Pythonによる統計検定
（使い分け用カタログ）

データ形式に応じて検定を使い分けできることが目標

※二項検定以外では、ばらつきに正規分布を仮定します

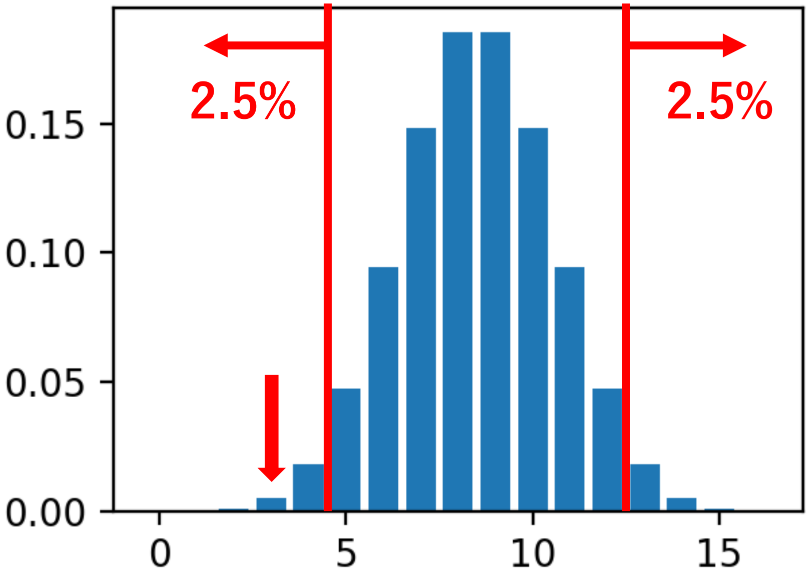
二項検定：ある種Aの雄と雌の性比は等しいか?

(Sokal & Rohlf, “生物統計学”, 1983)

[データ] 野外において1腹から採取された子どもの数

雄	3
雌	14

[仮説] 性比1:1を仮定した時の二項分布

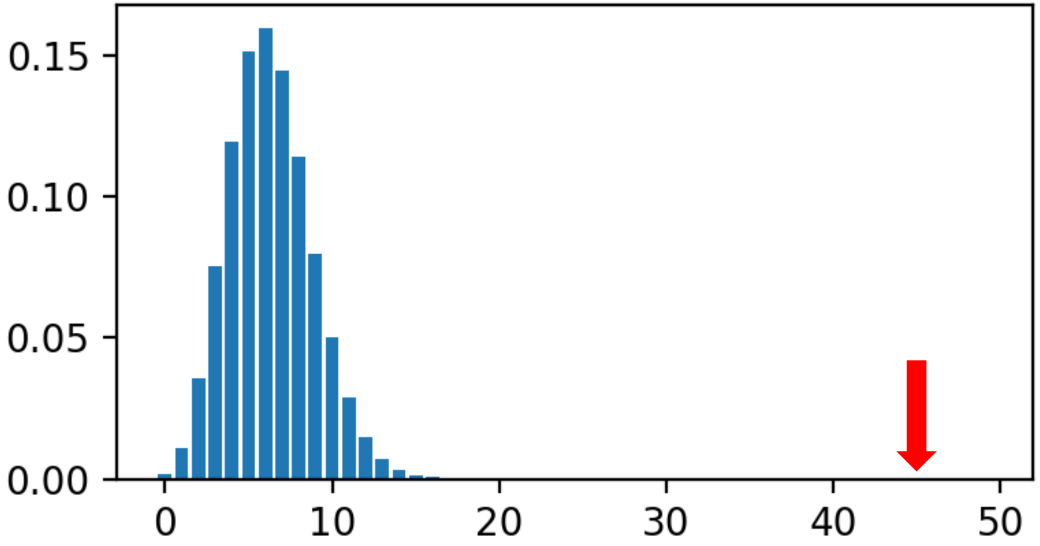



```
# 二項検定（p=0.5を仮定）
from scipy.stats import binomtest
binomtest(3, 17, 0.5).pvalue

# 結論（p値と考察）
# p=0.0127(<0.05)なので、有意に0.5から異なっている

# 図のプロット
from scipy.stats import binom
import matplotlib.pyplot as plt
x = range(17)
y = binom(17, 0.5).pmf(x)
plt.bar(x, y)
plt.show()
```

二項検定2: 20年以上アスベストに暴露された労働者の肺がん発症率(45/632)は、全国平均0.01と同じか?

(Selikoff, Churg, Hammond, JAMA 188:22-6, 1964)

肺がん発症	45
肺がん非発症	587




```
# 二項検定（p=0.01を仮定）
from scipy.stats import binomtest
binomtest(45, 632, 0.01).pvalue

# 結論（p値と考察）
# p=5.7*10-24 (<0.05)なので、全国平均よりも有意に高い

# 図のプロット
from scipy.stats import binom
import matplotlib.pyplot as plt
x = range(50)
y = binom(632, 0.01).pmf(x)
plt.bar(x, y)
plt.show()
```

独立性の検定：母親のサリドマイド服用は奇形児の割合を高めるか?

(W. Lenz & K. Knapp, Arch Environ Health, 1962; 柳川堯, “観察データの多変量解析”, 2016)

	サ剤服用	非服用
Yes	90	22
No	2	186

```
# 左のテーブルの2x2データを行列の形で入力する
count = [[90, 22], [2, 186]]

# χ²検定（独立を仮定）
from scipy.stats import chi2_contingency
chi2_contingency(count).pvalue

# 結論（p値と考察）
# p=3.0*10^-46 (<0.05)なので、有意な差が見られる
```

独立性の検定2：新抗生物質ストレプトマイシンは結核に効くか？

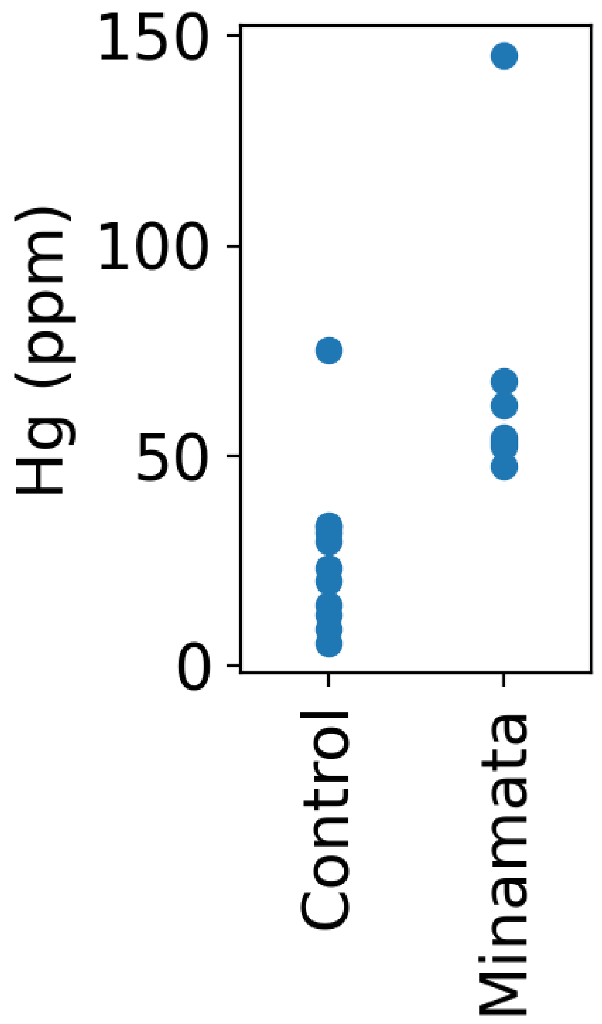
(Tuberc. Chemotherapy Trial Committee, BMRC, *Brit Med J*, **2**:769-82, 1948; 鶴田陽和, “独習統計学24講”, 2013)

	投与	安静療法
生存	51	38
死亡	4	14

```
# 左のテーブルの2x2データを行列の形で入力する
count = [[51, 38], [4, 14]]

# χ^2検定（独立を仮定）
from scipy.stats import chi2_contingency
chi2_contingency(count).pvalue

# 結論（p値と考察）
# p=0.014(<0.05)なので、有意な差が見られる
```

対応無しt検定： 水俣病罹患ネコと健康なネコの間で、肝臓中の総水銀量に差はありそうか?

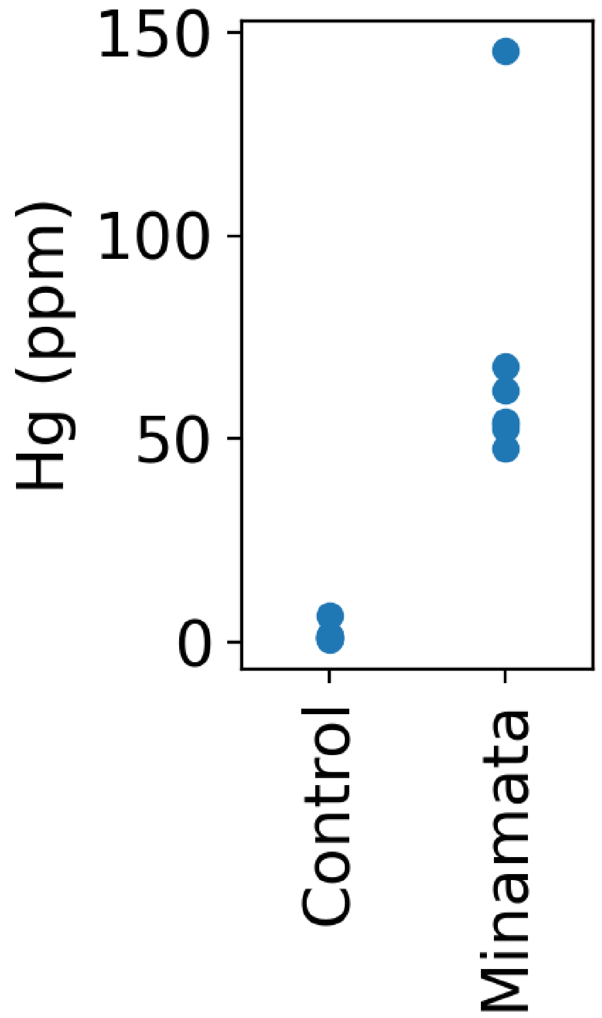
（喜田村, 1966; 土井&清水, *科学*, **43**, 436-442, 1973; 石居進, “生物統計学入門”, 1975）




```
# 水俣病、及び、不知火海沿岸の健康なネコの肝臓中の総水銀量（ppm）
Hg_M = [54.5, 68.0, 145.5, 53.5, 62.0, 47.6, 52.5]
Hg_c = [31.8, 14.5, 23.5, 33.3, 33.4, 9.0, 20.2, 75.2, 12.3, 29.7, 5.4]

# t検定（母平均に差無しを仮定）
from scipy.stats import ttest_ind
ttest_ind(Hg_c, Hg_M).pvalue

# 結論（p値と考察）
# p=0.0035(<0.05)なので有意な差が見られる

# 描画
import matplotlib.pyplot as plt
plt.figure(figsize=(2.5,4));  plt.rcParams['font.size']=16
plt.plot([1]*len(Hg_c)+[2]*len(Hg_M), Hg_c + Hg_M, 'o');
plt.xlim(0.5, 2.5);  plt.xticks([1, 2], ['Control', 'Minamata']);  plt.ylabel('Hg (ppm)')
plt.xticks(rotation=90); plt.tight_layout();  plt.show()
```

対応無しt検定2：水俣病罹患ネコと対照地区の健康なネコの間で、肝臓中の総水銀量に差はありそうか？

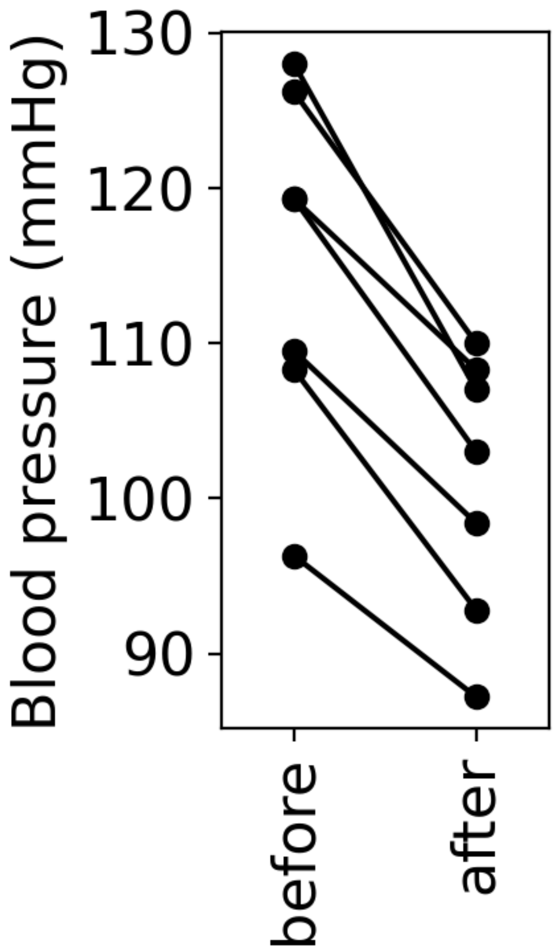
（喜田村, 1966; 土井&清水, *科学*, **43**, 436-442, 1973; 石居進, “生物統計学入門”, 1975）




```
# 水俣病、及び、不知火海沿岸以外の対照地区の健康ネコの肝臓中の総水銀量（ppm）
Hg_M = [54.5, 68.0, 145.5, 53.5, 62.0, 47.6, 52.5]
Hg_c = [1.18, 1.28, 1.56, 1.64, 0.99, 1.25, 0.64, 6.58]

# t検定（母平均に差無しを仮定）
from scipy.stats import ttest_ind
ttest_ind(Hg_c, Hg_M).pvalue

# 結論（p値と考察）
# p=9.3*10^-5 (<0.05)なので有意な差が見られる

# 描画
import matplotlib.pyplot as plt
plt.figure(figsize=(2.5,4));  plt.rcParams['font.size']=16
plt.plot([1]*len(Hg_c)+[2]*len(Hg_M), Hg_c + Hg_M, 'o');
plt.xlim(0.5, 2.5);  plt.xticks([1, 2], ['Control', 'Minamata']);  plt.ylabel('Hg (ppm)')
plt.xticks(rotation=90); plt.tight_layout();  plt.show()
```

対応有りt検定：レセルピンは血圧を下げるか?

（Cooper & Cranston, *Lancet*, 1, 396, 1957; ダン, “医歯系・生物系の統計学入門”, 1986）



\# レセルピン使用前後の最低血圧（mmHg）

```
before = [96.3, 119.3, 119.3, 108.3, 126.2, 128.0, 109.5]
after = [87.2, 103.0, 108.3, 92.8, 110.0, 107.0, 98.4]
```

\# t検定（母平均に差無しを仮定）

```
from scipy.stats import ttest_rel
ttest_rel(before, after).pvalue
```

\# 結論（p値と考察）

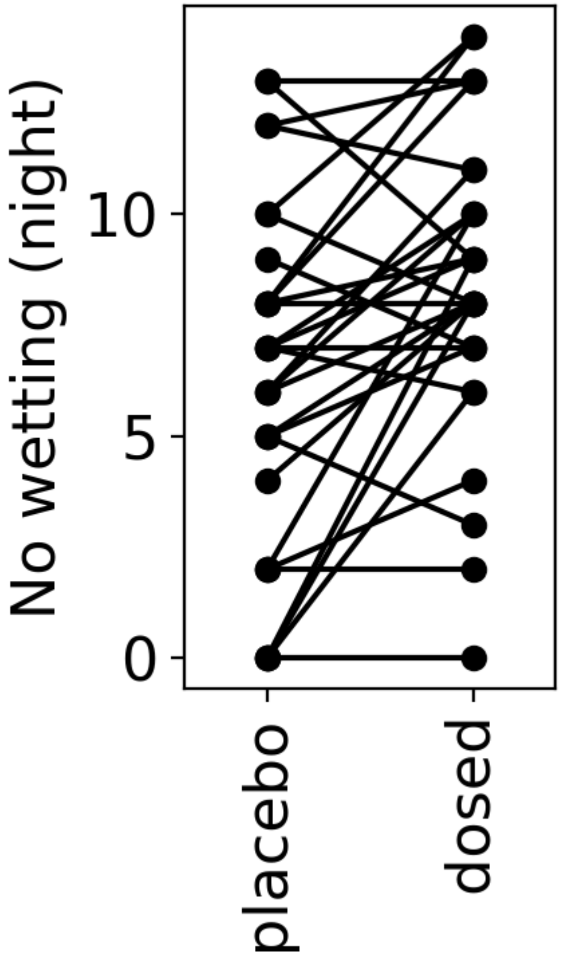
\# $p=9.4*10^{-5}$(<0.05)なので有意な低下が見られる

\# 描画

```
import matplotlib.pyplot as plt
plt.figure(figsize=(2.5,4));  plt.rcParams['font.size']=16
plt.plot([[1]*len(before), [2]*len(after)], [before, after], marker="o", color="k");
plt.xlim(0.6, 2.4); plt.ylabel('Blood pressure (mmHg)')
plt.xticks([1, 2], ['before', 'after']); plt.xticks(rotation=90); plt.tight_layout(); plt.show()
```

対応有りt検定2: 夜尿の新薬は夜尿回数を減らすか?

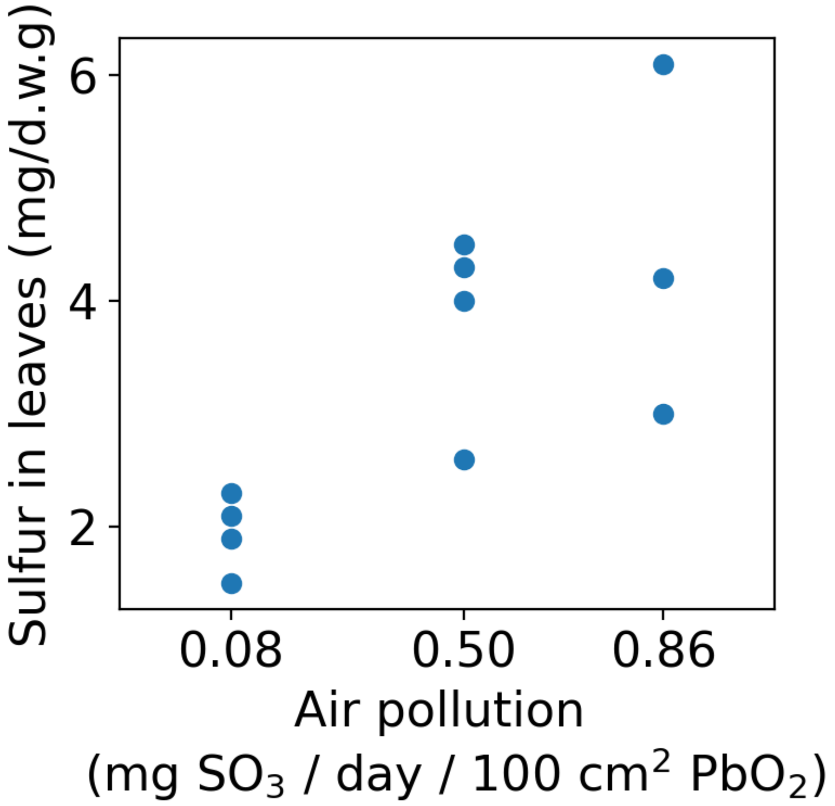
(Armitage & Berry, “医学研究のための統計的方法”, 2001)




```
# 新薬(プラセボ)14日内の夜尿の無かった日数（14日→休薬→14日とクロスオーバー）
placebo = [5, 10, 0, 7, 6, 5, 0, 0, 12, 2, 5, 13, 10, 7, 0, 6, 2] + [12, 6, 13, 8, 8, 4, 8, 2, 8, 9, 7, 7]
dosed = [8, 14, 8, 9, 11, 3, 6, 0, 13, 10, 7, 13, 8, 7, 9, 10, 2] + [11, 8, 9, 8, 9, 8, 14, 4, 13, 7, 10, 6]

# t検定（母平均に差無しを仮定）
from scipy.stats import ttest_rel
ttest_rel(placebo, dosed).pvalue

# 結論（p値と考察）
# p=0.0015(<0.05)なので有意な改善が見られる

# 描画
import matplotlib.pyplot as plt
plt.figure(figsize=(2.5,4));  plt.rcParams['font.size']=16
plt.plot([[1]*len(placebo), [2]*len(dosed)], [placebo, dosed], marker="o", color="k");
plt.xlim(0.6, 2.4); plt.ylabel('No wetting (night)'); plt.xticks(rotation=90);
plt.xticks([1, 2], ['placebo', ' dosed']); plt.tight_layout(); plt.show()
```

分散分析：大気汚染の被害を受けたケヤキの葉の中のイオウ含有量と、大気のSO_2汚染度とは、関係があるか?

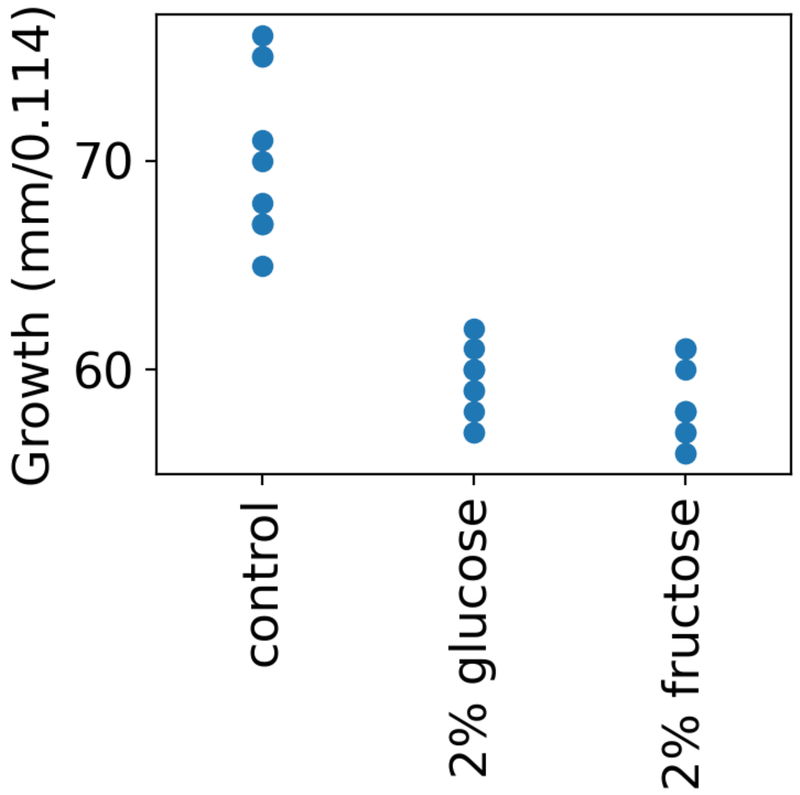
（埣田宏, *日本生態学会誌*, **23(2)**, 81–89, 1973; 石居進, “生物統計学入門”, 1975）




```
# 東京都内の1971年度大気汚染度別にした、ケヤキ葉中のイオウ含有量（mg/d.w.g）
S08 = [1.5, 1.9, 2.1, 2.3]
S50 = [2.6, 4.0, 4.3, 4.5]
S86 = [3.0, 4.2, 6.1]

# F検定（1元分散分析、母平均に差無しを仮定）
from scipy.stats import f_oneway
f_oneway(S08, S50, S86).pvalue

# 結論（p値と考察）
# p=0.02(<0.05)なので大気汚染度による有意な差が見られる

# 描画
import matplotlib.pyplot as plt
plt.figure(figsize=(4, 4));  plt.rcParams['font.size']=16;
plt.plot([0.08]*len(S08)+[0.5]*len(S50)+[0.86]*len(S86), S08 + S50 + S86, 'o');
plt.xticks([0.08, 0.50, 0.86], ['0.08', '0.50', '0.86']);  plt.xlim(-0.12, 1.06);
plt.xlabel('Air pollution \n (mg SO$_3$ / day / 100 cm$^2$ PbO$_2$)')
plt.ylabel('Sulfur in leaves (mg/d.w.g)');  plt.tight_layout();  plt.show()
```

分散分析2：エンドウ豆の成長に色々な糖は影響を及ぼすか?

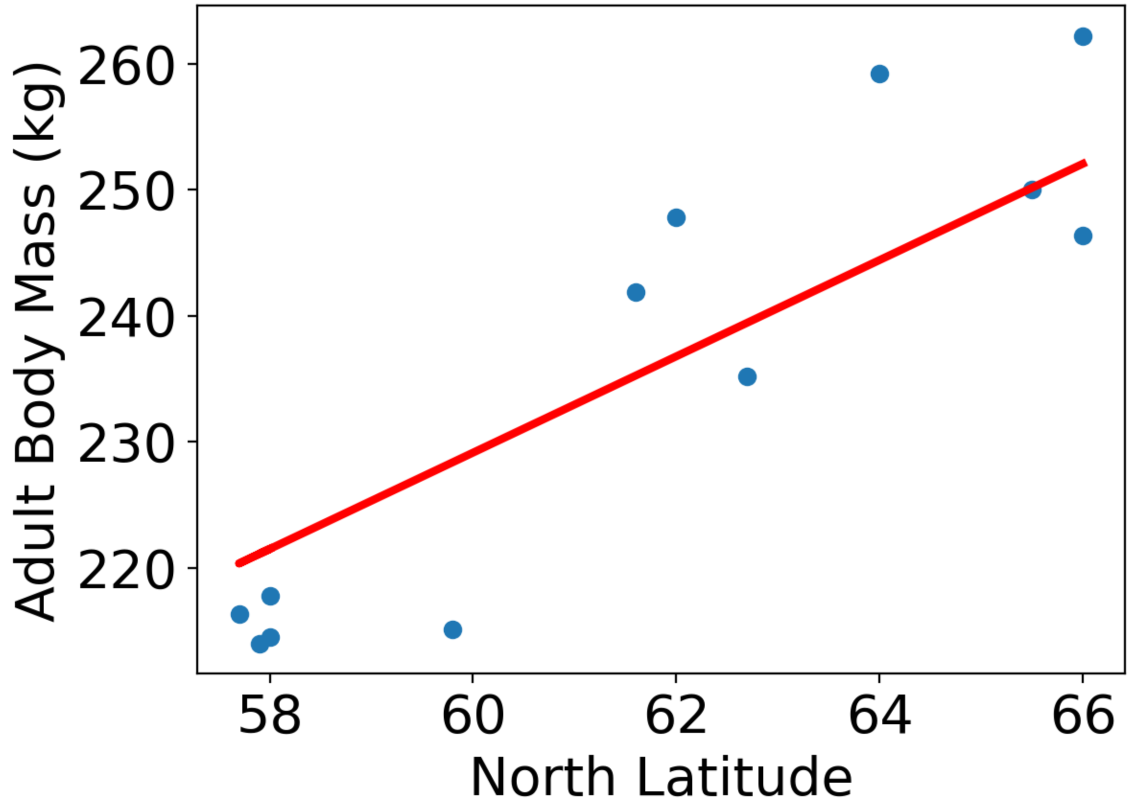
(Sokal & Rohlf, “生物統計学”, 1983)




```
# オーキシン存在下の組織培養におけるエンドウ豆の切片の成長の長さ（x0.114=mm）
L = [75, 67, 70, 75, 65, 71, 67, 67, 76, 68] # コントロール
L_glu = [57, 58, 60, 59, 62, 60, 60, 57, 59, 61] # 2%ブドウ糖追加
L_fru = [58, 61, 56, 58, 57, 56, 61, 60, 57, 58] # 2%果糖追加

# F検定（1元分散分析、母平均に差無しを仮定）
from scipy.stats import f_oneway
f_oneway(L, L_glu, L_fru).pvalue

# 結論（p値と考察）
# p=1.5*10^-10(<0.05)なので追加された糖による有意な差が見られる

# 描画
import matplotlib.pyplot as plt
plt.figure(figsize=(4, 4));  plt.rcParams['font.size']=16;
plt.plot([1]*len(L)+[2]*len(L_glu)+[3]*len(L_fru), L + L_glu + L_fru, 'o');
plt.xticks([1, 2, 3], ['control', '2% glucose', '2% fructose']);  plt.xticks(rotation=90);
plt.xlim(0.5, 3.5);  plt.ylabel('Growth (mm/0.114)');
plt.tight_layout();  plt.show()
```

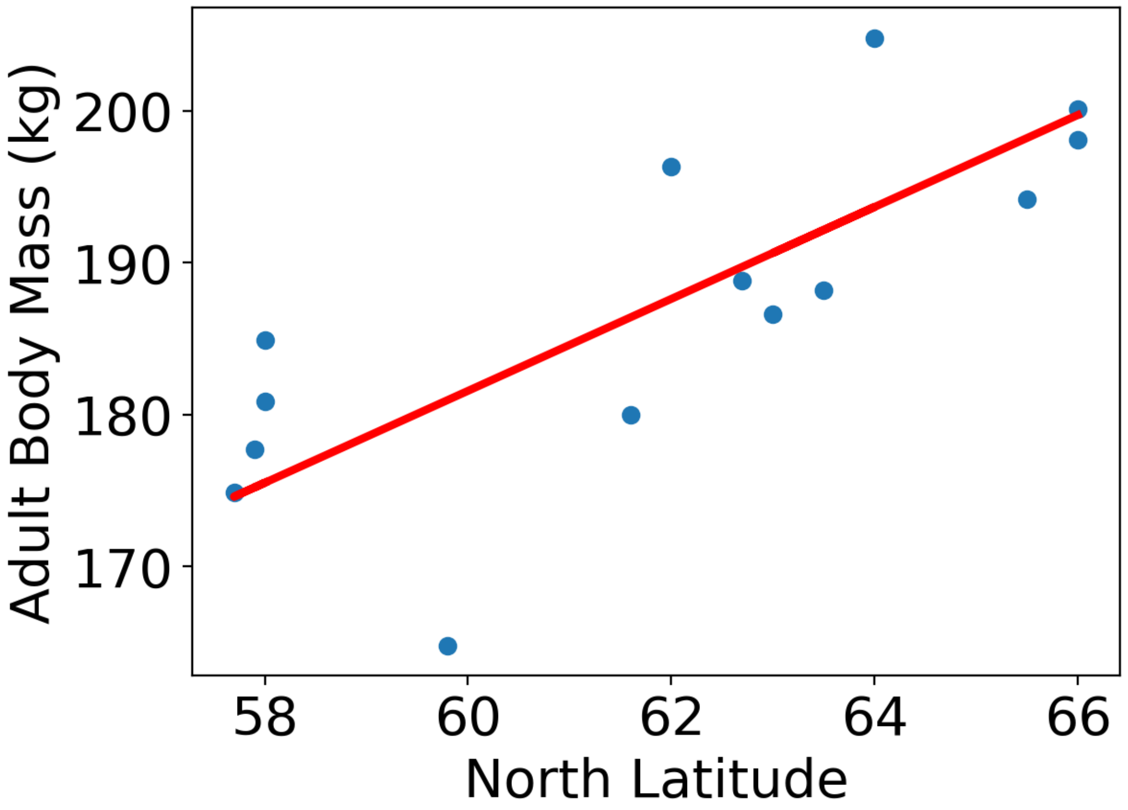
回帰分析：緯度が高いほどスウェーデンの成体の雄ヘラジカは大きくなるか?(ベルクマンの法則は正しいか?)

(Sand, Cederlund, Danell, 1995; 尾畑・荒木, “Pythonで学ぶ確率統計”, 2023)




```
# 緯度、及び、スウェーデンの成体の雄ヘラジカの体重(kg)のデータ
latitude = [58, 57.7, 58, 57.9, 59.8, 61.6, 62, 62.7, 64, 65.5, 66, 66]
BM_male = [214.5, 216.3, 217.8, 214.0, 215.1, 241.9, 247.8,
235.2, 259.2, 250.0, 246.4, 262.2]

# t検定（傾き0を仮説とする）
import statsmodels.api as sm
model = sm.OLS(BM_male, latitude).fit()
model.pvalues[0]

# 結論（p値と考察）
# p=2.7*10-17(<0.05)なので傾きは有意に0から異なる

# 描画
import matplotlib.pyplot as plt
plt.rcParams['font.size']=20
plt.plot(latitude, BM_male, 'o');
plt.plot(latitude, model.predict(latitude), color='red', lw=3);
plt.xlabel('North Latitude'); plt.ylabel('Adult Body Mass (kg)')
plt.tight_layout();  plt.show()
```

回帰分析2：緯度が高いほどスウェーデンの成体の雌ヘラジカは大きくなるか?（ベルクマンの法則は正しいか?）




```
# 緯度、及び、スウェーデンの成体の雌ヘラジカの体重(kg)のデータ
latitude = [58, 57.7, 58, 57.9, 59.8, 61.6, 62, 62.7, 64, 63.0, 63.5, 65.5, 66, 66]
BM_female = [180.9, 174.9, 184.9, 177.7, 164.8, 180.0, 196.3, 188.8, 204.8, 186.6, 188.2, 194.2, 198.1, 200.1]

# t検定（傾き0を仮説とする）
import statsmodels.api as sm
model = sm.OLS(BM_female, latitude).fit()
model.pvalues[0]
# 結論（p値と考察）
# p=5.4*10^-20(<0.05)なので傾きは有意に0から異なる

# 描画
import matplotlib.pyplot as plt
plt.rcParams['font.size']=20
plt.plot(latitude, BM_female, 'o');
plt.plot(latitude, model.predict(latitude), color='red', lw=3);
plt.xlabel('North Latitude'); plt.ylabel('Adult Body Mass (kg)')
plt.tight_layout();  plt.show()
```