

第1回 実験計画学とは？

A. 授業の枠組み

1. 講義について

実験計画学は統計学の一分野である。その理論そのものはあまりおもしろいとは感じられないだろう。理論を実際のデータに当てはめていくことで統計的な考え方を身につけるようになるのが望ましい。したがって、自分でデータを集め、そのデータを使った計算を必ずこなし、授業に積極的に参加して、統計的考え方を育てることが必要である。

毎回、できるだけたくさんの学生に質問する。

ノートパソコン（授業ではエクセルおよびエクセルに付属する分析ツールの使い方を説明する）、これまでに配ったプリント（あるいはホームページ上にあるファイルをパソコンに入れておいてもよい）を持参のこと

授業のホームページ <http://www.ipc.shimane-u.ac.jp/food/kobayasi/>
に資料などをのせておく。

授業で説明に使うエクセルのバージョンは Excel2003 であるが、それ以外のバージョン (Excel2007) でもかまわない。

2. 宿題について

次回につながる宿題を出す。レポートは原則として翌週の月曜日午後1時まで生物資源科学部2号館204号室の前にあるかごに提出すること。

配ったプリントは毎回必ず持ってくること。

レポートはノートに書くか、プリントアウトをノートに貼るか、バインダーやクリアファイルなどに綴じて、データの入ったエクセルファイルを保存したフロッピーディスクかCDを添えて、提出する (USBメモリーはウイルスの可能性があるので受け取りません)。ファイル名は実験計画学_氏名とする (例: 実験計画学_豊臣秀吉)。各回の宿題はそれぞれエクセルファイルに新しいシートを追加し、データを入力すること。レポートには毎回継続的な課題があるので、これまでの宿題も一緒に提出する。

授業の前にレポートを返却する。再提出の指示があるときは、再提出となった元のレポート、レポートにつけた採点表とあわせて、再提出のレポートを提出すること。再提出の場合、次のレポートと同時に提出してかまわない。

3. 成績について

以下の3つから成績を判定する。

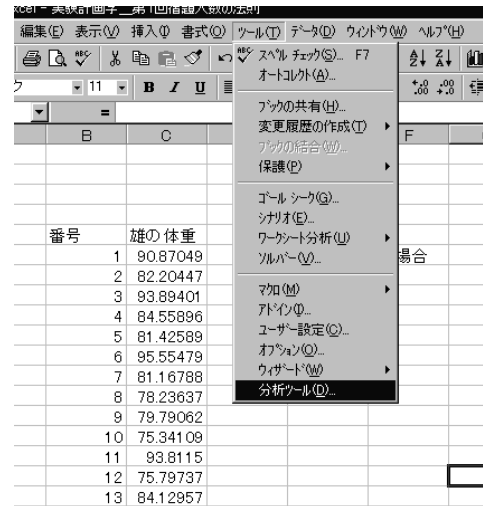
- ① 授業参加 質問への受け答え、疑問点を質問したかなど
- ② レポート
- ③ 期末試験 (ノート, 参考資料, ノートパソコンの持ち込み可, 通信機能の使用は不可)

B. エクセルの分析ツールの組み込み

1. Excel2003 の場合 (あるいはそれ以前のバージョン)

① 分析ツールが組み込まれているかを確認する

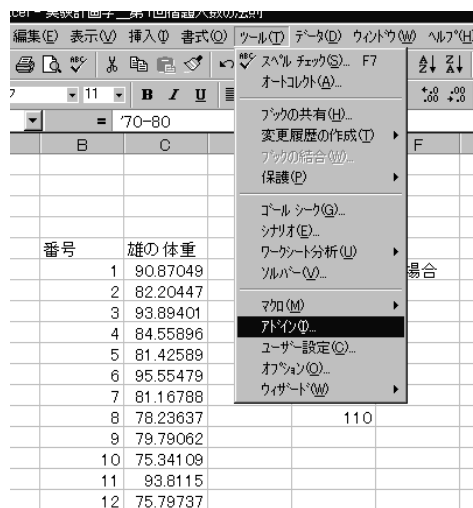
ツール→分析ツールで分析ツールを起動する. ツールをクリックしたときに出てくるプルダウンメニューに分析ツールがないときは, ②の方法で分析ツールを組み込む.



② 分析ツールの組み込み

ツールでプルダウンメニューを引き出し, アドインを選ぶ. 右下のようにアドイン (組み込み) するべきプログラムにチェックを入れる. ここでは分析ツールにだけチェックを入れる.

このとき場合によってはエクセルの入ったCDを要求してくる場合がある. このときは自分のパソコンに付属した, あるいは自分のパソコンにインストールしたエクセルのCDを用意する.



2. Excel2007 の場合

① 分析ツールが組み込まれているかを確認する

データをクリックし, データ分析があるかを確認する. データ分析がないときは, ②の方法で分析ツールを組み込む.

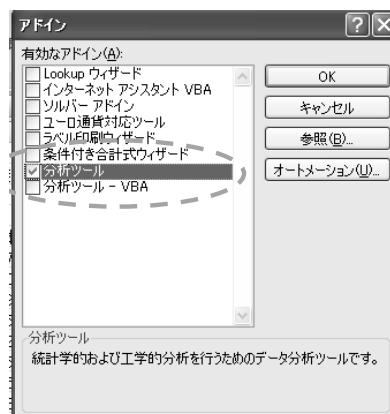
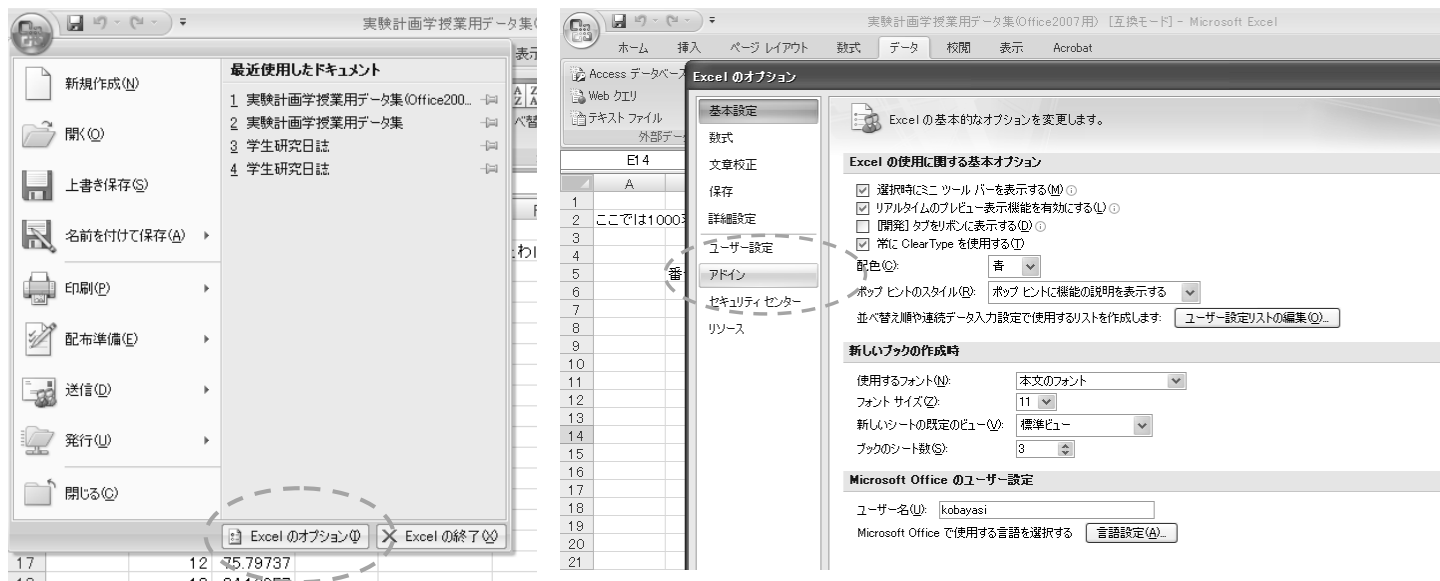


② 分析ツールの組み込み

一番左上のボタンを押し, 現れたメニューの下にあるエクセルのオプションを押す. オプションのメニューが出たら, 左のフレームメニューからアドインを選ぶ. アドインできるアプリケーションのリストが並び, アクティブでないアプリケーションアドインのリストの中に分析ツールをクリックし, 設定をクリックする. 次に表示されたアドインのリストの中から, アドイン (組

み込み)するべきプログラムにチェックを入れる. ここでは分析ツールにだけチェックを入れる.

このとき場合によってはエクセルの入ったCDを要求してくる場合がある. このときは自分のパソコンに付属した, あるいは自分のパソコンにインストールしたエクセルのCDを用意する.



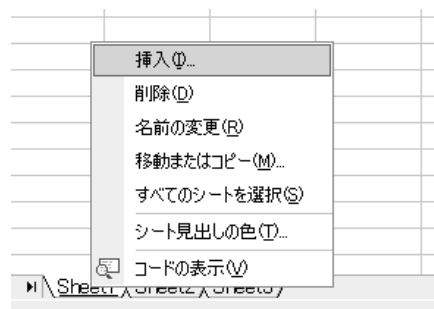
C. 宿題のデータを入れるエクセルファイルの作成

1. シートの追加

宿題は各回ごとに別のシートにデータや計算結果などを入力し, それを1つのファイルにまとめる.

2. シート名

下のように, シート名は第1回から第13回までを作っておく. データ集の入ったフロッピーあるいはCDには「実験計画学_豊臣秀吉」というファイルがある. このファイルはすでにシート名を変更してあり, ファイル名だけ変えれば, 宿題提出用になる.



57		53	73.2		53	73.2
58		54	73.3		54	73.3
59		55	73.4		55	73.4
60		56	73.4		56	73.4
61		57	73.5		57	73.5

第1回 / 第2回 / 第3回 / 第4回 / 第5回 / 第6回 / 第7回 / 第8回 / 第9回 / 第11回 / 第12回 /

D. 経験から科学へ

1. 実験して得たデータの解釈

実験をして得たデータの解釈をどうするか？

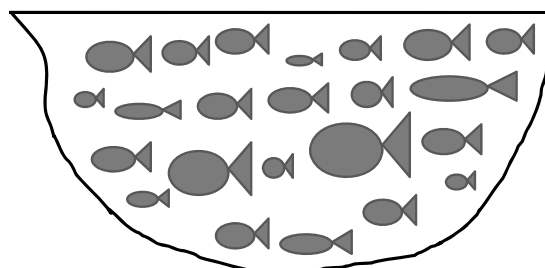
- A, B, Cの3種類の餌を与えたヒツジの成長を調べた. どの餌がいちばんよいか？
気温が上昇するとブドウの糖度はどうなるか？

データにはばらつきがあるので, 1つや2つだけ調べてもそれが本当に正しいかは確信を持って
ない. ではデータをどのようにとったら確信を持ってもよいのだろうか？

★ 考えてみよう.

池の中の魚の大きさを知りたい. 魚は池に百匹以上はいて, しかも正確な数はわからない. 魚
の大きさは図のようにかなりまちまちである.

1) 平均が知りたいなら何匹調べたらよいだろうか？



2) 同じ形の池が2つあった. しかし, 一方は富栄養化していて魚が大きくなったようだ. この
仮説を証明するには2つの池からそれぞれ何匹を調べたらよいだろうか？

1) は第5回の講義で学ぶ統計的推定, 2) は第6回の講義で学ぶ統計的検定と関連がある.

2. () の法則

昔の人は彗星は前触れもなくやってきたので不吉なことの前兆ととらえた.

- ★ 科学の始まり ケプラーの法則 ブラーエの集めた膨大な天文学データから
ケプラーの法則から万有引力の法則へ
ニュートンはケプラーの法則から万有引力の法則を導いた
万有引力の法則ですべての天体の運動を説明できる

多数のデータから少数の法則を導くことができた→少数の法則から予測・発見ができた
海王星の存在の予言, ハレー彗星が周期的に地球の近くに来ることを予言など

たくさんデータを集めれば集めるほど確実である.

保険のデータ 人間の死や事故はわからないことだらけ

統計を集めて, 確率的に掛け金を決める (生命表 車の事故)

★ 統計をみて考えよう.

① 日本の出生率と出生性比(女子新生児100人に対する男子新生児数)の右の表を見て気づいたことを書け.

	西暦	出生率	出生性比
明治38年	1905	31.2	102.7
明治39年	1906	29.6	108.7
明治40年	1907	34.0	102.7
明治41年	1908	34.7	104.6
明治42年	1909	34.9	104.1
昭和39年	1964	17.7	105.9
昭和40年	1965	18.6	105.3
昭和41年	1966	13.7	107.6
昭和42年	1967	19.4	105.3
昭和43年	1968	18.6	107.1
昭和44年	1969	18.5	107.2
昭和45年	1970	18.8	107.1

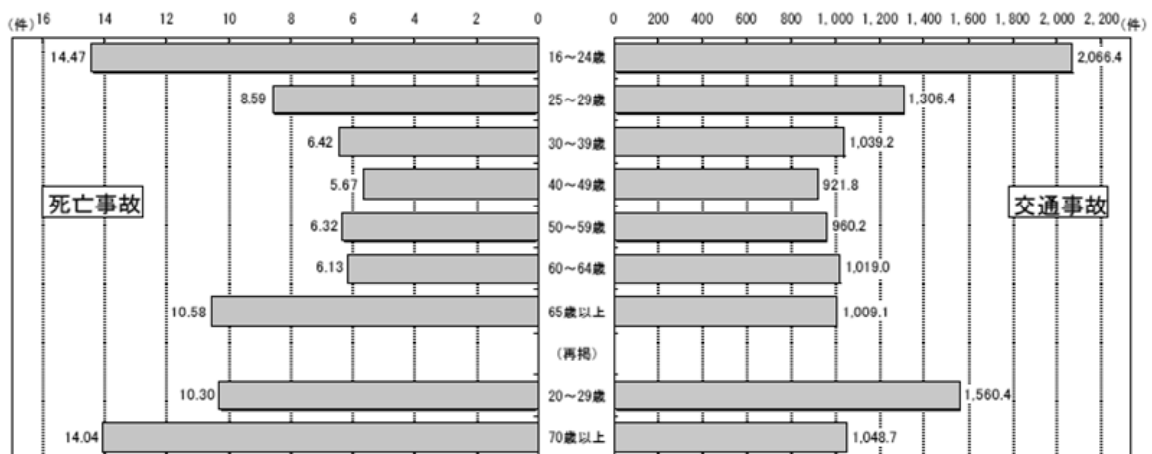
② 人の寿命はそれぞれ異なる. しかし, たくさんのデータを集めると傾向がわかる. 日本人の平均余命が年々伸びてきていることを示す右の表を見て気づいたこと, さらにこのようなデータはどういうことに利用されているかを書け.

(単位: %)

暦年	男			女		
	40歳	65歳	80歳	40歳	65歳	80歳
昭和22	68.0	39.8	9.5	70.9	49.1	17.3
25-27	81.8	55.1	16.6	83.2	62.8	26.1
30	87.0	61.8	20.0	89.0	70.6	31.9
35	89.7	64.8	20.1	92.2	75.2	33.8
40	92.6	69.1	22.6	95.0	80.0	38.4
45	93.7	72.1	26.1	96.1	82.6	43.0
50	95.1	76.8	33.2	96.9	86.1	50.7
55	96.1	79.4	37.8	97.6	88.5	57.0
60	96.7	81.1	42.8	98.0	90.1	63.0
平成2	97.1	82.6	46.9	98.3	91.3	67.8
7	97.2	83.3	48.2	98.4	91.6	70.2
12	97.5	84.7	52.5	98.6	92.6	74.5
13	97.6	85.1	53.5	98.6	92.8	75.3
14	97.7	85.4	54.2	98.6	92.9	75.9
15	97.6	85.3	54.5	98.6	93.0	76.3
16	97.7	85.7	55.2	98.7	93.0	76.8

注1) 平成12年までは, 完全生命表による.
 2) 昭和45年以前は, 沖縄県を除く値である.
 3) 生命表作成時点における死亡状況を一定不変とした場合の状況を表しており, 現実の生存者の割合とは異なっている.

④ 交通事故死者の年齢層別・状態別分類. 交通事故の実態を調べることによって, 自動車保険などの保険料を決めたり, 交通事故を防ぐための措置を考えたりする.
 若者の自動車保険料は高いが, どう考えるか?



2. 少数例で何かいえないか？ 近代的な統計学の登場

大数の法則とはいうけれど、たくさんのデータを集めるのは大変である
できるだけ少ない・多くのデータでできるだけ少ない・多くの法則（結果）を得たい・・・

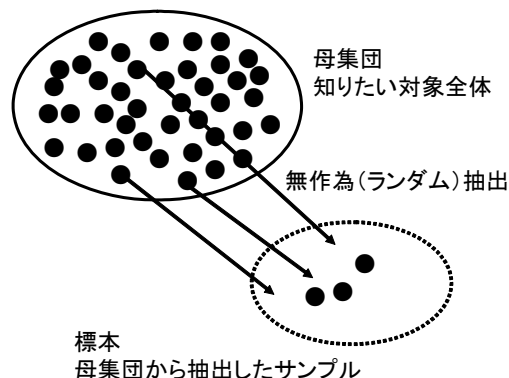
大数の法則にかなうぐらいのデータを集めることは困難なことも多い

自動車の耐久性テスト 少ない・多くの台数を破壊試験して、しかも正確なデータがほしい
オオサンショウウオの食性を知りたい 数自体が少ない・いくらでもいる
猫のエイズの治療薬 効かないかもしれない副作用の強い薬は動物虐待かも？

それゆえできるだけ少数の・多数のデータでできるだけ少数の・多数の法則（結果）を得たい

標本（サンプル）から母集団を推定する 近代的な統計学の手法

母集団（調査対象すべて）からランダム（無作為）に取ったサンプル（母集団からの代表）であれば、統計学からどの程度の精度・誤差で判断できるかがわかる。



正しく抽出した標本から母集団を推定する実例と誤差

視聴率 標本サイズと誤差 関東地区で 1455 万世帯から 600 世帯
山陰地方で 45 万世帯 200 世帯をサンプリングする
視聴率が 20%と出たら、その誤差はどのくらい？ どちらの精度が高い？

視聴率の誤差はサンプルの大きさに依存し、母集団の大きさとサンプルの大きさの比はほとんど関与しない。誤差を大まかに見積もると右の表ようになる。真の視聴率と測定された視聴率との誤差は 95%の確率で以下の範囲内に収まることを右表は示す。

視聴率	標本数600	標本数200
5%・95%	±1.8%	±3.1%
10%・90%	±2.4%	±4.2%
20%・80%	±3.3%	±5.7%
30%・70%	±3.7%	±6.5%
40%・60%	±4.0%	±6.9%
50%	±4.1%	±7.1%

3. 正確なデータを取るために：標本の選び方は難しい

標本に要求されること

母集団を代表しているか（統計的な手法を適用するための条件）

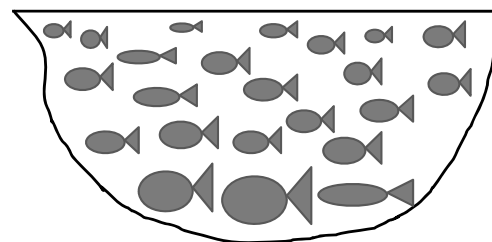
精度が必要な程度あるか（サンプルの数が多ければ精度は高くなる）

★ 母集団を代表しているか インターネットを使って、世論調査したら・・・？

元気そうなウシばかりサンプリングしたら？

愛煙家の夫の副流煙を受けた妻の例

★ 池の魚を調査する例をもう一度考える．この池の魚で、大きな魚は底に住む性質がある．もし調査のときに、面倒だからと水面近くの魚ばかり調査したらどうなるだろうか？



無作為（ランダム）なサンプリングであることが要請される

無作為標本（ランダムサンプル）ならば

標本数が少数であっても母集団について代表値、精度などを統計的に推定できる。

講義の第2, 3, 4, 5, 6, 7, 8, 12, 13回はこのような無作為標本についての基本的な統計学を学ぶ。

E. 実験計画学とは

大数の法則に従って、データをたくさん取ればより正確なことがわかるとはいえ、実験を何度も繰り返すことはたいへん手間がかかる

したがって、なるべく少ないサンプルで結論を得たい。しかし、サンプルが少ないと誤差が大きくなる・・・

1. どのようにデータを取れば、精度よく、しかも少ない実験ですむのか？

不均一なところ（圃場、牛の集団）からどのように実験し、データを得るのか？

① 圃場によって地力が違うので、同じ品種でも収量が異なる。

② ヒトに栄養剤を与えた効果を知りたいが、体の大きさによって効果が違いそうだ。

不均一さの克服のためには 1. 精度を知る どの程度ばらついているのか？

2. 傾向のある不均一さを偶然誤差に転化する

3. 可能ならば不均一さによる誤差を除去して、精度を高める

2. 実験計画学の3つの柱

- ① 誤差の定量と制御（精度がよくなる・精度がわかる）
- ② 実験回数を少なくできる
- ③ 実験のデータの変動のうち、処理による意味のある部分と偶然誤差を分けることができる。
さらに処理の主効果と交互作用を検出できる。

講義の第9-11, 14回は実験計画学の基本を学ぶ。

F. 宿題

1. 大数の法則が成り立つくらいのデータを集める。最低でも100以上のデータを集める。同じことを2人以上が調べるようなことがないように、自分が調べることを204室前のホワイトボードにある用紙に調査テーマを事前を書く（早い者勝ちとする）。テーマの合格を得てから、調査を始めること。

- ① 何か法則性がありそうなデータを集める。データを集める前に、どんな法則が出てくるかを予想する。
- ② 100以上のデータを集め、大きさの順に並べる。
- ③ 次の値を見つけだすあるいは計算する。
最大値, 最小値, レンジ（最大値と最小値の差）
平均
メジアン（中央値：大きい順に並べたときまん中に来る値。データが偶数のときはまん中に位置する2つの値の平均）
- ④ 分布をグラフにする（ヒストグラムの作成）
最大値と最小値の間を5, 10, 15に分級してそれぞれ頻度分布を書く。
- ⑤ ここまでの時点で観察の結果、わかったことを、最初の予想と照らし合わせながら書く。箇条書きでよいから、なるべくたくさん書くのがよい。なお上の②～④の手順に基づいてデータをまとめていくとき、どの手順の時にわかったかをなるべく書くこと。

調査テーマが大数の法則にはふさわしくないものなど以下の1)～8)を満たさないものであればテーマをもう一度考え直させる。承認を得ないテーマで調べてもレポートは受け取らない。

次のような条件を満たしているデータでなければならない。

- 1) 個々のデータには誤差が含まれている。
車のナンバープレートなどはそれぞれの値がそれぞれ固有の値であり、誤差がありえない。
- 2) 2, 3のデータではデータ全体の特徴はよくわからないが、データ数を増やしていくとデータ全体の特徴が次第に明らかになっていく。
- 3) もしすべてのデータを調べるとデータ数が1万以上あること。
世界各国のGDP, 市町村の人口などは1万以上はない。

4) 得られたデータが母集団を代表すること.

世界の山の高さなど・・・得られたデータは標高の高い山ばかりだから世界の山を代表しない.
上位 100 人のホームラン数なども同じく母集団を代表しない.

5) 大数の法則に従わない時系列のデータではないこと.

1901 年から 2000 年までの日本の人口, 物価, 交通事故数などは時間とともに変化するので,
データを増やしてもその平均で何かを論じることはできない. 地震の回数などは大数の法則に従
うと考えることができるので調べてもよい.

6) 離散データか連続データであること.

色, 性別などのデータは, 第 2 回目の授業で計算する平均, 標準偏差などの計算をすることが
できない. したがって, 以下の 5 種類のデータのうち, 離散データもしくは連続データに属する
ものでなければならない.

データの種類は大きく, 5 つに分けられる.

名目データ: 順序, 大きさのない属性データ. 例えば, 性別, 色, 血液型など.

順序データ: 順序, 段階だけを示したデータ. 大, 中, 小あるいは数字で 1 ~ 5 (成績) など.

順位データ: 順位を付けたデータ.

離散データ: とびとびの値をとるデータ. 人数, 個数, 値段など.

連続データ: 連続的な値を取るデータ. 身長, 速度, 濃度など.

7) データの値について少なくとも 50 以上の異なる値をとりうること. したがって, 松江市内の
建物の階数, 家の部屋の数, 所持する靴の数などはだめである.

8) ほかの学生のテーマや授業やこのプリントで示した例に似ているものはだめである. 賭け事
および相場に関するものもだめである (麻雀, 競馬, 株価など). できるだけ目新しいテーマを
考えるように.

調査例: イネの穂の長さ (連続データ)

ニワトリの卵の重さ (連続データ)

大橋川に生えているヨシの草丈 (連続データ)

映画に出演するエキストラの数 (離散データ)

宍道湖のシジミの収穫個数 (離散データ)

松江のスーパーで売られる豚肉の値段 (離散データ)

宿題の見本はホームページ <http://www.ipc.shimane-u.ac.jp/food/kobayasi/> を見てください.

宿題はノートに書くか, あるいはプリントアウトをノートに貼るか, バインダーやクリアフ
ァイルなどに入れて, データの入ったエクセルファイルを保存したフロッピーディスクか CD を添
えて, 10 月 14 日 (火) 午前 9 時までに生物資源科学部 2 号館 204 室に提出のこと (メールで
の提出は認めない).

実験計画学予定表

1. 10月 7日 実験計画学とは何か？
 2. 10月14日 平均と分散・データの要約
 3. 10月21日 母集団と標本，確率分布
 4. 10月28日 二項分布，ポアソン分布，正規分布
 5. 11月11日 統計的推定
 6. 11月18日 統計的検定
 7. 11月25日 t分布とt検定
 8. 12月 2日 カイ二乗分布，F分布とその応用
 9. 12月 9日 分散分析その1 一元配置
 10. 12月16日 分散分析その2 二元配置
 11. 1月 6日 分散分析その3 実験計画法
 12. 1月13日 相関分析
 13. 1月20日 回帰分析
 14. 1月27日 実験計画法と統計解析の実際
 15. 2月 3日 期末試験
- 11月4日（火）は月曜日の授業に振替になります。