公衆衛生学実習テキスト（2025 年版）

中澤　港

2025 年 9 月 24 日（水）

このテキストは公衆衛生学実習のため作成したものである。実験デザインや統計解析についての詳細は、大学院保健学研究共通特講 IV/VIII のテキスト（https://minato.sip21c.org/ebhc/ebhc-text.pdf）を参照されたい。また、本実習についての情報や統計解析の練習用のサンプルファイルは、https://minato.sip21c.org/pubhealthpractice/ に適宜掲載するので参照されたい（本テキストの pdf ファイル自体その web サイトに掲載する）。

Contents

問い合わせ先：神戸大学大学院保健学研究科パブリックヘルス領域・教授　中澤　港
e-mail: minato-nakazawa@people.kobe-u.ac.jp

1　本実習の位置づけ

本実習は、カリキュラム上、前期の講義「環境・食品・産業衛生学」に対応する実習であるが、これまで伝統的に学生がグループ実習として自主的に実習課題を決めて実施し、最終発表を相互評価するという形で進められてきた。また、水、大気・環境問題、産業衛生についての課題調べ、発表、討論と吸光度測定と検量線作成といった、環境試料の定量分析に関する基本技術の習得を目指した実験・統計解析も実施する。

事前学習やグループ学習、提出物がかなり多いので、覚悟して臨まれたい。後述するように、グループ実習では、（予算の制約はあるが）かなり多様な課題が許容されるので、各自の自主性がこの実習を成功させるかどうかの鍵である。

2　実習の進め方

前半に課題別討論と基本実験を行う際は A、B、C の 3 班に分かれる。後半で自主的に決めた課題についてのグループ実習を行う際は各班をさらに 2 つずつに分ける。班分けと日程は BEEFPLUS に掲載する。基本実験が終了した次の回にデータを統計解析する方法を説明するが、その際、自分の PC に EZR という統計ソフトをインストールして持参すること。

2.1　実習室について

原則として対面で実施するが、感染症流行状況によっては変更もありうるので、BEEFPLUS を適宜参照すること。

初回、2 回目、6 回目（統計解析の説明）は F306 に集合すること。

大気の実験は主に屋外と E501 と E802 で行う。水の実験は E501 で行う。産業保健の実験は E503 で行う。グループ別実習は時間をずらすなどして密にならないように工夫し、F306、E802、E503、E501 で実施する。詳細は BEEFPLUS 参照。

2.2　担当教員

- 中澤　港 教授（E707、内線 4551、minato-nakazawa@people.kobe-u.ac.jp）
- 靱　千恵 准教授（B304、nu_chie@people.kobe-u.ac.jp）

2.3　これまでのグループ実習テーマの例

商品試験のようなテーマを選ぶ班も多いが、環境・食品・産業衛生学の講義で学んだことを生かせるようなテーマを選んで欲しい。

2024 年度　「生田川の流域の違いにおける水質調査」「涙活とストレスの関係」「食品が湿気にくい方法とは」「回転による細菌への影響」「おにぎりの適切な保存方法」「手洗い選手権」

2023 年度　「消費期限日を過ぎても食べられるのか？」「机を清潔にする方法」「マグロの時間経過と温度変化によるヒスタミン量の変化」「緊急時における飲料水の確保」「除菌方法の比較」「コンタクト洗浄選手権」

2022 年度　「日焼け止めが UV-A にどのくらい遮断効果があるか」「マスクの通気性」「六甲川および都賀川の水質調査」「調理器具と細菌」「日焼け止めの効果実験」「素材と UV カット機能の関係性」

2021 年度　「添加物による食品保存への影響」「素材・繊維による紫外線透過度の違い」「水の硬度と泡立ち」「パソコンの作業環境とストレス・集中力・視力の関係」「手指衛生と食品の安全性」「ハンカチと菌」

2020 年度　「一番長持ちするおにぎりの条件とは」「空気濃度測定から見る大気汚染」「マスク着用時の影響」「様々なアルコールによる手指消毒」「アルコール濃度の違いによる殺菌効果を比較する」「繰り返し使用可能なマスクと洗剤」

2019 年度　「菌と金」「材質によって紫外線の反射率は異なるのか」「ペットボトル飲料の細菌繁殖」「家庭における食器洗浄の除菌効果」「食器用スポンジの汚れについて」「メガネの汚れ」

2018 年度　「睡眠時間とストレス値、作業効率の変動」「音楽と騒音の境目」「人間と色」「手洗いの種類について」「物質の抗菌作用」「日傘の UV カット」

2017 年度　「MASK」「保湿効果、カネに頼るか？　知恵に頼むか？」「電磁波と共に〜スマートフォンに潜む危険〜」「ハンドクリームの保湿効果〜ハンドクリーム使用による水分と油分の関係について〜」「手洗いの効果」「髪は女の命！〜ヘアケアで髪質はどう変わる？〜」

2016 年度　「朝食を抜くと能力は落ちるか」「さまざまな方法の違いによる歯垢除去効果の比較」「中身と容器の材質によって飲料容器を放置した場合の細菌繁殖は変わるか」「周囲の音で暗記力と集中力は変わるか」「手を洗った後どうすれば手を清潔に保てるか」「日焼け止めクリームは本当に表示されているほど UV-A と UV-B をカットできるのか」

2015 年度　「スマホ vs トイレ汚れ」「ヒートテックの保温効果」「賞味期限と消費期限」「水の浄化」「コンタクトレンズの汚れ」「水素水の肌保湿効果」

2014 年度　「布団の除菌方法の有効性」「ファブリーズの消臭効果」「手の洗い方」「何度で菌はいなくなるか？」「外干しと部屋干し」「電磁波の測定」

2013 年度 「不快指数とは？」、「光の種類による消毒効果」、「川の水質調査」、「3 秒ルールの有効性」、「JINS PC の効果について」、「身の回りの細菌と食中毒」

2012 年度 「体温上昇」、「真の清潔とは？」、「匂いとストレスの関係」、「集中力の高め方」、「紫外線から肌を守れ!!〜UV カット繊維は万能 !?　乙女の敵紫外線〜」、「ドローイング法と EMS による腹部脂肪減少効果の比較」

2011 年度 「身近な電磁波の危険性」、「上から化粧水」、「放射能汚染の実態」、「食品の抗菌効果」、「油の酸化度」、「Mystery of Sleepiness」

2001 年度 「嗚呼電磁波」、「暖房と換気」、「細菌の分布」、「紫外線の恐怖」、「衣服内気候の変化」、「生活排水の汚染について」

2000 年度 「ペットボトルの細菌」、「うがい・手洗いの効果について」、「住宅と騒音」、「身の周りの照度」、「手洗いの効果について」、「換気の重要性について」

2.4　提出物と成績評価の方法

第 3 回から第 5 回は、3 限に 3 つのテーマについての討論会を行う。大気、水、産業保健に関わる問題（以下提示するテーマから各班内で分担）について自主的に文献調べを行い、A4 で 1 枚の資料をまとめて BEEFPLUS にアップロードすること。各人約 5 分を持ち時間として発表し、討論を行う。討論のモデレータは大学院生 TA (Teaching Assistant) が行い、各人について 5 段階で評価する。

討論会に引き続き、各テーマについての基本実験を行う（出席を取る）。実習室は広くないので、マスクを着用し、３密にならないよう換気に注意すること。また、2024 年度から、試薬を扱うときは保護メガネとディスポグローブを使用することになったので、忘れずに正しく装着すること。

基本実験結果についての個人レポートを（全部まとめて最終日までに）BEEFPLUS へのアップロードにより提出すること。これについても教員が５段階で評価する。

後半はグループ実習となる。実習内容によっては正規の実習時間ではできない場合もあるため（COVID-19 クラスタ発生予防のためには、実験室に人が密集することを避けねばならないため、時間をずらして実験室を使う必要がある）、出席等はとらないが、積極的に参加して欲しい。

最終日は、パワーポイントなどを用い、班ごとに成果発表を行う。各班の持ち時間は 25 分以内とする。プレゼンテーションファイル（各班員がどの部分に貢献したかを明示すること[1]）と配付資料ファイル（あれば）を発表会の前日までに中澤教授にメール送信（または、可能なら BEEFPLUS にアップロード）するこ

[1] 最近は学術論文でも共著者一人一人がその論文のどの部分に貢献したのかを明記する必要がある。

と。グループ実習についての評価は、(1) グループ実習への貢献度、(2) 配付資料と成果発表、討論について教員が評価（班単位の評価）する。

成績評価は以上を総合して行う。

2.5 基礎実験の個人レポート作成上の注意

基礎実験の個人レポートを作成する際、目的・方法については「実習テキストの通り」と書いて省略して差し支えない。ただし、何らかの事情でテキストと異なることをした場合は明記する。

生データは班内で共有するのでレポートに載せる必要は無い。そのデータをテキストに記載されている方法に従って統計解析を各自実行して得られた結果を求め、図や表の形にまとめ、その結果の意味を解釈し考察することが最も重要である。

考察に際しては、このテキストだけではなく、他の教科書や資料、できれば論文などを探して比較することを推奨する。その際、丸写しにせず適切に引用し、引用文献としてレポートに書誌情報を適切に記載すること（文献の書き方は `https://jipsti.jst.go.jp/sist/pdf/SIST_booklet2011.pdf` などが参考になる）。

なお、基礎実験の結果は班ごとに共通になるはずだが（ただし、計算結果をシェアするのではなく、生データをシェアして、計算も各自やる方が、卒業研究などでも役に立つはずである）、そこからの**考察は一人ずつ異なるはず**なので、他人のレポートを絶対に丸写ししないこと。

3 討論について

以下の討論テーマから 1 人 1 つのトピックを選んで調べ[2]、レジュメを作って発表し、相互に討論する。人数が 13 人以下の場合は、「水」「大気・環境問題」「産業保健」のそれぞれについて、担当者がいないトピックができることになるが、それは無視して良い。

3.1 水

1. 水質汚染 1（代表的な水質汚染事例と健康被害）
2. 水質汚染 2（日本の飲料水の水質基準とその現況）
3. 水質汚染 3（日本の河川水の水質基準とその現況）
4. 水の BOD の意味と測定法・測定原理

[2]法制などは毎年のように変化するので、なるべく新しい資料を調べること。

5. 水の COD の意味と測定法・測定原理
6. 水の電気伝導度の意味と測定法・測定原理
7. 水の溶存酸素の意味と測定法・測定原理
8. 水の pH の意味と測定法・測定原理
9. 水の糞便性大腸菌の意味と測定法・測定原理
10. 水質管理 1（おいしい水）
11. 水質管理 2（途上国における浄水）
12. 水質管理 3（水道水の配水管理）
13. 日本の水事情（水源、用水、ダム等）
14. 世界の水事情（黄河の枯渇、旱魃等）

3.2 大気・環境問題

1. 大気汚染 1（NOx の測定法と NOx がもたらす健康被害）
2. 大気汚染 2（SOx の測定法と SOx がもたらす健康被害）
3. 大気汚染 3（二酸化炭素の測定法と二酸化炭素がもたらす健康被害）
4. 大気汚染 4（浮遊粒子状物質とは何か。その測定法と浮遊粒子状物質がもたらす健康被害）
5. 大気汚染 5（アスベストの測定法とアスベストがもたらす健康被害）
6. 土壌の有害化学物質 1（内分泌攪乱化学物質について、発生原因、健康被害、測定上の難点）
7. 土壌の有害化学物質 2（PCB について、発生原因、健康被害、測定上の難点）
8. 土壌の有害化学物質 3（ヒ素について、発生原因、健康被害、測定上の難点）
9. 放射線 1（紫外線について、健康影響と規制、測定法）
10. 放射線 2（電磁波について、健康影響と規制、測定法）
11. 放射線 3（電離放射線について、健康影響と規制、測定法）
12. 地球環境問題 1（森林伐採がもたらす健康影響と対策）
13. 地球環境問題 2（気候変動がもたらす健康影響と対策）
14. 地球環境問題 3（マイクロプラスティック問題とは何か、どう対策するか）

3.3 産業保健

1. メンタルヘルス 1（ストレスの評価法について、厚労省が事業者に義務づけているストレスチェックも含めて）
2. メンタルヘルス 2（過労死について、現状と原因と対策）
3. メンタルヘルス 3（気分障害〜とくに抑鬱について、現状と原因と対策）
4. 金属中毒 1（鉛中毒について、産業現場における原因、症状、対処、検査法）
5. 金属中毒 2（水銀中毒について、産業現場における原因、症状、対処、検査法）
6. 金属中毒 3（カドミウム中毒について、原因、症状、対処、検査法）
7. 毒物による中毒（シアン化合物、VX ガス、サリンについて、それぞれ症状、対処方法、確定診断のための検査項目）
8. 有機溶剤による急性中毒（代表的な物質ごとに症状、検査項目、とくに代謝産物）
9. 有機溶剤による慢性中毒（代表的な物質ごとに症状、検査項目、とくに代謝産物）
10. 農薬による急性中毒（代表的な物質ごとに症状と検査項目）
11. 農薬による慢性中毒（代表的な物質ごとに症状と検査項目）
12. 化学物質取扱者の健康管理 1（特定業務従事者健康診断）
13. 化学物質取扱者の健康管理 2（特定化学物質障害予防規則）
14. 作業環境管理における許容濃度・抑制濃度・管理濃度

4 水の基本実験

4.1 概要

3 種類の検水（いくつかのミネラルウォータ及び水道水から予め 200 $m\ell$ ずつ三角フラスコに入れてある）を用い、キレート滴定と簡易キット（パックテストまたはドロップテスト）により 3 回ずつ総硬度を求め、結果を比較する。硬度以外の指標としては、pH、電気伝導度、遊離残留塩素を測定する（どれも 3 回以上繰り返して測定する）。統計解析は、総硬度については、測定方法の違いと検水の違いという 2 つの要因によって測定値を説明する二元配置分散分析を行う。pH や電気伝導度については、検水の違いによって測定値に有意差が出るか一元配置分散分析を行う。

すべての結果から、どの検水がどのミネラルウォータなのかを推測してみよう。

4.2 硬度とは

硬度 (hardness) とは、水中の Ca^{2+} 及び Mg^{2+} の量を表す尺度である。日本と米国では対応する $CaCO_3$ の mg/ℓ に換算[3]して表す。ドイツでは対応する CaO の 10 mg/ℓ に対して硬度 1 度、フランスでは対応する $CaCO_3$ の 10 mg/ℓ に対して硬度 1 度とする。イギリスでは水 1 ガロン（=4.546 ℓ）に $CaCO_3$ を 1 グレイン（=0.0648 g）含む濃度に相当する場合に硬度 1 度とする。日本でも昭和 25 年までドイツ式の硬度だったので、古いデータには要注意。

総硬度（水中の Ca^{2+} 及び Mg^{2+} の総量によって示される硬度）、カルシウム硬度（Ca^{2+} の総量によって示される硬度）、マグネシウム硬度（Mg^{2+} の総量によって示される硬度）、永久硬度（硫酸塩、硝酸塩、塩化物などのような、煮沸によって析出しない Ca 塩及び Mg 塩による硬度）、一時硬度（重炭酸塩のような、煮沸によって析出する Ca 塩及び Mg 塩による硬度）といった区別がある。

自然水中の濃度は、日本の河川水では Ca が 10.4〜13.0(mg/ℓ)、Mg が 3.8〜4.8(mg/ℓ)、地下水では Ca が 17.6〜22.0(mg/ℓ)、Mg が 6.1〜7.3(mg/ℓ)。地下水の方が若干高い。海水中の濃度は、これらより 2 桁ほど高い。海外のミネラルウォータには海水並みに硬度が高いものも存在する。

4.3 キレート滴定法による総硬度の測定

最近のラボでは ICP-MS などによって Ca^{2+} と Mg^{2+} イオンの濃度を直接定量し、そこから硬度を計算するのが普通である。ただし、滴定は化学実験技法の基礎技術として重要なので、本実習で取り上げている。

（原理） EBT は pH10 付近で青色を呈するが、Ca^{2+} や Mg^{2+} などの金属イオンが存在する場合には、キレート生成定数の大きさにしたがって、まず Mg^{2+}、ついで Ca^{2+} と反応してキレート化合物を生成してブドウ赤色を呈する。

このキレート化合物を含む水溶液に EDTA・2Na 溶液を滴下すると EDTA の方が EBT よりもこれらの金属へのキレート生成定数が大きいため Ca^{2+}, Mg^{2+} の順に EDTA が反応し、無色のキレート化合物が生成する。

反応終了（滴定終末点）とともに液の色は遊離した EBT によって青色に変化する。反応終末点は Mg-EBT と EDTA の反応の方が Ca-EBT と EDTA との反応よりも識別しやすいことと、Ca-EBT と Mg-EBT が共存した場合に反応終末点近くでは Mg-EBT のみになっていることから、本法では予め Mg を添加して滴定終末点の識別を鋭敏にする。

（器具・薬品）
- 100 $m\ell$ メスフラスコ 3 つ
- 1 ℓ メスフラスコ 1 つ

[3] 水中のカルシウムイオンとマグネシウムイオンのモル濃度の和に相当するモル濃度の炭酸カルシウムの 1 ℓ 中のミリグラム数を計算する。

- 褐色瓶 1 つ
- 100 $m\ell$ メスシリンダー 2 本
- パラフィルム
- 100 $m\ell$ の三角フラスコ 2 つ
- ビュレット 2 つ
- 漏斗 2 つ
- 500 $m\ell$ の洗瓶 1 つ

（薬品）
1. 0.01 mol/ℓ $MgCl_2$ 溶液: 塩化マグネシウム六水和物, 99.9%グレードの 0.2033 g を水に溶かし、全量を 100 $m\ell$ とする。[4]
2. EBT 試液: エリオクロムブラック T($C_{20}H_{12}N_3O_7SNa$) 0.5 g および塩酸ヒドロキシルアミン ($NH_2OH \cdot HCl$) 4.5 g を 90%(v/v) エタノール 100 $m\ell$ に溶かし、褐色瓶に入れて冷暗所に保存する（有効期間約 1 ヶ月）[5]。
3. アンモニア緩衝液: 塩化アンモニウム 6.75 g に 28%アンモニア水 57 $m\ell$ を加えて溶かし、水を加えて全量を 100 $m\ell$ とする。滴定の至適 pH は 10 $\pm$ 0.1 ときわめて狭いので、この緩衝液を加える必要がある[6]。
4. 0.01 mol/ℓ EDTA 溶液: **和光純薬 0.01M 滴定液を購入して用いる**。力価 F は表示されているので標定不要である [7]。

（方法）以下、1 検水の操作手順を箇条書きで記す。各検水について 3 回の測定を行う。

1. 検水 50 $m\ell$[8]をメスシリンダーで三角フラスコに量りとる[9]。
2. 0.01 mol/ℓ $MgCl_2$ 溶液を 0.5 $m\ell$ 正確に加える。
3. アンモニア緩衝液を 1 $m\ell$ 加える。

---

[4]衛生試験法では酸化マグネシウムと塩酸から作成することになっているが、現在では高純度の塩化マグネシウムが購入可能なのでそれでも良いし、実習では調整済み 0.01 mol/ℓ 塩化マグネシウム水溶液を購入して用いる。

[5]教員または TA が調製済みである。

[6]和光純薬の炭酸塩緩衝液第 2 種 pH10.01 JCSS 認定品 037-16145 を購入してもよい。

[7]滴定液を購入せず調製する場合は、エチレンジアミン四酢酸二ナトリウム二水和物（EDTA2Na·$2H_2O$, MW=372.24；同仁、特級）の約 3.8 g をメスフラスコにとり、水に溶かして全量を 1000 $m\ell$ とする。EDTA を量って水に溶かす場合は正確な調製が困難なため、標定により力価 F を求める。調製した EDTA 水溶液 10.0 $m\ell$ を三角フラスコにとり水を加えて 100 $m\ell$ とし、これに上記アンモニア緩衝液を 2 $m\ell$、EBT 試液を 7〜8 滴加え、溶液が微紅色を呈するまで 0.01 mol/ℓ $MgCl_2$ 溶液で滴定し、要した $m\ell$ 数 a を求め、F=a/10 によって力価 F を算定する。

[8]硬度が高い水を測定する場合は、EDTA 溶液を節約するため、もっと少量で測定しても差し支えない。

[9]衛生試験法では、この後、他の金属イオンのマスキング剤として 10%シアン化カリウム (KCN) 水溶液を数滴加えることになっているが、危険なので実習では省く。そのため滴定終末点が若干不明瞭になるがやむをえない。

4. EBT 試液を 5〜6 滴加える。
5. 0.01 mol/ℓ EDTA 溶液で試験溶液の色が青色を呈するまで滴定し、ここに要した量 b ($m\ell$) を求める。
6. 総硬度 ($CaCO_3$ mg/ℓ) を、

$$\frac{1000 \times (\mathrm{b} \times \mathrm{F} - 0.5)}{50}$$

により算出する[10]。

4.4 その他の測定

パックテストまたはドロップテストの総硬度測定キットにより、キレート滴定と同様に各検水について 3 回ずつ総硬度の測定を行う。方法はキットの説明書を参照すること。

pH は pH メータ、電気伝導度はポータブル電気伝導度計（TwinCond）、遊離残留塩素はハンナインスツルメンツ残留塩素計（HI 96701C）により、各検水を 3 回ずつ測定する。方法は機器付属の説明書を参照すること。

データは総硬度の分析用のファイルと、遊離残留塩素、電気伝導度、pH の分析用のファイルに分けて保存する。総硬度の分析用のファイルは、1 行目に変数名として SAMPLE、METHOD、HARDNESS と入力し、1 列目には検水の種類として A または B または C を入力し、2 列目は測定方法が滴定なのかパックテストなのかを入力し、3 列目は総硬度測定値を入力する。3 回ずつ測定するので、変数名を含めて 19 行 3 列の表になるはずである。もう 1 つのファイルは、1 行目に変数名として SAMPLE、RC、EC、pH と入力し、1 列目には検水種類、2 列目に遊離残留塩素測定値、3 列目に電気伝導度測定値、4 列目に pH 測定値を入力する。こちらは変数名を含めて 10 行 4 列の表になるはずである。

5 大気・環境問題

以下 3 種類の測定を行うので、班内を 3 つのグループに分ける。

ただし、レポートは全部について書くので、自分が直接やっていない分も班内で情報をシェアすること。

5.1 （A）二酸化炭素測定グループ（3〜4 人）

1. 室内空気（F306 で討論会が終わった直後と、窓を 2 箇所開けて 5 分間の換気をした後の 2 回、それぞれ室内 3ヶ所で）と屋外空気（植え込み近くと道

[10] 炭酸カルシウムの分子量は 100.09 なので、約 100 と見なしてこの式は立てられている。b×F から引く 0.5 は予め加える 0.01 mol/ℓ $MgCl_2$ 溶液の量、1000 を掛けるのは g から mg への換算、最後に 50 で割るのは検水量が 50 $m\ell$ だからである。

路脇歩道の 2ヶ所）を、自動計測器で計測すると同時に 30 リットルのポリ袋に採取する。

2. 2LC の検知管を用いて、ポリ袋に採取した空気の二酸化炭素測定を行う。測定はそれぞれのポリ袋について 3 回以上行う。

3. 室内空気と屋外空気で二酸化炭素濃度に有意な差がないかどうかを 2 群の平均値の差の t 検定により検討する。検知管の測定値の正しさを自動計測器の測定値と比較検討する。参考までに、**E501** の CO_2 濃度は常に自動計測していて `https://ambidata.io/bd/board.html?id=25506` に公開されているので、F306 の室内空気の測定値はそれとも比較することができる。

必要な消耗品は以下の通りである。

- 30 ℓ のポリ袋 8 枚× 3 班
- ガステック検知管 2LC 24 本× 3 班

5.2 （B）粉塵・騒音グループ（2〜3人）

道路沿いでデジタル粉塵計と騒音計で連続的に同時測定をし、目視により 1 分ごとの交通量も記録し、それらの経時的な相関関係を検討することとした。なお、雨天時は道路沿いでの計測ができないため、埃っぽい廊下を歩く人が経時的に変化する状況を設定し、屋内で測定することとする。

実習で用いるデジタル粉塵計のうち、最も信頼性が高い機械（DC1700）では個数濃度しか測定できないが、PM2.5 や PM10 の環境基準は環境省の文書[11]に示されているように質量濃度で定められており、論文[12]に示されているように個数濃度と質量濃度の関係は単純ではないので、この機械での測定値の絶対レベルを評価することはできない。しかし相対値としては十分使えると考えられる[13]。廉価な機械として、RATOC のほこりセンサー REX-BTPM25 と WiFi 環境センサー RS-WFEVS1 は $\mu g/m^3$ 単位で質量濃度が測定できるが、信頼性がそれほど高くない（隣に置いても表示値が 1 桁異なることが珍しくない）。RS-WFEVS1 は AC 接続していないと動作しないが、REX-BTPM25 はバッテリー駆動する。ともに測定結果を見るにはスマートフォンにアプリを入れて接続する必要があり、RS-WEVS1 には「エアミル 2」、REX-BTPM25 には「RATOC PM2.5 対応 ほこりセンサー」というアプリをインストールせねばならない。さらに、「エアミル 2」はインターネット上のサーバにアカウント登録してログインしないと使えないので、本実習

[11] `http://www.env.go.jp/council/former2013/07air/y070-25/mat04-2.pdf`

[12] `https://www.jrias.or.jp/report/pdf/21J1.2.16.pdf` や `http://www.jsae-net.org/Pro-study/H28ERCA_170307.pdf`

[13] ちなみに、室内空気の参考値として、small（粒径 0.5μm 以上の小微粒子の 0.01ft^3 当たりの個数）の値が 0-75 で EXCELLENT、75-150 で VERY GOOD、150-300 で GOOD、300-1050 で FAIR、1050-3000 で POOR、3000 以上で VERY POOR という値が示されているが、あくまで目安である。

では、DR-1700 と並べて REX-BTPM25 を使って測定する。自分のスマートフォンに予めアプリをインストールしておくと良い。

また、交通量は粉塵や騒音のレベルに影響を与えると考えられるので、それらの相関関係を検討するのが、この項目の実習目的である。

1. デジタル粉塵計（内蔵バッテリーの充電が十分であることを確認する）と騒音計と外付けバッテリー（充電済であることを確認する）を持って測定場所（上図参照）に行く。デジタル粉塵計も騒音源となるため、約 3 メートルの間をおいて測定準備する。

2. 「RATOC PM2.5 対応 ほこりセンサー」の電源スイッチを ON にし、スマホアプリを起動する。デジタル粉塵計 DC1700 は、電源ボタンを入れると自動的に 1 分ごとに粉塵濃度の記録が開始される。騒音計は外付けバッテリーを USB ケーブルでつなげば電源が入り（ただし、POWER ボタンを長押しして液晶が点滅する状態にしないと 2 分半でオートパワーオフしてしまう）、$\frac{\text{FAST/SLOW}}{\text{RECORD}}$ と書かれているボタンを長押しすると記録が開始される。同時に、手押しカウンター（スマートフォンのカウンターアプリで良い）を使って 1 分ずつ目の前を通る車の通過台数を記録する。その 1 分間のどこかで、「RATOC PM2.5 対応 ほこりセンサー」が示す PM2.5 も記録する。20 分間、なるべく同時に測定開始し、同時に終了する（デジタル粉塵計 DC1700 は電源ボタンをオフするだけで良い。騒音計は $\frac{\text{FAST/SLOW}}{\text{RECORD}}$ ボタンを長押しする。長押しする前に電源が切れると何も記録されないので注意）。

3. デジタル粉塵計 DC1700 を E501 入口脇のデスクトップ機に付属の RS232C-USB ケーブルで接続し、Dylos Logger というソフトを起動する。COM ポー

トを正しく設定し、記録先のファイルを設定し、Download History のアイコンをクリックしてから出てくるウィンドウで Download ボタンをクリックするとデジタル粉塵計に記録されたデータがダウンロードされて表示されるので、それをファイルに保存する。テキストファイルなので、エディタで開いて最初の 7 行を消し、拡張子を csv に変えて保存すれば Excel で開ける。または、エディタ上で必要な部分を選択してコピーし、「貼り付け」の「テキストファイルウィザード」で区切り文字としてコンマを指定しても良い。

4. 騒音計のデータは microSD カードに wsm という拡張子が付いたバイナリファイルとして記録されているので、E501 入口脇のデスクトップ機に microSD カードを差し込み、SoundLink というソフトを起動して、左端のアイコン（Open File）をクリックして microSD カードからデータを読み込む。1 秒ごとの A 特性での騒音レベル測定値 L_A が表示される。ここで左から 2 番目のアイコンをクリックするとタブ区切りテキスト形式で保存できるので（ファイル形式を選択するだけで良い。なお、Excel 形式でも保存できるはずだがエラーが発生するのでテキスト形式を選ぶこと）、1 分ごとの L_{Aeq}（等価騒音レベル）を計算する。計算式は以下である。

$$L_{Aeq} = 10 \log_{10} \left(\frac{1}{60} \sum_{i=1}^{60} 10^{L_A/10} \right)$$

5. 粉塵の生データを Sheet1、騒音の生データを Sheet2 にした Excel 形式のファイルをウェブサイトにアップロードするので、各自ダウンロードしてその後の計算をする。

6. Excel の C 列に 1 秒ごとの測定データが入っているとすると、D2 のセルに

   ```
   =10^(C2/10)
   ```

   と入力してコピーし、D3 から D1201 までのセルに貼り付ける。次に E2 のセルに

   ```
   =10*LOG10(AVERAGE(D2:D61))
   ```

   と入力し、それをコピーして、E62、E122、E182、... と貼り付けていくと、E 列に 1 分ごとの等価騒音レベルが得られる。

7. 最終的に、1 列目が記録時刻、2 列目が small（粒径 0.5μm 以上の小微粒子の 1ft^3、つまり約 28ℓ 当たりの個数。ディスプレイ表示に比べると 100 倍されているので注意）、3 列目が large（粒径 2.5μm 以上の大微粒子の 1ft^3、つまり約 28ℓ 当たりの個数。こちらもディスプレイ表示の 100 倍）、4 列目は PM2.5 の個数濃度（0.01ft^3 当たり）の計算値として small の値から large の

値を引いた結果を 100 で割るという式を入れてコピーし、5 列目はアプリの表示を記録した PM2.5 の質量濃度を入力し、6 列目は開始時刻を合わせて 1 分ごとの L_{Aeq} の値を貼り付け（形式を選択して貼り付け、で値のみ貼り付けるか、数値をキーボードから入力する）、7 列目に車の通過台数（雨天時は人の通過人数）を入力した表を Excel 上に作ってテキスト（タブ区切り）またはテキスト（csv）形式で保存する。このファイルを使って相関関係を検討する他、回帰分析を行っても良い。

消耗品は無い。

5.3 （C）ザルツマン法により大気中二酸化窒素を測定するグループ（残りの人）

この項目では、プレートリーダーにより吸光度を測定することと、検量線から濃度を求めるという技術の習得を目的とする。また、複数の地点間での二酸化窒素濃度を比較し、もし統計学的に有意な違いがあれば、その理由を考察する。

なお、窒素酸化物のサンプリングと測定については、https://www.jstage.jst.go.jp/article/jriet1972/17/4/17_4_236/_pdf や https://www.env.go.jp/earth/coop/coop/materials/02-apctmj1/02-apctmj1-0910.pdf が参考になるので参照されたい。

1. 予め用意された**「亜硝酸ナトリウム溶液原液」**を 1 $m\ell$ とり、イオン交換水を加えて 100 $m\ell$ とし、亜硝酸ナトリウム溶液とする[14]。つまり、**「原液」**を 100 倍希釈する。この「亜硝酸ナトリウム溶液」1.0 $m\ell$ ＝ 0.01 $m\ell$ NO_2（標準状態）である。

2. 「亜硝酸ナトリウム溶液」を 0（ブランク）、0.2、0.4、0.8、1.2、1.6 $m\ell$ とって試験管にいれ、それぞれイオン交換水を加えて全量を 10 $m\ell$ とする（標準希釈系列）[16]。

[14] この方法ではザルツマン係数 0.84 を採用しているので亜硝酸ナトリウム ($NaNO_2$) 2.59 g が 0 ℃、1 気圧での NO_2 ガスの 1 ℓ に相当する。標準状態（0 ℃、1.03kPa）で気体 1 mol は 22.4 ℓ なので、気体 1 ℓ は 1/22.4 mol となる。この量の NO_2 ガスを $NaNO_2$ (MW=69) の重量に換算するには、上記のザルツマン係数を考慮して（1/22.4）× 0.84 × 69 ＝ 2.59 g となる。2.59 g/ℓ の水溶液を作るため、0.259 g を 100 $m\ell$ に溶かすのが標準的な方法だが、0.5 M 水溶液[15]を 7.5 $m\ell$ とってメスフラスコで 100 $m\ell$ まで純水を加えるのと同等なので、本実習ではそちらを作っておいて**「亜硝酸ナトリウム溶液原液」**とする。この「原液」は、３つの班が使う分を最初の班がまとめて調製するか、あるいは教員か TA が予め調製しておく。ザルツマン係数 0.84 の意味は、1 mol の NO_2 ガスが水溶液中で NO_2^- イオンを 0.84 mol 生じ、ザルツマン試薬と反応するのが NO_2^- イオンということである。

[16] 標準状態での二酸化窒素の体積としては、0, 2, 4, 8, 12, 16 $\mu\ell$ に相当する。これをボイル＝シャルルの法則を使ってサンプリング時の気温・気圧に合わせて換算しても良いが、せいぜい 1%しか変わらないので無視しても良い（70 ℓ の空気がポリ袋に採取できているかどうか、二酸化窒素のどのくらいが炭酸カリウム水溶液に溶かし込めたか、といった誤差の方がずっと大きいため）

3. 70 ℓ のポリ袋を 5 枚使い、いろいろな場所（ボイラーの近く、バイクや自動車のエンジンの近くなど高温になるところや、排気ガスなどは濃度が高いと考えられるので、そういった場所を含める）で大気を捕集してくる。

4. 10 $m\ell$ のメスピペットかホールピペットを使い（安全ピペッターを使うと楽である）、0.1%アスコルビン酸水溶液[17]を 10 $m\ell$ 加えて 10 分以上良く振ってから、袋の中身を薬皿（大きめのもの）に取り出す（このとき重さを量れば回収率が計算できるが、今回は 100%と仮定する）。これがサンプル溶液となる[18]。

5. サンプル溶液と標準希釈系列から 2 $m\ell$ ずつ試験管にとり、ザルツマン試薬を 3 $m\ell$ 加えて 10〜15 分放置後、ピペットでマイクロプレートに移す。（少なくともマイクロプレート上では）Triplicate する。

6. プレートリーダーにより 545 nm 付近の吸光度を測定し、検量線[19]により各サンプル溶液中の二酸化窒素量 ($\mu\ell$) を求める。捕集した大気サンプルの容積（今回は 70 ℓ と仮定する）で割って 100 万を掛ければ[20]、ppm 単位の濃度を求めることができる。

7. データファイルとしては、1 列目にサンプリング場所、2 列目に二酸化窒素濃度を入力し（1 行目は変数名として、1 列目を SAMPLE、2 列目を NO2 とする）、タブ区切りまたはコンマ区切りテキスト形式で保存する。16 行 2 列のファイルになるはずである。一元配置分散分析を使ってサンプル間に二酸化窒素濃度に差がないかどうかを調べる。統計的に有意な差があった場合は、検定の多重性を調整した上で、どのサンプルとどのサンプルの間で有意な差があるのかを検定する。

必要な器具と消耗品は以下の通り。ザルツマン試薬を含む廃液は重曹で中和して廃棄する。

- 100 $m\ell$ のメスフラスコ 1 つ
- 50〜300 $m\ell$ 程度の大きさの三角フラスコ 2 つ（亜硝酸ナトリウム溶液用とザルツマン試薬用）

[17] トリエタノールアミンの 10%アセトン溶液の方が良いが、炭酸カリウム水溶液やアスコルビン酸水溶液でも代用可能。水でも良いはずだが、これらの溶液の方が回収率が良いはず。

[18] （注）サンプル溶液の採取については、予算措置次第では変更する可能性がある。TEA 含浸シリカゲルチューブをミニポンプに装着し 200$m\ell$/min で 30 分間サンプリング後、蒸留水 15$m\ell$ に 1 時間溶出させ、3000rpm で 5 分間遠心分離して上澄みをサンプル溶液とする方法をとる可能性もある。

[19] Triplicate した中で明らかに測定ミスと思われる外れ値が 1 つあるときは、そのデータのみ分析から外す。3 つが互いにバラバラな場合は再測定（再実験ではない）する。EZR を使って計算する方法は後述するが、実験時点では方眼紙を使い、著しく直線性が悪くないか確認する。あまりに直線性が悪い場合は再実験となる。

[20] $\mu\ell$ は ℓ の 100 万分の 1 だから、検量線の横軸を $\mu\ell$ 単位の二酸化窒素量にしておけば、単純に 70 で割るだけで良い。

- 100 $\mu\ell$ のピペットを奥まで差し込める太さの試験管多数（最低限、22 本は必要。本当はザルツマン試薬を反応させること自体 triplicate したいので 44 本欲しい）
- P1000 と P100 のピペット及びチップ多数
- 70 ℓ のポリ袋 5 枚× 3 班
- 96 穴マイクロプレート× 3 班
- 0.5 M 亜硝酸ナトリウム水溶液（毎年 7.5 $m\ell$ だけ使う）
- 0.1%アスコルビン酸水溶液 10 $m\ell$ × 5 × 3 班＝ 150 $m\ell$（あまり失敗しないので 200 $m\ell$ もあれば十分）
- ザルツマン試薬 3 $m\ell$ × 11 × 3 班＝ 99 $m\ell$（ただしやり直しなどあるので、毎年 200 $m\ell$ 程度使う）
- ザルツマン試薬の中和用重曹

6 産業保健

照度の異なる作業環境下で 100 マス計算などの単純作業を行い、連続単純作業にともなう作業効率と正確さの低下（及びストレスの増加）が照度の影響を受けるかを調べる。記録者はデジタル温湿度計により温度と湿度も計測し、温度や湿度の影響がないかも考慮する（温度や湿度があまり変わらない条件で実験すべきである）。

短期的なストレスを簡便に測定するには、唾液をサンプルとして使う方法がいくつか開発されている。ストレスマーカーとしてはコルチゾール、クロモグラニン A、アミラーゼなどが有名であるが、コルチゾールやクロモグラニン A を測定するには ELISA や RIA が必要であり高価なので、本実習ではニプロから販売されている唾液アミラーゼモニタ[21]を用いる。この測定はきわめて簡単に操作でき、測定値もデジタルで表示されるので、チップを正しく舌下に当てさえすれば測定ミスの可能性は低いのが利点である。

なお、野村ら (2010)[22]によれば、コルチゾールは精神ストレスに伴う視床下部-下垂体-副腎系のストレス反応を反映する物質であるため、強い緊張や不安を伴うストレッサーに応じて濃度が上昇するが、認知課題や単純な精神作業に対しては増加しないが、唾液アミラーゼやクロモグラニン A は交感神経-副腎髄質系の賦活化を反映するため、スピーチ、ワープロ作業、高速道路走行、認知テスト、騒

[21]測定原理等については、東朋幸、山口昌樹、出口満生、若杉純一、水野康文：唾液アミラーゼ活性を利用した交感神経活動モニタと運転ストレスの評価. 信学技報, 2004-12: 35-40, 2004. を参照。

[22]野村収作、水野統太、野澤昭雄、浅野裕俊、井出秀人：短期精神ストレスマーカーとしての唾液中クロモグラニン A の特性評価. 生体医工学, 48(2): 207-212.

音刺激等に対して一過性の増加を示すことが知られている。したがって、本実験のバイオマーカーとしては唾液アミラーゼやクロモグラニン A の方がコルチゾールより適している。

本項の目的はクロスオーバー研究のデザインで実験し、その解析法を学ぶことである。

1. テンキーがついたノートパソコンが 6 台用意してあり、Ten2Ten という 100 マス計算のソフトウェアが入っている（１桁の足し算を 100 回する。答えが２桁になる場合は１の位のみ入力する。正解しないと先に進まず、かかった時間と誤答の回数が自動的に測定される。５回分まで別々の記録として保存できる）ので、これを作業と見なす（ノートパソコンがない場合は紙で作業しストップウォッチで計時。答え合わせが必要なので面倒）。

2. 各班パソコン台数×２名が被験者となり、残りの人が測定・記録者となる（予め記録用紙を作ると便利）。被験者はランダムに２群に分かれる（例：被験者人数を X とすると、名簿順に 1〜X として、R のコンソールウィンドウに

   ```
   sample(1:X, as.integer(X/2), rep=FALSE)
   ```

   と入力して実行し、表示された番号を第１群とする）。第１群は先に暗条件、後で明条件で作業する。第２群はその逆順にする。

3. 原則として暗条件は部屋の照明を付けず、明条件は部屋の照明を付けることで実現する。室内の位置によって明るさがあまり変わらないように工夫する。その他、椅子のセッティングなども揃える。各条件において、測定・記録者は照度計により照度を、デジタル温湿度計により温度と湿度を測定し記録する。全体の実験順序を、第１群暗、第２群明、第１群明、第２群暗とすることにより、実験の連続による carry-over 効果を避ける。

4. 各条件において、被験者がすることは以下の通りである。まず唾液アミラーゼモニタによりストレスを測定する。次いで、作業を３回実行する。終了後に再びストレスを測定する。測定・記録者は、ストレスと、作業結果のうち所要時間と誤答回数をそれぞれ記録する（所要時間と誤答数は、各条件ごとの３回の作業の合計を統計解析に用いる）。

5. 測定・記録者はすべての結果を Excel に入力し、ファイルを全員に配布すること。結果の分析は一人ずつ EZR などを用いて行い、最終レポートに記載する。

必要な器具・消耗品は以下の通り。

- 照度計

- 温湿度計
- テンキー付きの Windows ノートパソコン（6 台）
- ニプロ唾液アミラーゼモニタ（2 台）
- 唾液アミラーゼ測定チップ（12 × 4 × 3 班＝ 144 ＜ 160 枚＝ 8 袋）

7　統計解析の方法

本実習では、得られた結果について、さまざまな統計解析が必要となる。既に自分が使い慣れたソフトウェアがあれば、それで解析しても良いが、とくにそういうソフトウェアがなければ、EZR を用いることをお勧めする[23]。

EZR とは自治医科大学の神田善伸教授が開発した、保健医療分野で良く使われる統計解析をメニューから選んで実行できるようにしたフリーソフトである。統計解析自体は R というシステムを利用して行い、EZR は R の機能をメニューから呼び出すためのフロントエンドなので、得られる結果はの信頼性は十分に高く、論文投稿にも問題なく利用できる。

R も EZR もフリーソフトなので、実習用コンピュータ 6 台にインストールしてあるが、自分のコンピュータにインストールすることも自由にできる。R 関連のソフトウェアは CRAN (The Comprehensive R Archive Network) からダウンロードすることができる。CRAN のミラーサイトが各国に存在し、ダウンロードは国内のミラーサイトからすることが推奨されているので、日本では統計数理研究所のサイト（`https://cran.ism.ac.jp/`）を利用すると良い。Windows で EZR を利用するなら、最初から EZR まで一体になったインストーラファイルを自治医大の web サイト（`https://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html`）からダウンロードし、ダブルクリックしてインストールするのが最も簡単である。R をインストールしてから EZR のパッケージをインストールするには、R を起動して表示されるプロンプトに対して、

```
install.packages("Rcmdr", dep=TRUE)
install.packages("RcmdrPlugin.EZR\index{EZR}", dep=TRUE)
```

と打てば良い。ここから EZR を起動するには、いったん `library(Rcmdr)` と打って R コマンダーを起動し、その「ツール」メニューから「Rcmdr プラグインのロード」を選び、プラグインとして RcmdrPlugin.EZR を選んで OK ボタンをクリックする。少し待つと、「R コマンダーを再起動しないとプラグインを利用できません。再起動しますか？」と尋ねるダイアログが表示されるので、「はい (Y)」

[23] jamovi でも良い。EZR とはまた別のフリーソフトだが、計算のバックボーンは EZR と同じく R である。jamovi については、`https://www.jamovi.org/` や `https://bookdown.org/sbtseiji/jamovi_complete_guide/` を参照されたい。

をクリックすると EZR が起動する。この際、オリジナルの Rcmdr メニューはメニュー右端の「標準メニュー」に残っている。

なお、Windows 環境で自治医大のサイトからダウンロードしたインストーラで EZR 組み込み済みの R をインストールした場合は、デスクトップにできている EZR on R（64 ビット版の OS では 64 ビット版と 32 ビット版の両方が入るが、動作する環境なら 64 ビット版がお勧め）のアイコンから起動すると、R 本体が起動した後で自動的に EZR のメニューまで起動させてくれる。

コラム: 3 つ以上の値を測る 3 つの理由

triplicate の理由 本実習の二酸化窒素測定では、プレートリーダーには同じサンプルを 3 箇所のウェルに入れて、3 つずつの値を測定した。ウェルの中に空気が入ってしまった場合など、2 つしか測っていなくて 2 つの値がかけ離れていると、どちらが正しいかわからないが、3 つあれば、2 つ以上のウェルに空気が混入してしまう可能性は低いので、おそらく近い 2 つの値の方が正しいだろうと判定できる。duplicate でなく triplicate するのはそのためである。もちろん、triplicate でも 96 穴のプレートならばいっぺんに全部測れるくらいしかサンプルがなく、コストが変わらないというのも大きな理由である。

サンプルサイズが 3 以上でなくてはいけない理由 滴定や二酸化炭素濃度測定を 1 つのサンプルについて 3 回以上した理由は、2 つ以上のサンプル間で平均値を比べる際のバラツキの信頼性を担保するためである。2 つしか値がないと、個々の測定値が平均値から相対的にどれくらい外れているかを評価できない。偏差の絶対値が標準偏差と一致してしまうため、実測値が何であろうと、偏差値を計算すると大きい方が 60、小さい方が 40 になってしまう。

検量線を書くための標準試料が 3 つ以上の濃度でなくてはいけない理由 検量線は、予め濃度または絶対量がわかっている標準試料を横軸に取り、縦軸に吸光度などをとって、両者の関係を直線として作図し（統計学的には後述する回帰直線）、測定したい未知試料の吸光度から濃度または絶対量を逆算するために用いる。2 種類の標準試料しかないと、直線は必ず 2 つの測定点を通ってしまうので、標準試料調製に失敗していても気づくことすらできない。液体の塩分濃度の簡易測定器であるカーディソルト（堀場製作所）などでも、純水でゼロ点調整後のスロープの調整には 0.1%と 5%の標準液で校正する、3 点校正を行う。標準液の保存状態が悪くて濃度が狂ってしまっていると校正がうまくできないが、そのことに気づくことができるのが 3 点校正のメリットである（その場合は未開封の標準液を用意して校正をやり直す）。

7.1 実験計画法

実験的研究の肝はデザインである。デザインが悪い研究から良いデータは生まれない。実験計画法は、R.A. Fisher がロザムステッドで行った農学研究に始まるが、保健医療分野では、この種のデザインは、新規に合成した化学物質の毒性試験や、

新薬の臨床試験で、用量反応関係を分析するために必須である。もちろん、ヒトを対象にした研究は、実施前に倫理審査を通らねばならず、倫理審査に提出する書類には、適切なサンプルサイズの設定を含む、適切なデザインが記述されねばならない。実験計画法には、目的に応じて様々なデザインがある。本稿では以下の３つについて説明する。

なお、実習では不可能だが、実験デザインの中では、サンプルサイズの決定も非常に重要な問題である。サンプルサイズを大きくすれば、医学生物学的な意味がない微小な差であっても、統計学的には「ゼロでない差がある」という検定結果を出せてしまうし、逆にサンプルサイズが小さすぎると、本当は意味のある大きさの差があっても、検定結果は有意にならないこともありうる（そういう状況を、専門用語では「検出力が不十分」という）。EZR では、メニューの「統計解析」＞「必要サンプルサイズの計算」から項目を選んで、先行研究などから得た事前情報を入力することで、サンプルサイズを計算することができる。動物実験などを計画する場合は、必ずこのプロセスを踏むべきである。

- 単一群、事前-事後デザイン
- 平行群間比較試験（完全無作為化法）
- クロスオーバーデザイン

7.2 単一群、事前-事後デザイン

このデザインを使うと、研究者は、個々の対象者に対して同じ精度で測定された、何かの処理の前後で測定値に変化がないかどうかを評価することができる。通常使われる検定手法は、対応のある t 検定や、ウィルコクソンの符号順位検定になる（事前や事後の測定点が 2 時点以上ある場合は、反復測定分散分析やフリードマンの検定になる）。なお、対応のある t 検定は、個人ごとに算出した変化量の平均値がゼロという帰無仮説を検定する一標本 t 検定と、数学的には同値である。以下のような研究が典型的な例であるが、本実習には、これに該当するデザインの項目は含まれていない。

- 慢性関節リウマチ (RA) 患者の手術前後の血清コルチゾールレベルを比較することで、手術の効果を判定
- うつ病患者の音楽療法の前後で、質問紙によるうつ得点を比較することで、音楽療法の効果を判定
- 珈琲を飲む前後で単純計算にかかる時間や正答率を比較することで、珈琲を飲むことが計算能力や集中力に影響するかを判定

7.3　平行群間比較試験（完全無作為化法）

これは非常に単純である。研究参加に同意した対象者各人に対して、完全にランダムに（行き当たりばったりに、ではなく）、いくつかの処理（曝露）の１つを割り付け、処理間での比較をするというものである。

無作為化 (randomization) の方法にはいくつかある。Fleiss JL (1986)“The design and analysis of clinical experiments” は、乱数表 (random number table) の代わりに乱数順列表 (random permutation table) を使うことを推奨している。もっとも、今ではコンピュータソフトが使えるので、紙の表を使わなくても、簡単にランダム割り付けができる。

結果の解析は 2 群間での平均値の比較なら Welch の方法による t 検定、3 群以上の間での平均値の比較なら一元配置分散分析 (One-way ANOVA) になる。

本実習項目の中で、換気前後の屋内空気、屋内と屋外、2ヶ所の屋外空気の間といった 2 条件間の二酸化炭素濃度の比較は、2 群間の平均値の比較なので（無作為割り付けした処理の異なる 2 群間ではないが）、Welch の方法による t 検定を適用することができる[24]。3ヶ所の屋内空気間の二酸化炭素濃度の比較なら一元配置分散分析となる。ただし、屋内空気に関しては、換気前後という要因と、採取場所という要因を組み合わせた、二元配置分散分析にする方が筋が良い。また、5 地点間で二酸化窒素濃度に差がないかどうかを分析するのは一元配置分散分析となる。

二酸化炭素濃度の t 検定

二酸化炭素濃度データは、`https://minato.sip21c.org/pubhealthpractice/airco2n.csv` のような形で入力する。変数 SAMPLE はサンプルの種類を示し、ROOM が室内空気、OUTSIDE が屋外空気を意味する。CONDITION はサンプリング時の条件を示し、室内空気については NOVENTI が換気前、VENTI が換気後を意味し、屋外空気については NONE としている。PLACE は測定地点で、屋内について R1、R2、R3 の 3ヶ所、屋外について O1 と O2 の 2ヶ所である。MEASURE はそれぞれの測定が何回目かを意味する（検知管の場合、1 回測定するごとに検知管は変わるし、測定によるサンプルへの影響もほとんど考えられないので、無くても良い変数である）。KENCHIKAN という変数が検知管で測定した二酸化炭素濃度（ppm 単位）、MACHINE という変数が自動測定器（ヤガミ YCO-L［CO2・O2・温・湿度・気圧データロガー］[25]）で測定した二酸化炭素濃度（ppm 単位）を示す。Excel などでデータを入力し、「形式を選択して保存」から「テキスト（コンマ区切り）」で保存すればできる。データを EZR に読み込むには、「ファイル」の「データのインポート」から「ファイルまたはクリップボード、URL からテキストデータを読み込む」を選ぶ。表示されるウィンドウの中で、「データセット名

[24] 換気前後で屋内で採取する空気は、たとえ同じ場所で測定しても厳密な対応関係はないので、独立 2 群間の平均値の差の検定を行う。

[25] マニュアル：`https://ict.yagami-inc.co.jp/images/download_img/YCO-L_users_manual.pdf`

を入力：」に対して、読み込んだデータに対してつける名前を入力し（デフォルトは Dataset だが、例えば CO2 にする）、「データファイルの場所」が「ローカルファイルシステム」になっているのを確認し、「フィールドの区切り記号」が「カンマ」になっているのを確認して「OK」ボタンをクリックし、データが入っている CSV ファイルを選んで「OK」すると、EZR の中に CO2 という名前を付けてデータを読み込むことができる[26]。

まずサンプルの種類ごとの平均値と標準偏差を計算させるため、「グラフと表」の「サンプルの背景データのサマリー表の出力」を選ぶ。「群別する変数（0〜1 つ選択）」として「CONDITION」を選び、「連続変数（正規分布）」として「KENCHIKAN」と「MACHINE」を選び、「OK」をクリックすると下表が表示される。

```
-------------------------------------------------------------------------------
              |   CONDITION                                       |
    Factor    |      NONE           NOVENTI           VENTI       |    p.value
--------------+---------------+-----------------------------------+---------------
            n |       6       |       9       |       9           |
--------------+---------------+---------------+-------------------+---------------
     KENCHIKAN | 427.50 (20.43) | 550.00 (43.30) | 538.89 (31.00) |         <0.001
       MACHINE | 484.17 (14.96) | 599.22 (35.26) | 546.44 (30.41) |         <0.001
-------------------------------------------------------------------------------
```

数字は平均±標準偏差である[27]。換気前の屋内空気が検知管では 550±43.30 ppm、自動測定器では 599.22±35.26 ppm、換気後の屋内空気が検知管では 538.89±31.00 ppm、自動測定器では 546.44±30.41 ppm、屋外空気が検知管では 427.50±20.43 ppm、自動測定器では 484.17±14.96 ppm である。この `p.value` は一元配置分散分析の結果の有意確率であり、< 0.001 という値は、屋外と屋内で二酸化炭素濃度に統計学的に有意水準 5 %で有意な差があるという当たり前のことを意味しているだけであるが、一度にすべての条件の平均と標準偏差が得られるのは便利である。

サンプル種類ごとに 2 群間で t 検定をするには、以下のようにする。「統計解析」の「連続変数の解析」メニューから「2 群間の平均値の比較 (t 検定)」を選んで表示されるウィンドウで、「目的変数」として「KENCHIKAN」または「MACHINE」、「比較する群」として「CONDITION」を選び、「等分散と考えますか？」を「いいえ (Welch 検定)」を選んでから、下の方の「↓一部のサンプルだけを解析対象にする場合の条件式」の枠内に、

```
SAMPLE=="ROOM"
```

[26]jamovi の場合は、起動画面上部左端の横三本線記号をクリックし、OPEN を選んで表示されるダイアログにデータが存在する URL またはファイルを選べば読み込める。

[27]書くまでもないと思うが、標準偏差は、個々の値から平均を引いて二乗した値の和をとってサンプルサイズから 1 を引いた自由度で割った値である不偏分散の平方根であり、母集団におけるデータのばらつきの大きさの指標である。

と打って「OK」ボタンをクリックする。エラーバー付きの棒グラフが表示されるのと同時に、出力ウィンドウに、例えば、「MACHINE」で「SAMPLE=="ROOM"」の場合だと、

```
                    平均      標準偏差      P 値
CONDITION=NOVENTI 599.2222 35.25896 0.00375
CONDITION=VENTI   546.4444 30.40605
```

と表示されて、Welch の t 検定の結果の有意確率が p=0.00375 であり、屋内空気の二酸化炭素濃度には換気前後で有意水準 5%で統計学的な有意差がある。

屋外 2 地点間の比較をするには、「目的変数」として「KENCHIKAN」または「MACHINE」、「比較する群」として「PLACE」を選び、「↓一部のサンプルだけを解析対象にする場合の条件式」の枠内に、

```
SAMPLE=="OUTSIDE"
```

と打って「OK」ボタンをクリックすると、「KENCHIKAN」を選んだ場合だと、次の結果が表示される。p 値が 0.801 なので、統計学的に有意な差があるとはいえない。

```
          平均 標準偏差   P 値
PLACE=O1  425     25    0.801
PLACE=O2  430     20
```

なお、屋内空気について、採取場所と換気前後という２要因の影響を調べるには、「統計解析＞連続変数の解析＞複数の因子での平均値の比較（多元配置分散分析）」と選び、「目的変数（１つ選択）」で「KENCHIKAN」または「MACHINE」を選び、「因子（１つ以上選択）」で「CONDITION」と「PLACE」を選び、「↓一部のサンプルだけを解析対象にする場合の条件式」の枠内に、

```
SAMPLE=="ROOM"
```

と打って「OK」ボタンをクリックすると、グラフとともに次の結果が表示される。採取場所による影響や採取場所と換気前後との交互作用は有意でなく、換気前後では統計的な有意差が見られている。

```
Anova Table (Type III tests)

Response: MACHINE
                                 Sum Sq Df   F value    Pr(>F)
(Intercept)                     5906485  1 4772.4883 < 2.2e-16 ***
Factor1.CONDITION                 12535  1   10.1282  0.007885 **
Factor2.PLACE                       786  2    0.3177  0.733772
Factor1.CONDITION:Factor2.PLACE    1704  2    0.6885  0.521132
Residuals                         14851 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

検知管と自動測定器の測定値の対応関係

この二酸化炭素濃度測定では、室内空気と屋外空気について、検知管と自動測定器という 2 つの方法で**ほぼ同じ**サンプルを測定しているので、測定法の比較をすることができる。通常、安価で迅速にできる測定法を新しく開発した場合に、その測定値が信頼できるかどうか示すためには、同一サンプルを既に信頼性が確立している方法と同時に測定し、(1) 相関が良く（大小関係をともにする関係であること、詳しくは後述するが、「グラフと表＞散布図」メニューで散布図を描くのは容易である）、(2) 絶対値に差がなく（通常は「統計解析＞連続変数の解析＞対応のある 2 群間の平均値の比較 (paired t 検定)」で調べる[28]）、(3) 絶対値の大小と 2 つの測定法の差の間に一定の傾向がない（2 つの変数の平均と、2 つの変数の差の間に相関関係がないかを見る[29]）、という 3 つの条件を満たす必要がある。

単純に散布図を書いて直線的な強い相関関係が見られるとか、相関係数が高いとか、t 検定で有意差がないというだけでは不十分であり、Bland-Altman プロット（Bland and Altman, 1999[30]）が推奨される。

EZR のメニューには含まれていないが、MethComp パッケージや brandr パッケージで実行できる（詳細は大学院の保健学研究共通特講 IV のテキスト https://minato.sip21c.org/ebhc/ebhc-text.pdf の Chapter 15 を参照されたい）。下記のコードを EZR の「R スクリプト」ウィンドウに打ち、マウスを使って選択した状態で「実行」ボタンをクリックすると Bland-Altman プロットが実行できる。

[28] このデータの場合、厳密に同じサンプルとはいえないので、それで良いかどうかは微妙だが

[29] このデータの場合、厳密な対応関係はないので、それで良いかどうかは微妙だが

[30] Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2): 135-160https://doi.org/10.1177/096228029900800204

```
if (require(blandr)==FALSE) {
install.packages("blandr", dep=TRUE)
library(blandr)
}
blandr.draw(CO2$KENCHIKAN, CO2$MACHINE, plotter="rplot")
blandr.method.comparison(CO2$KENCHIKAN, CO2$MACHINE)
blandr.output.text(CO2$KENCHIKAN, CO2$MACHINE)
```

二酸化窒素濃度の一元配置分散分析

二酸化窒素濃度データは、https://minato.sip21c.org/pubhealthpractice/airno2.csv のような形で入力する。PLACE はサンプリング場所、NO2 は二酸化窒素濃度を示す。1 つのサンプリング場所について NO2 の値が 3 行あるのは、triplicate したプレート上の 3 つの穴の吸光度から計算される NO_2 濃度を、それぞれ入力するためである。通常はサンプル自体を triplicate し、さらにプレート上でも triplicate し、データとして入力するのは、サンプルごとの平均値（極端な外れ値があればそれを除外した平均値）となるが、本実習ではプレート上でのみ triplicate するので、このように処理する。

データの読み込みは二酸化炭素濃度のときと同様に、「ファイル」＞「データのインポート」＞「ファイルまたはクリップボード、URL からテキストデータを読み込む」を選んで、「データセット名を入力：」を NO2 とし、「データファイルの場所」が「ローカルファイルシステム」、「フィールドの区切り記号」が「カンマ」になっているのを確認して「OK」ボタンをクリックし、データが入っている CSV ファイルを選んで「OK」すると、EZR の中に NO2 という名前を付けてデータを読み込むことができる。

一元配置分散分析を実行するには、「統計解析」＞「連続変数の解析」＞「3 群以上の間の平均値の比較（一元配置分散分析 one-way ANOVA）」を選択し、「目的変数（1 つ選択）」は NO2 が選ばれているのでそのまま、「比較する群（1 つ以上選択）」は PLACE を選ぶ。一元配置分散分析だけで良ければ、「等分散と考えますか？」は「いいえ（Welch 検定）」を選ぶべきだが、どの場所とどの場所で二酸化窒素濃度に差があったのかという 2 地点ずつの対比較を繰り返すためには、Holm や Tukey などの方法で検定の多重性の調整をする必要があり[31]、その場合

[31]Neyman-Pearson 流の仮説検定では、「差がない」という帰無仮説の下で、実際に得られた値より大きな差が偶然得られる確率である P 値を計算し、この P 値が一定の基準—通常は 5%にするが、有意水準と呼ばれる—より小さかったら帰無仮説が間違っていると考えるが、検定を繰り返すと、本当は差がないのに P 値が偶然に有意水準より小さくなってしまう確率が 5%よりずっと大きくなってしまうため、P 値を調整しなければならない（厳密には P 値ではなく、本当は差がないのに間違って差があると判定してしまう第一種の過誤を 5%以内にするために有意水準を調整するのだが、有意水準を 3 分の 1 にすることと P 値を 3 倍にすることは表裏一体なので、大抵の統計ソフトは P 値を調整した結果を表示することになっており、有意水準は最初に設定したまま有意性を判定すれば良い）。調整の方法がいろいろあり、Holm の方法や FDR 法が良く使われる。平均値の比較の場合に限っては、古典的に良く使われてきた Tukey の HSD 法がいまでも使われている。

は一元配置分散分析も等分散を仮定するしかない設定になっている。

サンプルデータを「いいえ（Welch 検定）」で解析すると、出力ウィンドウに以下が表示される。

```
平均 標準偏差      P 値
PLACE=A 3.333333 1.154701 0.00923
PLACE=B 4.333333 1.527525
PLACE=C 5.000000 1.000000
PLACE=D 3.000000 1.000000
PLACE=E 9.000000 1.000000
```

P 値の 0.00923 は 0.05 より小さいので、有意水準 5%で Welch の方法による一元配置分散分析を実行した結果、場所の違いが二酸化窒素濃度に統計学的に有意に影響していると結論することができる。ちなみに等分散性の仮定について「はい」を選んで、Tukey の多重比較のオプションをチェックしてから実行すると、出力ウィンドウに以下が表示される。p adj は Tukey の方法で検定の多重性を調整したあとの P 値であり、この結果で p adj が 0.05 より小さいのは、E-A、E-B、E-C、E-D だけなので、E 地点の二酸化窒素濃度は他のどの地点よりも有意に高く、その他の地点間では統計学的に有意な差は認められなかったといえる（ただし、サンプルサイズが小さいので検出力が大きくないため、E 以外の地点間の差についてはあるともないとも言えないと考えるべきである）。

```
diff        lwr      upr     p adj
B-A  1.0000000 -2.1028621 4.102862 0.8219440
C-A  1.6666667 -1.4361954 4.769529 0.4400298
D-A -0.3333333 -3.4361954 2.769529 0.9960810
E-A  5.6666667  2.5638046 8.769529 0.0009542
C-B  0.6666667 -2.4361954 3.769529 0.9501740
D-B -1.3333333 -4.4361954 1.769529 0.6330823
E-B  4.6666667  1.5638046 7.769529 0.0040788
D-C -2.0000000 -5.1028621 1.102862 0.2829865
E-C  4.0000000  0.8971379 7.102862 0.0115580
E-D  6.0000000  2.8971379 9.102862 0.0006061
```

※水の測定における電気伝導度や pH や残留塩素の分析も、これと同様（ただし検水が 3 種類）に実行すること。

7.4 クロスオーバー法 (cross-over design)

クロスオーバー法では、それぞれの対象者が 2 種類の処理を受ける。このとき、適当な間隔（ウォッシュアウト期間と呼ぶ。前の処理の影響のキャリーオーバーを避けるために設ける）をおくことと、処理の順番が違う 2 つのグループを設定

することが必要である。

Hilman BC et al. “Intracutaneous immune serum globulin therapy in allergic children.”, JAMA. 1969; 207(5): 902-906. を例として説明しよう。

この研究では、同意が得られた 574 人から、まず研究目的に対して不適格な 43 人を除外し、531 人をランダムに 2 群に分けた。グループ 1 が 266 人、グループ 2 が 265 人となった。グループ 1 に処理 A を行い、34 人が脱落した。同時にグループ 2 には処理 B を行い、脱落が 15 人であった。その後、2ヶ月のウォッシュアウト期間をおき、グループ 2 の 250 人に処理 A を、グループ 1 の 232 人に処理 B を行ったところ、それぞれ 45 人、29 人が脱落したので、2 回の処理を完了したのは合計 408 人となった。

本実習の産業衛生の実験では、クロスオーバー法で作業環境がストレス増加に与える影響を評価した。このデザインは少々複雑である。処理の順番が結果に影響しなければクロスオーバーにしなくても良いのだが、今回の実験デザインでは、作業前後での唾液アミラーゼ濃度増加が、作業環境の照度によって影響を受けないかどうか、照度条件が、先に明条件、後に暗条件か、その逆かという順番による影響がないかどうか、それらの交互作用がないかどうかをすべて調べる必要がある。仮に下表のようなデータが得られたとする（分析用のデータは、このような形式で表計算ソフトに入力する（`https://minato.sip21c.org/pubhealthpractice/crossover.csv`）。変数 PID は被験者の識別番号を意味する。変数 Ord.LD は照度条件の順番を意味し、LD が先に明条件の群、DL が先に暗条件の群である。変数 LD は照度条件を意味し、L が明条件、D が暗条件である。変数 T0A と T1A は作業前後の唾液アミラーゼ値、TIME が 3 回の作業の合計時間、ERROR が 3 回の計算の誤答数の合計を意味する）。

PID	Ord.LD	LD	T0A	T1A	TIME	ERROR
1	LD	L	15	21	100	1
2	LD	L	17	19	70	2
3	LD	L	19	25	75	3
4	LD	L	21	23	120	4
5	LD	L	23	27	80	6
6	DL	D	15	34	150	2
7	DL	D	17	27	160	3
8	DL	D	19	39	140	5
9	DL	D	21	28	130	2
10	DL	D	23	41	130	3
1	LD	D	15	35	140	5
2	LD	D	17	28	150	6
3	LD	D	19	40	145	3
4	LD	D	21	29	130	5
5	LD	D	23	42	135	6
6	DL	L	15	22	120	2
7	DL	L	17	20	100	4
8	DL	L	19	26	110	5
9	DL	L	21	24	90	3
10	DL	L	23	28	85	2

分析には（他のやり方もあるが、ここでは最も簡単な方法のみ示す)、「ファイル」の「データのインポート」メニューからこのデータを読み込ませた後、「統計解析」の「連続変数の解析」から「対応のある 2 群以上の間の平均値の比較（反復測定分散分析)」を選んで、「反復測定したデータを示す変数（2 つ以上選択)」の枠から T0A と T1A を選び、「群別する変数を選択 (0〜複数選択可)」の下の枠から LD と Ord.LD を選んでから OK をクリックする。結果は下の枠内のように得られる（２つのグラフが自動的に描画される)。

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                              Sum Sq num Df Error SS den Df  F value    Pr(>F)
(Intercept)                  22944.1      1    402.8     16 911.3843 1.554e-15 ***
Factor1.LD                     291.6      1    402.8     16  11.5829  0.003634 **
Factor2.Ord.LD                   0.0      1    402.8     16   0.0000  1.000000
Factor1.LD:Factor2.Ord.LD        2.5      1    402.8     16   0.0993  0.756738
Time                           980.1      1    154.8     16 101.3023 2.510e-08 ***
Factor1.LD:Time                291.6      1    154.8     16  30.1395 4.942e-05 ***
Factor2.Ord.LD:Time              0.0      1    154.8     16   0.0000  1.000000
Factor1.LD:Factor2.Ord.LD:Time   2.5      1    154.8     16   0.2584  0.618160
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

右端の Pr(>F) という列をみると、Factor1.LD の行が 5%より小さいので明

条件と暗条件ではアミラーゼ値が有意に異なり、Time の行も 5%より小さいので作業前後でアミラーゼ値が有意に異なり、Factor1.LD:Time の行も 5%より小さいので作業条件と時間の交互作用効果が有意である（即ち、**明条件か暗条件かで、作業前後でのアミラーゼ値の変化の仕方が有意に異なる**）ことがわかる。この結果では、Factor2.Ord.LD や、それを含む行は Pr(>F) が大きい値なので、クロスオーバーデザインにはしたけれども、明暗条件の順序は結果に影響しなかったといえる。

なお、群別がない場合は、球面性検定の結果が表示され、その結果が統計学的に有意だった場合に用いられる 2 種類の補正の結果も表示されるが、このように群別変数を指定して交互作用効果を見る場合は計算されない。

計算に掛かった時間と誤答数については、通常の二元配置分散分析で良い。「統計解析＞連続変数の解析＞複数の因子での平均値の比較（多元配置分散分析）」と選び、「目的変数（１つ選択）」として「TIME」または「ERROR」、「因子（１つ以上選択）」として「LD」と「Ord.LD」を選んで「OK」をクリックすると、例えば「ERROR」の場合、次の結果が得られる。「LD」の主効果も「Ord.LD」の主効果も、それらの交互作用効果のいずれも統計的に有意ではないことがわかる。

```
Anova Table (Type III tests)

Response: ERROR
                          Sum Sq Df  F value         Pr(>F)
(Intercept)                259.2  1 123.4286 0.000000006232 ***
Factor1.LD                   3.2  1   1.5238         0.2349
Factor2.Ord.LD               5.0  1   2.3810         0.1424
Factor1.LD:Factor2.Ord.LD    5.0  1   2.3810         0.1424
Residuals                   33.6 16
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

コラム：測定の正確さ・精度・妥当性

測定の正しさを考えるとき、少なくとも 3 種類の異なる正しさが存在することに注意したい。Validity (妥当性)、Accurary（正しさ・正確性)、Precision（精度）の 3 つである。

Validity（妥当性） 測りたいものを正しく測れていること。例えば、ELISA で抗体の特異性が低いと、測定対象でないものまで測ってしまうことになるので、測定の妥当性が低くなる。途上国で子供の体重を調べる研究において、靴を履いた子供がいてもそのまま体重計に載せて、表示された値を使ったという研究があったが、そういうときの靴は往々にしてブーツなので 1 kg 近くの重さがある場合もあり、表示値そのままでは体重が正しく測れていない。

Accuracy（正確さ） バイアス（系統誤差）が小さいこと。モノサシの原点が狂っていると（目盛幅が狂っているときも)、正しさが損なわれる。例えば、何も載っていない状態で 1 kg と表示されている体重計で体重を測定すると、真の体重が 60 kg の人が載ったときの表示は 61 kg となるだろう。すべての人の体重が 1 kg 重く表示されたら、正しいデータは得られない。

Precision（精度） 確率的な誤差（ランダムな誤差）が小さいこと。信頼区間が狭いことや CV（変動係数）が小さいことも、精度の高さに相当する。測定値の有効数字の桁数も精度を示す値である。例えば、手の大きさを測るのに、通常のものさし（最小表示目盛が 1 mm）で測るのと、ノギス（最小表示目盛 0.1 mm）で測るのでは、最小目盛の 1/10 まで読むとしても、精度が 1 桁異なる。CV は、本来的には同じ値になるはずの複数の測定値の標準偏差（不偏標準偏差ではない）を平均値で割った値を百分率表示で示した値だが、CV が小さい測定は精度が高いといえる。

正確さと精度について、もう少し丁寧に考えてみよう。筆者が学生の頃に所属していたラボでは、ピペッティングが正しくできていることを保証するために、純水を 1 $m\ell$ マイクロピペットで吸って、その重さを電子天秤（校正済みで十分信頼できるとする）で量る操作を 5 回とか 10 回繰り返すことが行われていた。熟練した実験者であれば、5 回ともほぼ 1 グラムになるだろう。仮に 5 回の測定値が、0.998, 0.997, 1.001, 1.002, 1.000 グラムだったとすると、平均値は 0.9996 グラムである。CV を以下のように R で計算すると[a]、0.2% 未満であり、この実験者のピペッティングの精度は十分に高いといえる（当時、1%未満であることが 1 つの基準であった)。

```
> x <- c(0.998, 0.997, 1.001, 1.002, 1.000)
> mx <- mean(x)
> sqrt(sum((x-mx)^2)/5)/mx*100
[1] 0.1855466
```

[a] CV を計算する際に sd(x)/mx としないのは、sd(x) では 5 ではなく 4 で割った不偏標準偏差になってしまって、この目的には不適切なためである。

正確さの方は、真値から系統的にずれていないことが大事なので、例えば、この 5 つの測定値について、母平均が 1 という帰無仮説を 1 標本 t 検定するとよい。R では、`t.test(x, mu=1)` で結果が得られ、`p=0.6885` なので、統計的に有意な差があるとはいえず、この実験者のピペッティングは正確さも高いといえる。
実験を始めたばかりの学部生がピペッティングをすると、量るたびに全然違う値になったり、似たような値だけれども常に少ないといったことが珍しくなかった。例えば、0.950, 0.954, 0.952, 0.955, 0.953 グラムという測定値を出した学生は、CV は先の実験者と同じく 0.2%未満であり、精度は高い。けれども、母平均が 1 という帰無仮説について 1 標本 t 検定をすると、`p=6.605e-07` と、100 万分の 1 よりも小さな p 値が得られ、真値よりも統計的に有意に低い測定値が得られてしまっていることがわかり、この学生のピペッティングは、精度は高いけれども正確さが低いと言える。逆に、1.100, 0.950, 0.910, 1.150, 1.005 という結果を出した学生は、CV が 8.8%なので精度は低いが、母平均を 1 とする 1 標本 t 検定の p 値は 0.6361 となるので平均値と真値の統計的な有意差はなく、正確さは低いとはいえない。**この場合、どちらの学生のピペッティングも信用できない。精度と正確さの両方が十分な水準に到達するまで、ピペッティングを訓練する必要がある。**

7.5 データ入力

研究によって得られたデータをコンピュータを使って統計的に分析するためには、まず、コンピュータにデータを入力する必要がある。データの規模や利用するソ

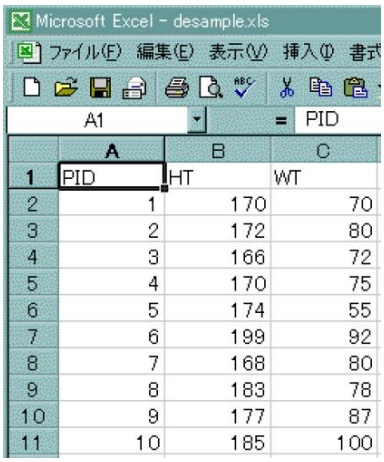
フトウェアによって、どういう入力方法が適当か（正しく入力でき、かつ効率が良いか）は異なってくる。

ごく小さな規模のデータについて単純な分析だけ行う場合、電卓で計算してもよいし、分析する手続きの中で直接数値を入れてしまってもよい。例えば、60 kg, 66 kg, 75 kg という３人の平均体重を R を使って求めるには、プロンプトに対して `mean(c(60,66,75))` または `(60+66+75)/3` と打てばいい。

しかし実際にはもっとサイズの大きなデータについて、いろいろな分析を行う場合が多いので、データ入力と分析は別々に行うのが普通である。そのためには、同じ調査を繰り返しするとか、きわめて大きなデータであるとかでなければ、Microsoft Excel のような表計算ソフトで入力するのが手軽であろう。きわめて単純な例として、10 人の対象者についての身長と体重のデータが次の表のように得られているとする。

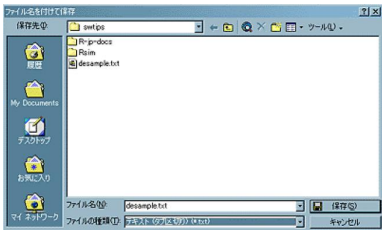
対象者 ID	身長 (cm)	体重 (kg)
1	170	70
2	172	80
3	166	72
4	170	75
5	174	55
6	199	92
7	168	80
8	183	78
9	177	87
10	185	100

まずこれを Microsoft Excel などの表計算ソフトに入力する。一番上の行には変数名を入れる。日本語対応 R なら漢字やカタカナ、ひらがなも使えるが、半角英数字（半角ピリオドも使える）にしておくのが無難である。入力が終わったら、一旦、そのソフトの標準の形式で保存しておく（Excel ならば*.xls 形式、OpenOffice.org の calc ならば*.ods 形式）。入力完了した状態は、次の画面のようになる。



	A	B	C
1	PID	HT	WT
2	1	170	70
3	2	172	80
4	3	166	72
5	4	170	75
6	5	174	55
7	6	199	92
8	7	168	80
9	8	183	78
10	9	177	87
11	10	185	100

次に、この表をタブ区切りテキスト形式で保存する。Microsoft Excel の場合、メニューバーの「ファイル (F)」から「名前を付けて保存」を選び、現れるウィンドウの一番下の「ファイルの種類 (T)」のプルダウンメニューから「テキスト（タブ区切り）(*.txt)」を選ぶと、自動的にその上の行のファイル名の拡張子も xls から txt に変わるので、「保存 (S)」ボタンを押せば OK である（下のスクリーンショットを参照）。複数のシートを含むブックの保存をサポートした形式でないとかいう警告が表示されるが無視して「はい」を選んでよい。その直後に Excel を終了しようとすると、何も変更していないのに「保存しますか」と聞く警告ウィンドウが現れるが、既に保存してあるので「いいえ」と答えてよい（「はい」を選んでも同じ内容が上書きされるだけであり問題はない）。この例では、desample.txt ができる。



R コンソールを使って、このデータを `Dataset` という名前のデータフレームに読み込むのは簡単で、次の 1 行を入力するだけでいい（ただしテキストファイルが保存されているディレクトリが作業ディレクトリになっていなくてはいけない）。

```
Dataset <- read.delim("desample.txt")
```

EZR では「ファイル」「データのインポート」の「ファイルまたはクリップボード、URL からテキストデータを読み込む」を開き、「データセット名を入力：」の欄に適当な参照名をつけ（変数名として使える文字列なら何でもよいのだが、デフォルトでは Dataset となっている）、「フィールドの区切り記号」を「空白」から「タブ」に変えて（「タブ」の右にある○をクリックすればよい）、OK ボタンをクリックしてからデータファイルを選べばよい。
なお、データをファイル保存せず、Excel 上で範囲を選択して「コピー」した直後であれば、「データのインポート」の「テキストファイル又はクリップボードから」を開いてデータセット名を付けた後、「クリップボードからデータを読み込む」の右のチェックボックスにチェックを入れておけば、（データファイルを選ばずに）OK ボタンを押しただけでデータが読み込める。
「ファイル」「データのインポート」の「Excel、Access、dBase のデータをインポート」を開き、「データセット名を入力：」の欄に適当な参照名をつけ、Excel ファイルを開くとシートを選ぶウィンドウが出てくるので、データが入っているシートを選べば、Excel のファイルも読み込める。

7.6 データの図示

データの大局的性質を把握するには、図示をするのが便利である。人間の視覚的認識能力は、パターン認識に関してはコンピュータより遥かに優れていると言われているから、それを生かさない手はない。また、入力ミスをチェックする上でも有効である。つまり、データ入力が終わったら、**何よりも先に図示をすべき**といえる。

変数が表す尺度の種類によって、さまざまな図示の方法がある。離散変数の場合は、度数分布図、積み上げ棒グラフ、帯グラフ、円グラフが代表的である。

survey というデータを使って例示する。EZR では「ツール」の「パッケージの読み込み」から MASS を選んで OK し、「ファイル」「パッケージに含まれるデータを読み込む」から、パッケージとして MASS をダブルクリックし、データセットとして survey をダブルクリックしてから「OK」ボタンをクリックする。

“survey” というデータは、アデレード大学の学生 237 人の調査結果であり、含まれている変数は、Sex（性別：要因型）、Wr.Hnd（字を書く利き手の親指と小指の間隔、cm 単位：数値型）、NW.Hnd（利き手でない方の親指と小指の間隔、cm 単位：数値型）、W.Hnd（利き手：要因型）、Fold（腕を組んだときにどちらが上になるか?: 右が上、左が上、どちらでもない、の 3 水準からなる要因型）、Pulse（心拍数／分：整数型）、Clap（両手を叩き合わせた時、どちらが上にくるか?: 右、左、どちらでもない、の 3 水準からなる要因型）、Exer（運動頻度：頻繁に、

時々、しない、の 3 水準からなる要因型)、Smoke（喫煙習慣：ヘビースモーカー、定期的に吸う、時々吸う、決して吸わない、の 4 水準からなる要因型)、Height（身長：cm 単位の数値型)、M.I（身長の回答がインペリアル（フィート／インチ）でなされたか、メトリック（cm ／ m）でなされたかを示す要因型)、Age（年齢：年単位の数値型）である。

このデータフレームを使って、いくつかのグラフを描いてみよう。

カテゴリ変数の場合

度数分布図 値ごとの頻度を縦棒として、異なる値ごとに、この縦棒を横に並べた図である。離散変数の名前を X とすれば、R では barplot(table(X)) で描画される。

> **EZR では「グラフ」＞「棒グラフ（頻度)」を選び、変数として Smoke を選ぶと、喫煙習慣ごとの人数がプロットされる。**

積み上げ棒グラフ 値ごとの頻度の縦棒を積み上げた図である。R では

```
fx <- table(X)
barplot(matrix(fx,NROW(fx)),beside=F)
```

で描画される。**Rcmdr や EZR では描けない。**

帯グラフ 横棒を全体を 100%として各値の割合にしたがって区切って塗り分けた図である。R では

```
px <- table(X)/NROW(X)
barplot(matrix(pc,NROW(pc)),horiz=T,beside=F)
```

で描画される。**これも Rcmdr や EZR では描けない。**

7.7 連続変数の場合

ヒストグラム 変数値を適当に区切って度数分布を求め、分布の様子を見るものである。R では hist() 関数を用いる。デフォルトでは「適当な」区切り方として “Sturges” というアルゴリズムが使われるが、明示的に区切りを与えることもできる。また、デフォルトでは区間が「〜を超えて〜以下」であり、日本で普通に用いられる「〜以上〜未満」ではないことにも注意されたい。「〜以上〜未満」にしたいときは、right=FALSE というオプションを付ければ良い。R コンソールで年齢 (Age) のヒストグラムを描かせるには、hist(survey$Age) だが、「10 歳以上 20 歳未満」から 10 歳ごとの区切りでヒストグラムを描くよう

に指定するには、hist(survey$Age, breaks=1:8*10, right=FALSE) とする。

Rcmdr や EZR では「グラフ」の「ヒストグラム」を選ぶ。survey データでは、変数として Age を選べば、年齢のヒストグラムが描ける（アデレード大学の学生のデータのはずだが、70 歳以上の人や 16.75 歳など、大学生らしくない年齢の人も含まれている）。Rcmdr や EZR では「〜以上〜未満」にはできない（裏技的には、予め「〜以上〜未満」のカテゴリデータに変換しておき、「棒グラフ（頻度）」で描画することはできる）。

正規確率プロット 連続変数が正規分布しているかどうかを見るものである（正規分布に当てはまっていれば点が直線上に並ぶ）。R では qqnorm() 関数を用いる。例えば、survey データフレームの心拍数 (Pulse) について正規確率プロットを描くには、qqnorm(survey$Pulse) とする。

EZR では QQ プロットはできないが、「統計解析」「連続変数の解析」「正規性の検定 (Kolmogorov-Smirnov 検定)」で検定と同時にヒストグラムに正規分布の曲線を重ね描きしてくれるので、正規分布と見なせるかどうかは見当がつく。

幹葉表示 (stem and leaf plot) 大体の概数（整数区切りとか 5 の倍数とか 10 の倍数にすることが多い）を縦に並べて幹とし、それぞれの概数に相当する値の細かい部分を葉として横に並べて作成する図。R では stem() 関数を用いる。同じデータで心拍数の幹葉表示をするには、stem(survey$Pulse) とする。

EZR では「グラフ」「幹葉表示」を選び、「変数（１つ選択）」の枠内から Pulse を選んで「OK」ボタンをクリックすれば Output ウィンドウにテキスト出力される。さまざまなオプションが指定可能である。

箱ヒゲ図 (box and whisker plot) 縦軸に変数値をとって、第１四分位を下に、第３四分位を上にした箱を書き、中央値の位置にも線を引いて、さらに第１四分位と第３四分位の差（四分位範囲）を 1.5 倍した線分をヒゲとして第１四分位の下と第３四分位の上に伸ばし、ヒゲの先より外れた値を外れ値として○をプロットした図である。カテゴリによって層別した箱ヒゲ図を横に並べて描くと、大体の分布の様子と外れ値の様子が同時に比較できるので便利である。R では boxplot() 関数を用いる。例えば、sur-

vey データで喫煙状況 (Smoke) 別に心拍数 (Pulse) の箱ヒゲ図を描くには、`boxplot(survey$Pulse ~ survey$Smoke)` とする。

EZR では「グラフ」の「箱ひげ図」を選ぶ。survey データで喫煙状況別に心拍数の箱ひげ図を描かせるには、「群別する変数（0〜1 つ選択）」で Smoke を選び、上下のひげの位置として「第 1 四分位数-1.5x 四分位範囲、第 3 四分位数+1.5x 四分位範囲」の左のラジオボタンをチェックして「OK」をクリックするだけでよい。似た用途のグラフとして、層別の平均とエラーバーを表示して折れ線で結ぶことも「グラフ」の「棒グラフ（平均値）」または「折れ線グラフ（平均値）」でできる。survey データで喫煙習慣ごとに心拍数の平均値とエラーバーを表示して折れ線で結びたいなら、「因子」として Smoke、「目的変数」として Pulse を選べばよい。エラーバーとしては標準誤差（デフォルト）、標準偏差、信頼区間から選択できる。

散布図 (scatter plot) ２つの連続変数の関係を２次元の平面上の点として示した図である。R では `plot()` 関数を用いる。異なる群ごとに別々のプロットをしたい場合は `plot()` の `pch` オプションで塗り分けたり、`points()` 関数を使って重ね打ちしたりできる。点ごとに異なる情報を示したい場合は `symbols()` 関数を用いることができるし、複数の連続変数間の関係を調べるために、重ね描きしたい場合は `matplot()` 関数と `matpoints()` 関数を、別々のグラフとして並べて同時に示したい場合は `pairs()` 関数を用いることができる。データ点に文字列を付記したい場合は `text()` 関数が使えるし、マウスで選んだデータ点にだけ文字列を付記したい場合は `identify()` 関数が使える。survey データで、R コンソールを使って、横軸に年齢 (Age)、縦軸に身長 (Height) をとってプロットしたいときは、`plot(Height ~ Age, data=survey)` と打てば良い。

EZR では「グラフ」の「散布図」で描ける（「散布図行列」メニューで `pairs()` の機能も実装されている）。層別にマークを変えることもできる。

7.8 記述統計・分布の正規性・外れ値

記述統計は、(1) データの特徴を把握する目的、また (2) データ入力ミスの可能性をチェックする目的で計算する。あまりにも妙な最大値や最小値、大きすぎる標準偏差などが得られた場合は、入力ミスを疑って、元データに立ち返ってみるべきである。

記述統計量には、大雑把にいって、分布の位置を示す「中心傾向」と分布の広がりを示す「ばらつき」があり、中心傾向としては平均値、中央値、最頻値がよく用いられ、ばらつきとしては分散、標準偏差、四分位範囲、四分位偏差がよく用いられる。

中心傾向の代表的なものは以下の３つである。

平均値（mean）分布の位置を示す指標として、もっとも頻繁に用いられる。実験的仮説検証のためにデザインされた式の中でも、頻繁に用いられる。記述的な指標の１つとして、平均値は、いくつかの利点と欠点をもっている。日常生活の中でも平均をとるという操作は普通に行われるから説明不要かもしれないが、数式で書くと以下の通りである。

母集団の平均値 μ（ミューと発音する）は、

$$\mu = \frac{\sum X}{N}$$

である。X はその分布における個々の値であり、N は値の総数である。$\sum$（シグマと発音する）は、一群の値の和を求める記号である。すなわち、$\sum X = X_1 + X_2 + X_3 + ... + X_N$ である。

標本についての平均値を求める式も、母集団についての式と同一である。ただし、数式で使う記号が若干異なっている。標本平均 $\bar{X}$（エックスバーと発音する）は、

$$\bar{X} = \frac{\sum X}{n}$$

である。n は、もちろん標本サイズである[32]。

ちなみに、重み付き平均は、各々の値にある重みをかけて合計したものを、重みの合計で割った値である。式で書くと、

$$\bar{X} = \frac{n_1(\bar{X}_1) + n_2(\bar{X}_2) + ... + n_n(\bar{X}_n)}{n_1 + n_2 + ... + n_n}$$

中央値（median）中央値は、全体の半分がその値より小さく、半分がその値より大きい、という意味で、分布の中央である。言い換えると、中央値は、頻度あるいは値の数に基づいて分布を２つに等分割する値である。中央値を求めるには式は使わない（決まった手続き＝アルゴリズムとして、並べ替え (sorting) は必要）。極端な外れ値の影響を受けにくい（言い換えると、外れ値に対して頑健である）。歪んだ分布に対する最も重要な central tendency の指標が中央値である。R で中央値を計算するには、`median()` という関数

[32]記号について注記しておくと、集合論では $\bar{X}$ は集合 X の補集合の意味で使われるが、代数では確率変数 X の標本平均が $\bar{X}$ で表されるということである。同じような記号が別の意味で使われるので混乱しないように注意されたい。補集合は X^C という表記がなされる場合も多いようである。標本平均は $\bar{X}$ と表すのが普通である。

を使う。なお、データが偶数個の場合は、普通は中央にもっとも近い２つの値を平均した値を中央値として使うことになっている。

最頻値（Mode） 最頻値はもっとも度数が多い値である。すべての値の出現頻度が等しい場合は、最頻値は存在しない。R では `table(X)[which.max(table(X))]` で得られる（ただし、複数の最頻値がある場合は、これだと最も小さい値しか表示されないので要注意）。

平均値は、(1) 分布のすべての値を考慮した値である、(2) 同じ母集団からサンプリングを繰り返した場合に一定の値となる、(3) 多くの統計量や検定で使われている、という特長をもつ。標本調査値から母集団の因果関係を推論したい場合に、もっとも普通に使われる。しかし、(1) 極端な外れ値の影響を受けやすい、(2) 打ち切りのある分布では代表性を失う場合がある[33]、という欠点があり、外れ値があったり打ち切りがあったりする分布では位置の指標として中央値の方が優れている。最頻値は、標本をとったときの偶然性の影響を受けやすいし、もっとも頻度が高い値以外の情報はまったく使われない。しかし、試験の点で何点の人が多かったかを見たい場合は最頻値が役に立つし、名義尺度については最頻値しか使えない。

ここで上げた 3 つの他に、幾何平均 (geometric mean) や調和平均 (harmonic mean) も、分布の位置の指標として使われることがある。幾何平均はデータの積の累乗根（対数をとって平均値を出して元に戻したもの）、調和平均はデータの逆数の平均値の逆数であり、どちらもゼロを含むデータには使えない。大きな外れ値の影響を受けにくいという利点があり、幾何平均は、とくにデータの分布が対数正規分布に近い場合によく用いられる。

一方、分布のばらつき（Variability）の指標として代表的なものは、以下の 4 つである。

四分位範囲 (Inter-Quartile Range; IQR) 四分位範囲について説明する前に、分位数について説明する。値を小さい方から順番に並べ替えて、４つの等しい数の群に分けたときの 1/4, 2/4, 3/4 にあたる値を、四分位数 (quartile) という。1/4 の点が第１四分位、3/4 の点が第３四分位である（つまり全体の 25% の値が第１四分位より小さく、全体の 75%の値が第３四分位より小さい）。2/4 の点というのは、ちょうど順番が真中ということだから、第２四分位は中央値に等しい。ちょっと考えればわかるように、ちょうど４等分などできない場合がもちろんあって、上から数えた場合と下から数えた場合で四分位数がずれる可能性があるが、その場合はそれらを平均するのが普通である。また、最小値、最大値に、第１四分位、第３四分位と中央値を加えた

[33] 氷水で痛みがとれるまでにかかる時間とか、年収とか。無限に観察を続けるわけにはいかないし、年収は下限がゼロで上限はビル・ゲイツのそれのように極端に高い値があるから右すそを長く引いた分布になる。平均年収を出している統計表を見るときは注意が必要である。年収の平均的な水準は中央値で表示されるべきである。

５つの値を五数要約値と呼ぶことがある（R では `fivenum()` 関数で五数要約値を求めることができる）。第１四分位、第２四分位、第３四分位は、それぞれ Q1, Q2, Q3 と略記することがある。四分位範囲とは、第３四分位と第１四分位の間隔である。上と下の極端な値を排除して、全体の中央付近の 50%（つまり代表性が高いと考えられる半数）が含まれる範囲を示すことができる。

四分位偏差 (Semi Inter-Quartile Range; SIQR) 四分位範囲を２で割った値を四分位偏差と呼ぶ。もし分布が左右対称型の正規分布であれば、中央値マイナス四分位偏差から中央値プラス四分位偏差までの幅に全データの半分が含まれるという意味で、四分位偏差は重要な指標である。IQR も SIQR も少数の極端な外れ値の影響を受けにくいし、分布が歪んでいても使える指標である。

分散 (variance) データの個々の値と平均値との差を偏差というが、マイナス側の偏差とプラス側の偏差を同等に扱うために、偏差を二乗して、その平均をとると、分散という値になる。分散 V は、

$$V = \frac{\sum (X - \mu)^2}{N}$$

で定義される[34]。標本数 n で割る代わりに自由度 $n-1$ で割って、不偏分散 (unbiased variance) という値にすると、標本データから母集団の分散を推定するのに使える。即ち、不偏分散 V_{ub} は、

$$V_{ub} = \frac{\sum (X - \bar{X})^2}{n - 1}$$

である（R では `var()` で得られる）。

標準偏差 (standard deviation) 分散の平方根をとったものが標準偏差である。**平均値と次元を揃える意味**をもつ。不偏分散の平方根をとったものは、不偏標準偏差と呼ばれる（R では `sd()` で得られる）[35]。もし分布が正規分布ならば、Mean±2SD[36]の範囲にデータの 95%が含まれるという意味で、標準偏差は便利な指標である。なお、名前は似ているが、「標準誤差」はデータのばらつきでなくて、推定値のばらつきを示す値なので混同しないように注意されたい。例えば、平均値の標準誤差は、サンプルの不偏標準偏差をサンプルサイズの平方根で割れば得られるが、意味は、「もし標本抽出を何度も繰り返して行ったとしたら、得られる標本平均のばらつきは、一定の確率で標準誤差の範囲におさまる」ということである。

[34]実際に計算するときは２乗の平均から平均の２乗を引くとよい。

[35]不偏分散は母分散の不偏推定量だが、不偏標準偏差は不偏分散の平方根なので分散の平方根と区別する意味で不偏標準偏差と呼ばれるだけであって、一般に母標準偏差の不偏推定量ではない。

[36]普通このように 2SD と書かれるが、正規分布の 97.5 パーセント点は 1.959964... なので、この 2 は、だいたい 2 くらいという意味である。

上記の記述統計量を計算するには、EZR では、「統計解析＞連続変数の解析＞連続変数の要約」を選べばよいが、新しいバージョンでは「グラフと表＞サンプルの背景データのサマリー表の出力」の方が便利かもしれない。

分布の正規性の検定は、EZR では、「統計解析」の「連続変数の解析」の「正規性の検定」を選び、変数として wbc を選んで OK ボタンをクリックするだけでできる。自動的にコルモゴロフ＝スミルノフ検定とシャピロ＝ウィルク検定の結果が表示される。この例では結果が異なるが、これらの検定は分布の正規性へのアプローチが異なるので、結果が一致しないこともある。多くの検定手法がデータの分布の正規性を仮定しているが、この検定で正規性が棄却されたからといって機械的に変数変換やノンパラメトリック検定でなくてはいけないとは限らない。独立 2 標本の平均値の差がないという仮説を検定するための t 検定は、かなり頑健な手法なので、正規分布に従っているといえなくても、そのまま実行してもいい場合も多い。

外れ値の検定は、EZR では、「統計解析」の「連続変数の解析」の「外れ値の検定」を選んで、変数として wbc を指定して OK ボタンをクリックするだけでいい。外れ値を NA で置き換えた新しい変数を作成することも右のオプション指定から容易にできる。しかし、既に書いた通り、この結果だけで機械的に外れ値を除外することはお薦めできない。また、この例では “No outliers were identified.” と表示されるので、統計学的に外れ値があるとはいえない。

7.9 独立２標本の差の検定

既に屋外空気の採取場所間での二酸化炭素濃度の差の検定の例などを示したが、独立 2 群間の統計的仮説検定の方法は、以下のようにまとめられる。

1. 量的変数の場合
   (a) 正規分布に近い場合（`shapiro.test()` で Shapiro-Wilk の検定ができるが、その結果を機械的に適用して判断すべきではない）：Welch の検定（R では `t.test(x,y)`）[37]
   (b) 正規分布とかけ離れている場合:Wilcoxon の順位和検定（R では `wilcox.test(x,y)`）
2. カテゴリ変数の場合：母比率の差の検定（R では `prop.test()`）。ただし実は、群別変数とカテゴリ変数が独立かどうかという検定と同等なので、

[37] それに先立って 2 群の間で分散に差がないという帰無仮説で F 検定し、あまりに分散が違いすぎる場合は、平均値の差の検定をするまでもなく、2 群が異なる母集団からのサンプルと考えられるので、平均値の差の検定には意味がないとする考え方もある。また、かつては、まず F 検定して 2 群間で分散に差がないときは通常の t 検定、差があれば Welch の検定、と使い分けるべきという考え方が主流だったが、群馬大学社会情報学部青木繁伸教授や三重大学奥村晴彦教授のシミュレーション結果により、F 検定の結果によらず、平均値の差の検定をしたいときは常に Welch の検定をすればよいことがわかっている。

フィッシャーの直接確率検定（R では fisher.test()）でも分析できる。

等分散性についての F 検定

まず、標本調査によって得られた独立した 2 つの量的変数 X と Y（サンプルサイズが各々 n_X と n_Y とする）について、平均値に差があるかどうかを検定することを考える。

2 つの量的変数 X と Y の不偏分散 SX<-var(X) と SY<-var(Y) の大きい方を小さい方で（以下の説明では SX>SY だったとする）割った F0<-SX/SY が第 1 自由度 DFX<-length(X)-1、第 2 自由度 DFY<-length(Y)-1 の F 分布に従うことを使って検定する。有意確率は 1-pf(F0,DFX,DFY) で得られる。しかし、F0 を手計算しなくても、var.test(X,Y) で等分散かどうかの検定が実行できる。また、1 つの量的変数 X と 1 つの群分け変数 C があって、C の 2 群間で X の分散が等しいかどうか検定するというスタイルでデータを入力してある場合は、var.test(X~C) とすればよい。

EZR では、「統計解析」の「連続変数の解析」から「2 群の等分散性の検定（F 検定）」を選び、目的変数（1 つ選択）の枠から X を、グループ（1 つ選択）の枠から C を選び（EZR では要因型にしなくても選べる）, OK ボタンをクリックする。survey データで、「男女間で身長の分散に差がない」という帰無仮説を検定するには、目的変数として Height を、グループとして Sex を選び、OK ボタンをクリックすると、男女それぞれの分散と検定結果が Output ウィンドウに表示される。

Welch の方法による t 検定

$t_0 = |E(X) - E(Y)| / \sqrt{S_X/n_X + S_Y/n_Y}$ が自由度 ϕ の t 分布に従うことを使って検定する。但し、ϕ は下式による。

$$\phi = \frac{(S_X/n_X + S_Y/n_Y)^2}{\{(S_X/n_X)^2/(n_X - 1) + (S_Y/n_Y)^2/(n_Y - 1)\}}$$

R では、t.test(X,Y,var.equal=F) だが、var.equal の指定を省略した時は等分散でないと仮定して Welch の検定がなされるので省略して t.test(X,Y) でいい。量的変数 X と群分け変数 C という入力の仕方の場合は、t.test(X~C) とする。survey データで「男女間で平均身長に差がない」という帰無仮説を検定したいときは、t.test(Height ~ Sex, data=survey) とする。

EZR の場合は「統計解析」「連続変数の解析」から「2 群間の平均値の比較（t 検定）」を選び、目的変数として Height を、比較する群として Sex を選び、「等分散と考えますか？」の下のラジオボタンを「No (Welch test)」の方をチェックして、「OK」ボタンをクリックすると、結果が Output ウィンドウに表示される。男女それぞれの平均、不偏標準偏差と検定結果の p 値が示され、エラーバーが上下に付いた平均値を黒丸でプロットし、それを直線で結んだグラフも自動的に描かれる。

なお、既に平均値と不偏標準偏差が計算されている場合の図示は、エラーバー付きの棒グラフがよく使われるが（R では、barplot() 関数で棒グラフを描画してから、arrows() 関数でエラーバーを付ける）、棒グラフを描く時は基線をゼロにしなくてはいけないことに注意されたい。生データがあれば、stripchart() 関数を用いて、生データのストリップチャートを描き、その脇に平均値とエラーバーを付け足す方がよい。そのためには、量的変数と群別変数という形にしなくてはいけないので、たとえば、2つの量的変数 V <- rnorm(100,10,2) と W <- rnorm(60,12,3) があったら、予め

```
X <- c(V, W)
C <- as.factor(c(rep("V", length(V)), rep("W", length(W))))
x <- data.frame(X, C)
```

または

```
x <- stack(list(V=V, W=W))
names(x) <- c("X", "C")
```

のように変換しておく必要がある[38]。プロットするには次のように入力すればよい[39]。

```
stripchart(X~C, data=x, method="jitter", vert=TRUE)
Mx <- tapply(x$X, x$C, mean)
Sx <- tapply(x$X, x$C, sd)
Ix <- c(1.1, 2.1)
points(Ix, Mx, pch=18, cex=2)
arrows(Ix, Mx-Sx, Ix, Mx+Sx, angle=90, code=3)
```

[38] この操作は、EZR でも「アクティブデータセット」の「変数の操作」から「複数の変数を縦に積み重ねたデータセットを作成する」を選べば簡単に実行できる。

[39] EZR では「グラフ」「ドットチャート」でプロットする変数として X、群分け変数として C を選ぶことで、生データについては jitter ではないが似たグラフを描くことができる。また、平均値とエラーバーを線で結んだグラフは「グラフ」「折れ線グラフ（平均値）」で描くことができる。ただし、両者を重ね合わせることは、2013 年 8 月時点の EZR のメニューからはできない。

対応のある２標本の平均値の差の検定

各対象について２つずつの値があるときは、それらを独立２標本とみなすよりも、対応のある２標本とみなす方が切れ味がよい。全体の平均に差があるかないかだけをみるのではなく、個人ごとの違いを見るほうが情報量が失われないのは当然である。

対応のある２標本の差の検定は、paired-*t* 検定と呼ばれ、意味合いとしてはペア間の値の差を計算して値の差の母平均が 0 であるかどうかを調べることになる。R で対応のある変数 `X` と `Y` の paired-*t* 検定をするには、`t.test(X,Y,paired=T)` で実行できるし、それは `t.test(X-Y,mu=0)` と等価である。

survey データで「親指と小指の間隔が利き手とそうでない手の間で差がない」という帰無仮説を検定するには、R コンソールでは、

```
t.test(survey$Wr.Hnd, survey$NW.Hnd, paired=TRUE)
```

と打てばよい。グラフは通常、同じ人のデータは線で結ぶので、例えば次のように打てば、差が 1 cm 以内の人は黒、利き手が 1 cm 以上非利き手より大きい人は赤、利き手が 1 cm 以上非利き手より小さい人は緑で、人数分の線分が描かれる。

```
Diff.Hnd <- survey$Wr.Hnd - survey$NW.Hnd
C.Hnd <- ifelse(abs(Diff.Hnd)<1, 1, ifelse(Diff.Hnd>0, 2, 3))
matplot(rbind(survey$Wr.Hnd, survey$NW.Hnd),
 type="l", lty=1, col=C.Hnd, xaxt="n")
axis(1, 1:2, c("Wr.Hnd", "NW.Hnd"))
```

EZR では「統計解析」「連続変数の解析」から「対応のある 2 群間の平均値の検定（paired t 検定）」を選ぶ。第１の変数として `Wr.Hnd` を、第２の変数として `NW.Hnd` を選び、[OK] ボタンをクリックすると、Output ウィンドウに結果が得られる。有意水準 5% で帰無仮説は棄却され、利き手の方がそうでない手よりも親指と小指の間隔が有意に広いといえる。

Wilcoxon の順位和検定

Wilcoxon の順位和検定は、データが外れ値を含んでいる場合や正規分布から大きく外れた分布である場合に用いられる、連続分布であるという以外にデータの分布についての仮定を置かない（ノンパラメトリックな）検定の代表的な方法である。パラメトリックな検定でいえば、*t* 検定を使うような状況、つまり、独立２標本の分布の位置に差がないかどうかを調べるために用いられる。Mann-Whitney の U 検定と（これら２つほど有名ではないが、Kendall の S 検定とも）数学的に等価である。データがもつ情報の中で、単調変換に対して頑健なのは順位なので、これを使って検定しようという発想であるが、本稿では詳細な原理の説明は省略

する。

例として、survey データで、身長 (Height) の分布の位置が男女間で差がないという帰無仮説を検定してみよう。R コンソールでは以下の通り簡単である。

```
wilcox.test(Height ~ Sex, data=survey)
```

EZR では「統計解析」の「ノンパラメトリック検定」から「2 群間の比較（Mann-Whitney U 検定）」を選び、目的変数として Height を、比較する群として Sex を選んで「OK」ボタンをクリックする。検定結果が Output ウィンドウに表示されるだけでなく、箱ひげ図も同時に描かれる。

Wilcoxon の符号付き順位検定

Wilcoxon の符号付き順位検定は、対応のある t 検定のノンパラメトリック版である。ここでは説明しないが、多くの統計学の教科書に載っている。

実例だけ出しておく。survey データには、利き手の大きさ（親指と小指の先端の距離）を意味する Wr.Hnd という変数と、利き手でない方の大きさを意味する NW.Hnd という変数が含まれているので、これらの分布の位置に差が無いという帰無仮説を有意水準 5%で検定してみよう。

同じ人について利き手と利き手でない方の手の両方のデータがあるので対応のある検定が可能になる。R コンソールでは以下の通り。

```
wilcox.test(survey$Wr.Hnd, survey$NW.Hnd, paired=TRUE)
```

EZR では、「統計解析」「ノンパラメトリック検定」から「対応のある 2 群間の比較（Wilcoxon の符号付き順位和検定）」を選び、第 1 の変数として左側のリストから Wr.Hnd を選び、第 2 の変数として右側のリストから NW.Hnd を選んで [OK] ボタンをクリックするだけである。順位和検定のときと同じく検定方法のオプションを指定できるが、通常はデフォルトで問題ない。

7.10 3 群以上の比較

実験計画の二酸化窒素濃度の比較の例で既に説明した通り、3 群以上を比較するために、**単純に 2 群間の差の検定を繰り返すことは誤り**である。なぜなら、n 群から 2 群を抽出するやりかたは ${}_nC_2$ 通りあって、1 回あたりの第 1 種の過誤（本当は差がないのに、誤って差があると判定してしまう確率）を 5%未満にしたとしても、3 群以上の比較全体として「少なくとも 1 組の差のある群がある」というと、全体としての第 1 種の過誤が 5%よりずっと大きくなってしまうからである。

この問題を解消するには、**多群間の比較という捉え方をやめて、群分け変数が注目している量の変数に与える効果があるかどうかという捉え方にする**のが一つの

方法であり、具体的には一元配置分散分析やクラスカル＝ウォリス (Kruskal-Wallis) の検定がこれに当たる。

そうでなければ、有意水準 5%の 2 群間の検定を繰り返すことによって全体として第 1 種の過誤が大きくなってしまうことが問題なので、**第 1 種の過誤を調整することによって全体としての検定の有意水準を 5%に抑える**方法もある。このやり方は「多重比較法」と呼ばれる。

一元配置分散分析

一元配置分散分析では、データのばらつき（変動）を、群間の違いという意味のはっきりしているばらつき（群間変動）と、各データが群ごとの平均からどれくらいばらついているか（誤差）をすべての群について合計したもの（誤差変動）に分解して、前者が後者よりもどれくらい大きいかを検討することによって、群分け変数がデータの変数に与える効果があるかどうかを調べる。

例えば、南太平洋の３つの村 X, Y, Z で健診をやって、成人男性の身長や体重を測ったとしよう。このとき、データは例えば次のようになる（架空のものである）[40]。

ID 番号	村落 (VG)	身長 (cm)(HEIGHT)
1	X	161.5
2	X	167.0
（中略）		
22	Z	166.0
（中略）		
37	Y	155.5

村落によって身長に差があるかどうかを検定したいならば、HEIGHT という量的変数に対して、VG という群分け変数の効果があるかどうかを一元配置分散分析することになる。R コンソールでは以下のように入力する。

```
> sp <- read.delim("http://minato.sip21c.org/grad/sample2.dat")
> summary(aov(HEIGHT ~ VG, data=sp))
```

すると、次の枠内に示す「分散分析表」が得られる。

```
            Df  Sum Sq Mean Sq F value   Pr(>F)
VG           2  422.72  211.36  5.7777 0.006918 **
Residuals   34 1243.80   36.58
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

右端の*の数は有意性を示す目安だが、確率そのものに注目してみるほうがよい。Sum Sq のカラムは偏差平方和を意味する。VG の Sum Sq の値 422.72 は、村

[40] http://minato.sip21c.org/grad/sample2.dat として公開してあり、R から read.delim() 関数で読み込み可能な筈である。

ごとの平均値から総平均を引いて二乗した値を村ごとの人数で重み付けした和であり、群間変動または級間変動と呼ばれ、VG 間でのばらつきの程度を意味する。Residuals の Sum Sq の値 1243.80 は各個人の身長からその個人が属する村の平均身長を引いて二乗したものの総和であり、誤差変動と呼ばれ、村によらない（それ以外の要因がないとすれば偶然の）ばらつきの程度を意味する。Mean Sq は平均平方和と呼ばれ、偏差平方和を自由度 (Df) で割ったものである。平均平方和は分散なので、VG の Mean Sq の値 211.36 は群間分散または級間分散と呼ばれることがあり、Residuals の Mean Sq の値 36.58 は誤差分散と呼ばれることがある。F value は分散比と呼ばれ、群間分散の誤差分散に対する比である。この場合の分散比は第 1 自由度 2、第 2 自由度 34 の F 分布に従うことがわかっているので、それを使った検定の結果、分散比がこの実現値よりも偶然大きくなる確率（Pr(>F) に得られる）が得られる。この例では 0.006918 なので、VG の効果は 5% 水準で有意であり、帰無仮説は棄却される。つまり、身長は村落によって有意に異なることになる。

EZR で sample2.dat **を** sp **というデータフレームに読み込むには、[ファイル] の [データのインポート] から [テキストファイル、クリップボードまたは URL から] と進んで、[データフレーム名を入力:] のところに** sp **と打ち、[インターネット URL] の右側のラジオボタンをチェックし、フィールド区切りを [タブ] として [OK] をクリックして表示されるダイアログに** http://minato.sip21c.org/grad/sample2.dat **と入力して [OK] する。**
ANOVA を実行するには、[統計解析] の [連続変数の解析] で [三群以上の間の平均値の比較（一元配置分散分析 one-way ANOVA)] を選び、「目的変数」として HEIGHT **を、「比較する群」として** VG **を選び、[OK] をクリックすればよい。エラーバー付きの棒グラフが自動的に描かれ、アウトプットウィンドウには分散分析表に続いて、村ごとの平均値と標準偏差の一覧表が表示される。右端の p 値は一元配置分散分析における** VG **の効果の検定結果を再掲したものになっている。なお、実行前に、「等分散と考えますか？」というオプション項目について、「いいえ（Welch 検定）」の方をチェックしておくと、等分散性を仮定する通常の一元配置分散分析でなく、Welch の拡張による一元配置分散分析の関数である** oneway.test() **を実行してくれる。**

古典的な統計解析では、各群の母分散が等しいことを確認しないと一元配置分散分析の前提となる仮定が満たされない。母分散が等しいという帰無仮説を検定するには、バートレット (Bartlett) の検定と呼ばれる方法がある。R では、量的変数を Y、群分け変数を C とすると、bartlett.test(Y~C) で実行できる（もちろん、これらがデータフレーム dat に含まれる変数ならば、bartlett.test(Y~C, data=dat) とする）。この結果得られる p 値は 0.5785 なので、母分散が等しいという帰無仮説は有意水準 5%で棄却されない。これを確認できると、安心して一元配置分散分析が実行できる。

EZR では、メニューバーの「統計解析」から「連続変数の解析」の「三群以上の等分散性の検定 (Bartlett 検定)」を選び、「目的変数」として HEIGHT、「グループ」として VG を選んで [OK] する。

しかし、このような 2 段階の検定は、検定の多重性の問題を起こす可能性がある。群馬大学の青木繁伸教授や三重大学の奥村晴彦教授の数値実験によると、等分散であるかどうかにかかわらず、2 群の平均値の差の Welch の方法を多群に拡張した方法を用いるのが最適である。R では `oneway.test()` で実行できる（EZR でもできることは既に述べた）。上記、村落の身長への効果をみる例では、`oneway.test(HEIGHT ~ VG, data=sp)` と打てば、Welch の拡張による一元配置分散分析ができて、以下の結果が得られる。

```
> oneway.test(HEIGHT ~ VG, data=sp)

One-way analysis of means (not assuming equal variances)

data:  HEIGHT and VG
F = 7.5163, num df = 2.00, denom df = 18.77, p-value = 0.004002
```

クラスカル＝ウォリス (Kruskal-Wallis) の検定と Fligner-Killeen の検定

多群間の差を調べるためのノンパラメトリックな方法としては、クラスカル＝ウォリス (Kruskal-Wallis) の検定が有名である。R では、量的変数を Y、群分け変数を C とすると、`kruskal.test(Y~C)` で実行できる。以下、Kruskal-Wallis の検定の仕組みを箇条書きで説明する。

- 「少なくともどれか 1 組の群間で大小の差がある」という対立仮説に対する「すべての群の間で大小の差がない」という帰無仮説を検定する。
- まず 2 群の比較の場合の順位和検定と同じく、すべてのデータを込みにして小さい方から順に順位をつける（同順位がある場合は平均順位を与える）。
- 次に、各群ごとに順位を足し合わせて、順位和 $R_i (i = 1, 2, ..., k; k$ は群の数) を求める。
- 各群のオブザーベーションの数をそれぞれ n_i とし、全オブザーベーション数を N としたとき、各群について統計量 B_i を $B_i = n_i\{R_i/n_i - (N+1)/2\}^2$ として計算し、

$$B = \sum_{i=1}^{k} B_i$$

として B を求め、$H = 12 \cdot B/\{N(N+1)\}$ として H を求める。同順位を含むときは、すべての同順位の値について、その個数に個数の 2 乗から 1 を引

いた値を掛けたものを計算し、その総和を A として、

$$H' = \frac{H}{1 - \frac{A}{N(N^2-1)}}$$

により H を補正した値 H' を求める。

- H または H' から表を使って（データ数が少なければ並べかえ検定によって）有意確率を求めるのが普通だが、$k \geq 4$ で各群のオブザーベーション数が最低でも 4 以上か、または $k = 3$ で各群のオブザーベーション数が最低でも 5 以上なら、H や H' が自由度 $k - 1$ のカイ二乗分布に従うものとして検定できる。

上の例で村落の身長への効果をみるには、R コンソールでは以下の通り。

```
kruskal.test(HEIGHT ~ VG, data=sp)
```

EZR では、「統計解析」、「ノンパラメトリック検定」、「3 群以上の間の比較 (Kruskal-Wallis の検定)」と選び、「グループ」として VG を、「目的変数」として HEIGHT を選び、[OK] をクリックするだけである。

Fligner-Killeen の検定は、グループごとのばらつきに差が無いという帰無仮説を検定するためのノンパラメトリックな方法である。Bartlett の検定のノンパラメトリック版といえる。上の例で、身長のばらつきに村落による差が無いという帰無仮説を検定するには、R コンソールでは、`fligner.test(HEIGHT ~ VG, data=sp)` とすればよい。

EZR のメニューには入っていないので、必要な場合はスクリプトウィンドウにコマンドを打ち、選択した上で「実行」ボタンをクリックする。

検定の多重性の調整を伴う対比較

多重比較の方法にはいろいろあるが、良く使われているものとして、ボンフェローニ (Bonferroni) の方法、ホルム (Holm) の方法、シェフェ(Scheffé) の方法、テューキー (Tukey) の HSD、ダネット (Dunnett) の方法、ウィリアムズ (Williams) の方法がある。最近は、FDR(False Discovery Rate) 法もかなり使われるようになった。ボンフェローニの方法とシェフェの方法は検出力が悪いので、特別な場合を除いては使わない方が良い。テューキーの HSD またはホルムの方法が薦められる。なお、ダネットの方法は対照群が存在する場合に対照群と他の群との比較に使われるので、適用場面が限定されている[41]。ウィリアムズの方法は対照群があって

[41]ただし、対照群が他の群との比較のすべての場合において差があるといいたい場合は、多重比較をするのではなくて、t 検定を繰り返して使うのが正しいので、注意が必要である。もちろんそういう場合は多くはない。

他の群にも一定の傾向が仮定される場合には最高の検出力を発揮するが、ダネットの方法よりもさらに限られた場合にしか使えない。

テューキーの HSD は平均値の差の比較にしか使えないが、ボンフェロー二の方法、ホルムの方法、FDR 法は位置母数のノンパラメトリックな比較にも、割合の差の検定にも使える。R コンソールでは、`pairwise.t.test()`、`pairwise.wilcox.test()`、`pairwise.prop.test()` という関数で、ボンフェローニの方法、ホルムの方法、FDR 法による検定の多重性の調整ができる。

fmsb ライブラリを使えば、`pairwise.fisher.test()` により、Fisher の直接確率法で対比較をした場合の検定の多重性の調整も可能である。

なお、Bonferroni のような多重比較法で p 値を調整して表示するのは表示上の都合であって、本当は帰無仮説族レベルでの有意水準を変えているのだし、p.adjust.method="fdr"でも、p 値も有意水準も調整せず、帰無仮説の下で偶然 p 値が有意水準未満になって棄却されてしまう確率（誤検出率）を計算し、帰無仮説ごとに有意水準に誤検出率を掛けて p 値との大小を比較して検定するということになっているが、これは弱い意味で帰無仮説族レベルでの有意水準の調整を意味する、と原論文に書かれているので、統計ソフトが p 値を調整した値を出してくるのはやはり表示上の都合で、本当は有意水準を調整している（参照：Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Royal Stat. Soc. B, 57: 289-300, 1995.）。

Bonferroni、Holm、FDR という 3 つの多重比較の考え方はシンプルでわかりやすいので、ここで簡単にまとめておく。k 個の帰無仮説について検定して得られた p 値が p(1)<p(2)<...<p(k) だとすると、有意水準 α で帰無仮説族の検定をするために、Bonferroni は p(1) から順番に α /k と比較し、p(i) ≧ α /k になったところ以降判定保留、Holm は p(i) ≧ α /i となったところ以降判定保留とする。有意水準 α で fdr をするには、まず p(k) を α と比較し、次に p(k-1) を α × (k-1)/k と比較し、と p 値が大きい方から比較していき、p(i)< α × i/k となったところ以降、i 個の帰無仮説を棄却する。R の pairwise.*.test() では、Bonferroni ならすべての p 値が k 倍されて表示、Holm では小さい方から i 番目の p 値が i 倍されて表示、fdr では小さい方から i 番目の p 値が k/i 倍されて表示されることによって、表示された p 値を共通の α との大小で有意性判定ができるわけだが、これは表示上の都合である。**（残念ながら、FDR 法はまだ Rcmdr や EZR のメニューには含まれていない）**

実例を示そう。先に示した 3 村落の身長データについて、どの村落とどの村落の間で身長に差があるのかを調べたい場合、R コンソールでは、

```
pairwise.t.test(sp$HEIGHT, sp$VG, p.adjust.method="bonferroni")
```

とすれば、2 村落ずつのすべての組み合わせについてボンフェローニの方法で有意水準を調整した p 値が表示される（"bonferroni"は"bon"でも良い。また、`pairwise.*`系の関数では `data=`というオプションが使えないので、データフレーム内の変数を使いたい場合は、予めデータフレームを `attach()` しておくか、またはここで示したように、変数指定の際に一々、“データフレーム名＄” を付ける

必要がある)。

ボンフェローニの方法で有意水準を調整した、すべての村落間での身長の差を順位和検定した結果は、以下で得られる。

```
pairwise.wilcox.test(sp$HEIGHT,sp$VG,p.adjust.method="bonferroni")
```

これらの関数で、p.adjust.method を指定しなければホルムの方法になるが、明示したければ、p.adjust.method="holm"とすればよい。FDR 法を使うには、p.adjust.method="fdr"とすればよい。R でもボンフェローニが可能なのは、一番単純な方法であるという理由と、ホルムの方法に必要な計算がボンフェローニの計算を含むからだと思われる。なお、R を使って分析するのだけれども、データがきれいな正規分布をしていて、かつ古典的な方法の論文しか accept しない雑誌に対してどうしても投稿したい、という場合は、TukeyHSD(aov(HEIGHT ~ VG, data=sp)) などとして、テューキーの HSD を行うことも可能である。

EZR では、一元配置分散分析メニューのオプションとして実行できる。「統計解析」「連続変数の解析」「3 群以上の平均値の比較（一元配置分散分析 one-way ANOVA)」を選んで、「目的変数」として HEIGHT を、「比較する群」として VG を選んでから、下の方の「↓ 2 組ずつの比較 (post-hoc 検定) は比較する群が１つの場合のみ実施される」から欲しい多重比較法の左側のボックスにチェックを入れてから「OK」ボタンをクリックする。

いずれのやり方をしても、TukeyHSD の場合だと、2 群ずつの対比較の結果として、差の推定値と 95%同時信頼区間に加え、Tukey の方法で検定の多重性を調整した p 値が下記のように表示され、検定の有意水準が 5%だったとすると、Z と Y の差だけが有意であることがわかる。

```
> TukeyHSD(AnovaModel.3, "factor(VG)")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = HEIGHT ~ factor(VG), data = sp, na.action = na.omit)

$`factor(VG)`
         diff        lwr      upr     p adj
Y-X -2.538889 -8.3843982  3.30662 0.5423397
Z-X  5.850000 -0.9598123 12.65981 0.1038094
Z-Y  8.388889  2.3382119 14.43957 0.0048525
```

Dunnett の多重比較法

Dunnett の多重比較は、コントロールと複数の実験群の比較というデザインで用いられる。以下、簡単な例で示す。例えば、5 人ずつ 3 群にランダムに分けた

高血圧患者がいて、他の条件（食事療法、運動療法など）には差をつけずに、プラセボを 1ヶ月服用した群の収縮期血圧 (mmHg 単位) の低下が 5, 8, 3, 10, 15 で、代表的な薬を 1ヶ月服用した群の低下は 20, 12, 30, 16, 24 で、新薬を 1ヶ月服用した群の低下が 31, 25, 17, 40, 23 だったとしよう。このとき、プラセボ群を対照として、代表的な薬での治療及び新薬での治療に有意な血圧降下作用の差が出るかどうかを見たい（悪くなるかもしれないので両側検定で）という場合に、Dunnett の多重比較を使う。R でこのデータを bpdown というデータフレームに入力して Dunnett の多重比較をするためには、次のコードを実行する。

```
bpdown <- data.frame(
 medicine=factor(c(rep(1,5),rep(2,5),rep(3,5)), labels=c("プラセボ","
代表薬","新薬")),
 sbpchange=c(5, 8, 3, 10, 15, 20, 12, 30, 16, 24, 31, 25, 17, 40, 23))
summary(res1 <- aov(sbpchange ~ medicine, data=bpdown))
library(multcomp)
res2 <- glht(res1, linfct = mcp(medicine = "Dunnett"))
confint(res2, level=0.95)
summary(res2)
```

つまり、基本的には、multcomp ライブラリを読み込んでから、分散分析の結果を glht() 関数に渡し、linfct オプションで、Dunnett の多重比較をするという指定を与えるだけである。multcomp ライブラリのバージョン 0.993 まで使えた simtest() 関数は、0.994 から使えなくなったので注意されたい。

EZR では、まず「ファイル」＞「データのインポート」＞「ファイルまたはクリップボード、URL からテキストデータを読み込む」として、「データセット名を入力」の右側のボックスに bpdown **と入力し、「データファイルの場所」として「インターネットの URL」の右側のラジオボタンをクリックし、「フィールドの区切り記号」を「タブ」にして「OK」ボタンをクリックする。表示される URL 入力ウィンドウに** http://minato.sip21c.org/bpdown.txt **と打って「OK」ボタンをクリックすれば、上記データを読み込むことができる。**
そこで「統計解析」の「連続変数の解析」から「3 群以上の平均値の比較（一元配置分散分析 one-way ANOVA)」を選んで、「目的変数」として sbpchange**、「比較する群」として** medicine **を選び、「2 組ずつの比較 (Dunnett の多重比較)」の左のチェックボックスをチェックしてから「OK」ボタンをクリックすればいい。**
なお、このデータで処理名を示す変数 medicine **の値として** o.placebo**、**1.usual**、**2.newdrug **のように先頭に数字付けた理由は、それがないと水準がアルファベット順になってしまい、Dunnett の解析において新薬群がコントロールとして扱われてしまうからである。**
ノンパラメトリック検定の場合は、「統計解析」の「ノンパラメトリック検定」から「3 群以上の間の比較 (Kruskal-Wallis 検定)」と選び、「目的変数」を sbpchange**、「グループ」を** medicine **にし、「2 組ずつの比較 (post-hoc 検定、Steel の多重比較)」の左のチェックボックスをチェックして「OK」ボタンをクリックすれば、Steel の多重比較が実行できる。**

7.11 二元配置分散分析

既に室内空気の二酸化炭素濃度を採取場所と換気前後という２要因で説明する場合や、産業保健のクロスオーバーデザイン実験での合計作業時間や誤答数への明暗条件とその順序という２要因の効果を調べる例で説明したが、水の総硬度の分析も、３種類の検水をパックテストと滴定という２通りの方法で測定するので、総硬度の値を、検水の種類と測定方法という２要因によって説明することができる。共存物質によって測定方法が総硬度に与える影響は変わる可能性があるので、検水の種類と測定方法の交互作用効果も検証する必要がある。

サンプルデータが用意してある[42]のでダウンロードする。EZR では、「ファイル」＞「データのインポート」＞「ファイルまたはクリップボード、URL からテキストデータを読み込む」として、「データセット名を入力」の右側のボックスに hardness と入力し、「データファイルの場所」として「ローカルファイルシステム」の左側のラジオボタンがチェックされていることを確認し、「フィールドの区切り記号」を「カンマ」にして「OK」ボタンをクリックする。ファイル選択ウィ

[42] http://minato.sip21c.org/pubhealthpractice/waterhardness.csv

ンドウが開くので、予めダウンロードしておいた waterhardness.csv を選んで OK する。

分析は、「統計解析」＞「量的な変数の解析」＞「複数の因子での平均値の比較（多元配置分散分析 multi-way ANOVA）」を選んで表示されるウィンドウで、「目的変数（1 つ選択）」で HARDNESS が選択されていることを確認し、「因子（1 つ以上選択）」で METHOD と SAMPLE を選択する（上の方に書かれている通り、Ctrl キーを押しながらクリックすれば 2 つ以上選択できる）。「交互作用の解析も行う（群別変数が 3 個以下の場合）」の左にチェックが入っていることを確認し、「OK」をクリックすると、自動的に標準偏差のエラーバー付きの棒グラフが描かれ、出力ウィンドウには以下の結果がでる。

```
Anova Table (Type III tests)

Response: HARDNESS
                             Sum Sq Df   F value    Pr(>F)
(Intercept)                  484128  1 1097.5194 3.628e-13 ***
Factor1.METHOD                  128  1    0.2902    0.6000
Factor2.SAMPLE                222629  2  252.3506 1.569e-10 ***
Factor1.METHOD:Factor2.SAMPLE    229  2    0.2599    0.7753
Residuals                       5293 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Factor1.METHOD と Factor1.METHOD:Factor2.SAMPLE の Pr(>F) の値は 0.05 より大きいので総硬度の測定値は、測定方法も測定方法と検水の交互作用も有意には影響しておらず、Factor2.SAMPLE の Pr(>F) は 1.569e-10（e-**はコンピュータの浮動小数点表示で、10^{-**} を意味する。つまり、1.569×10^{-10}）と、0.05 よりずっと小さいので、検水間では有意な差があることがわかる。

7.12　２つの量的な変数間の関係

２つの量的な変数間の関係を調べるための、良く知られた方法が２つある。相関と回帰である。いずれにせよ、まず散布図を描くことは必須である。

本実習では、大気・環境の粉塵濃度と騒音と車の通過台数（雨天時は人の通過人数）について、お互いの関係を調べる際に、これらの方法を用いる。

相関と回帰の違い

大雑把に言えば、相関が変数間の関連の強さを表すのに対して、回帰はある変数の値のばらつきがどの程度他の変数の値のばらつきによって説明されるかを示す。回帰の際に、説明される変数を（従属変数または）目的変数、説明するための変数を（独立変数または）説明変数と呼ぶ。２つの変数間の関係を予測に使うためには、回帰を用いる。

相関分析

一般に、2 個以上の変量が「かなりの程度の規則正しさをもって、増減をともにする関係」のことを相関関係 (correlation) という。相関には正の相関 (positive correlation) と負の相関 (negative correlation) があり、一方が増えれば他方も増える場合を正の相関、一方が増えると他方は減る場合を負の相関と呼ぶ。例えば、身長と体重の関係は正の相関である。

散布図で相関関係があるように見えても、見かけの相関関係 (apparent correlation) であったり[43]、擬似相関 (spurious correlation) であったり[44]することがあるので、注意が必要である。

相関関係は増減をともにすればいいので、直線的な関係である必要はなく、二次式でも指数関数でもシグモイドでもよいが、通常、直線的な関係をいうことが多い（指標はピアソンの積率相関係数）。曲線的な関係の場合、直線的になるように変換したり、ノンパラメトリックな相関の指標（順位相関係数）を計算する。順位相関係数としてはスピアマンの順位相関係数が有名である。

ピアソンの積率相関係数 (Pearson's Product Moment Correlation Coefficient) は、r という記号で表し、2 つの変数 X と Y の共分散を X の分散と Y の分散の積の平方根で割った値であり、範囲は $[-1, 1]$ である。最も強い負の相関があるとき $r = -1$、最も強い正の相関があるとき $r = 1$、まったく相関がないとき（2 つの変数が独立なとき）、$r = 0$ となることが期待される。X の平均を $\bar{X}$、Y の平均を $\bar{Y}$ と書けば、次の式で定義される。

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

相関係数の有意性の検定においては、母相関係数がゼロ（＝相関が無い）という帰無仮説の下で、実際に得られている相関係数よりも絶対値が大きな相関係数が偶然得られる確率（これを「有意確率」という。通常、記号 p で表すので、「p 値」とも呼ばれる）の値を調べる。偶然ではありえないほど珍しいことが起こったと考えて、帰無仮説が間違っていたと判断するのは有意確率がいくつ以下のときか、という水準を有意水準といい、検定の際には予め有意水準を（例えば 5% と）決めておく必要がある。例えば $p = 0.034$ であれば、有意水準 5%で有意な相関があるという意思決定を行なうことができる。p 値は、検定統計量

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

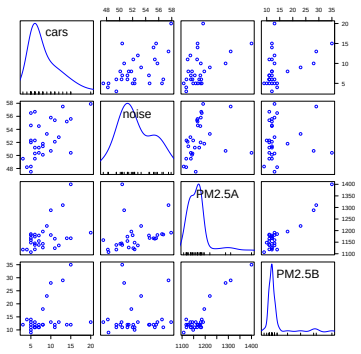
が自由度 $n-2$ の t 分布に従うことを利用して求められる。

[43]例）同業の労働者集団の血圧と所得。どちらも一般に加齢に伴って増加する。

[44]例）ある年に日本で植えた木の幹の太さと同じ年に英国で生まれた少年の身長を 15 年分、毎年 1 回測ったデータには相関があるようにみえるが、直接的な関係はなく、どちらも時間経過に伴って大きくなるために相関があるように見えているだけである。

本実習で相関を検討するのは、粉塵と騒音と交通量（雨天でなければ車の通過台数、屋内の実習になった場合は通行する人数）の関係である。まず、粉塵と騒音のサンプルデータをダウンロードする[45]。既に述べた通り、変数としては、TIME、small、large、PM2.5A、PM2.5B、noise、cars が入っている。EZR への読み込みは、「ファイル」＞「データのインポート」＞「ファイルまたはクリップボード、URL からテキストデータを読み込む」として、「データセット名を入力」の右側のボックスに dnc と入力し、「データファイルの場所」として「ローカルファイルシステム」の左側のラジオボタンがチェックされていることを確認し、「フィールドの区切り記号」を「カンマ」にして「OK」ボタンをクリックする。ファイル選択ウィンドウが開くので、予めダウンロードしておいた dust-noise-cars.csv を選んで OK する。

まず PM2.5 の個数濃度、質量濃度、騒音、交通量の散布図行列を描く。「グラフと表」＞「散布図行列」を選んで現れるウィンドウで、変数を選択（３つ以上）の中で、cars、noise、PM2.5A、PM2.5B を、Ctrl キーを押しながらクリックすることで選択する。最小 2 乗直線の左側のボックスにチェックが入っていたら外してから「OK」ボタンをクリックすると、図が表示される。互いに正の相関関係がありそうにみえる（既に述べた通り、PM2.5 の個数濃度と質量濃度の関係は単純ではないが、概ね正の相関関係はありそうにみえる）。



「統計解析」の「連続変数の解析」から「相関係数の検定（Pearson の積率相関係数）」を選び、変数として、例えば PM2.5A と noise を選ぶ[46]。検定については「対立仮説」の下に「両側」「相関＜０」「相関＞０」の３つから選べるようになっているが、通常は「両側」でよい。OK をクリックすると、出力ウィンドウに次の内容が表示される。

[45] http://minato.sip21c.org/pubhealthpractice/dust-noise-cars.csv

[46] Ctrl キーを押しながら変数名をクリックすれば複数選べるが、一度には 2 つずつしか検定できない仕様になっている。他の組合せも調べること。なお、そうすると厳密には検定の多重性の問題が生じるので、有意水準または有意確率を調整すべきである。

```
Pearson's product-moment correlation

data:  dnc$noise and dnc$PM2.5A
t = 1.4105, df = 28, p-value = 0.1694
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1132151  0.5653677
sample estimates:
      cor
0.2575594
（中略）
相関係数 = 0.258, 95%信頼区間 -0.113-0.565, P 値 = 0.169
```

最下行にまとめられているように、粉塵の個数濃度と騒音の関係について求めたピアソンの積率相関係数は、$r = 0.26$（95%信頼区間が $[-0.11, 0.57]$）であり[47]、p-value = 0.169 より、「相関が無い」可能性を棄却することはできない。なお、相関の強さは相関係数の絶対値の大きさによって判定し、伝統的に 0.7 より大きければ「強い相関」、0.4〜0.7 で「中程度の相関」、0.2〜0.4 で「弱い相関」とみなすのが目安なので、この結果は弱い相関があるかもしれないが統計学的に有意でないと解釈される。

なお、このデータは時系列で連続測定した値なので、測定値間が独立であることを仮定できず、本当は時系列的な変化の向きも検討しなくてはいけないが、そこまでは EZR のメニューには含まれていないので、本実習では扱わない。

[47] 95%信頼区間の桁を丸めて示す場合、真の区間を含むようにするために、四捨五入ではなく、下限は切り捨て、上限は切り上げにするのが普通である。

順位相関係数

スピアマンの順位相関係数 ρ は[a]、値を順位で置き換えた（同順位には平均順位を与えた）ピアソンの積率相関係数と同じである。X_i の順位を R_i、Y_i の順位を Q_i とかけば、

$$\rho = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - Q_i)^2$$

となる。スピアマンの順位相関係数がゼロと差がないことを帰無仮説とする両側検定は、サンプルサイズが 10 以上ならばピアソンの場合と同様に、

$$T = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}$$

が自由度 $n-2$ の t 分布に従うことを利用して行うことができる。ケンドールの順位相関係数 τ は、

$$\tau = \frac{(A-B)}{n(n-1)/2}$$

によって得られる。ここで A は順位の大小関係が一致する組の数、B は不一致数である。

R コンソールでスピアマンの順位相関係数を計算するには、`cor.test()` 関数の中で、`method="spearman"` または `method="kendall"` と指定すれば良い。EZR では、「統計解析」「ノンパラメトリック検定」「相関係数の検定（Spearman の順位相関係数）」を選んで表示されるウィンドウで相関を検討したい変数を選び、解析方法のところで Spearman の横のラジオボタンを選んで OK ボタンをクリックすれば計算できる。

[a] ピアソンの相関係数の母相関係数を ρ と書き、スピアマンの順位相関係数を r_s と書く流儀もある。

回帰モデルの当てはめ

回帰は、従属変数のばらつきを独立変数のばらつきで説明するというモデルの当てはめである。十分な説明ができるモデルであれば、そのモデルに独立変数の値を代入することによって、対応する従属変数の値が予測あるいは推定できるし、従属変数の値を代入すると、対応する独立変数の値が逆算できる。こうした回帰モデルの実用例の最たるものが**検量線**である。検量線とは、実験において予め濃度がわかっている標準物質を測ったときの吸光度のばらつきが、その濃度によってほぼ完全に（通常 98%以上）説明されるときに（そういう場合は、散布図を描くと、点々がだいたい直線上に乗るように見える）、その関係を利用して、サンプルを測ったときの吸光度からサンプルの濃度を逆算するための回帰直線である（曲線の場合もあるが、通常は何らかの変換をほどこし、線形回帰にして利用する）。

検量線の計算には、(A) 試薬ブランクでゼロ点調整をした場合の原点を通る回帰直線を用いる場合と、(B) 純水でゼロ点調整をした場合の切片のある回帰直線を用いる場合がある。例えば、濃度の決まった標準希釈系列 (0, 1, 2, 5, 10 $\mu g/\ell$) について、純水でゼロ点調整をしたときの吸光度が、(0.24, 0.33, 0.54, 0.83, 1.32)

だったとしよう。吸光度の変数を y、濃度を x と書けば、回帰モデルは $y = bx + a$ とおける。係数 a と b（a は切片、b は回帰係数と呼ばれる）は、次の偏差平方和を最小にするように、最小二乗法で推定される。

$$f(a,b) = \sum_{i=1}^{5} (y_i - bx_i - a)^2$$

この式を解くには、$f(a,b)$ を a ないし b で偏微分したものがゼロに等しいときを考えればいいので、次の２つの式が得られる。

$$b = \frac{\sum_{i=1}^{5} x_i y_i/5 - \sum_{i=1}^{5} x_i/5 \cdot \sum_{i=1}^{5} y_i/5}{\sum_{i=1}^{5} {x_i}^2/5 - \left(\sum_{i=1}^{5} x_i/5\right)^2}$$

$$a = \sum_{i=1}^{5} y_i/5 - b \cdot \sum_{i=1}^{5} x_i/5$$

これらの a と b の値と、未知の濃度のサンプルについて測定された吸光度（例えば 0.67 としよう）から、そのサンプルの濃度を求めることができる。注意すべきは、サンプルについて測定された吸光度が、標準希釈系列の吸光度の範囲内になければならないことである。回帰モデルが標準希釈系列の範囲外でも直線性を保っている保証は何もないのである[48]。

R コンソールでは、`lm()`（linear model の略で線形モデルの意味）を使って、次のようにデータに当てはめた回帰モデルを得ることができる。

```
y <- c(0.24, 0.33, 0.54, 0.83, 1.32)
x <- c(0, 1, 2, 5, 10)
# 線形回帰モデルを当てはめる
res <- lm(y ~ x)
# 詳しい結果表示
summary(res)
# 散布図と回帰直線を表示する
plot(y ~ x)
abline(res)
# 吸光度 0.67 に対応する濃度を計算する
(0.67 - res$coef[1])/res$coef[2]
```

結果は次のように得られる。

[48]回帰の外挿は薦められない。サンプルを希釈したり濃縮したりして吸光度を再測定し、標準希釈系列の範囲におさめることをお薦めする。

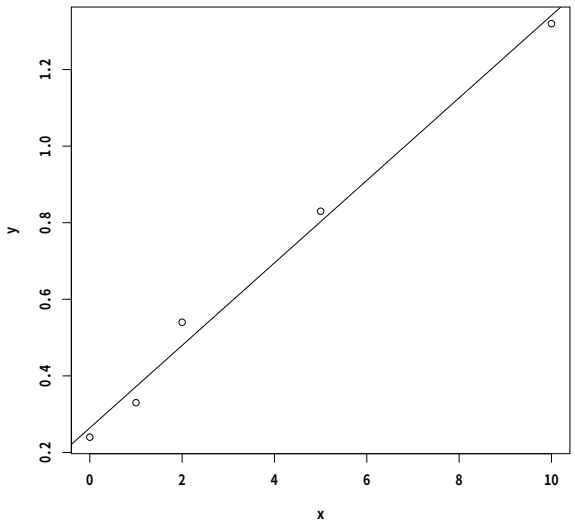
```
Call:
lm(formula = y ~ x)

Residuals:
       1        2        3        4        5
-0.02417 -0.04190  0.06037  0.02718 -0.02147

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.26417    0.03090   8.549 0.003363 **
x            0.10773    0.00606  17.776 0.000388 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04894 on 3 degrees of freedom
Multiple R-squared: 0.9906,     Adjusted R-squared: 0.9875
F-statistic:   316 on 1 and 3 DF,  p-value: 0.0003882
```

推定された切片は $a = 0.26417$、回帰係数は $b = 0.10773$ である。また、このモデルはデータの分散の 98.75% (0.9875) を説明していることが、`Adjusted R-squared` からわかる。また、`p-value` は、吸光度の分散がモデルによって説明される程度が誤差分散によって説明される程度と差が無いという帰無仮説の検定の有意確率である。



0.67 という吸光度に相当する濃度は、3.767084 となる。したがって、この溶液の濃度は、3.8 $\mu g/\ell$ だったと結論することができる。

本実習では、二酸化窒素濃度の計算で検量線が必要となる。上記のような計算をするためには、データ入力に工夫が必要である。Excel などの表計算ソフトで、まず 1 行目に SAMPLE、CONC、ABS と入力する。1 列目の最初の 18 行は S と入力し、次いで A、B、C、D、E を 3 行ずつ入力する。2 列目は最初の 18 行は標準希釈系列の濃度なので、0、2、4、8、12、16 を 3 行ずつ入力する。それ以降はサンプルであり濃度は不明なのでブランクのままで良い。3 列目はプレートリーダで印字された吸光度の測定値を入力する。ただし 3 列目の測定値で triplicate の中で 1 つだけ外れた値があるときは、1 列目を X とする。Excel で入力した場合は、「形式を選択して保存」で「テキスト（コンマ区切り）」でデータをファイルに保存する。サンプルデータを `http://minato.sip21c.org/pubhealthpractice/workingcurve.csv` に置くので、ダウンロードしておく。

EZR に読み込むには、「ファイル」＞「データのインポート」＞「ファイルまたはクリップボード、URL からテキストデータを読み込む」として、「データセット名を入力」の右側のボックスに `WC` と入力し、「データファイルの場所」として「ローカルファイルシステム」の左側のラジオボタンがチェックされていることを確認し、「フィールドの区切り記号」が「カンマ」であることを確認してから「OK」ボタンをクリックする。ファイル選択ウィンドウが開くので、予めダウンロードしておいた workingcurve.csv を選んで OK する。

まず、標準希釈系列のデータを使って、検量線を計算する。「統計解析」＞「連続変数の解析」＞「線形回帰（単回帰、重回帰）」と選び、「モデル名を入力」に `WC1` と入力し、「目的変数（１つ選択）」で `ABS`、「説明変数（１つ以上選択）」で `CONC` を選び、「モデル解析用に解析結果をアクティブモデルとして残す」の左のチェックボックスにチェックを入れる。さらに「↓一部のサンプルだけを解析対象にする場合の条件式。例：age > 50 & Sex == 0 や age < 50 | Sex == 1」の下のボックスに `SAMPLE=="S"`と入力して [OK] をクリックする。

出力ウィンドウに結果が表示される（後述する多重共線性をチェックするために VIF を計算しようとしてエラー: `model contains fewer than 2 terms` と表示されるが気にしなくて良い）。下の方に日本語でまとめた結果が表示されるが、自由度調整済重相関係数の二乗が含まれないので、上の方の結果を見る。自由度調整済重相関係数の二乗が 0.987 なので、それなりに当てはまりは良い。回帰式の切片は `WC1$coef[1]` か `coef(WC1)[1]` で参照でき、回帰係数は `WC1$coef[2]` か `coef(WC1)[2]` で参照できる。

```
Call:
lm(formula = ABS ~ CONC, data = WC, subset = SAMPLE == "S")

Residuals:
       Min         1Q     Median         3Q        Max
-0.0239223 -0.0080413 -0.0002198  0.0107207  0.0221966

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0027207  0.0055576   -0.49    0.632
CONC         0.0209703  0.0006014   34.87 8.97e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01369 on 15 degrees of freedom
Multiple R-squared:  0.9878,Adjusted R-squared:  0.987
F-statistic:  1216 on 1 and 15 DF,  p-value: 8.967e-16
```

次に、この検量線を使って、サンプルの吸光度から濃度を計算する。「アクティブデータセット」＞「変数の操作」＞「計算式を入力して新たな変数を作成する」と選んで表示されるウィンドウで、「新しい変数名」を CONC2 とし、「計算式」に (ABS-coef(WC1)[1])/coef(WC1)[2] と入力して「OK」ボタンをクリックすると、新しい変数 CONC2 に、各サンプルの二酸化窒素濃度が得られるので、それを一元配置分散分析に用いることができる。

このまま一元配置分散分析をするには、「統計解析」＞「連続変数の解析」＞「3 群以上の間の平均値の比較（一元配置分散分析 one-way ANOVA）」で「目的変数（1 つ選択）」で CONC2、「比較する群（1 つ以上選択）」で SAMPLE を選び、下の方の「↓一部のサンプルだけを……」の下の枠に、

```
SAMPLE %in% LETTERS[1:5]
```

と打ってから「OK」すれば良い。

検量線以外の状況でも、同じやり方で線形回帰モデルを当てはめることができる。

粉塵濃度が騒音と交通量の両方で説明されるという重回帰モデルを考えよう[49]。既に dnc という名前でサンプルデータは読み込み済なので、ロゴのすぐ右の “データセット:” の右側をクリックして dnc を指定して dnc データセットをアクティブにしてから、「統計解析」＞「連続変数の解析」＞「線形回帰（単回帰、重回帰）」

[49]騒音が粉塵に影響することは科学的に考えにくいので、このデータの場合は、むしろ交通量が粉塵濃度を説明するという単回帰分析や、交通量が騒音を説明するという単回帰分析の方が合理的だが、重回帰分析の使い方を説明するため許されたい。

と選び、目的変数として PM2.5A、説明変数として noise と cars を選び、「OK」ボタンをクリックすると出力ウィンドウに次の結果が得られる。重回帰モデルの解釈については後述するが、Adjusted R-squared の値が 0.168 なので、このモデルではデータのばらつきの 17%弱しか説明できていないことになり、モデルはデータにあまり適合していない（この 2 つ以外の要因が PM2.5 の個数濃度のばらつきの 80%以上を説明することになる）。

```
Call:
lm(formula = PM2.5A ~ cars + noise, data = dnc)

Residuals:
   Min     1Q Median     3Q    Max
-69.32 -35.50 -10.03  23.12 177.82

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1101.9891   213.1748   5.169 1.93e-05 ***
cars           7.3197     3.1093   2.354   0.0261 *
noise          0.2062     4.2964   0.048   0.9621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56.58 on 27 degrees of freedom
Multiple R-squared:  0.2253,Adjusted R-squared:  0.168
F-statistic: 3.927 on 2 and 27 DF,  p-value: 0.03184


> vif(RegModel.7)
    cars    noise
1.386756 1.386756

> multireg.table
            回帰係数推定値 95%信頼区間下限 95%信頼区間上限    標準誤差     t 統計量           P
値
(Intercept)   1101.9890907     664.5904779     1539.387704 213.174826 5.16941476 0.00001932295
cars             7.3197137       0.9400317       13.699396   3.109264 2.35416309 0.02609356489
noise            0.2061965      -8.6092118        9.021605   4.296363 0.04799326 0.96207486496
```

推定された係数の安定性を検定する

回帰直線のパラメータ（回帰係数 b と切片 a）の推定値の安定性を評価するためには、t 値が使われる。いま、Y と X の関係が $Y = a_0 + b_0X + e$ というモデルで表されるとして、誤差項 e が平均 0、分散 σ^2 の正規分布に従うものとすれば、切片の推定値 a も、平均 a_0、分散 $(\sigma^2/n)(1 + M^2/V)$（ただし M と V は x の平均と分散）の正規分布に従い、残差平方和 Q を誤差分散 σ^2 で割った Q/σ^2 が自由度 $(n - 2)$ のカイ二乗分布に従うことから、

$$t_0(a_0) = \frac{\sqrt{n(n-2)}(a - a_0)}{\sqrt{(1 + M^2/V)Q}}$$

が自由度 $(n - 2)$ の t 分布に従うことになる。

しかしこの値は a_0 がわからないと計算できない。a_0 が 0 に近ければこの式で

$a_0 = 0$ と置いた値（つまり $t_0(0)$。これを切片に関する t 値と呼ぶ）を観測データから計算した値が $t_0(a_0)$ とほぼ一致し、自由度 $(n-2)$ の t 分布に従うはずなので、その絶対値は 95%の確率で t 分布の 97.5%点（サンプルサイズが大きければ約 2 である）よりも小さくなる。つまり、データから計算された t 値がそれより大きければ、切片は 0 でない可能性が高いことになるし、t 分布の分布関数を使えば、「切片が 0 である」という帰無仮説に対する有意確率が計算できる。

回帰係数についても同様に、

$$t_0(b) = \frac{\sqrt{n(n-2)V}b}{\sqrt{Q}}$$

が自由度 $(n-2)$ の t 分布に従うことを利用して、「回帰係数が 0」であるという帰無仮説に対する有意確率が計算できる。有意確率が充分小さければ、切片や回帰係数がゼロでない何かの値をとるといえるので、これらの推定値は安定していることになる。線形回帰をした結果の中の、`Pr(>|t|)` というカラムに、これらの有意確率が示されている。

重回帰モデルでは、個々の説明変数について推定される回帰係数は、他の説明変数の目的変数への影響を調整した上で、その変数独自の目的変数への影響を示す「偏回帰係数」である。しかし偏回帰係数の値は、各変数の絶対的な大きさに依存しているので、各説明変数の目的変数への影響の相対的な強さを示すものにはならない。そうした比較をしたければ、標準化偏回帰係数を計算する。EZR のメニューにはこの機能はないが、重回帰モデルを残してあるので、そのオブジェクトを使ったコマンドをスクリプトウィンドウに打つ。右上に「モデル：」とあるところの右側に、最後に生成されたモデル名が表示されている。これが `RegModel.7` だとすると、

```
csd <- sapply(RegModel.7$model, sd)
print(coef(RegModel.7)*csd/csd[1])
```

とスクリプトウィンドウに打ち、この 2 行を選んでから、「実行」ボタンをクリックすれば、出力ウィンドウに標準化偏回帰係数が表示される。

標準化偏回帰係数の代わりに、偏相関係数の二乗を使っても、それぞれの説明変数の目的変数への影響の相対的な寄与を示すことができる（Residuals の項を除く、すべての説明変数の偏相関係数の二乗を合計すると、次節で説明する重相関係数の二乗—ただし当然ながら自由度調整する前の値—になる）。最近の論文では、重回帰分析の結果の表としては、各説明変数に対して、偏回帰係数、標準誤差、95%信頼区間の下限と上限、偏相関係数の二乗を示すことが多い。偏相関係数の二乗を求めるには、重回帰分析のモデルが `RegModel.7` だとすると、下記の二行を実行すれば良い。

```
SS <- anova(RegModel.7)["Sum Sq"]
SS/sum(SS)
```

回帰モデルを当てはめる際の留意点

身長と体重のように、どちらも誤差を含んでいる可能性がある測定値である場合には、一方を説明変数、他方を目的変数とすることは妥当でないかもしれない（一般には、身長によって体重が決まるなど方向性が仮定できれば、身長を説明変数にしてもよいことになっている）。また、最小二乗推定の説明から自明なように、回帰式の両辺を入れ替えた回帰直線は一致しない。従って、どちらを目的変数とみなし、どちらを説明変数とみなすか、因果関係の方向性に基づいて（先行研究や臨床的知見を参照し）きちんと決めるべきである。

回帰を使って予測をするとき、外挿には注意が必要である。とくに検量線は外挿してはいけない。実際に測った濃度より濃かったり薄かったりするサンプルに対して、同じ関係が成り立つという保証はどこにもないからである（吸光度を y とする場合は、濃度が高くなると分子の重なりが増えるので飽和 (saturate) してしまい、吸光度の相対的な上がり方が小さくなっていき、直線から外れていく）。サンプルを希釈したり濃縮したりして、検量線の範囲内で定量しなくてはいけない。

重回帰モデルでは、説明変数同士に強い相関があると、係数推定が不安定になるのでうまくない。ごく単純な例でいえば、目的変数 `Y` に対して説明変数群 `X1` と `X2` が相加的に影響していると考えられる場合、`lm(Y ~ X1+X2)` という重回帰モデルを立てるとしよう。ここで、実は `X1` が `X2` と強い相関をもっているとすると、もし `X1` の標準化偏回帰係数の絶対値が大きければ、`X2` による効果もそちらで説明されてしまうので、`X2` の標準化偏回帰係数の絶対値は小さくなるだろう。まったくの偶然で、その逆のことが起こるかもしれない。したがって、係数推定は必然的に不安定になる。この現象は、説明変数群が目的変数に与える線型の効果を共有しているという意味で、多重共線性 (multicolinearity) と呼ばれる。

多重共線性があるかどうかを判定するには、説明変数間の散布図を 1 つずつ描いてみるなど、丁寧な吟味をすることが望ましいが、各々の説明変数を、それ以外の説明変数の目的変数として重回帰モデルを当てはめたときの重相関係数の 2 乗を 1 から引いた値の逆数を VIF（Variance Inflation Factor; 定訳は不明だが、分散増加因子と訳しておく）として、VIF が 10 を超えたら多重共線性を考えねばならないという基準を使う（Armitage et al., 2002）のが簡便である。多重共線性があるときは、拡張期血圧（DBP）と収縮期血圧（SBP）のように本質的に相関があっても不思議はないものだったら片方だけを説明変数に使うとか、2 つの変数を使う代わりに両者の差である脈圧を説明変数として使うのが 1 つの対処法だが、その相関関係自体に交絡が入る可能性はあるし、情報量が減るには違いない。変数を減らさずに調整する方法としては、centring という方法がある。

EZR で重回帰分析を実行すると、自動的に VIF は計算される。上の例では、どちらの変数についても小さい値なので、多重共線性は考えなくて良い。

7.13 文献

- 新谷歩 (2011) 今日から使える医療統計学講座【Lesson 3】サンプルサイズとパワー計算. 週刊医学界新聞, 2937 号[50]
- 青木繁伸 (2009) R による統計解析. オーム社
- 古川俊之（監修）, 丹後俊郎（著） (1983) 医学への統計学. 朝倉書店.
- 中澤　港 (2003) R による統計解析の基礎. ピアソン・エデュケーション.[51]
- 中澤　港 (2007) R による保健医療データ解析演習. ピアソン・エデュケーション.[52]
- 神田善伸 (2012) EZR でやさしく学ぶ統計学〜EBM の実践から臨床研究まで〜. 中外医学社[53].

[50] http://www.igaku-shoin.co.jp/paperDetail.do?id=PA02937_06
[51] http://minato.sip21c.org/statlib/stat-all-r9.pdf から全文ダウンロードできる。
[52] http://minato.sip21c.org/msb/medstatbookx.pdf から全文ダウンロードできる。
[53] http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html

Index