**Tugas 3 Analisis Data dan Visualisasi**

**Bagian 1: Analisis Variansi Satu Arah**

Masing-masing soal dikerjakan secara manual dan dengan program R.

Berikan jawaban yang lengkap dan sertakan program R dan output dalam satu file.

Dilarang copy paste jawaban dari kelompok lain. Pelanggaran ini mengakibatkan pengurangan skor nilai.

Lakukan pengambilan nomor soal secara acak dengan dua perintah berikut:

sample(x=1:5,size=2,replace=FALSE)

sample(x=6:10,size=2,replace=FALSE)

Kerjakan nomor soal yang muncul dari dua perintah tersebut. Pada masing-masing soal, jawab pertanyaan-pertanyaan berikut:

1. Buat plot rata-rata respons dan berikan interpretasi.
2. Tentukan hipotesis nol dan hipotesis alternatif dalam kalimat dan simbol.
3. Jika taraf signifikansi , maka tentukan kriteria keputusan yang bersesuaian.
4. Buat tabel ANOVA dan berikan kesimpulan.
5. Jika hipotesis pada one-way ANOVA ditolak, lakukan perbandingan berganda dengan prosedur Bonferroni. Berikan kesimpulan dan rekomendasi. Gunakan .
6. Jika hipotesis pada one-way ANOVA ditolak, lakukan perbandingan berganda dengan prosedur Tukey. Berikan kesimpulan dan rekomendasi. Gunakan .
7. Lakukan pengecekan asumsi normalitas dan homogenitas variansi.
8. Seorang peneliti menyelidiki banyaknya infeksi virus yang diderita orang sebagai fungsi dari level stres yang mereka alami selama periode 6 bulan. Data yang diperoleh adalah:

|  |  |  |  |
| --- | --- | --- | --- |
| Level | | | |
| Tidak stress | Stress minimal | Stress moderat | Stress berat |
| 2  1  4  1 | 4  3  2  3 | 6  5  7  5 | 5  7  8  4 |

1. Bunga Iris liar adalah bunga indah yang dapat ditemukan di Amerika Serikat, Kanada, dan Eropa Utara. Fokus pada studi ini adalah panjang sepal (bagian seperti daun yang menutupi bunga) dari berbagai jenis bunga iris liar. Data didasarkan pada informasi yang diambil dari artikel oleh R. A. Fisher dalam Annals of Eugenics (Vol. 7, bagian 2, hal. 179-188). Pengukuran panjang sepal dalam sentimeter dari sampel acak Iris setosa (I), Iris versicolor (II), dan Iris virginica (III) adalah sebagai berikut:

|  |  |  |
| --- | --- | --- |
| I | II | III |
| 5.4  4.9  5.0  5.4  4.4  5.8  5.7 | 5.5  6.5  6.3  4.9  5.2  6.7  5.5  6.1 | 6.3  5.8  4.9  7.2  5.7  6.4 |

1. Kuantitas oksigen terlarut adalah suatu ukuran pencemaran air di danau, sungai, dan sungai kecil. Sampel air diambil di empat lokasi berbeda di sungai dalam upaya untuk menentukan apakah polusi air bervariasi dari satu lokasi ke lokasi lainnya. Lokasi I adalah 500 meter di atas titik pembuangan air tanaman industri dan dekat pantai. Lokasi II adalah 200 meter di atas titik pembuangan dan di tengah sungai. Lokasi III adalah 50 meter hilir dari titik pembuangan dan dekat pantai. Lokasi IV adalah 200 meter di hilir dari titik pembuangan dan di tengah sungai. Kuantitas oksigen terlarut yang lebih rendah berarti lebih banyak polusi.

|  |  |  |  |
| --- | --- | --- | --- |
| Lokasi I | Lokasi II | Lokasi III | Lokasi IV |
| 7.3  6.9  7.5  6.8  6.2 | 6.6  7.1  7.7  8.0 | 4.2  5.9  4.9  5.1 | 4.4  5.1  6.2 |

1. Total natrium (dalam miligram) dalam satu porsi untuk sampel acak tiga jenis makanan yang berbeda disajikan dalam tabel berikut. Apakah perbedaan dalam total natrium rata-rata antara makanan berbumbu, sereal, dan makanan penutup.

|  |  |  |
| --- | --- | --- |
| Makanan berbumbu (Condiments) | Sereal (Cereals) | Makanan penutup (Desserts) |
| 270  130  230  180  80  70  200 | 260  220  290  290  200  320  140 | 100  180  250  250  300  360  300  160 |

1. Seorang peneliti ingin mencoba tiga teknik berbeda untuk menurunkan tekanan darah orang yang didiagnosis menderita tekanan darah tinggi. Subjek secara acak dibagi menjadi tiga kelompok; kelompok pertama minum obat, kelompok kedua berolahraga, dan kelompok ketiga mengikuti diet khusus. Setelah empat minggu, pengurangan tekanan darah setiap orang dicatat.

|  |  |  |
| --- | --- | --- |
| Obat | Berolahraga | Diet |
| 10  12  9  15  13 | 6  8  3  0  2 | 5  9  12  8  4 |

1. **Empty.stomach.csv.** Dalam publikasi “*How Often Do Fishes ‘Run on Empty’*?” (*Ecology*, Vol. 83, No 8, hlm. 2145–2151), D. Arrington dkk. memeriksa hampir 37.000 ikan dari 254 spesies di perairan Afrika, Amerika Selatan dan Tengah, dan Amerika Utara untuk menentukan persentase ikan dengan perut kosong. Ikan diklasifikasikan sebagai piscivores (pemakan ikan), invertivores (pemakan invertibrate), omnivora (pemakanan apa saja) dan algivora / detritivora (pemakan ganggang dan bahan organik lainnya). Proporsi masing-masing spesies ikan dengan perut kosong dicatat. Apakah data memberikan bukti yang cukup untuk menyimpulkan bahwa ada perbedaan dalam persentase rata-rata ikan dengan perut kosong di antara empat jenis ikan berbeda?
2. **Artificial.teeth.csv**. Dalam studi oleh J. Zeng et al., tiga bahan untuk membuat gigi tiruan — Endura, Duradent, dan Duracross — diuji untuk digunakan. Hasilnya dipublikasikan dengan judul paper “*In Vitro Wear Resistance of Three Types of Composite Resin Denture Teeth*” (*Journal of Prosthetic Dentistry*, Vol. 94, Edisi 5, hlm. 453–457). Dari mesin untuk mensimulasikan penggilingan oleh dua molar pertama kanan pada 60 stroke per menit dengan total 50.000 stroke, para peneliti mengukur volume material yang aus, dalam milimeter kubik. Enam pasang gigi diuji untuk setiap bahan. Apakah data memberikan bukti yang cukup untuk menyimpulkan bahwa ada perbedaan keausan rata-rata di antara ketiga bahan?
3. **Diet.populer.csv**. Dalam artikel "*Comparison of the Atkins, Ornish, Weight Watchers, and Zone Diets for Weight Loss and Heart Disease Risk Reduction*” (*Journal of the American Medical Association*, Vol. 293, No. 1, hlm. 43–53), M . Dansinger dkk telah melakukan uji coba secara acak untuk menilai efektivitas empat diet populer dalam menurunkan berat badan. Partisipan orang dewasa yang kelebihan berat badan dengan indeks massa tubuh rata-rata sebesar 35 dengan rentang usia 22-72 tahun diikutkan dalam uji coba acak selama 1 tahun. Penurunan berat badan, dalam kilogram dicatat. Kerugian bernilai negatif menyatakan pertambahan berat badan. WW = weight watchers.
4. **Prolonging.life.csv**. Vitamin C (askorbat) meningkatkan sistem kekebalan tubuh manusia dan efektif dalam mencegah berbagai penyakit. Dalam suatu studi oleh E. Cameron dan L. Pauling, yang hasil penelitiannya diterbitkan dengan judul artikel “*Supplemental Ascorbate in the Supportive Treatment of Cancer: Reevaluation of Prolongation of Survival Times in Terminal Human Cancer*” (*Proceedings of the National Academy of Science* USA, Vol. 75) , No. 9, hlm. 4538–4542), pasien kanker stadium lanjut diberi suplemen vitamin C. Pasien dikelompokkan sesuai dengan organ yang terkena kanker: lambung, bronkus, usus besar, ovarium, atau payudara. Studi ini menghasilkan data waktu bertahan hidup, dalam beberapa hari.
5. **Magazine.ads.csv**. Peneliti periklanan F. Shuptrine dan D. McVicker ingin menentukan apakah ada perbedaan yang signifikan dalam keterbacaan iklan majalah. Tiga puluh majalah diklasifikasikan berdasarkan tingkat pendidikannya — tinggi, sedang, atau rendah — dan kemudian tiga majalah dipilih secara acak dari setiap tingkat. Dari masing-masing majalah, enam iklan dipilih secara acak dan diperiksa untuk dibaca. Dalam kasus khusus ini, keterbacaan ditandai oleh jumlah kata, kalimat, dan kata dari tiga suku kata atau lebih di setiap iklan. Para peneliti menerbitkan temuan mereka dalam makalah “*Readability Levels of Magazine Ads*” (*Journal of Advertising Research*, Vol. 21, No. 5, hlm. 45–51). Data jumlah kata dari tiga suku kata atau lebih dalam setiap iklan dicatat.

**Bagian 2: Analisis Variansi Dua Arah**

Masing-masing soal dikerjakan secara manual dan dengan program R. Pengerjaan manual hanya pada perhitungan jumlah kuadrat, derajat bebas, kuadrat tengah, F hitung di tabel ANOVA.

Catatan bahwa soal nomor 1 tidak memerlukan perhitungan yang rumit, hanya melengkapi tabel.

Berikan jawaban yang lengkap dan sertakan program R dan output dalam satu file.

Dilarang copy paste jawaban dari kelompok lain. Pelanggaran ini mengakibatkan pengurangan skor nilai.

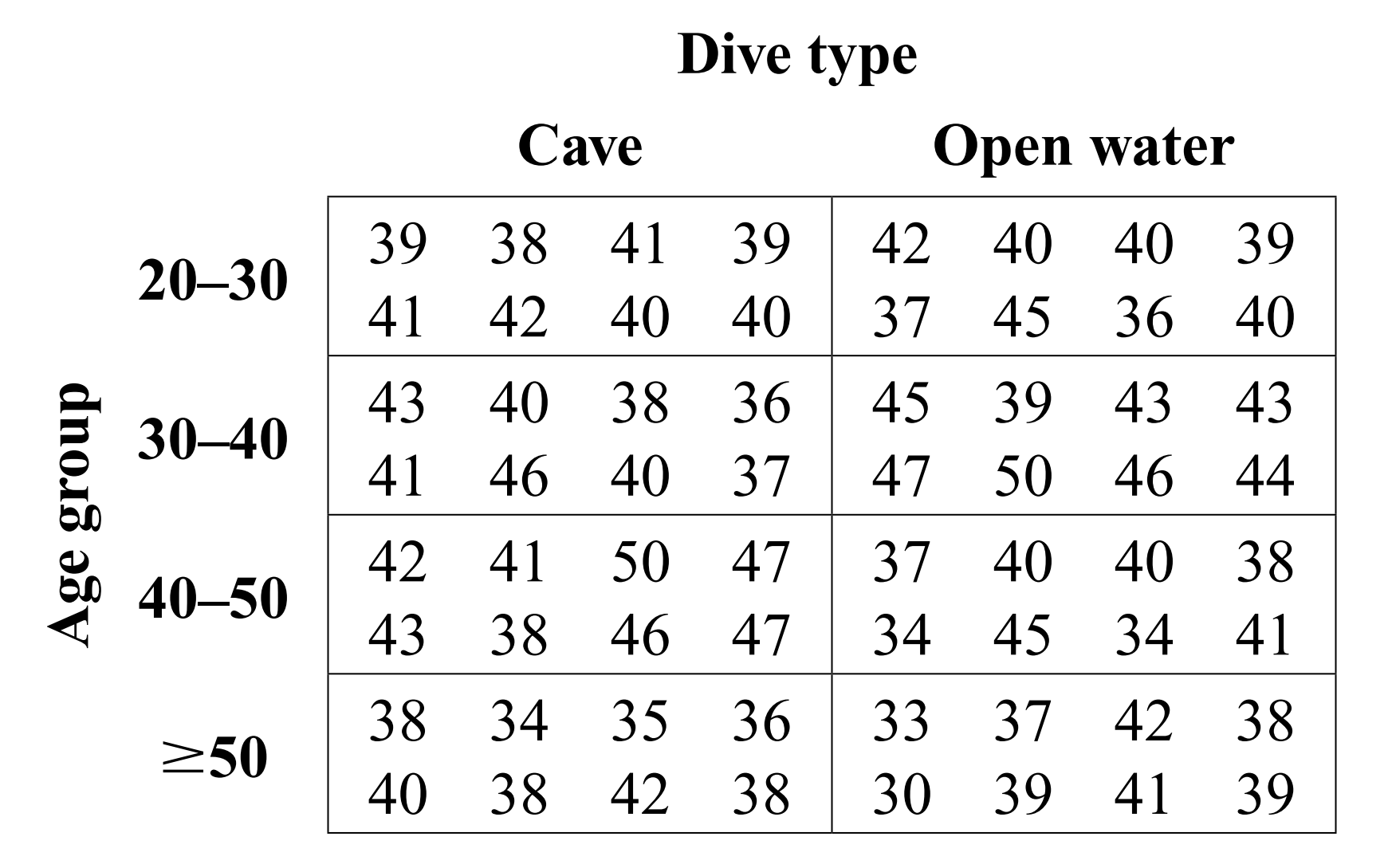
Kerjakan soal nomor 1, 2 dan dua nomor soal secara acak dari perintah berikut

sample(x=3:6,size=2,replace=FALSE)

1. Suatu agen kerja telah melakukan penelitian untuk menentukan pengaruh dari jenis pekerjaan dan jenis kelamin pada lama kerja. Sampel acak independen dari karyawan yang bekerja dalam layanan, teknologi, penjualan, keamanan, dan tenaga kerja diperoleh. Lama kerja (dalam minggu) dari setiap penugasan dicatat. Berikut adalah bagian dari tabel ringkasan ANOVA.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sumber variasi | Jumlah kuadrat | Derajat bebas | Kuadrat tengah | F | p-value |
| Jenis kelamin | 16.33 | 1 |  |  |  |
| Jenis pekerjaan | 184.39 |  |  |  |  |
| Interaksi |  | 4 |  |  |  |
| Galat | 202.42 | 50 |  |  |  |
| Total | 422.48 |  |  |  |  |

1. Lengkapi tabel ANOVA.
2. Berapa total banyaknya pengamatan?
3. Apakah ada bukti pengaruh interaksi? Lakukan pengujian hipotesis. Gunakan .
4. Apakah ada bukti pengaruh jenis kelamin atau jenis pekerjaan terhadap lama kerja? Lakukan pengujian hipotesis. Gunakan .
5. The Genesis Diving Institute of Florida menyatakan penyelam scuba menggunakan berbagai sistem pendidikan yang berbeda. Organisasi ini telah mempelajari waktu yang dihabiskan di bawah air oleh penyelam scuba dalam menjelajahi gua dan perairan terbuka. Setiap penyelam juga diklasifikasikan berdasarkan kelompok umur. Penyelaman mandiri dipilih secara acak untuk setiap kombinasi kelompok umur dan tipe penyelaman. Data (dalam menit) diberikan dalam tabel berikut.



Lakukan analisis variansi dua arah. Interpretasikan hasilnya. Gunakan untuk setiap pengujian hipotesis.

Kerjakan nomor soal yang muncul dari perintah R. Pada masing-masing soal, jawab pertanyaan-pertanyaan berikut:

1. Buat plot rata-rata respons dan berikan interpretasi.
2. Tentukan hipotesis nol dan hipotesis alternatif dalam kalimat dan simbol.
3. Jika taraf signifikansi , maka tentukan kriteria keputusan yang bersesuaian.
4. Buat tabel ANOVA dan berikan kesimpulan.
5. Jika hipotesis pada **pengaruh interaksi ditolak**, lakukan perbandingan berganda dengan prosedur Tukey. Berikan kesimpulan dan rekomendasi. Gunakan .
6. Jika hipotesis pada **pengaruh interaksi tidak ditolak**, lakukan perbandingan berganda dengan prosedur Tukey pada pengaruh faktor utama yang signifikan. Berikan kesimpulan dan rekomendasi. Gunakan .
7. Lakukan pengecekan asumsi normalitas dan homogenitas variansi.
8. **potato.dat.txt**. Dari dataset ini lakukan analisis variansi dua arah untuk mengetahui pengaruh antara area dan temperatur terhadap skor rasa kentang (*flavor*).

Dataset: potato.dat

Source: A. Mackey and J. Stockman (1958). "Cooking Quality of Oregon-Grown

Russet Potatoes", American Potato Journal, Vol.35, pp395-407

Description: Texture, Flavor, and Moistness Scores for potatoes

of 2 sizes, from two areas, two holding temps, 4 holding periods,

and 5 cooking methods. Each score is an average of 20 scores

(5 judges x 4 replications).

Variables/Columns

Growing Area 8 /\* 1=Southern OR, 2=Central OR \*/

Two week holding temp 16 /\* 1=75F 2=40F \*/

Size of Potato 24 /\* 1=Large, 2=Medium \*/

Storage Period 32 /\* 1=0 months, 2=2 mths 3=4mths, 4=6mths \*/

Cooking method 40 /\* 1=Boil, 2=Steam, 3=Mash, 4=Bake@350, 5=Bake@450 \*/

Texture score 46-48

Flavor score 54-56

Moistness score 62-64

> potato <- read.table("C:/data/potato.dat.txt",header=F,

col.names=c("area","temp","size","storage","cooking","texture","flavor","moistness"))

> potato$area <- factor(potato$area)

> levels(potato$area) <- c("Southern OR","Central OR")

> potato$temp <- factor(potato$temp)

> levels(potato$temp) <- c("75F","40F")

> potato$size <- factor(potato$size)

> levels(potato$size) <- c("Large","Medium")

> potato$storage <- factor(potato$storage)

> levels(potato$storage) <- c("0months","2months","4months","6months")

> potato$cooking <- factor(potato$cooking)

> levels(potato$cooking) <- c("Boil","Steam","Mash","Bake@350","Bake@450")

> str(potato)

'data.frame': 160 obs. of 8 variables:

$ area : Factor w/ 2 levels "Southern OR",..: 1 1 1 1 1 1 1 1 1 1 ...

$ temp : Factor w/ 2 levels "75F","40F": 1 1 1 1 1 1 1 1 1 1 ...

$ size : Factor w/ 2 levels "Large","Medium": 1 1 1 1 1 1 1 1 1 1 ...

$ storage : Factor w/ 4 levels "0months","2months",..: 1 1 1 1 1 2 2 2 2 2 ...

$ cooking : Factor w/ 5 levels "Boil","Steam",..: 1 2 3 4 5 1 2 3 4 5 ...

$ texture : num 2.9 2.3 2.5 2.1 1.9 1.8 2.6 3 2.2 2 ...

$ flavor : num 3.2 2.5 2.8 2.9 2.8 3 3.1 3 3.2 2.8 ...

$ moistness: num 3 2.6 2.8 2.4 2.2 1.7 2.4 2.9 2.5 1.9 ...

> head(potato,n=3)

area temp size storage cooking texture flavor moistness

1 Southern OR 75F Large 0months Boil 2.9 3.2 3.0

2 Southern OR 75F Large 0months Steam 2.3 2.5 2.6

3 Southern OR 75F Large 0months Mash 2.5 2.8 2.8

1. **apple1.dat.txt**. Dari dataset ini lakukan analisis variansi dua arah untuk mengetahui pengaruh antara varietas apel (**variety**) dan fusarium strain (**strain**) terhadap *rate of fungal invasion* (**rate**).

Dataset: apple1.dat.txt

Source: A.S. Horne, F.G. Gregory (1928). "A Quantitaive Study of Fungal

Invasion of the Apple Fruit, and Its Bearing on the Nature of Disease

Resistance. - Part II. The Application of the Statistical Method to Certain

Specific Problems", Proceedings of the Royal Society of London. Series B,

Containing Papers of a Biological Character, Vol.102,#719,pp.444-466

Description: Rate of fungal invasion of 5 varieties of apples for 7 fusarium

strains.

Varieties:

1=Bramley's Seedling, 1925-26 @12C. 70 Days

2=Bramley's Seedling, 1924-25 @12C. 103 Days

3=Cox's Orange Pippin, 1924-25 @12C. 54 Days

4=Cox's Orange Pippin, 1924-25 @3C. 138 Days

5=Cox's Orange Pippin, 1925-26 @12C. 89 Days

Fusarium Strains:

1=A

2=B11

3=B111

4=C1

5=C21

6=C3

7=D

Variables/Columns

Variety 8

Fusarium Strain 16

Days 22-24

Apple Weight (grams) 27-32

Radius (cm) 36-40

Fungal Radial Advance (cm) 44-48

Rate of advance (cm/day) 51-56

> apel <- read.table("C:/data/apple1.dat.txt",header=F,

col.names=c("variety","strain","days","weight","radius","fra","rate"))

> apel$variety <- factor(apel$variety)

> levels(apel$variety) <- c("Bramleys1","Bramleys2","Coxs1","Coxs2","Coxs3")

> apel$strain <- factor(apel$strain)

> levels(apel$strain) <- c("A","B11","B111","C1","C21","C3","D")

> apel$days <- factor(apel$days)

> str(apel)

'data.frame': 35 obs. of 7 variables:

$ variety: Factor w/ 5 levels "Bramleys1","Bramleys2",..: 1 1 1 1 1 1 1 2 2 2 ...

$ strain : Factor w/ 7 levels "A","B11","B111",..: 1 2 3 4 5 6 7 1 2 3 ...

$ days : Factor w/ 5 levels "54","70","89",..: 2 2 2 2 2 2 2 4 4 4 ...

$ weight : num 156 154 138 141 146 ...

$ radius : num 3.66 3.65 3.51 3.53 3.58 3.6 3.81 3.07 3.05 3.15 ...

$ fra : num 2.04 1.23 1.25 1.66 1.74 1.88 2.26 1.16 1.26 0.41 ...

$ rate : num 0.0291 0.0176 0.0179 0.0237 0.0249 0.0269 0.0323 0.0113 0.0122 0.004 ...

> head(apel,n=3)

variety strain days weight radius fra rate

1 Bramleys1 A 70 156.2 3.66 2.04 0.0291

2 Bramleys1 B11 70 153.9 3.65 1.23 0.0176

3 Bramleys1 B111 70 137.8 3.51 1.25 0.0179

1. **thaltb.dat.txt**. Dari dataset ini lakukan analisis variansi dua arah untuk mengetahui pengaruh antara perlakuan HIV tipe 1 dan M (**treatment**) dan status tuberculosis (**tb**) terhadap pertumbuhan berat badan (**weight**).

Dataset: thaltb.dat.txt

Source: J.D. Klausner, S.Makonkawkeyoon, P.Akarasewi, et al (1996). "The

Effect of Thalidomide on the Pathogenesis of Human Immunodeficiency

Virus Type 1 and M. Tuberculosis Infection", Journal of Acquired

Immune Deficiency Syndroms and Human Retrovirology. Vol. 11, pp247-257.

Description: Effect of Thalidomide versus Placebo on weight gain in

HIV+ subjects with and without TB.

Variables/Columns:

Treatment 7 /\* 1=Thalidomide, 0=Placebo \*/

TB status 14 /\* 1=Positive, 0=Negative \*/

Weight gain 18-21 /\* Kilograms over 21-day period \*/

> thaltb <- read.table("C:/data/thaltb.dat.txt",header=F,

col.names=c("treatment","tb","weight"))

> thaltb$treatment <- factor(thaltb$treatment)

> levels(thaltb$treatment) <- c("Placebo","Thalidomide")

> thaltb$tb <- factor(thaltb$tb)

> levels(thaltb$tb) <- c("Negative","Positive")

> str(thaltb)

'data.frame': 32 obs. of 3 variables:

$ treatment: Factor w/ 2 levels "Placebo","Thalidomide": 2 2 2 2 2 2 2 2 1 1 ...

$ tb : Factor w/ 2 levels "Negative","Positive": 2 2 2 2 2 2 2 2 2 2 ...

$ weight : num 9 6 4.5 2 2.5 3 1 1.5 2.5 3.5 ...

> head(thaltb,n=3)

treatment tb weight

1 Thalidomide Positive 9.0

2 Thalidomide Positive 6.0

3 Thalidomide Positive 4.5

1. ingots.dat.txt. Dari dataset eksperimen konstruksi plat baja, lakukan analisis variansi dua arah untuk mengetahui pengaruh pre-treatment dan metode slabbing terhadap *breaking strain*.

Dataset: ingots.dat

Source: Riley (1887). Some Investigations as to the Effects of Different Methods

of Treatment of Mild Steel in the Manufacture of Plates. Journal of the Iron and

Steel Institute, Vol. 30 #1, pp. 121-162

Description: Lengthwise breaking strain and contraction of area in plates that

were rolled from slabs made from ingots of varying types.

Variables/Columns

Pre treatment 8 (1=Reheated,0=Soaked)

Slabbing Method 16 (1=Hammered, 0=Cogged)

Slab Thickness 24 (4ò,8ò)

Plate Thickness 29-32 (1ò,0.5ò,0.25ò)

Annealing 40 (1=Yes,0=No)

Breaking Strain 45-48

Contraction of Area 53-56

> ingot <- read.table("C:/data/ingots.dat.txt",header=F,

col.names=c("pretrt","method","slabthick","platethick","annealing","breaking","constraction"))

> ingot$pretrt <- factor(ingot$pretrt)

> levels(ingot$pretrt) <- c("Soaked","Reheated")

> ingot$method <- factor(ingot$method)

> levels(ingot$method) <- c("Cogged","Hammered")

> ingot$slabthick <- factor(ingot$slabthick)

> ingot$platethick <- factor(ingot$platethick)

> ingot$annealing <- factor(ingot$annealing)

> levels(ingot$annealing) <- c("No","Yes")

> str(ingot)

'data.frame': 48 obs. of 7 variables:

$ pretrt : Factor w/ 2 levels "Soaked","Reheated": 2 2 2 2 1 1 1 1 2 2 ...

$ method : Factor w/ 2 levels "Cogged","Hammered": 2 2 2 2 2 2 2 2 1 1 ...

$ slabthick : Factor w/ 2 levels "4","8": 2 2 1 1 2 2 1 1 2 2 ...

$ platethick : Factor w/ 3 levels "0.25","0.5","1": 3 3 3 3 3 3 3 3 3 3 ...

$ annealing : Factor w/ 2 levels "No","Yes": 1 2 1 2 1 2 1 2 1 2 ...

$ breaking : num 28.4 26 28.2 27 28 26.5 27.1 27.2 27.5 26.1 ...

$ constraction: num 32.5 51.5 33 41.3 50.5 56 48.5 52 36.5 42 ...

> head(ingot,n=3)

pretrt method slabthick platethick annealing breaking constraction

1 Reheated Hammered 8 1 No 28.4 32.5

2 Reheated Hammered 8 1 Yes 26.0 51.5

1. Reheated Hammered 4 1 No 28.2 33.0