

確率と統計

佐賀大学 新井康平

クロス集計

- 1年生100匹に、おやつに何を持っていきたいかを聞いてみました。次の表はその結果です。

持っていきたいおやつ（オス）	度数	持っていきたいおやつ（メス）	度数
かつおぶし	25	かつおぶし	20
にぼし	28	にぼし	12
ささみ	3	ささみ	7
チーズ	4	チーズ	1
合計	60	合計	40

度数分布

- ねこ第1小学校の新入猫100匹のデータを使って量的変数を集計してみましょう。量的変数の集計には「度数分布表」というものがよく使われます。例えば、体長のデータを使って度数分布表を作ると次のようになります。

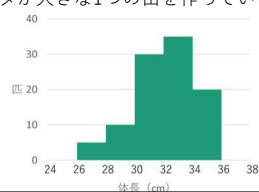
階級 (cm以上～cm未満)	階級値 (cm)	度数	相対度数	累積相対度数
26～28	27	5	5%	5%
28～30	29	10	10%	15%
30～32	31	30	30%	45%
32～34	33	35	35%	80%
34～36	35	20	20%	100%

- この2つの集計表をまとめると「クロス集計表」というものができあがります。クロス集計表は、2つのカテゴリーに属するデータをそれぞれのカテゴリーで同時に分類し、その度数を集計したものです。

	かつおぶし	にぼし	ささみ	チーズ	合計
オス	25	28	3	4	60
メス	20	12	7	1	40
合計	45	40	10	5	100

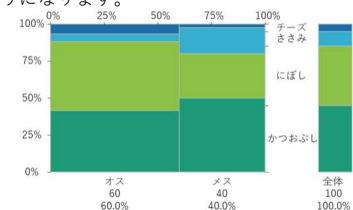
ヒストグラム

- 度数分布表が出来上がると、「ヒストグラム」を使ってグラフを描くことができます。横軸が体長の階級を、縦軸が度数を表しています。ヒストグラムを見ると、体長32cm～34cmあたりを中心にデータが大きな1つの山を作っていることが分ります。

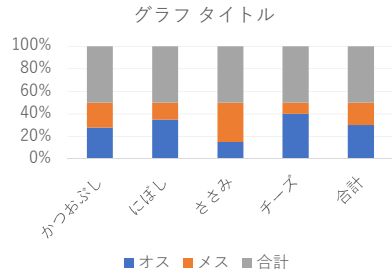


モザイク図

- 「モザイク図」を使うと、それぞれの度数の割合をより分かりやすく表現することができます。上のクロス集計表からモザイク図を作ると次のようになります。



積み上げ棒グラフ



平均値

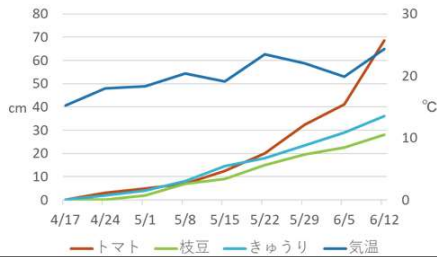
$$\frac{1000 + 500 + 700 + 1200 + 800}{5} = 840$$

名前	お小遣いの金額 (円)
ちび	1000
はな	500
たま	700
そら	1200
くろ	800

時系列データ

日付	トマト (cm)	枝豆 (cm)	きゅうり (cm)	気温 (℃)	測定係
4月17日	0	0	0	15.2	たま、ぶち
4月24日	3	0	2	18.0	ほす、ひげ
5月1日	5	2	4	18.3	しろ、みけ
5月8日	7	7	8	20.4	もち、みみ
5月15日	12.5	9	14.5	19.1	くろ、とら
5月22日	20	15	18	23.5	そら、うみ
5月29日	32.5	19.5	23.5	22.0	はる、あき
6月5日	41	22.5	29	19.9	りん、まる
6月12日	68.5	28	36	24.3	じゅん、ちよび

出店	並んでいる猫の数 (匹)	・中央値	出店	並んでいる猫の数 (匹)
わたがし	3		射的	0
焼き鳥	1		焼き鳥	1
金魚すくい	10		焼きトウモロコシ	2
たこ焼き	6		わたあめ	3
射的	0		たこ焼き	6
リンゴ飴	7		リンゴ飴	7
焼きトウモロコシ	2		くじ引き	7
かき氷	12		水ヨーヨー	8
水ヨーヨー	8		金魚すくい	10
くじ引き	7		かき氷	12



最頻値200円

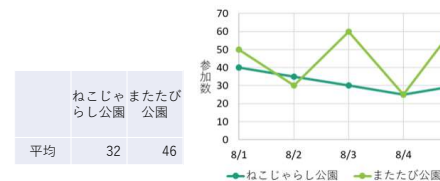
値段 (円)	度数
50	1
100	3
150	2
200	4
300	2
400	1

出店	値段 (円)
はし巻き	300
焼き鳥	100
焼きトウモロコシ	200
わたあめ	100
たこ焼き	400
りんご飴	150
たい焼き	100
チョコバナナ	200
わらび餅	200
ラムネ	150
ポップコーン	200
水あめ	50
アユの塩焼き	300

四分位数

- 「四分位数（しぶんいすう）」とはデータを小さい順に並び替えたときに、データの数で4等分した時の区切り値のことです。4等分すると3つの区切りの値が得られ、小さいほうから「25パーセンタイル（第一四分位数）」、「50パーセンタイル（中央値）」、「75パーセンタイル（第三四分位数）」とよびます。

順番	曲目	楽曲の時間（分）
1	cats celebrate you	3.0
2	猫ダンス	4.0
3	TSUNAKAN	5.5
4	雲の上ではディセンバー	3.5
5	ルビーの首輪	4.2
6	恋するフォーチュンカカリ	3.4
7	WAIになって眠ろう	2.8
8	海も泳げるはず	4.2
9	かつおぶしだよ人生は	4.7
10	破れかけのfusuma	2.2
11	髪をこめてねこじゃらしを	3.8



分散の求め方

日付	ねこじゃらし公園	1. 各データと平均値の差	2. (各データと平均値の差)の2乗
8月1日	40	40-32=8	8 ² =64
8月2日	35	35-32=3	3 ² =9
8月3日	30	30-32=-2	(-2) ² =4
8月4日	25	25-32=-7	(-7) ² =49
8月5日	30	30-32=-2	(-2) ² =4

	ねこじゃらし公園	またたび公園
分散	26	254
標準偏差	5.1	15.9

データのばらつき

- ラジオ体操は、校区内の様々な場所で行われています。次の表は山のふもとと町内の2つの公園で行われているラジオ体操の参加数をまとめたものです。

日付	ねこじゃらし公園	またたび公園
8月1日	40	50
8月2日	35	30
8月3日	30	60
8月4日	25	25
8月5日	30	65

変動係数 = 標準偏差 ÷ 平均値

枝が立派な木(m)	赤い実がなる木(m)	大きな葉っぱの木(m)
12	4	0.8
15	5	1.2
15	3	1.0
12	5	1.6
13	—	0.7
14	—	—

	枝が立派な木(m)	赤い実がなる木(m)	大きな葉っぱの木(m)
平均	13.5	4.3	1.1
標準偏差	1.4	1.0	0.4
変動係数	0.1	0.2	0.3