

Additional sources of agricultural data

Kevin Wright

2017-11-29

Books

Die Landwirtschaftlichen Versuchs-Stationen

<http://catalog.hathitrust.org/Record/000549685>

Full view of research station reports 1859-1920. In German.

S. J. Welham et al. (2015). *Statistical Methods In Biology*.

The online-supplements contain many small datasets for the examples and exercises.

D. F. Andrews and A. M. Herzberg (1985). *Data*.

<http://www.stat.duke.edu/courses/Spring01/sta114/data/andrews.html>

Table 2.1: `agridat::darwin.maize`

Table 5.1: `agridat::broadbalk.wheat`

Table 6.1: `agridat::mercer.wheat.uniformity`

Table 6.2: `agridat::wiebe.wheat.uniformity`

Table 58.1: `agridat::caribbean.maize`

D. Bayisa (2010). *Application of Spatial Mixed Model in Agricultural Field Experiment*.

Master thesis. Department of Statistics, Addis Ababa University. At least one dataset from wheat, RCB, with field coordinates. See also http://www.ethstat.org.et/esa_publications.html, Volume 19

M. N. Das & Narayan C. Giri (1987). *Design and Analysis of Experiments*.

31 Wool from 24 ewes, 6 cuttings

116 NPK factorial, 3 years

116 2^5 factorial, 1 rep

117 2^3 factorial, 3 rep

117 3^3 factorial

149 NPK split-plot, xy

150 nitro-variety split-plot

201 variety balanced lattice

279 covariate, yield & plant count

Peter Diggle, Patrick Heagerty, Kung-Yee Liang, Scott Zeger. *Analysis of Longitudinal Data.*

<http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/datasets.html>

Pig weight data is found in SemiPar::pig.weights

Sitka spruce data is found in: geepack::spruce

Milk protein data is found in: nlme::Milk. A thorough description of this data can be found in Molenberghs & Kenward, /Missing Data in Clinical Studies/, p. 377. Original source: A. P. Verbyla and B. R. Cullis, Modelling in Repeated Measures Experiments. <http://www.jstor.org/stable/2347384>

Federer, Walt (1955). *Experimental Design.*

192 3x3 factorial

204 3x2 factorial

236 2x2x2 factorial with confounding

257 2x3x2 factorial with confounding

276 split-plot with layout

285 nested multi-loc (Also problems page 22)

350 cubic lattice

420 balanced inc block

491 Latin square with covariate

Finney 1972. *An Introduction to Statistical Science in Agriculture.*

Not reviewed.

Galwey, N.W. (2014). *Introduction to Mixed Modelling*, 2nd ed.

<http://www.wiley.com/WileyCDA/WileyTitle/productCd-1119945496.html>

83 agridat::yates.oats

104 doubled-haploid barley

135 wheat/rye competition, heritability

190 chickpea flowering in families

250 canola oil gxe, sowing date, rainfall, oil Si & Walton

290 wheat anoxia, McDonald

291 wool fibers

370 alphalpha design (row-column inc block)

434 hollamby wheat trial, layout

Kwanchai A. Gomez & Gomez (1984). *Statistical Procedures for Agricultural Research.*

Extensive collection of datasets from rice experiments.

Cyril H. Goulden, *Methods of Statistical Analysis.*

First edition: <http://archive.org/details/methodsofstatist031744mbp>

18 Uniformity trial: agridat::goulden.barley.uniformity
153 Split-split plot with factorial sub-plot treatment: agridat::goulden.splitsplit
194 Incomplete block
197 Inc block
205 Latin square
208 Inc block 255 Covariates in feeding trial: agridat::crampton.pig

Second edition: <http://krishikosh.egranth.ac.in/handle/1/2034118>

216 Latin square: agridat::goulden.latin
423 Control chart with egg weights: agridat::goulden.eggs

Harry Love (1936). *Applications of Statistical Methods to Agricultural Research.*

379 MET 4 year, 2 field, 5 block, 5 gen

Kang, Manjit (2003). *Handbook of Formulas and Software for Plant Geneticists and Breeders*

Kuehl, Robert. *Design of Experiments*, 2nd ed.

357 alfalfa quadruple lattice
358 alpha design
488 split-plot sorghum hybrid,density
516 alfalfa rcb, two-year
521 crossover design cattle feedstuff

Erwin LeClerg, Warren Leonard, Andrew Clark (1962). *Field Plot Technique*

Many small datasets.

27 uniformity agridat::goulden.barley.uniformity
213 split-plot
234 immer multi-environment
260 lattice pinto-bean
276 triple lattice cotton
280 lattice sugar beet
289 balanced lattice
336 repeated wheat

Thomas M Little & F. Jackson Hills (1978). *Agricultural Experimentation.*

79 Latin square
89 Split-plot
103 Split-split
117 Split-block: agridat::little.splitblock
126 Repeated harvests

Roger Mead, Robert N. Curnow, Anne M. Hasted (2002). *Statistical Methods in Agriculture and Experimental Biology*, 3rd ed.

10 weekly milk yields
24 carrot weight
96 cabbage fertilizer
143 intercropping cowpea maize
177 honeybee repellent non-normal
251 cauliflower poisson agridat::mead.cauliflower
273 rhubarb RCB covariate
296 onion density
316 lambs
341 germination
350 germination factorial agridat::mead.germination
352 poppy
359 lamb loglinear agridat::mead.lambs
375 rats
386 intercrop
390 intercrop cowpea maize agridat::mead.cowpeamaize
404 apple characteristics (incomplete)

Roger Mead (1988). *The Design of Experiments*

<https://books.google.com/books?id=CaFZPbCllrMC&pg=PA323>

323 Turnip spacing data

Bernard Ostle (1963). *Statistics in Research*, 2nd ed.

<https://archive.org/details/secondeditionsta001000mbp>

455 2 factors, 1 covariate
458 1 factor, 2 covariates: agridat::crampton.pig

D. D. Paterson (1939). *Statistical Technique in Agricultural Research*.

<https://archive.org/details/statisticaltechn031729mbp>

84 Distribution of purple/white starchy/sweet seeds from 11 ears
190 Sugar cane MET: 2 year, 5 block, 5 variety
199 Tea MET: 3 year, 2² factorial fertilizer
206 Grass: 4 rep, 2 gen, 4 cutting treatments
211 Cotton: 4 dates, 3 spacings, 3 irrigation, 2 nitro. agridat::gregory.cotton

Arthur Asquith Rayner (1969). *A First Course In Biometry For Agriculture Students*.

19/456. 2x2x4 Factorial, 2 rep
19/466. 2x4 factorial, layout, plot size, kale (from Rothamsted)
19/466. 3x5 factorial, 3 rep, potato
20/494. 3x4 Split-plot with layout

21/505. 2x2x2 Factorial, 5 rep
21/515. 2x2x2x2 Factorial, 3 rep, with layout. (Evaluated, rejected as too variable)
22/537. 2x2x2 factorial, 6 rep, potato
22/537. 2x2x2x2 factorial, 2 rep, wheat, layout

G. W. Snedecor & W. G. Cochran. *Statistical Methods*.

Robert G. D. Steel & James Hiram Torrie. *Principles and Procedures of Statistics*, 2nd ed.

154 Mint plant growth, 2-way + pot + plant
244 Trivariate data
319 Regression with three predictors
384 Split-plot yield
387 Split-plot row spacing
400 Soybean 3 loc
423 Pig weight gain
429 Guinea pig weight gain
434 Soybean lodging

Oliver Schabenberger and Francis J. Pierce. *Contemporary Statistical Models for the Plant and Soil Sciences*.

Many datasets

Pesticides in the Nation's Streams and Ground Water, 1992-2001

Extensive data for detection of pesticides in water samples. See Appendix 5 and Appendix 6 of the supporting info. https://water.usgs.gov/nawqa/pnsp/pubs/circ1291/supporting_info.php

Journals - Data

Ag Data Commons

<https://data.nal.usda.gov/about-ag-data-commons>

CyVerse Data Commons

<http://datacommons.cyverse.org/>

http://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated

CyVerse: Genomes by environment dataset now publicly accessible

<http://www.cyverse.org/news/genomes-environment-dataset-now-publicly-accessible>

http://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Carolyn_Lawrence_Dill_G2F_Nov_2016

Very large GxE data here. Currently 2014, but more years planned. Hybrid & inbred phenotype data, weather data, genomic data. Very nice.

DataDryad

<http://datadryad.org/>

Data In Brief

<http://www.sciencedirect.com/science/journal/23523409>

Nature Scientific Data

<http://www.nature.com/sdata/>

Open Data Journal for Agricultural Research

<http://library.wur.nl/ojs/index.php/odjar/>

Plant Genomics and Phenomics Research Data Repository

<http://edal.ipk-gatersleben.de/repos/pgp/>

Journals - Bulletins

Iowa State Agricultural Research Bulletins

http://lib.dr.iastate.edu/ag_researchbulletins/

Vol 26/ 281. Cox: Analysis of Lattice and Triple Lattice.

Page 11: Lattice, 81 hybs, 4 reps

Page 24: Triple lattice, 81 hybs, 6 reps

Vol 29/347. Homeyer. Punched Card and Calculating Machine Methods for Analyzing Lattice Experiments Including Lattice Squares and the Cubic Lattice.

Page 37: Triple lattice (9 blocks * 9 hybrids) with 6 reps.

Page 60: Simple lattice, 8 blocks * 8 hybrids, 4 reps.

Page 76: Balanced lattice, 25 hybrids

Page 87: Lattice square with $(k+1)/2$ reps, 121 hybrids, 6 rep

Page 109: Lattice square with $k+1$ reps, 7 blocks * 7 hyb, 8 reps

Page 126: Cubic lattice, 16 blocks * 4 plots = 64 varieties, 9 reps, cotton

Vol 32/396. Wassom. Bromegrass Uniformity Trial: agridat::wassom.bromegrass.uniformity

Vol 33/424. Heady. Crop Response Surfaces and Economic Optima in Fertilizer agridat::heady.fertilizer

Vol 34/358. Schwab. Research on Irrigation of Corn and Soybeans At Conesville.

Page 257. 2 year, 2 loc, 4 rep, 2 nitro. Stand & yield.

Nice graph of soil moisture deficit (fig 9)

Vol. 34/463. Doll. Fertilizer Production Functions for Corn and Oats.

Table 1, 1954 Clarion Loam. N,P,K.
Table 14, 1955 McPaul Silt Loam. N,P.
Table 25, 1955 corn. K,P,N.
Table 31, 1956 oats, K,P,N. Trends difficult to establish.

Vol 34/472. Pesek. Production Surfaces and Economic Optima For Corn Yields.
Same data published in SSA journal?

Vol 34/488. Walker. Application of Game Theory Models to Decisions.

Vol 35/494. North Central Regional Potassium Studies with Alfalfa.
Page 176. Two years, several locs per state, multiple states, multiple fertilizer levels, multiple cuttings. Soil test attributes.
Page 183. Yield and %K.

Vol 35/503. North Central Regional Potassium Studies with Corn.

Papers

Barrero, Ivan D. et al. (2013). A multi-environment trial analysis shows slight grain yield improvement in Texas commercial maize. *Field Crops Research*, 149, Pages 167-176. <http://doi.org/10.1016/j.fcr.2013.04.017>

Note: This is a large (14500 records), multi-year, multi-location, 10-trait data. Sent a note encouraging the authors to formally publish the data. Source: <http://maizeandgenetics.tamu.edu/CTP/CTP.html>

Cleveland, M.A. and John M. Hickey, Selma Forni (2012). A Common Dataset for Genomic Analysis of Livestock Populations. *G3*, 2, 429-435. <http://doi.org/10.1534/g3.111.001453>

Note: The supplemental information for this paper contains data for 3534 pigs with high-density genotypes (50000 SNPs), and a pedigree including parents and grandparents of the animals.

Daillant-Spinnler (1996). Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. *Food Quality and Preference*, 7(2), 113-126. [http://dx.doi.org/10.1016/0950-3293\(95\)00043-7](http://dx.doi.org/10.1016/0950-3293(95)00043-7)

Note: The data are available in ClustVarLV package.

Gregory, Crowther & Lambert (1932). The interrelation of factors controlling the production of cotton under irrigation in the Sudan. *Jour Agric Sci*, 22, p. 617.

Hedrick (1920). Twenty years of fertilizers in an apple orchard. <https://books.google.com/books?hl=en&lr=&id=SqlJAAAAMAAJ&oi=fnd&pg=PA446> Note: No significant differences between fertilizer treatments.

Kayad et al. (2016). Assessing the Spatial Variability of Alfalfa Yield Using Satellite Imagery and Ground-Based Data. *PLOS One*. <https://doi.org/10.1371/journal.pone.0157166> Supporting information contains yield monitor data for 4 hay harvests on a center-pivot field.

Roger W. Hexem, Earl O. Heady, Metin Caglar (1974) A compendium of experimental data for corn, wheat, cotton and sugar beets grown at selected sites in the western United States and alternative production functions fitted to these data. Technical report: Center for Agricultural and Rural Development, Iowa State University. <https://babel.hathitrust.org/cgi/pt?id=wu.89031116783;view=1up;seq=3>

The technical report provides data from experiments on corn, wheat, cotton & sugar beets, each crop tested at several locations over two years, with a factorial structure on irrigation and nitrogen treatments, with replications. Three polynomial functions were fit to the data for each location (quadratic, square root, three-halves).

Kenward, Michael G. (1987). A Method for Comparing Profiles of Repeated Measurements. *Applied Statistics*, 36, 296-308.

Note: An ante-dependence model is fit to repeated measures of cattle weight. Data available here: https://faculty.biostat.ucla.edu/robweiss/filedepot_download/87/495

Klumper & Qaim (2015). A Meta-Analysis of the Impacts of Genetically Modified Crops. <http://doi.org/10.1371/journal.pone.0111629>

Note: Nice meta-analysis dataset. Published data only include differences, not standard-errors. See the comments on PLOS article for some peculiarities in the data.

Lado, B. et al. (2013). *Increased Genomic Prediction Accuracy in Wheat Breeding Through Spatial Adjustment of Field Trial Data*. *G3*, 3, 2105-2114. <http://doi.org/10.1534/g3.113.007807>

Note: Has a large haplotype data (83 MB) and two-year phenotype data with multiple traits.

Payne, Roger (2015). The Design and Analysis of Long-Term Rotation Experiments. *Agronomy Journal*, 107, 772-784. <http://doi.org/10.2134/agronj2012.0411>

Note: The data and R code appeared in the paper. Free access, but closed copyright.

Snedecor, George and E. S. Haber (1946). Statistical Methods For an Incomplete Experiment on a Perennial Crop. *Biometrics Bulletin*, 2, 61-67. <http://doi.org/10.2307/3001959>

Harvest of asparagus over 10 years, three cutting dates per year, 6 blocks.

Technow, Frank, et al. (2014). Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. August 1, 2014 vol. 197 no. 4 1343-1355. <http://doi.org/10.1534/genetics.114.165860>

Note: Genotype and phenotype data appears in the sommer package.

Tian, Ting (2015). Application of Multiple Imputation for Missing Values in Three-Way Three-Mode Multi-Environment Trial Data. <http://doi.org/10.1371/journal.pone.0144370>

Note: Uses agridat::australia.soybean data and one other real dataset with 4 traits that are not identified. All data and code available.

Randall J. Wisser et al. (2011). Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. *PNAS*. <http://doi.org/10.1073/pnas.1011739108>

Note: Supplement contains genotype data, but no phenotype data.

Yan, Weiwei (2002). Singular value partitioning in biplots. *Agron Journal*.

Note: Winter wheat, 31 gen in 8 loc. This data is different from Yan's earlier papers. Unfortunately, the data given in the paper are missing two rows.

Papers - Uniformity trials

Baker, G.A. (1953). Strawberry uniformity yield trials. *Biometrics*, Vol. 9, No. 3, pp. 412-421. <http://doi.org/10.2307/3001713>

Note: Two small uniformity trials with yield per plant for 200, 500 plants. CVs fairly high and correlation between neighboring plants very low.

Niels Anton Hansen, 1914. Prøvedyrkning paa Forsøgsstationen ved Aarslev. *Tidsskrift for landbrugets planteavl*. <http://dca.au.dk/publikationer/historiske/planteavl/> Bind 21, page 553. In Danish.

See comments by Eden & Maskell found here: <https://archive.org/stream/in.ernet.dli.2015.25895/2015.25895.Journal-Of-Agricultural-Science-Vol-xviii-1928#page/n175> Page 165: "One of the trials, comprising 35 plots, extended over the years 1907-11 and included oats, rye, barley, roots, barley." Oats and rye 1907 and 1908 show marked positive correlation. rye and barley 1908 and 1909 no correlation of significance. barley and

roots 1909 and 1910 no correlation. roots and barley 1910 and 1911 a significant correlation. rye and roots 1908 and 1910 satisfactory.

Edmund B. Lambert (1934). Size and arrangement of plots for yield tests with cultivated mushrooms. Journal of Agricultural Research, Vol 48, 971-980.

Note: Uniformity trial at three locations for mushrooms in growth houses. <https://naldc.nal.usda.gov/naldc/download.xhtml?id=IND43968493>

Moore, John Francis (1952). A study of field plot technique with sprouting broccoli. Proc Amer Soc Hort Sci: 1-474.

Note: This paper is not available, but a summary can be found here: <http://eurekamag.com/research/013/624/013624967.php>

R packages

aml - Adaptive Mixed LASSO

Has 'wheat' data with genetic and phenotypic data for wheat.

agricolae

Has assorted data and functions for analysis of agricultural data.

BGLR - Bayesian Generalized Linear Regression.

Has an A matrix (but no pedigree) for 499 genotypes at 4 locations.

BLR - Bayesian Linear Regression.

Has an A matrix (but no pedigree) for 499 genotypes at 4 locations.

BSagri

Safety assessment in agriculture trials

ClustVarLV

The apples_sh data has sensory attributes and preference scores for 12 apple varieties.

cropcc - Climate change on crops

<https://r-forge.r-project.org/projects/cropcc/>

drc - Dose response curves

Has nice herbicide dose response curves and germination data for mungbean, rice, wheat.

FW - Finlay-Wilkinson regression

<https://github.com/lian0090/FW/> Has phenotype data and marker data for 599 wheat lines in 4 environments.

lmtest

Data `lmtest::ChickEgg` is a time series of annual chicken and egg production in the United States 1930-1983.

missMDA

The referenced source for 'geno' data does NOT contain the data.

NADA

The 'Atra' and 'Recon' datasets contain measurements of Atrazine in water samples.

nlraa

Miguez. Non-linear models in agriculture. http://r-forge.r-project.org/R/?group_id=1599

nlme

Datasets for growth of Orange trees and Soybean plants. Soybean data is from the book "Nonlinear Models for Repeated Measurement Data".

plantbreeding

<https://r-forge.r-project.org/projects/plantbreeding/>

Data: `fulldial`

Data: `linetester`

Data: `peanut` same as `agridat::kang.peanut`

SDaA - Survey Data and Analysis

This package has county-level data from the United States Census of Agriculture, along with a vignette to illustrate survey sampling analyses.

SemiPar

Data: `SemiPar::onions` is same as `agridat::ratkowski.onions`

sommer - Solving mixed model equations in R

Data: h2. Modest-sized GxE experiment in potato Data: cornHybrid. Yield/PLTHT for 100 hybrids from 20 inbred * 20 inbred, 4 locs. Phenotype and relationship matrix.

Data: wheatLines CIMMYT wheat data for 599 lines. Phenotype and relationship data.

Data: RICE

Data: FDdata taken from agridat::bond.diallel

Data: Technow_data. AF=Additive Flint. AD=Additive Dent. MF=Marker Flint. MD=Marker Dent. pheno=phenotype data for 1254 hybrids (GY=yield, GM=moisture). This data is from Technow et al: <http://www.genetics.org/content/197/4/1343.supplemental>

spdep

Has a vignette ‘The Problem of Spatial Autocorrelation: forty years on’ that examines agriculture in Irish counties. See also the data ade4::irishdata.

spuRs

Data: spuRs::trees has data for 107 trees that were cut into cross sections with the volume calculated at roughly 10-year increments. This is a subset of the much-larger original data from Guttenberg: <https://archive.org/stream/wachstumundertra00gutt>

SoyNAM - Soybean nested association mapping

Dataset with phenotype data 3 yr, 9 locations, 18 environments, 60 thousand observations for height, maturity, lodging, moisture, protein, oil, fiber, seed size. There are 5000+ strains, 40 families.

st4gi - Stat for genetic improvement

<https://github.com/reyzaguirre/st4gi>

Web sites

ARS oat trials

<http://www.ars.usda.gov/Main/docs.htm?docid=8419&page=4>

Rothamsted Electronic Archive

<http://www.era.rothamsted.ac.uk/index.php> Data from Broadbalk and other long-term experiments.

Rothamsted Documents Archive

<http://www.era.rothamsted.ac.uk/eradoc/collections.php> Annual reports from Rothamsted 1908-1987. Many have data, especially in the early years (before WWII) there are data given for the 'Classical Experiments'.

Year, page

1908-1926

1926-1927 agridat::sawyer.multi.uniformity

1927-1928 agridat::sawyer.multi.uniformity

1929-1930

1931,143 agridat::yates.oats

1932

1933

1934,215-222 Sugar beet multi-environment trial with 3³ fertilizer

treatments at each site Roots, SugarPercent, SugarWeight, PlantNumber, Tops, Purity.

1935

1936,241 Similar to the 1934 experiment, but only gives the main effects,

not the actual data.

1937-1939

1946-1953

Germplasm Enhancement of Maize (ISU)

<http://www.public.iastate.edu/~usda-gem/index.htm> Data from trials at multiple locations. Yield, mst, plght, lodging. No field plans.

Google datasets search engine

<https://cse.google.com/cse/publicurl?cx=002720237717066476899:v2wv26idk7m>

Illinois Corn Hybrid Variety Trials

<http://vt.cropsci.illinois.edu/corn.html>

ILRI International Livestock Research Institute

<http://www.ilri.org/biometrics/CS/> Case study 4 is a nice dialel example with sheep data. Available as agridat::ilri.sheep

IRRI Biometrics and Breeding Informatics

<http://bbi.irri.org/products> STAR, PBTools, CropStat. The STAR user guide has well-documented data (even using 2 from agridat), but the PBTools user guide does not document the data.

SolCAP Solanaceae Coordinated Agricultural Project

<http://www.solcap.msu.edu/> Potato and Tomato genotype and phenotype data.

Statistical Analysis of Agricultural Experiments with R

<http://rstats4ag.org/> Datasets for mixed models, ancova, dose response curves, competition.

USDA National Agricultural Statistics Service

<http://www.nass.usda.gov> <http://quickstats.nass.usda.gov/> Group: Field Crops Commodity: Corn Category: Area Harvested, Yield Data Item: Corn grain Acres Harvested, Yield Bu/Ac Domain: Total Geography: State See agridat::nass.corn, nass.wheat, etc.