE similar to those in CRD without subsampling, except averaging is now also over subsamples.
- These calculations make sense because of the balance.

Testing treatment and plot effects with F tests

Variance estimates

$$\hat{\sigma}_e^2 = \text{MSSSE}, \qquad \hat{\sigma}_\beta^2 = \frac{\text{MSPE} - \text{MSSSE}}{s}$$

F test for treatment effect

To test $H_0 : \alpha_j = 0$ for all j, use

$$F = \frac{\text{MSTrt}}{\text{MSPE}} \quad \text{on df} = a - 1, a(n-1).$$

Beware! we should *not* use $\dfrac{\text{MSTrt}}{\text{MSSSE}}$.

F test for plot differences

To test $H_0 : \sigma_\beta^2 = 0$, use

$$F = \frac{\text{MSPE}}{\text{MSSSE}} \quad \text{on df} = a(n-1), an(s-1).$$

Irrigation data

```
> irrig = read.table("irrigation.txt", header=T)
> str(irrig)
'data.frame':   48 obs. of  4 variables:
 $ irrigation: int  1 1 1 1 1 1 1 1 1 1 ...
 $ plot      : int  1 1 1 2 2 2 3 3 3 4 ...
 $ subplot   : int  1 2 3 1 2 3 1 2 3 1 ...
 $ moisture  : num  12.6 11.9 12.3 13 12.4 12.4 11.3 11.9 10.9 12.5 ..

# Change 'irrigation' treatment and 'plot' to factors:

> irrig$irrigation = factor(irrig$irrigation)
> irrig$plot       = factor(irrig$plot)
```

F test for irrigation with `lmer`

```
> library(lme4)


> fit.lmer = lmer(moisture ~ irrigation + (1|plot), data=irrig)
> summary(fit.lmer)
...
Random effects:
 Groups   Name        Variance Std.Dev.
 plot     (Intercept) 0.44023  0.66350
 Residual             0.12812  0.35795
Number of obs: 48, groups: plot, 16

> anova(fit.lmer)
Analysis of Variance Table
           Df Sum Sq Mean Sq F value
irrigation  3 2.5990 0.86635  6.7617
```

With `lmer`, the denominator df is not available. To get the p-value: we know df's should be #irrigation-1 and #irrigation(#plots per irrig. $-$ 1):

```
> pf(6.7617, df1=3, df2=4*(4-1), lower.tail=F)
[1] 0.006379641
```

Testing irrigation effects with aov

- `aov(formula, data)` is another function to fit models with random effects. Output geared toward ANOVA tables.
- Limitations: needs perfectly balanced data.
  If several random effects, these need to be nested.
- Use `Error()` in aov to indicate random effects

```
> fit.aovR = aov(moisture ~ irrigation + Error(plot), data=irrig)
> summary(fit.aovR)
Error: plot
           Df  Sum Sq Mean Sq F value   Pr(>F)
irrigation  3 29.4073  9.8024  6.7658 0.006366 **
Residuals  12 17.3858  1.4488                    # this is  'plot Error'

Error: Within
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 32 4.1000  0.1281
```

▸ design cartoon ▸ ANOVA table

Testing irrigation effects with aov

- MS-based variance estimates: $\hat{\sigma}_e^2 = 0.128$ between subplots, $\hat{\sigma}_\beta^2 = (1.4488 + 0.1281)/3 = 0.44$ between plots. Same as REML-based estimates from `lmer`.
- Overall ANOVA table:

Source	df	SS	MS
Trt	3	29.4073	9.8024
Plot Err	12	17.3858	1.4488
SubSamp Err	32	4.1000	0.1281
Total	47	50.8931	

- Testing irrigation: $H_0 : \alpha_j = 0$ for all j vs. H_a : not all $\alpha_j = 0$.

$$f = \frac{9.8024}{1.4488} = 6.77 \quad \text{on df} = 3, 12.$$

- We could also test plot effects: $H_0 : \sigma_\beta^2 = 0$ vs. $H_0 : \sigma_\beta^2 > 0$

$$f = \frac{1.4488}{0.1281} = 11.308 \quad \text{on df} = 12, 32.$$

$p = 2.10^{-08}$ (with `pf`): significant variability between plots.

Compare with LRTest of irrigation effects

```
> fit.lmer0 = update(fit.lmer, .~.-irrigation)
> anova(fit.lmer, fit.lmer0)

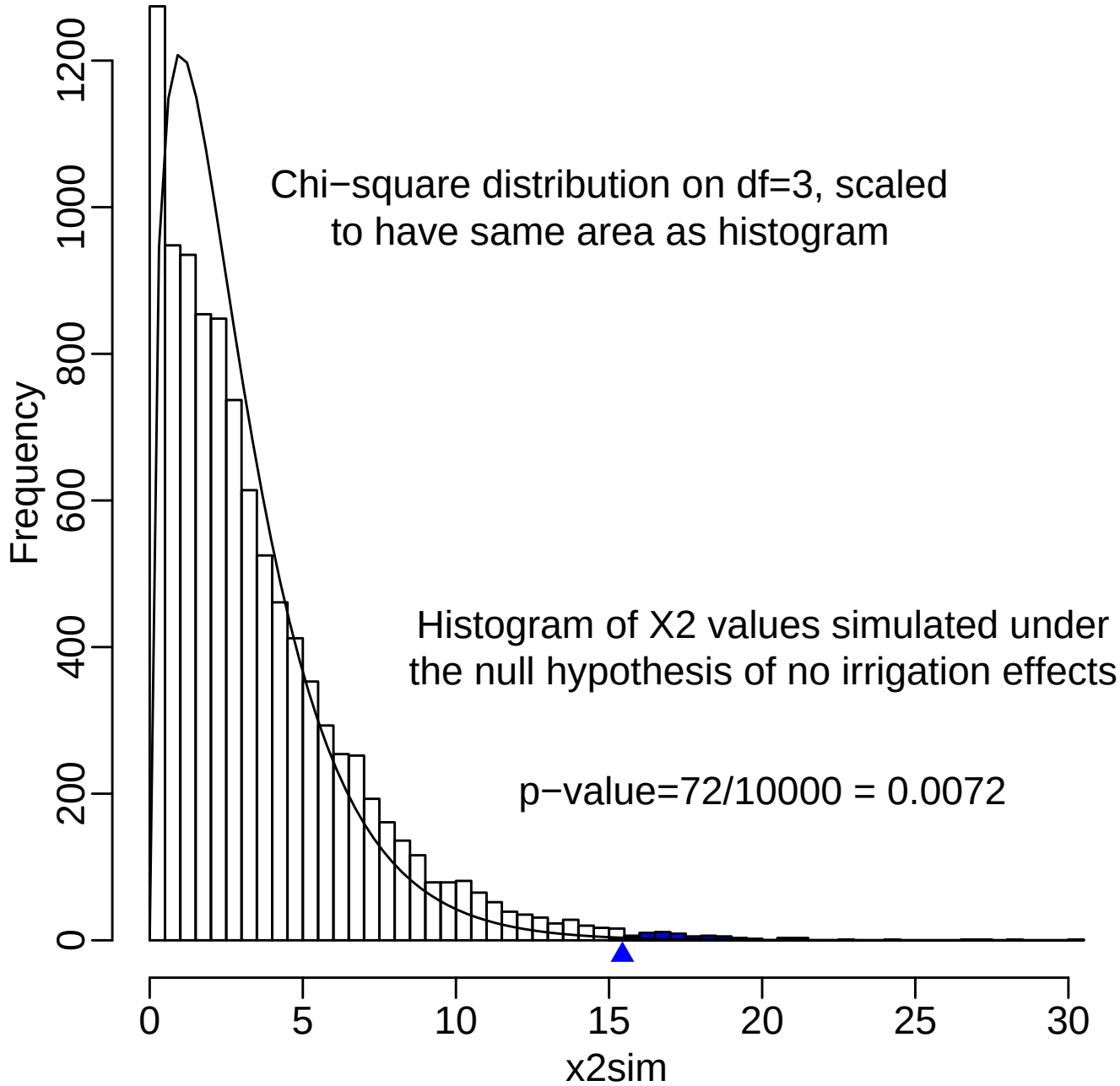
          Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fit.lmer0  3  93.659  99.272 -43.829
fit.lmer   6  84.221  95.449 -36.111 15.437      3   0.001479 **
```

The p-value from comparing $X^2 = 15.437$ to χ^2_3 distribution can be **anti-conservative** (too small). Simulation approach:

```
oneX2 = function(){
 y = simulate(fit.lmer0)
 f0 = lmer(y~(1|plot), data=irrig)
 f1 = lmer(y~irrigation+(1|plot), data=irrig)
 anova(f1,f0)$Chisq[2]
}
oneX2() # just to check
x2sim = replicate(10000, oneX2()) # takes a little while!
sum(x2sim>15.437) / 10000  # got 0.0072
```

LRT p-value $= 0.0072$. Was 0.0064 with the f-test: very close.

Testing irrigation effects with `lmer` and LR test



Danger: of using fixed plot effects

```
> fit.lm = lm(moisture ~ irrigation + plot, data=irrig)
> anova(fit.lm)
Analysis of Variance Table

Response: moisture
           Df Sum Sq Mean Sq F value    Pr(>F)
irrigation  3 29.407  9.8024  76.507 1.097e-14 ***
plot       12 17.386  1.4488  11.308 2.342e-08 ***
Residuals  32  4.100  0.1281
```

Wrong irrigation p-value here!! because anova makes sequential tests: compares a model with just an intercept to a model with just irrigation effects –ignoring plot effects.

df and MS values are correct in this anova table, but they are wrongly utilized: wrong F values, wrong p-values.

How to diagnose the problem?

```
> summary(fit.lm)
Coefficients: (3 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.2667     0.2067  59.357  < 2e-16 ***
irrigation2   0.9667     0.2923   3.308 0.002332 **
irrigation3  -1.2333     0.2923  -4.220 0.000188 ***
irrigation4   0.9667     0.2923   3.308 0.002332 **
plot2         0.3333     0.2923   1.141 0.262529
plot3        -0.9000     0.2923  -3.079 0.004237 **
plot4        -0.2000     0.2923  -0.684 0.498700
plot5        -0.8667     0.2923  -2.965 0.005673 **
plot6        -1.3333     0.2923  -4.562 7.06e-05 ***
plot7         0.5333     0.2923   1.825 0.077372 .
plot8             NA         NA      NA       NA
plot9         0.8667     0.2923   2.965 0.005673 **
plot10        0.3667     0.2923   1.255 0.218716
plot11       -0.9667     0.2923  -3.308 0.002332 **
plot12            NA         NA      NA       NA
plot13       -0.6000     0.2923  -2.053 0.048328 *
plot14        0.7000     0.2923   2.395 0.022640 *
plot15       -0.5000     0.2923  -1.711 0.096798 .
plot16            NA         NA      NA       NA
```

Plots are nested in irrigation methods: only 16 plot means are observed, but trying to estimate $1 + 3 + 15 = 19$ coefficients. 3 coefficients cannot be estimated.

How to diagnose the problem?

Which 3 coefficients are missing depend on the order in which predictors are provided:

```
> fit.lm =  lm(moisture ~ plot + irrigation, data=irrig)
> summary(fit.lm)
Coefficients: (3 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.2667     0.2067  59.357  < 2e-16 ***
plot2         0.3333     0.2923   1.141 0.262529
plot3        -0.9000     0.2923  -3.079 0.004237 **
plot4        -0.2000     0.2923  -0.684 0.498700
plot5         0.1000     0.2923   0.342 0.734468
plot6        -0.3667     0.2923  -1.255 0.218716
plot7         1.5000     0.2923   5.132 1.35e-05 ***
plot8         0.9667     0.2923   3.308 0.002332 **
plot9        -0.3667     0.2923  -1.255 0.218716
plot10       -0.8667     0.2923  -2.965 0.005673 **
plot11       -2.2000     0.2923  -7.528 1.43e-08 ***
plot12       -1.2333     0.2923  -4.220 0.000188 ***
plot13        0.3667     0.2923   1.255 0.218716
plot14        1.6667     0.2923   5.703 2.57e-06 ***
plot15        0.4667     0.2923   1.597 0.120152
plot16        0.9667     0.2923   3.308 0.002332 **
irrigation2       NA         NA      NA       NA
irrigation3       NA         NA      NA       NA
irrigation4       NA         NA      NA       NA
```

How to diagnose the problem?

Another diagnostic: use `drop1`. Does non-sequential f-tests:

```
> drop1(fit.lm)
Single term deletions

Model:
moisture ~ irrigation + plot
           Df Sum of Sq     RSS     AIC F value     Pr(F)
<none>                    4.100 -86.090
irrigation  0     0.000   4.100 -86.090
plot       12    17.386  21.486 -30.583  11.308 2.342e-08 ***
```

- Irrigation effects are absorbed into plot effects when irrigation is removed, because plot effects are fixed.
- At least, the output of `drop1` is correct: with fixed plot effects, the models with or without irrigation have the exact same df: Identical models with same RSS, same AIC.
- Correct f-test for plot effects.

Testing irrigation effects - in practice

We must use random plot effects to test treatment effects. In R:

- Use `aov` with plots as random effect: `Error(plot)`
  Only works with balanced data.
- Use `lmer` and get f value with `anova` but obtain df ourselves.
- Use `lmer` and a likelihood ratio test.

Simpler ways?

- We **cannot** use the following ANOVA and f-test for $a = 4$ trts and $n = 12$ samples per trt, from:

  `lm( moisture ~ irrigation )`

Source	df
Trt	3
Plot Err	44
Total	47

  This analysis is wrong because it assumes that the design is CRD without subsampling. Pseudoreplication.

- Alternatively: average the 3 subsamples to get 1 observation per plot. The usual one-way ANOVA using averaged plot values gives the **same** f-value and p-value for testing irrigation.

Pairwise treatment comparisons: LSD

1. Average over subsamples: one-way ANOVA with trt df = 3 and error df = 12. Recall LSD was $t_{\alpha/2,k(n-1)} * \sqrt{\frac{2\text{MSErr}}{n}}$.
2. Without averaging over subsamples:

$$\text{var}(\bar{Y}_{j..} - \bar{Y}_{j'..}) = 2\,\text{var}(\bar{Y}_{i..}) = 2\,\text{var}(\bar{\beta}_{k.} + \bar{e}_{i..}) = \frac{2}{ns}(s\sigma^2_\beta + \sigma^2_e)$$

which we estimate by $\frac{2\text{MSPE}}{ns}$ on df PE. Now:

Testing treatment differences: after protection by f-test

$$\text{var}(\hat{\alpha}_j - \hat{\alpha}_{j'}) = \frac{2\text{MSPE}}{ns} \text{ on df} = \text{dfPE} = a(n-1)$$

can be used to perform a t-test. Alternatively, use

$$\text{LSD} = t_{\alpha/2,a(n-1)} \times \sqrt{\frac{2\text{MSPE}}{ns}}$$

Pairwise comparisons of irrigation methods, aov

SE of differences between irrigation effects $\sqrt{\frac{\text{MSPE}}{ns}} = 0.491$:

```
> summary(fit.aovR)
Error: plot
           Df Sum Sq Mean Sq F value   Pr(>F)
irrigation  3 29.407  9.8024  6.7658 0.006366 **
Residuals  12 17.386  1.4488

Error: Within
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 32    4.1 0.12812

> sqrt(2*1.4488/(4*3))  # 4*3 = #plots * #subsamples
[1] 0.4913926           # SE for comparing irrigation tmts

> qt(0.025, df=12, lower.tail=F)
[1] 2.178813            # df_plot error: 12=#irrigation(#plots -1)

> lsd = 2.178813 * 0.4913926
> lsd
[1] 1.070653
```

Pairwise comparisons of irrigation methods, aov

Compare differences between irrigation means to LSD= 1.07 and using `model.tables` with our aov fit:

```
> model.tables(fit.aovR, type="means")
Tables of means
Grand mean
12.28125

 irrigation
irrigation
     1      2      3      4
12.075 12.817 11.100 13.133

> 12.817 - 12.075
[1] 0.742
# smaller than LSD: methods 1 and 2 are not sign. different
> 2*pt(0.742 / 0.491, df=12, lower.tail=F)
[1] 0.1566148     # p-value for comparing methods 1 and 2

> 13.133 - 11.100
[1] 2.033         # > LSD: methods 3 and 4 are different
> 2*pt(2.033 / 0.491, df=12, lower.tail=F)
[1] 0.001369526   # p-value for comparing methods 3 and 4
```

Pairwise comparisons of irrigation methods, aov

Use `model.tables` with option `type="effects"`:

```
> model.tables(fit.aovR,"effects", se=T)
Tables of effects
 irrigation
      1       2       3       4
-0.2062  0.5354 -1.1813  0.8521

Standard errors of effects
        irrigation
            0.3475
replic.         12
```

Beware: 0.3475 is the SE of one effect, not the SE of the difference between 2 effects.

```
> sqrt(2)*0.3475 # this is SE of estimated differences
[1] 0.4914392    # same as before. Pfew!

> 0.5354 - (-0.2062)
[1] 0.7416          # difference between methods 1 and 2: <LSD
> 0.8521 - (-1.181)
[1] 2.0331          # difference between methods 3 and 4: >LSD
```

Pairwise comparisons of irrigation methods, `lmer`

```
> fit.lmer  # SE for comparing 2 methods is readily available:
Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.0750     0.3475   34.75
irrigation2   0.7417     0.4914    1.51
irrigation3  -0.9750     0.4914   -1.98
irrigation4   1.0583     0.4914    2.15
```

We know the df of the appropriate t distribution is $a(n-1) = 12$.

```
> qt(0.025, df=12, lower.tail=F)
[1] 2.178813
> 2.178813 * 0.4913926
[1] 1.070653                # this is LSD
```

- No evidence (at $\alpha = .05$) that method 1 is different from any of the other methods.
- How about other differences? Method 3 versus method 4?

```
> 1.0583 - (-0.9750)
[1] 2.0333              # >LSD: significant
> (1.0583 - (-0.9750)) / 0.4914
[1] 4.13777             # t-value associated with LSD method
> 2*pt(4.13777, df=12, lower.tail=F)
[1] 0.001376287         # p-value associated with LSD method
```

Impact of sample and subsample size on the f test

$$F = \frac{\text{MSTrt}}{\text{MSPE}} \quad \text{and} \quad \frac{\mathbb{E}(\text{MSTrt})}{\mathbb{E}(\text{MSPE})} = \frac{\sigma_e^2 + s\sigma_\beta^2 + \frac{ns\sum_{i=1}^{a}\alpha_j^2}{a-1}}{\sigma_e^2 + s\sigma_\beta^2}$$

We want this ratio as large as possible, for more power to detect treatment effects. The power of the f-test depends on this ratio, which is also:

$$\frac{\mathbb{E}(\text{MSTrt})}{\mathbb{E}(\text{MSPE})} = 1 + \frac{\frac{ns\sum_{i=1}^{a}\alpha_j^2}{a-1}}{\sigma_e^2 + s\sigma_\beta^2} = 1 + \frac{\frac{\sum_{i=1}^{a}\alpha_j^2}{a-1}}{\sigma_e^2/(ns) + \sigma_\beta^2/n}$$

The larger, the better. More power with:

- smaller variances σ_e^2 or σ_β^2
- larger n or larger s,
- larger actual true differences between treatments ($\frac{\sum_{i=1}^{a}\alpha_j^2}{a-1} \approx \hat{\sigma}_\alpha^2$).

Impact of sample and subsample size on the f test

- If preliminary data are available, we can determine sample sizes n and s based on the Mean Square ratio.
- Suppose σ_e^2 (within) $< \sigma_\beta^2$ (between plots). Then subsampling does little.

Example:
$\sigma_e^2 = 1$,
$\sigma_\beta^2 = 4$,
$\frac{\sum_{j=1}^{a} \alpha_j^2}{a-1} = 1$.

	$n = 5$	$n = 10$	$n = 20$
s	$\mathbb{E}(\text{MS})$ ratio $= \frac{1+4s+ns}{1+4s}$		
1	2	3	5
2	2.11	3.22	5.44
4	2.18	3.35	5.70
100	2.25	3.49	5.99

- If $\sigma_e^2 \geq \sigma_\beta^2$, then subsampling is useful. Preliminary estimates of σ_e^2 and σ_β^2 are needed to determine (sub)sample sizes.