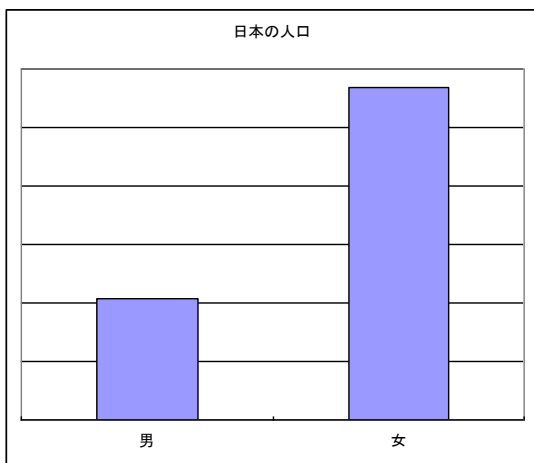


練習問題

1. 「サラリーマンの年収（給与所得）の平均は約 461 万円である」という記述は、情報としてはほとんど意味がない。絶対的に不足しているものは何か、また、詳しく検討するためには何が必要か。

いつの情報なのかが示されていない。地域が限定されていない（日本全体なのか特定の地域あるいは都道府県なのか）。
男女別、職種・業種別、年齢別、勤続年数別、など。

2. 次のグラフから分かることは何か。より良いグラフにするためには、何が不足しているか。



女性の人口の方が多いたことが示されている。
いつの情報なのかが示されていない。
単位が示されていない。
縦軸の数値目盛が示されていない。

3. それぞれの問いに答えよ。

- (1) ある企業の製品の出荷量は、2年目は1年目の2倍、3年目は2年目の4倍に増加した。1年あたりでは平均的に何倍ずつ増えたことになるか。

1年目を基準とすると、3年目には $2 \times 4 = 8$ 倍増えたことになるので、1年あたりの倍率を x とすると、 $x^2 = 8$ という関係になっている。これを x について解いて、
 $x = \sqrt{8} = 2\sqrt{2} \doteq 2.828$
(一般的には幾何平均を用いる)

- (2) ある場所に、行きは平均 30km/h、帰りは平均 40km/h で往復した。往復とも同じ道（同じ距離）を移動したとすると、全行程の移動速度は平均何 km/h となるか。

片道の距離を l [km] とすると、行き・帰りの移動時間はそれぞれ $\frac{l}{30}$ [h], $\frac{l}{40}$ [h] であるので、総移動時間 $\frac{l}{30} + \frac{l}{40}$ [h] で全行程の移動距離 $2l$ [km] を除して、
$$\frac{2l}{\frac{l}{30} + \frac{l}{40}} = \frac{1}{\frac{1}{2}(\frac{1}{30} + \frac{1}{40})} = \frac{2400}{70} \doteq 34.29$$
 [km/h] (一般的には調和平均を用いる)

(3) ふつうのサイコロを2つ同時に振った時に出る目の和の期待値はいくつか。

2つの目の和の出現パターンは、2, 3, ..., 12までの11通り。この時、それぞれのサイコロの目の組み合わせ方は、和が2→1と1, 和が3→1と2または2と1（の2通り）...のように考えて、それぞれの目が出る確率が1/6であることから、

$$2 \times (1/6 \times 1/6) + 3 \times (1/6 \times 1/6 + 1/6 \times 1/6) + \dots + 12 \times (1/6 \times 1/6) =$$

$$1/36 \times (2 \times 1 + 3 \times 2 + 4 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 6 + 8 \times 5 + 9 \times 4 + 10 \times 3 + 11 \times 2 + 12 \times 1) =$$

$$1/36 \times (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12) = 252/36 = 7 \quad (\text{一般的には加重平均})$$

4. ある小学校のクラス（30人学級）で、その年にもらったお年玉の額を調べたい。適当に子を選んで調べたところ、1人は2万円、もう1人は4万円であった。これらのことから、このクラスの小学生たちは、どの位のお年玉をもらったと推測できるか。

点推定では $(2+4)/2 = 3$ [万円] とみるのが妥当。
 詳細は省くが、抜き取ったデータの数を n 、それぞれの標本値を X_1, X_2, \dots, X_n とすると、標本分散 $s^2 = \sum (X_i - \bar{X})^2 / n$ を用いて、母集団の平均 μ は、

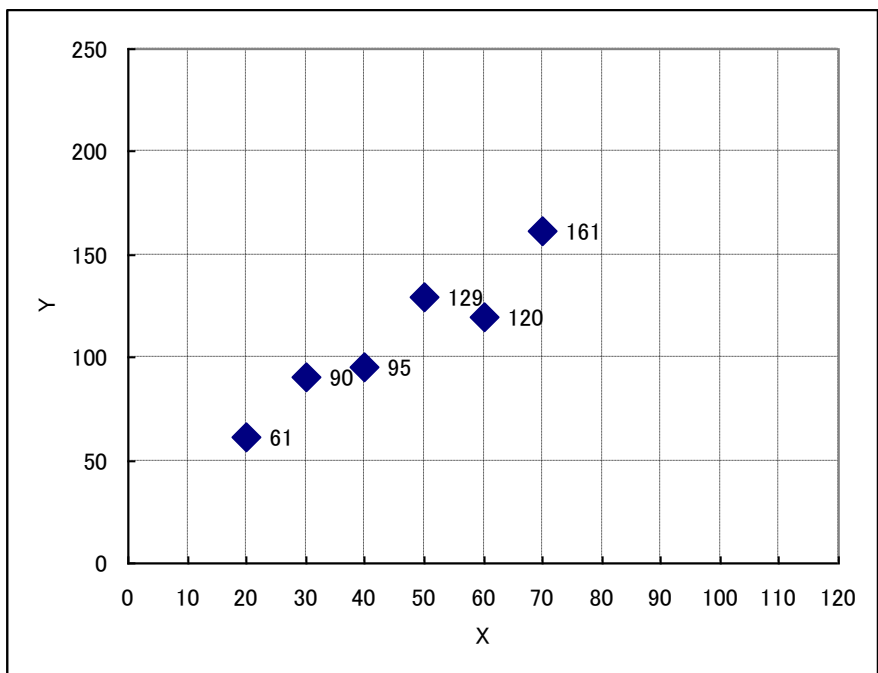
$$\bar{X} - t \frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t \frac{s}{\sqrt{n-1}}$$

を満たす範囲にあると示すことができる。ここで、 t は t 分布から得られる値で、標本数から得られる自由度 $\nu (= n - 1)$ と所定の確率（99%、95%、90%などが使用される）によって決まる。この問題の場合、

$$n = 2, \bar{X} = \frac{2+4}{2} = 3, s = \sqrt{\frac{1}{2}((2-3)^2 + (4-3)^2)} = 1$$

なので、95%の確率（信頼性）の場合、自由度 $\nu = 1 (= 2 - 1)$ の t 分布に従うので、その値 $t = 12.71$ を用いて、母集団の平均は 3 ± 12.71 の範囲であると導ける（区間推定）。

5. 2次元の量からなるデータを散布図上に描いた。X=100の時、Yはいくつになると予想できるか。



一般的には、回帰分析を用いて Y と X の関係を開数として表し、X=100 に対応する開数値を導けばよい。

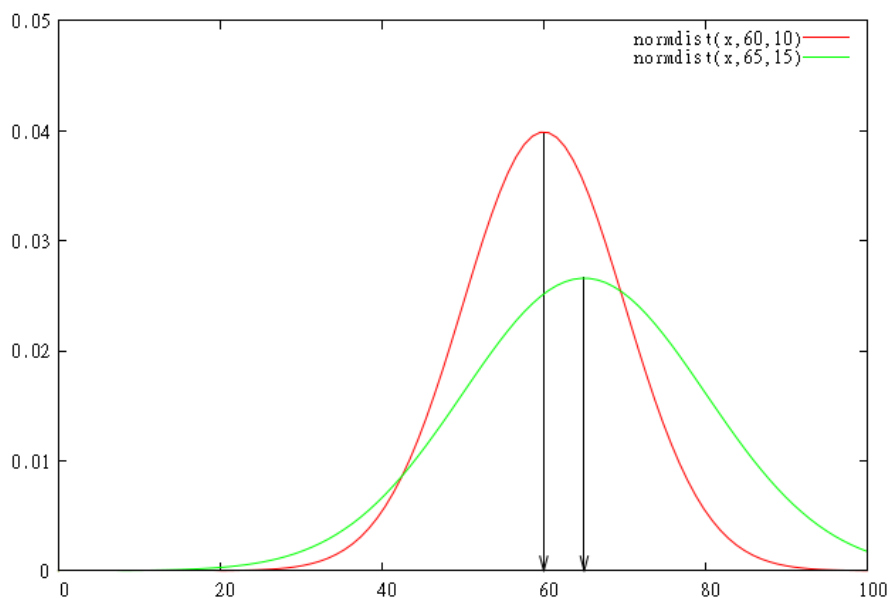
線形（1次）： $Y=a+b \cdot X$

2次： $Y=a+b \cdot X+c \cdot X^2$

など多項式や指数関数など、当てはめる関数によって結果は異なる。線形の場合 Excel では「分析ツール」の「回帰分析」を用いたり、簡便には散布図に「近似曲線」を数式付きで重ね書きすることで a や b などの値を具体的に知ることができる。

6. あるクラスの数学の試験の結果は、1学期は平均 60 点で、標準偏差は 10 点であり、2学期は、平均 65 点、標準偏差は 15 点であった。クラスの人気は 30 人で、どちらのテストも欠席者はいなかったとすると、平均点は伸びたと言えるか。

正規分布を仮定して、1学期・2学期の分布をグラフ表示すると下図のようになる。



平均や分散が標本で与えられる場合は、母分散の同値性を検定後、通常の t 検定 (母分散が同じ場合) または Welch の t 検定 (母分散が異なる場合) を行う。

7. 次のデータの尺度は何か。
 - (1) 電話番号 **名義尺度**
 - (2) 誕生日 **間隔尺度**
 - (3) パソコンの価格 **比例尺度**

8. 次ページのデータを Excel のワークシートに入力し、該当する基本的な統計量を求めなさい。
具体的な関数名はヘルプ機能や関数ウィザードを用いて調べてみなさい。

Excel 練習用のデータ例

	A	B	C	D	E
1	No.	データ群 A	データ群 B	データ群 C	データ群 D
2	1	1	1	1	1
3	2	1	2	1	1
4	3	2	2	1	1
5	4	2	2	2	2
6	5	2	2	2	2
7	6	3	2	2	2
8	7	3	2	3	2
9	8	3	3	3	2
10	9	3	3	3	3
11	10	3	3	4	3
12	11	4	3	4	4
13	12	4	4	4	4
14	13	4	4	4	5
15	14	4	4	5	5
16	15	4	5	5	5
17	16	5	5	5	5
18	17	5	5	5	5
19	18	5	6	5	6
20	19	6	6	5	6
21	20	6	6	6	6
22	算術平均 AVERAGE				
23	幾何平均 GEOMEAN				
24	調和平均 HMEAN				
25	最小値 MIN				
26	最大値 MAX				
27	レンジ(最大値-最小値) =B26-B25 などで計算				
28	分散 VAR				
29	標準偏差 STDEV				

データは 2~21 行 (A 列が数字のもの)