



情報科での『統計』へのアプローチ

宮城県泉館山高等学校
上杉 茂樹



生徒の実情

- **仙台市北部の2番手進学校**
 - 100%進学希望
 - センター試験受験率90%
- **毎年100名程度が、現役で国公立大学へ進学**
 - 文理半々
- **1学年7クラス**
- **部活動も盛ん**
 - 吹奏楽部・放送部
 - テニス部・弓道部・陸上部

情報科の授業

- **2時間連続**
 - 1時間目は解説と基本実習。T1:上杉にT2が入る。
 - 2時間目は実習、レポート作成。T2が担当
- **課題を重視した評価**
 - 課題60%、考査30%、その他10%の評価
- **ポイントを絞って、狭く、深く**
 - 情報を上手に集める。
 - 情報を上手に整理する。
 - 情報を正しく読み取る。
 - 情報を上手に伝える。





役にたつ『数学』への障害

- 数学の「理論値化」
現象を関数として把握、理論値の計算
- 現実の多様な結果
－確率的な把握，難解な数式
- 大量な情報－手作業の限界
- 「資料の整理」という方向
－ 平均・度数分布表・グラフ
- 「統計」＝入試に出ない単元＝軽視される傾向

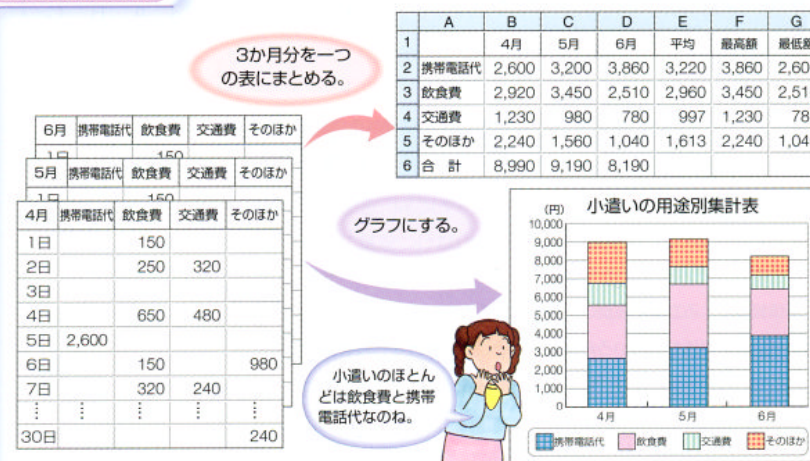
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

表計算ソフトの学習

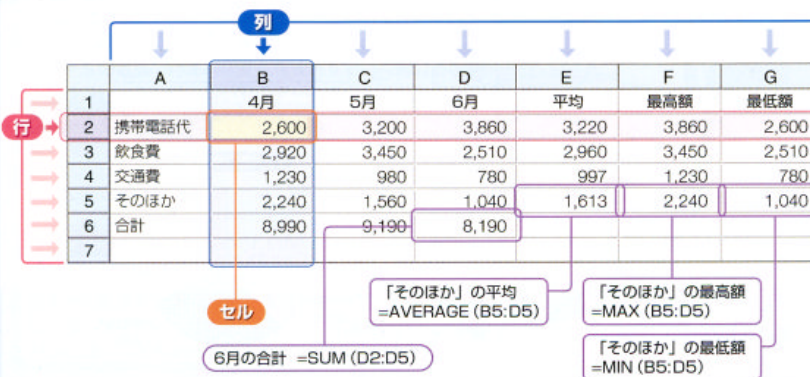
- 中学での学習を想定。
 - 起動と終了
 - 保存と呼び出し
 - セルへの入力
- 平易な使い方は教える必要を感じない。
 - データの入力と書式設定
 - 簡単な数式と計算

小遣い帳の分析



表計算ソフトウェアのワークシート

各セルは、列番号と行番号を組み合わせればよばれる。例 B列の2行目のセル → B2





表計算ソフトの学習

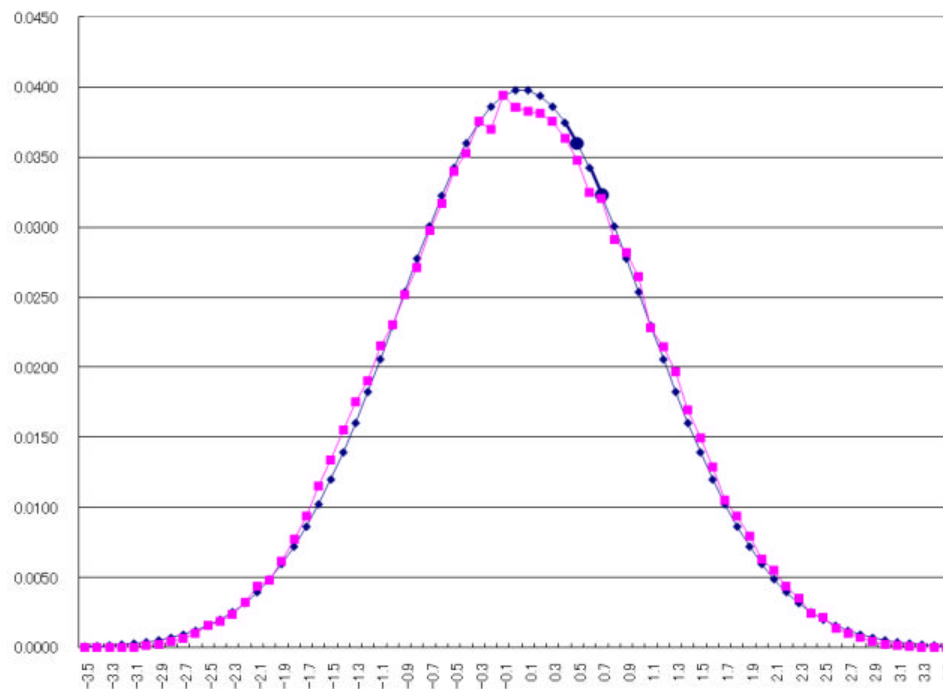
- 教わらないとわからない部分もある。
 - 絶対番地・相対番地
 - 関数式の使用方法
- 表計算ソフトだからできるところもある。
 - 大量のデータの処理
 - グラフ作成

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		1	2	3	4	5	6	7	8	9	10	11	12
2	1	1	2	3	4	5	6	7	8	9	10	11	12
3	2	2	4	6	8	10	12	14	16	18	20	22	24
4	3	3	6	9	12	15	18	21	24	27	30	33	36
5	4	4	8	12	16	20	24	28	32	36	40	44	48
6	5	5	10	15	20	25	30	35	40	45	50	55	60
7	6	6	12	18	24	30	36	42	48	54	60	66	72
8	7	7	14	21	28	35	42	49	56	63	70	77	84
9	8	8	16	24	32	40	48	56	64	72	80	88	96
10	9	9	18	27	36	45	54	63	72	81	90	99	108
11	10	10	20	30	40	50	60	70	80	90	100	110	120
12	11	11	22	33	44	55	66	77	88	99	110	121	132
13	12	12	24	36	48	60	72	84	96	108	120	132	144

数理統計の難点

- **膨大な計算**
 - 表計算ソフトの得意領域
- **正規分布へのアプローチ**
 - 「乱数」から生成可能
- **実際の問題への応用**
 - 課題を与えてレポート作成

中心極限定理が効く





指導の方向

- **表計算ソフトの使用方法**
 - 絶対番地・相対番地
 - 計算式の入力とコピー、計算結果、グラフ
- **数式の意味の理解＝「数学」**
 - 式表現としての Σ への習熟
 - 平均・分散・標準偏差・相関係数
- **データ分析への挑戦**
 - 「道具」を活用しての解析～レポート作成
 - 解析方法の思索、PCによる計算、結果の整理と考察
 - 難解。「正解」は期待できない。



大切なのは統計的(確率的)な感覚



統計の授業(座学 第1回)

■ 用語の確認と作業

- 全体の目標
 - 合計・平均・最大・最小
- 数値としてデータ理解
 - 合計・平均・最大・最小
- 分布としてデータ理解
 - 度数分布表
- テストデータで作業

辞書 A.0911.

数理統計 (1) .

- ローリスク・ハイリターンを求めて - .

Data1	Data2
56	70
73	53
46	50
50	37
56	45
54	41
56	81
69	55
55	36
60	40
56	66
52	49
74	44
74	44
78	62
58	94
47	47
37	77
51	58
50	50
71	25
97	69
71	72
62	66
37	59
94	57
73	75
81	73
51	79
47	65
32	51
60	83
91	65
77	73
57	96
66	65
58	64
41	91
44	71
76	73
82	62
42	54
55	74
46	69
59	41
54	47
81	45
63	47
58	42
63	41
38	49

【今日の用語】 .

数理統計・合計・平均値・最大値・最小値 .

階級・階級値・度数・度数分布表・相対度数・ヒストグラム .

1 数理統計の目標 .

2 データを理解する数値を換す . .

(1) 合計・平均・最大値・最小値 .

3 データ全体を把握する . .

(1) 度数分布表 .



統計の授業(表計算ソフト入門 第1回)

■ Excelを学ぶための準備

- 各部の名称
- 拡張子
- 簡単な計算式
- 簡単な関数

- Int()

■ 実験データの入力

- 最後のレポートの題材

四則演算の計算順序

- $() > ^ > * , / > + , -$
 - (カッコ)が最優先
 - * の省略はできない
- 計算例
 - 次の計算式を入力してみよ。
 - ① $30 - 20 / 5 + 5$
 - ② $30 - 20 / (5 + 5)$
 - ③ $(30 - 20) / 5 + 5$
 - ④ $(30 - 20) / (5 + 5)$
 - ①～④の計算の順序を確かめよ





統計の授業(座学 第2回)

- Σ を用いた式表現
 - Σ を用いた表現方法
 - 合計の表現
 - 平均の表現
- Σ の計算練習

数学Bの数列の単元で、 Σ は公式を用いた計算が強調されるので、ここでは「和」をまとめて表現していることを強調する。

情報A 月 日 課題

Σ に慣れよう。

提出期限：次回の「情報」の授業。
(冊番が予め決めておいて提出せよ)。

1 データが次の値をとるとき、 Σ で表された次の計算を、 Σ を一変和の形に書き直してから行え。計算は電卓を用いて良い。

$\{x_i\} = \{95, 23, 66, 87, 43, 51, 76, 40, 85, 79, 58, 69, 77, 99, 48, 72, 69, 58, 16, 38\}$

(1) $\sum_{i=1}^{20} x_i$

..

(2) $\sum_{i=1}^{20} x_i$

..

..

(3) $\sum_{i=1}^{20} x_i$

..

..

2 Σ であらわされた次の計算を、一変和の形に書き直してから行え。

(1) $\sum_{k=1}^n k$ (2) $\sum_{k=1}^n (k^2 - k)$

..

..

..

..

..

..

..

..

..

..

3 Σ であらわされた、次の計算を行え。

$\sum_{k=1}^n (k^2 - a)$



統計の授業(表計算ソフト入門 第2回)

■ 関数式の利用

- データファイル(TestData.xls)のオープン
- 範囲指定の方法
- 関数式を用いた計算の方法
 - sum(), avarage(), max(), min()
- TestData.xlsを用いて計算を行う。
 - Data1+Data2
 - sum(), avarage(), max(), min()
- データファイルを保存する。

	A	B	C	D	E
1		Data.1	Data.2	D1+D2	順位
2		56	70	126	20
3		73	53	126	20
4		46	50	96	41
5		50	37	87	48
38		41	91	132	15
39		44	71	115	27
40		76	73	149	8
41		82	62	144	10
42		42	54	96	41
43		55	74	129	18
44		46	69	115	27
45		59	41	100	37
46		54	47	101	34
47		81	45	126	20
48		63	47	110	31
49		58	42	100	37
50		63	41	104	33
51		38	49	87	48
52	合計	3005	2998		
53	平均	60.1	59.96		
54	最大	97	96		
55	最小	32	25		

統計の授業(座学 第3回)

- データの変換
 - $y_i = ax_i + b$ の平均 \bar{y}
- データの散らばり
 - ヒストグラムによる調査
 - 偏差と問題点
 - 分散・標準偏差

(課題)

- $\sum_{k=1}^n (x_k - \bar{x}) = 0$ の証明

0. 5 準備 $y_i = ax_i + b$ の平均 \bar{y}

データ $\{x_1, x_2, x_3, \dots, x_n\}$ が与えられたとき、新しいデータ $y_i = ax_i + b$ をつくる。

したがって、 $\{x_1, x_2, x_3, \dots, x_n\}$ の平均 \bar{x} は既知である。 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ のとき、

$$\begin{aligned} \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} (y_1 + y_2 + y_3 + \dots + y_n) \\ &= \frac{1}{n} \{ (ax_1 + b) + (ax_2 + b) + (ax_3 + b) + \dots + (ax_n + b) \} \\ &= a \cdot \frac{1}{n} (x_1 + x_2 + x_3 + \dots + x_n) + \frac{1}{n} (b + b + b + \dots + b) = a\bar{x} + b \end{aligned}$$

よって、 $y_i = ax_i + b$ の平均は、 $\bar{y} = a\bar{x} + b$ で与えられる。

そこで、2種類のデータの集まりを準備して考えを進めてみる。

1 データの散らばりを考える。(具体的に生徒に平均を計算させる)

例1 60, 45, 50, 40, 50, 55

例2 80, 10, 50, 20, 90

平均 $\bar{x} = 50$

平均 $\bar{x} = 50$



上図のように、データを整理したものをヒストグラムという。また、数値だけを整理した表を度数分布表という。

課題 次の式が一般に成立することを確かめよ。次回までレポートで提出。

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

【回答例】

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}) &= (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \dots + (x_n - \bar{x}) \\ &= (x_1 + x_2 + x_3 + \dots + x_n) - (\bar{x} + \bar{x} + \bar{x} + \dots + \bar{x}) = \sum_{i=1}^n x_i - n \cdot \frac{1}{n} \sum_{i=1}^n x_i = 0 \end{aligned}$$



統計の授業(表計算ソフト入門 第3回)

- **絶対番地と相対番地**
 - 絶対番地・総体番地の意味
 - 指定の方法
 - かけ算九九の表

- **順位の計算**
 - TestData.xlsでrank()を用いる。

- **課題提出**
 - これまでの作業の整理

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		1	2	3	4	5	6	7	8	9	10	11	12
2	1	1	2	3	4	5	6	7	8	9	10	11	12
3	2	2	4	6	8	10	12	14	16	18	20	22	24
4	3	3	6	9	12	15	18	21	24	27	30	33	36
5	4	4	8	12	16	20	24	28	32	36	40	44	48
6	5	5	10	15	20	25	30	35	40	45	50	55	60
7	6	6	12	18	24	30	36	42	48	54	60	66	72
8	7	7	14	21	28	35	42	49	56	63	70	77	84
9	8	8	16	24	32	40	48	56	64	72	80	88	96
10	9	9	18	27	36	45	54	63	72	81	90	99	108
11	10	10	20	30	40	50	60	70	80	90	100	110	120
12	11	11	22	33	44	55	66	77	88	99	110	121	132
13	12	12	24	36	48	60	72	84	96	108	120	132	144



統計の授業(座学 第4回)

■ データの標準化による統計値の変化

$$y_i = \frac{x_i - \bar{x}}{\sigma}$$

– 平均・分散・標準偏差の証明

■ 散布図

- 作り方と意味
- 作業は課題

データを基に散布図を書こう ～2項目間の関連の強さを調べる～

例

日	東京(x)	大阪(y)	日	東京(x)	大阪(y)
1	12.0	11.5	16	20.8	24.0
2	16.7	18.6	17	22.3	22.9
3	21.3	15.4	18	22.1	20.4
4	14.5	14.3	19	20.0	24.4
5	17.9	17.5	20	25.7	20.2
6	20.7	22.4	21	13.3	12.2
7	20.0	22.4	22	20.3	18.2
8	20.6	24.3	23	18.2	19.9
9	21.5	22.8	24	21.9	20.4
10	22.1	27.0	25	15.8	16.7
11	23.1	24.7	26	17.8	20.0
12	22.4	17.7	27	19.7	21.7
13	18.5	18.0	28	22.2	23.2
14	20.8	18.7	29	21.5	19.5
15	22.3	19.9	30	17.3	21.9
			平均	19.6	19.7

◎ 散布図から言えることを考えよう。↓



統計の授業(表計算ソフト入門 第4回)

■ 統計値の計算

- 分散・標準偏差の計算
- 度数分布表の作成
- グラフの作成

度数分布表を作る

■ =frequency(データ範囲, 階級範囲)

- セルA58は 度数分布表 の表示
- 階級を作る

例

50	50以下の値のデータ数
60	50より大きく60以下の値のデータ数
	60より大きい値のデータ数

- 関数を埋め込む範囲を選択 B59:B70
- データ範囲は B2:B51
- 階級範囲は A59:A69
- [Ctrl] + [Shift] + [Enter] で決定



☆ Frequency()は使い方の難しい関数

- ・ 先に関数を埋め込む範囲を指定しない。
- ・ [Ctrl]+[Shift]+[Enter] を [Enter] だけ入れるなどのミスが多発するので注意して指導する必要がある。



統計の授業(座学 第5回)

■ 散布図の作成と相関

- 点の分布による理解
- 正の相関・負の相関
- 相関がない

■ 相関係数

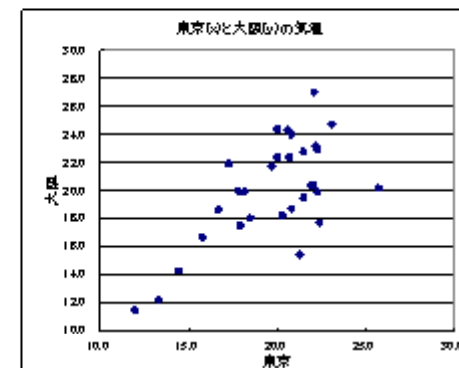
- 考え方
- 偏差積と共分散
- 相関係数

統計入門5.1

- 2変数の関係(散布図・相関係数) -

1 データを基に散布図を書こう ～2項目間の関連の強さを調べる～

日	東京(℃)	大阪(℃)	日	東京(℃)	大阪(℃)
1	12.0	11.5	18	20.8	24.0
2	18.7	18.8	17	22.3	22.9
3	21.3	15.4	18	22.1	20.4
4	14.5	14.3	19	20.0	24.4
5	17.9	17.5	20	23.7	20.2
6	20.7	22.4	21	13.3	12.2
7	20.0	22.4	22	20.3	18.2
8	20.8	24.3	23	18.2	19.9
9	21.5	22.8	24	21.9	20.4
10	22.1	27.0	25	15.8	18.7
11	23.1	24.7	26	17.8	20.0
12	22.4	17.7	27	19.7	21.7
13	18.5	18.0	28	22.2	23.2
14	20.8	18.7	29	21.5	19.5
15	22.3	19.9	30	17.3	21.9
			平均	19.8	19.7



2 正の相関・負の相関

(1) 負の相関

(2) 相関なし

(3) 正の相関

3 相関関係を数値化する。-相関係数-

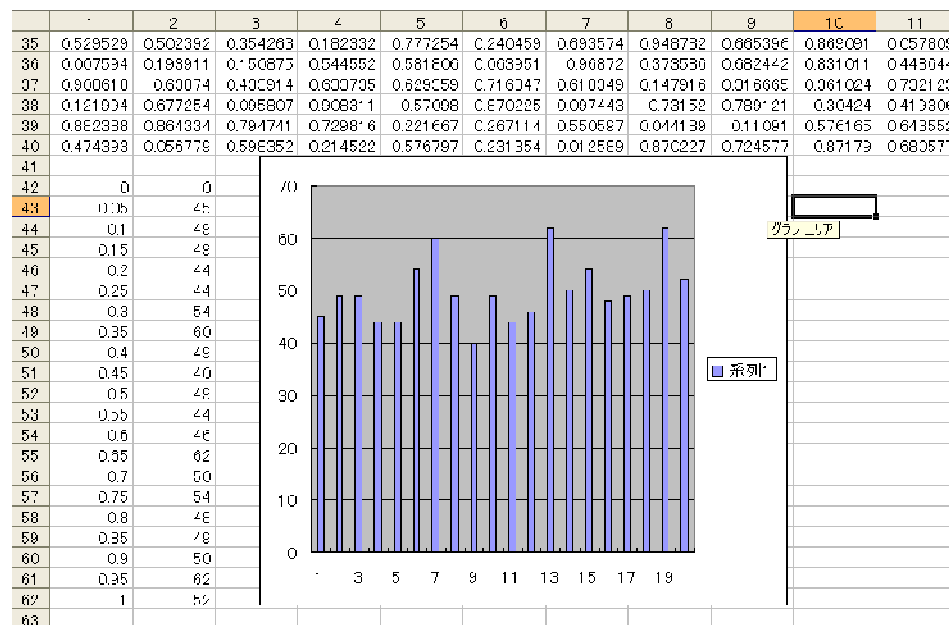
統計の授業(表計算ソフト入門 第5回)

■ 乱数

- 乱数とは
- 乱数の発生とデータの作成
- =rand()の問題
- 統計値
 - 平均・分散・標準偏差
- 分布
 - 一様分布

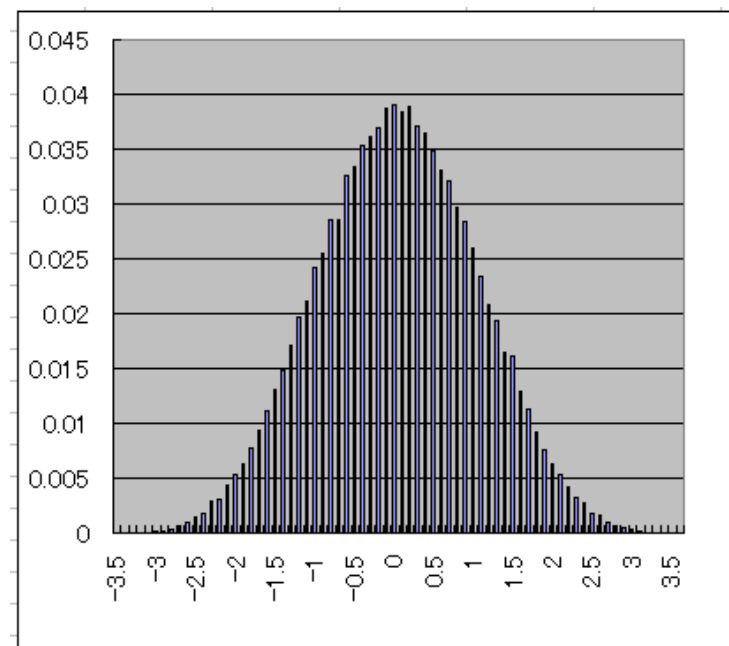
☆ きれいな分布を出すために大量の乱数を発生させる。

☆ 再計算の関係で、注意しないとハングする。
関数式はデータ作成後消去する事を徹底する。



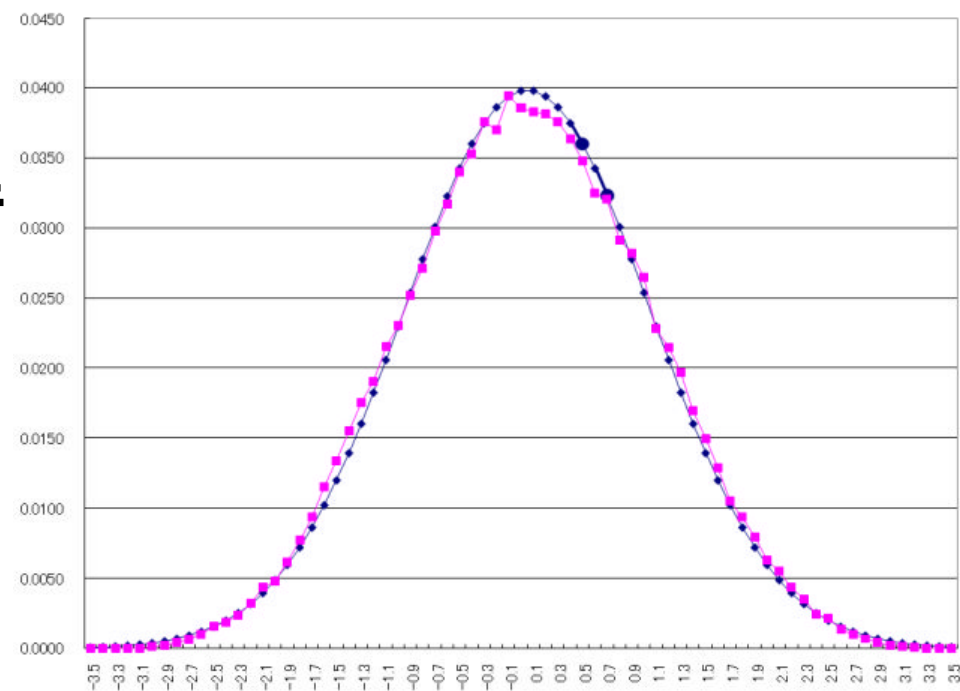
統計の授業(第6回)

- **相関係数の補足**
 - 「相関がない」時の判断
 - x, y を乱数で生成したときの相関係数を計算する。
- **正規分布を作成する。**
 - 1つのデータは乱数を3つ加えて作る。
 - データを100,000個発生する。
 - 度数分布表を作る。
 - ヒストグラムを作る。



統計の授業(第6回②)

- データの標準化 $y_i = \frac{x_i - \bar{x}}{\sigma}$
 - 平均 0
 - 標準偏差 1
 - 標準正規分布
- (課題)
これまでの作業の整理



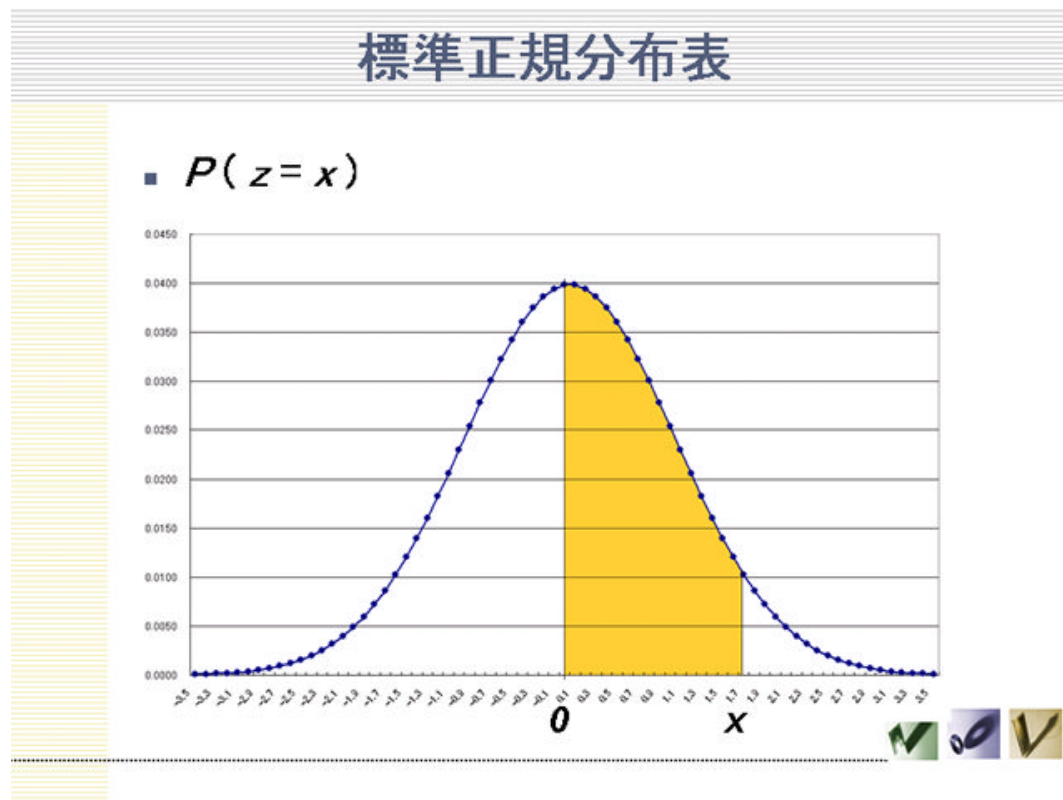
統計の授業(第7回①)

■ 標準正規分布

- 定義
- 性質
- 性質
- 表とその意味
- 表の活用

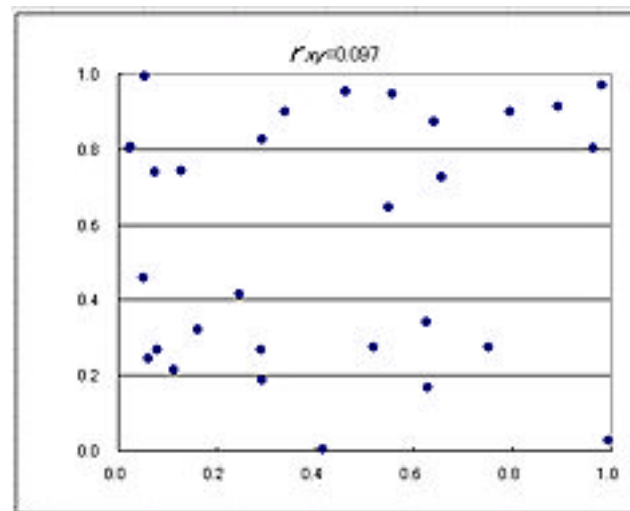
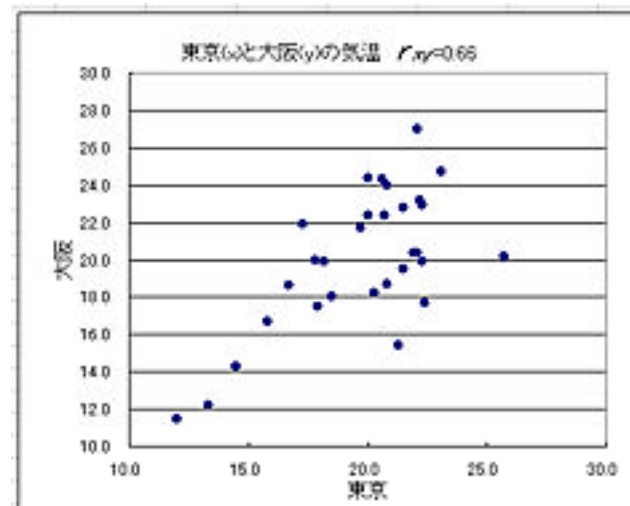
■ 偏差値

- 偏差値の意味
- 計算式
- 標準正規分布表を利用した計算方法と計算練習



統計の授業(第7回②)

- 相関係数の計算
 - 相関係数の性質
 - 相関係数の計算
 - 散布図の作成
 - 散布図と相関係数
- まとめ
 - これまでの知識の整理
 - データ解析の方法





統計の授業(第8回:レポート作成)

■ テーマ1

「平均値回帰の原則」は成り立つか

- 最初に学年の生徒全員の両親の身長・自分の身長を入力したファイルを準備しておき、このデータを元に論述する。

■ テーマ2

「泉館山高校生に最適な靴のサイズと生産量」

- 最初に学年の生徒全員の左右の足の大きさ、幅を計測してデータファイルを作成しておき、これを各自が加工して結論を導き出す。



授業を実施しての問題

- **難解になりがち**
 - 数学的思考も重要
 - Excelの操作も間違えやすい。
- **時間がかかる**
 - 速く進むと益々難解に
- **最後のレポートの処理**
 - 「分析方法」から考えることの難しさ。
 - レポートがまとめられない。
 - 追跡しての指導が不可能
 - 評価も難しい。



おまけ 表計算ソフトのおもしろい例

■ 住宅ローンの計算

- 定数エリア a1:f3
- 第4行・第5行に計算式入力
- 第5行を下方へコピー
- 定数部の数値を変化させて結果を観察

The screenshot shows a Microsoft Excel spreadsheet titled "Microsoft Excel - ローン". The spreadsheet contains a table with columns A through F and rows 1 through 12. The table is as follows:

	A	B	C	D	E	F
1	借入額	30000000	月返済額	65000	年間総返済額	1580000
2	金利(%)	3.5	ボーナス加	400000	返済回数	
3	返済回	借用額	利息	返済額	残額	総返済額
4	1	30000000	87500	65000	30022500	65000
5	2	30022500	87565	65000	30045065	130000
6	3	30045065	87631	465000	29667696	595000
7	4	29667696	86530	65000	29689226	660000
8	5	29689226	86593	65000	29710819	725000
9	6	29710819	86656	65000	29732475	790000
10	7	29732475	86719	65000	29754194	855000
11	8	29754194	86783	65000	29775977	920000
12	9	29775977	86846	465000	29397823	1385000



各種情報の提供

■ P.SITE

(宮城県高等学校情報科教育研究会運営)

<http://e-net3.edu-c.pref.miyagi.jp/psite/>

P.SITEでは、以下の場所に旧版を学習指導案、パワーポイントファイル付きで公開しています。

[授業事例]－[情報A]－[情報の統合的な処理]

■ Sige-Lab (上杉個人サイト)

- まとめサイト <http://www016.upp.so-net.ne.jp/sige-lab/>
- ブログ <http://sige-lab.blog.so-net.ne.jp/>
- メール s_uesugi@izumit-h.myswan.ne.jp