線形回帰（LINEAR REGRESSION）

Advertising data

- 宣伝費と売上に関係はあるか？
- 関係がある場合その強さは？
- どのメディアが売上に最も貢献するか？
- 各メディアの宣伝費を増やしたときの売上の増加をどれだけ正確に予測できるか？
- 将来の売上をどれだけ正確に予測できるか？
- 宣伝費と売上の関係は線形か？
- メディア間にシナジー効果はあるか？

線形単回帰（simple linear regression）

- ひとつの予測変数 X によって量的応答変数 Y を予測
- 仮定: X と Y の関係は線形

$$Y \approx \beta_0 + \beta_1 X$$

例）sales $\approx \beta_0 + \beta_1 \times$ TV

パラメータ:　切片（intercept）と傾き（slope）

- 学習データによってパラメータを推定し、出力を予測

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

パラメータ推定

- 学習データ

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

- （回帰）残差（residual）

$$e_i = y_i - \hat{y}_i$$ ← 予測のずれ

- 残差平方和（residual sum of squares, RSS）

$$\begin{aligned}\mathrm{RSS} &= e_1^2 + e_2^2 + \cdots + e_n^2 \\ &= \left(y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1\right)^2 + \left(y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2\right)^2 + \cdots + \left(y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n\right)^2\end{aligned}$$

パラメータ推定

- 残差平方和が最小になるようにパラメータを決める

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_0} = 0 \qquad \frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_1} = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

x の標本平均　　y の標本平均

参考

$\hat{\beta}_0, \hat{\beta}_1$の導出

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

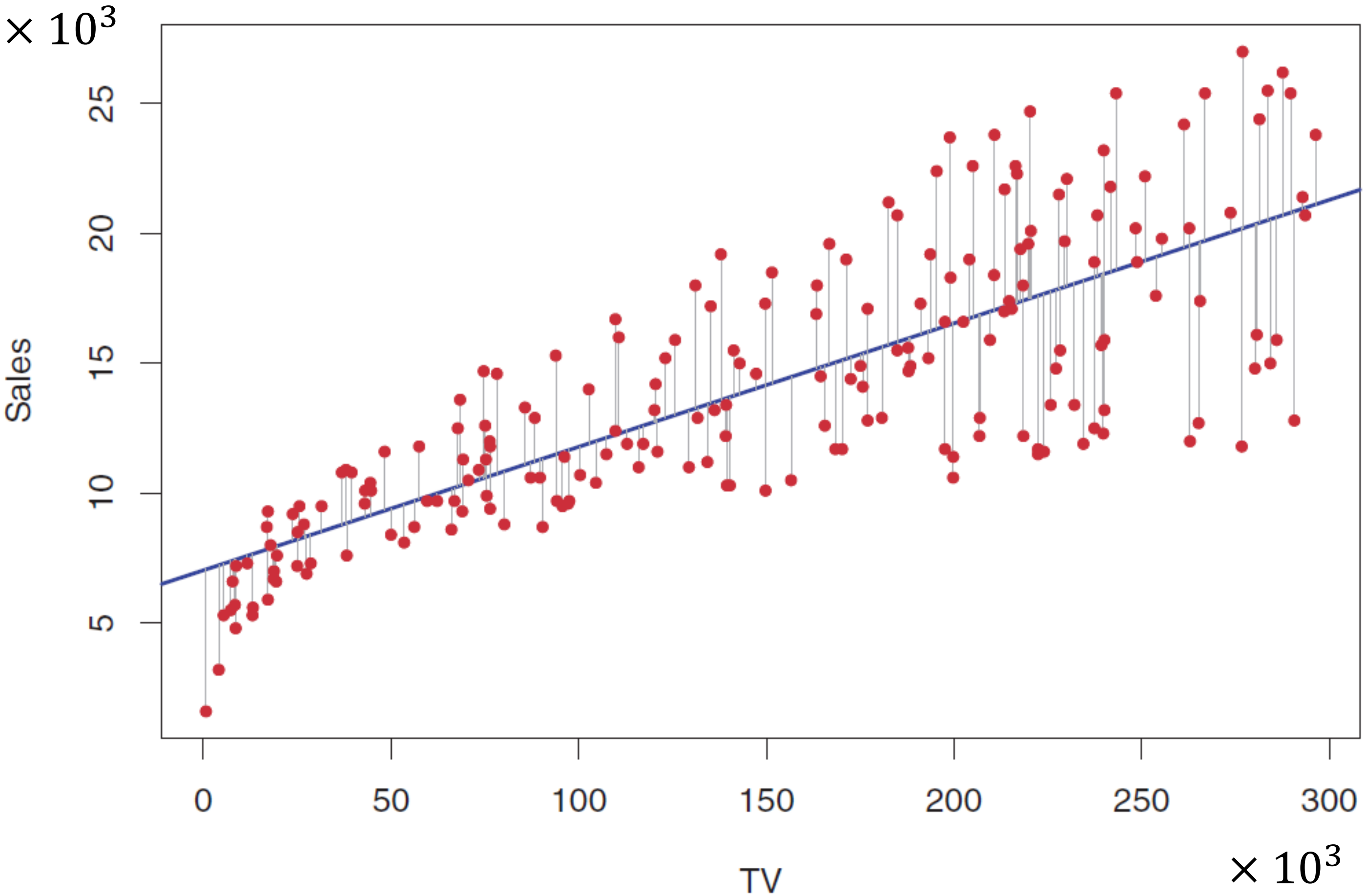
$$\begin{cases} \dfrac{\partial \mathrm{RSS}}{\partial \hat{\beta}_0} = -2 \displaystyle\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \\ \dfrac{\partial \mathrm{RSS}}{\partial \hat{\beta}_1} = -2 \displaystyle\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i = 0 \end{cases}$$

正規方程式

$$\begin{cases} n\hat{\beta}_0 + \left(\displaystyle\sum_{i=1}^{n} x_i\right)\hat{\beta}_1 = \displaystyle\sum_{i=1}^{n} y_i \\ \left(\displaystyle\sum_{i=1}^{n} x_i\right)\hat{\beta}_0 + \left(\displaystyle\sum_{i=1}^{n} {x_i}^2\right)\hat{\beta}_1 = \displaystyle\sum_{i=1}^{n} x_i y_i \end{cases}$$

例

- Advertising data



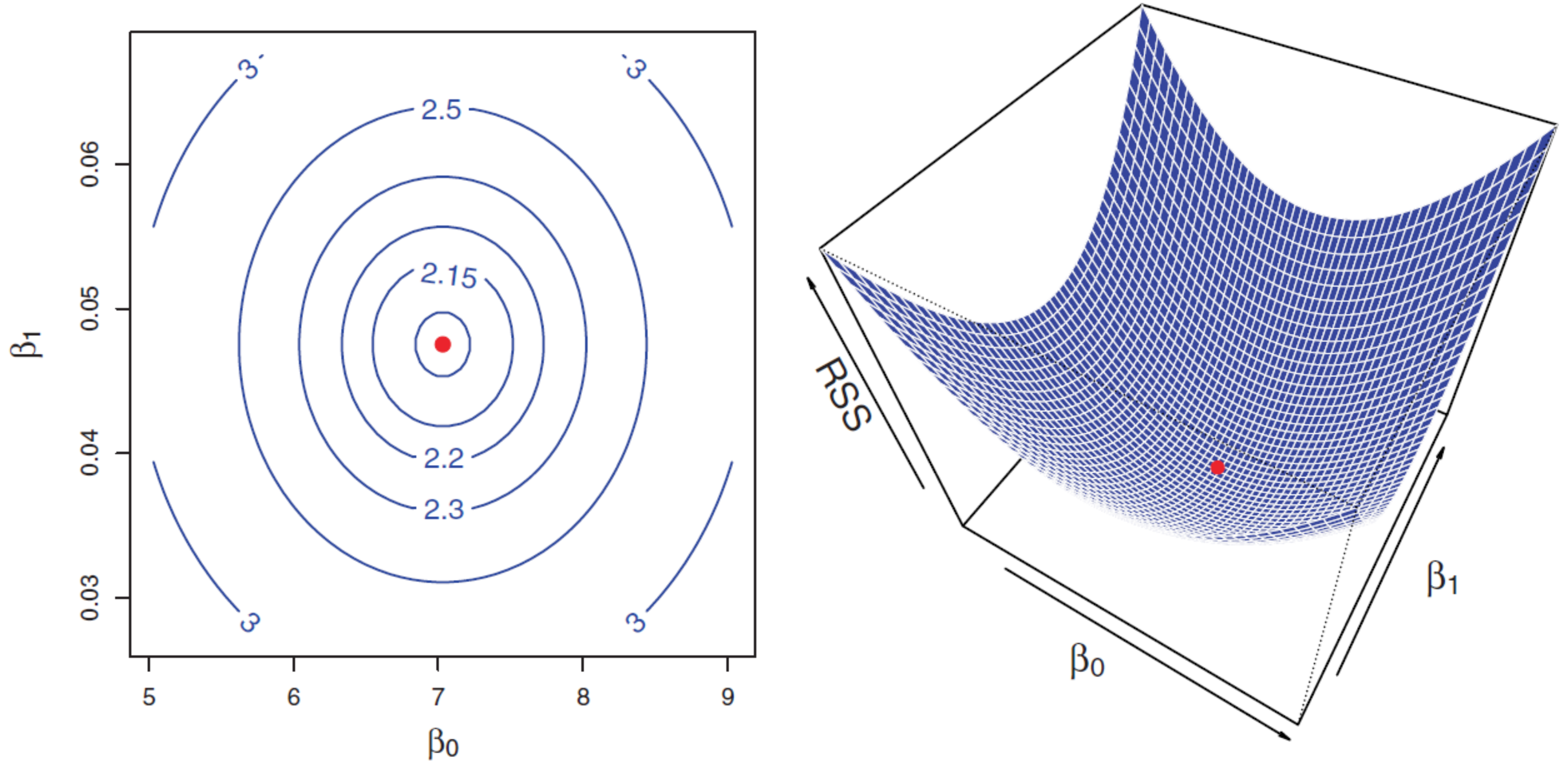
$$\hat{\beta}_0 = 7.03$$

$$\hat{\beta}_1 = 0.0475$$

宣伝費を$1000増やすと
売上が47.5増える

目的関数の形

- パラメータの値とRSS



下に凸な関数

残差

- 残差に関する性質

$$\mathrm{RSS} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} e_i = 0$$

残差の合計はゼロ

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) x_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} e_i x_i = 0$$

残差と **x** は直交

母回帰直線

- 母集団
  - f を線形な関数と仮定した場合 X と Y の関係は

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

平均がゼロ、分散がσ^2の正規分布による誤差

  - 母回帰直線(population regression line)
    - 母集団での回帰直線

$$Y = \beta_0 + \beta_1 X$$

復習

正規分布

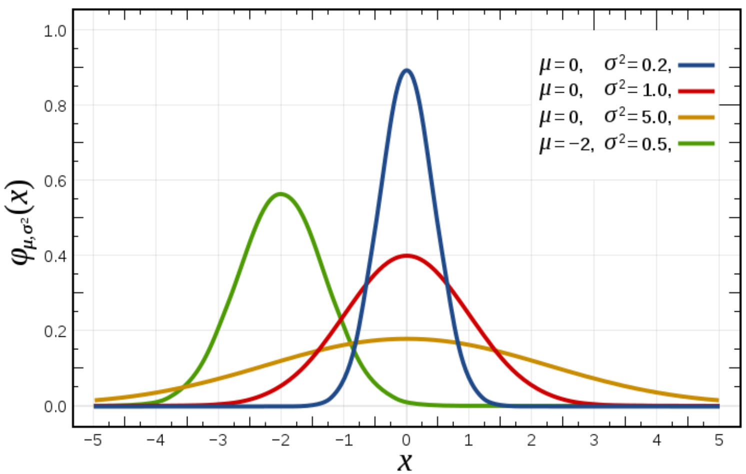
- 正規分布（normal distribution, Gaussian distribution）
  - 代表的な連続型の確率分布
  - 自然界の数多くの現象に対してあてはまる
  - 平均 μ 分散 σ^2 の正規分布を $N(\mu, \sigma^2)$ と書く
  - 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

  - 標準化変数

$$Z = (X-\mu)/\sigma$$

は標準正規分布$N(0,1)$に従う



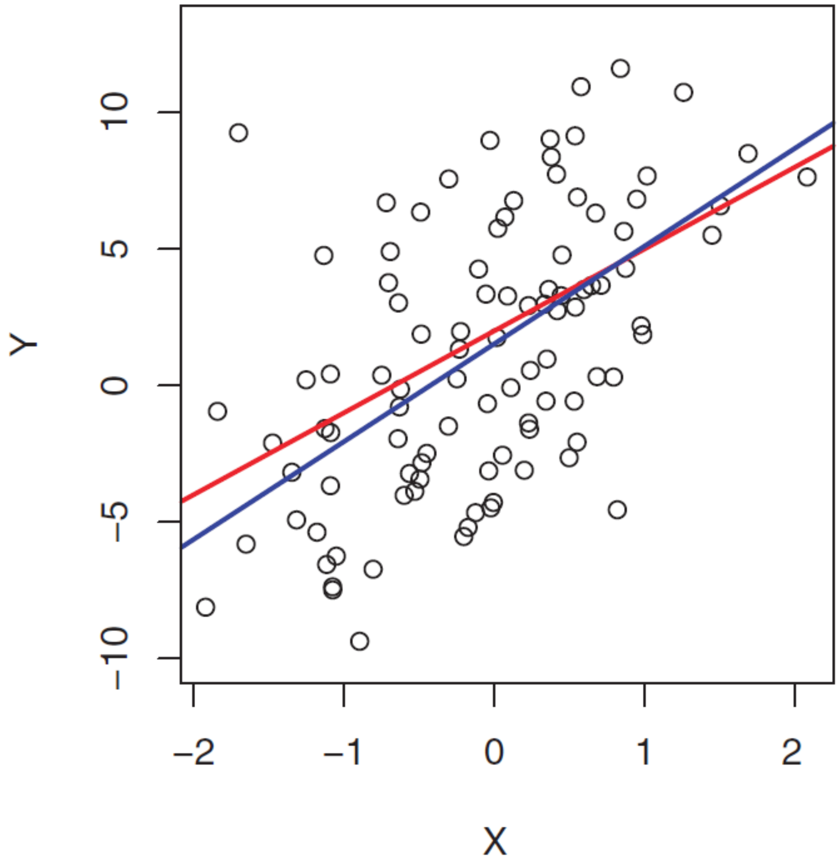
https://en.wikipedia.org/wiki/Normal_distribution

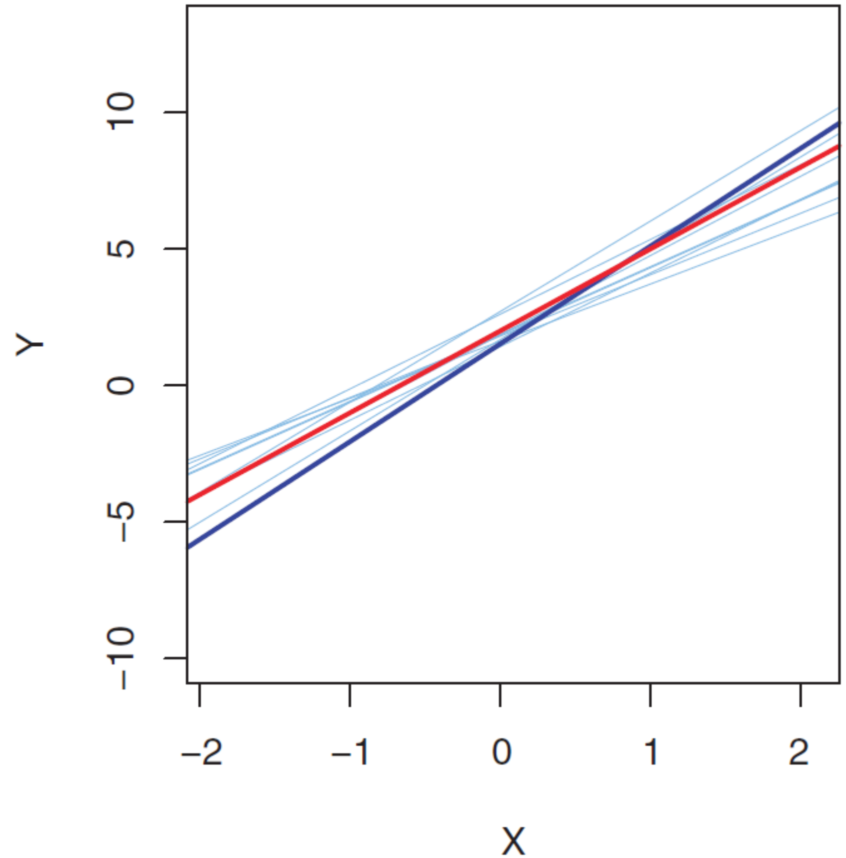
母回帰直線

- 人工データ: $Y = 2 + 3X + \varepsilon$

赤線:母回帰直線
青線:最小二乗による回帰線

薄い青線:異なる観測データによる最小二乗回帰線





傾きの推定量 $\hat{\beta}_1$ の諸性質

- 平均

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

を代入して整理すると

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$E(\varepsilon_i) = 0$ より

$$E\left[\hat{\beta}_1\right] = \beta_1$$

⇒ $\hat{\beta}_1$ は β_1 の不偏推定量

傾きの推定量 $\hat{\beta}_1$ の諸性質

- 分散

$$\mathrm{Var}\left(\hat{\beta}_1\right) = E\left[\left(\hat{\beta}_1 - \beta_1\right)^2\right]$$

$$= E\left[\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^2\right]$$

誤差項はすべて独立で分散は一定

$$E\left[\varepsilon_i \varepsilon_j\right] = 0 \qquad E\left[\varepsilon_i^2\right] = \sigma^2$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

← データがx方向に広がっていると分散が小さくなる

⇒ $\hat{\beta}_1$ は平均が β_1 、分散が $\dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ の正規分布

※独立な正規確率変数の和は正規確率変数

傾きの推定量 $\hat{\beta}_1$ の諸性質

- 誤差の分散 σ^2 がわかっているのであれば $\hat{\beta}_1$ の標準誤差（standard error）は以下のように得られる

$$\mathrm{SE}\left(\hat{\beta}_1\right) = \sqrt{\mathrm{Var}\left(\hat{\beta}_1\right)} = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}}$$

- しかし σ^2 は実際には未知なので、回帰残差から推定

$$\sigma^2 \approx s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 \quad \Rightarrow \quad \mathrm{SE}\left(\hat{\beta}_1\right) = \frac{s}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}}$$

※ $n-2$ で割るのは誤差項に関する前述の2つの制約条件により自由度が2つ失われているため

傾き $\hat{\beta}_1$ に関する推論

- 信頼区間（confidence interval）
  - 95%の確率でパラメータの真の値が含まれる区間
  - 近似的には

$$\hat{\beta}_1 \pm 2 \cdot \mathrm{SE}\left(\hat{\beta}_1\right)$$

- 例）Advertising data
  - $\hat{\beta}_1 = 0.0475$
  - 95%信頼区間は [0.042, 0.053]

傾き $\hat{\beta}_1$ に関する推論

- 検定
  - 帰無仮説: $\beta_1 = 0$ （ X で Y を説明することができない）
  - 対立仮説: $\beta_1 \neq 0$
  - 帰無仮説のもとで t 統計量（t-static）を計算

$$t_1 = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}\left(\hat{\beta}_1\right)}$$

← $\hat{\beta}_1$ が 0 から標準誤差何個分離れているか？

  - t 分布表から有意水準 α に応じたパーセント点 $t_{\alpha/2}$ を求め

$$\left|t_1\right| \geq t_{\alpha/2}$$

0.05 とか 0.01 とか

であれば帰無仮説を棄却（両側検定）

傾き $\hat{\beta}_1$ に関する推論

- t 統計量
  - $\hat{\beta}_1$を標準誤差で標準化した

$$t_1 = \frac{\hat{\beta}_1 - \beta_1}{\mathrm{SE}\left(\hat{\beta}_1\right)}$$

は、自由度 $n-2$ の t分布に従う

参考

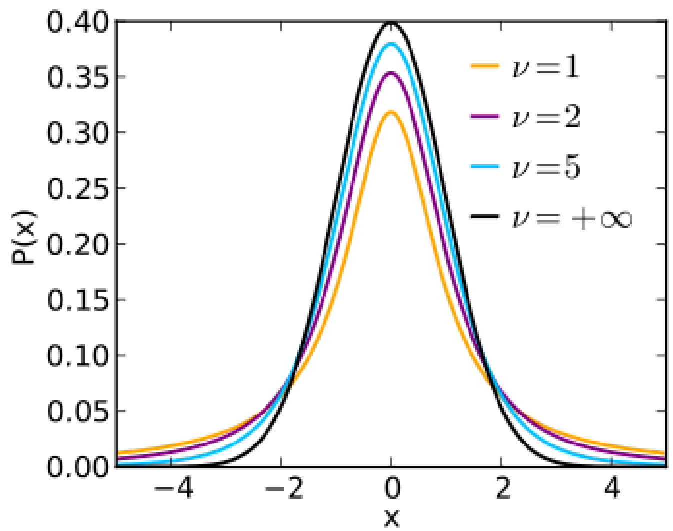
t分布（Student's t-distribution）

- 二つの確率変数 Y と Z が次の条件を満たす
  - Z は標準正規確率分布 $N(0,1)$ に従う
  - Y は自由度 ν の χ^2分布 $\chi^2(k)$ に従う
  - Y と Z は独立である

- このとき確率変数

$$t = \frac{Z}{\sqrt{Y/\nu}}$$

が従う確率分布を自由度 νの t分布という



自由度が30以上であれば正規分布とほぼ同じ

参考

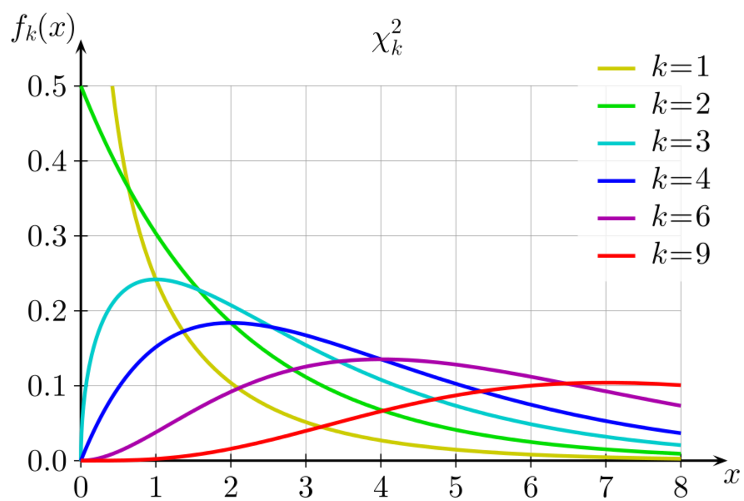
χ^2（カイ二乗）分布

- $Z_1, Z_2, \ldots, Z_k$を独立な、標準正規分布$N(0,1)$に従う確率変数とすると、確率変数

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

が従う確率分布を自由度 k の χ^2分布という
- 自由度 k の χ^2分布を $\chi^2(k)$ と書く

- 性質
  - $\mathrm{E}[\chi^2(k)] = k$
  - $\mathrm{Var}\left(\chi^2(k)\right) = 2k$



参考

t分布になる理由

$$t_1 = \frac{\hat{\beta}_1 - \beta_1}{\mathrm{SE}(\hat{\beta}_1)}$$

$$= \frac{\hat{\beta}_1 - \beta_1}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$= \frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \Bigg/ \sqrt{\frac{(n-2)s^2}{\sigma^2}/(n-2)}$$

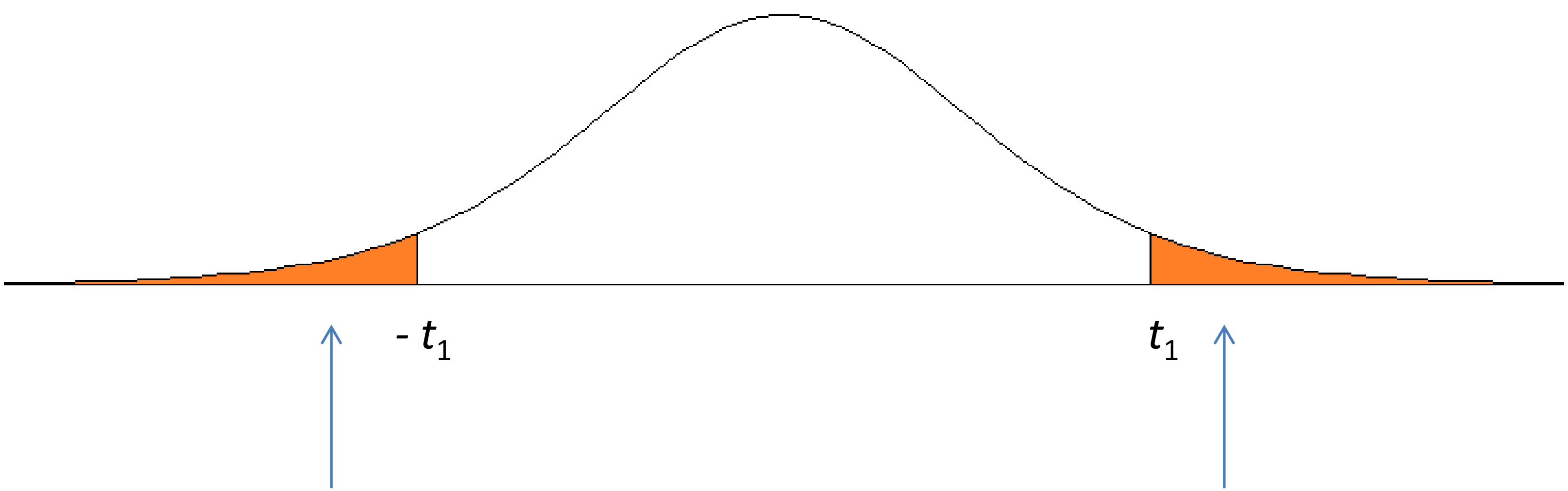
$$= \frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \Bigg/ \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{\sigma^2}/(n-2)}$$

標準正規分布 N(0,1)に従う

自由度 n-2 のχ^2分布に従う
（証明は「自然科学の統計学」第2章など）

傾き $\hat{\beta}_1$ に関する推論

- p値（p-value）
  - 帰無仮説が正しいとした場合に、t_1 よりも極端な t値が観測される確率
  - p値が**小さい**ほど統計的有意性が高い



（両側検定の場合）２つの領域を合わせた確率

パラメータに関する推論

- 切片 β_0 に関する推論
  - β_1 と似たような議論により、信頼区間、p値などが計算できる

- Advertising data での β_0, β_1 に関する推論

	Coefficient	Std. error	t-static	p-value
Intercept (β_0)	7.0325	0.4578	15.36	< 0.0001
TV (β_1)	0.0475	0.0027	17.67	< 0.0001

$\beta_1 = 0$ だとしたら $|\hat{\beta}_1| \geq 0.0475$ のような結果が得られる確率は 0.0001 より小さい
→ $\beta_1 = 0$ という仮説を棄却

モデルの評価

- モデルがどの程度データにフィットしているのかを定量評価したい
  - RSE
  - 決定係数R^2

- RSE（residual standard error）
  - 誤差 ε の標準偏差の推定量

$$\mathrm{RSE} = \sqrt{\frac{1}{n-2}\mathrm{RSS}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

例） 前出の Advertising data の例では RSE は 3.26

モデルの評価

- 決定係数R^2
  - 応答変数 Y の変動を X による回帰式で減らせた割合

回帰をしても残っている変動の和

$$R^2 = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

もともとの変動の総和

Total sum of squares: $\mathrm{TSS} = \sum_{y=1}^{n} \left(y_i - \bar{y} \right)^2$

例）前出の Advertising data の例では R^2 は 0.61

Python実習

- モジュールのインストール
  - Windows

```
> py -m pip install statsmodels
```

  - Linux

```
> python3 -m pip install statsmodels
```

- モジュールの読み込み

```
>>> import numpy as np
>>> import matplotlib.pyplot as plt
>>> import pandas as pd
>>> import math
>>> import statsmodels.api as sm
>>> import statsmodels.formula.api as smf
>>> from statsmodels.graphics.regressionplots import *
>>> from sklearn import datasets, linear_model
```

Python実習

- Bostonデータセットの読み込み

```
>>> from sklearn.datasets import load_boston
>>> b = load_boston()
>>> print(b.DESCR)
>>> Boston = pd.DataFrame(b.data, columns=b.feature_names)
>>> list(Boston)
['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS',
'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT']
>>> Boston.head()
...
>>> Boston.shape
(506, 13)
```

Python実習

- Boston データセット
  - ボストン郊外の住宅価格データ

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax
1	0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296
2	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242
3	0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242
4	0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222
5	0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222
6	0.02985	0	2.18	0	0.458	6.43	58.7	6.0622	3	222
7	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311
8	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311
9	0.21124	12.5	7.87	0	0.524	5.631	100	6.0821	5	311

  - CRIM: per capita crime rate by town
  - :
  - LSTAT: lower status of population (percent)
  - MEDV: median value of owner-occupied homes in $1000s

Python実習

- 線形単回帰
  - ols()
    - ols(y ~ x, data)で x を予測変数、 y を応答変数として線形モデル(linear model)をフィッティング(パラメータ推定)

```
>>> Boston['MEDV'] = b.target
>>> lm = smf.ols('MEDV ~ LSTAT', data=Boston).fit()
```

Python実習

- フィットさせたモデルの情報
  - summary()

```
>>> lm.summary()
 OLS Regression Results
==============================================================================
Dep. Variable:                   MEDV   R-squared:                       0.544
Model:                            OLS   Adj. R-squared:                  0.543
Method:                 Least Squares   F-statistic:                     601.6
Date:                Tue, 07 May 2019   Prob (F-statistic):           5.08e-88
Time:                        23:28:16   Log-Likelihood:                -1641.5
No. Observations:                 506   AIC:                             3287.
Df Residuals:                     504   BIC:                             3295.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     34.5538      0.563     61.415      0.000      33.448      35.659
LSTAT         -0.9500      0.039    -24.528      0.000      -1.026      -0.874
==============================================================================
Omnibus:                      137.043   Durbin-Watson:                   0.892
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              291.373
Skew:                           1.453   Prob(JB):                     5.36e-64
Kurtosis:                       5.319   Cond. No.                         29.7
==============================================================================
```

Python実習

- フィットさせたモデルの情報
  - パラメータの値

```
>>> lm.params
Intercept    34.553841
LSTAT        -0.950049
dtype: float64
```

  - 信頼区間

```
>>> lm.conf_int()
                   0          1
Intercept  33.448457  35.659225
LSTAT      -1.026148  -0.873951
```

Python実習

- フィットさせたモデルを用いて予測
  - predict()

```
>>> lm.predict(pd.DataFrame({'LSTAT':[5, 10, 15]}))
0    29.803594
1    25.053347
2    20.303101
dtype: float64
```

Python実習

- プロット

```
>>> Boston.plot(kind='scatter', x='LSTAT', y='MEDV')
>>> plt.show()
>>> range = pd.DataFrame({'LSTAT': [Boston.LSTAT.min(),
Boston.LSTAT.max()]})
>>> preds = lm.predict(range)
>>> plt.plot(range, preds, c='red', linewidth=2)
>>> Boston.plot(kind='scatter', x='LSTAT', y='MEDV')
>>> plt.show()
```

- 残差プロット

```
>>> fig, ((ax1, ax2), (ax3, ax4)) = plt.subplots(2, 2)
>>> ax1.plot(Boston.LSTAT, lm.predict(),'ro')
>>> ax2.plot(lm.predict(), lm.resid, 'go')
>>> ax3.plot(lm.predict(), lm.resid_pearson, 'bo')
>>> plt.show()
```

Python実習

- scikit-learn による方法

```
>>> x = pd.DataFrame(Boston.LSTAT)
>>> y = Boston.MEDV
>>> print(x.shape)
(506, 1)
>>> model = linear_model.LinearRegression()
>>> model.fit(x, y)
...
>>> print(model.intercept_)
...
>>> print(model.coef_)
...
```

多重線形回帰

- Advertising data
  - 予測変数: TV, radio, newspaper
  - 応答変数: sales
- 複数の変数を用いて予測したい
  - 個別に線形単回帰を行うと

	Coefficient	Std. error	t-static	p-value
Intercept	9.312	0.563	16.54	< 0.0001
radio	0.203	0.020	9.92	< 0.0001

	Coefficient	Std. error	t-static	p-value
Intercept	12.351	0.621	19.88	< 0.0001
newspaper	0.055	0.017	3.30	< 0.0001

多重線形回帰

- 個別に線形単回帰を行うことの問題点
  - 予測の際に複数の単回帰モデルをどのように組み合わせるべきかが明らかでない
  - 予測変数どうしの関係を考慮できない

多重線形回帰

- 多重線形回帰（multiple linear regression、重回帰）
  - 複数の予測変数を使う単一のモデルで回帰

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

β_{j}: 他の予測変数を固定して、X_j を1単位増やしたときに増える Y の量

例）

$$\text{sales} \approx \beta_0 + \beta_1 \times \text{TV} + \beta_2 \times \text{radio} + \beta_3 \times \text{newspaper} + \varepsilon$$

パラメータ推定

- 予測

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

- 残差平方和

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$$

$$= \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip} \right)^2$$

パラメータ推定

- 勾配がゼロになるパラメータを直接求める

$$\left.\begin{aligned} \frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_0} &= 0 \\ \frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_1} &= 0 \\ &\vdots \\ \frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_p} &= 0 \end{aligned}\right\} \Rightarrow$$

行列演算で解析的に解ける

参考

β の求め方

- RSS を行列で表すと

$$\mathrm{RSS} = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

$$\frac{\partial \mathrm{RSS}}{\partial \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

より

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

・パラメータ（ベクトル）

$$\beta = \left(\beta_0, \beta_1, \dots, \beta_p\right)^T$$

・学習データ（入力）

$$\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$$

・学習データ（出力）

$$\mathbf{y} \in \mathbb{R}^n$$

パラメータ推定

- 勾配法による逐次計算でも求められる

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n} e_i$$

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} e_i x_{i1}$$

$$\vdots$$

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_p} = -2\sum_{i=1}^{n} e_i x_{ip}$$

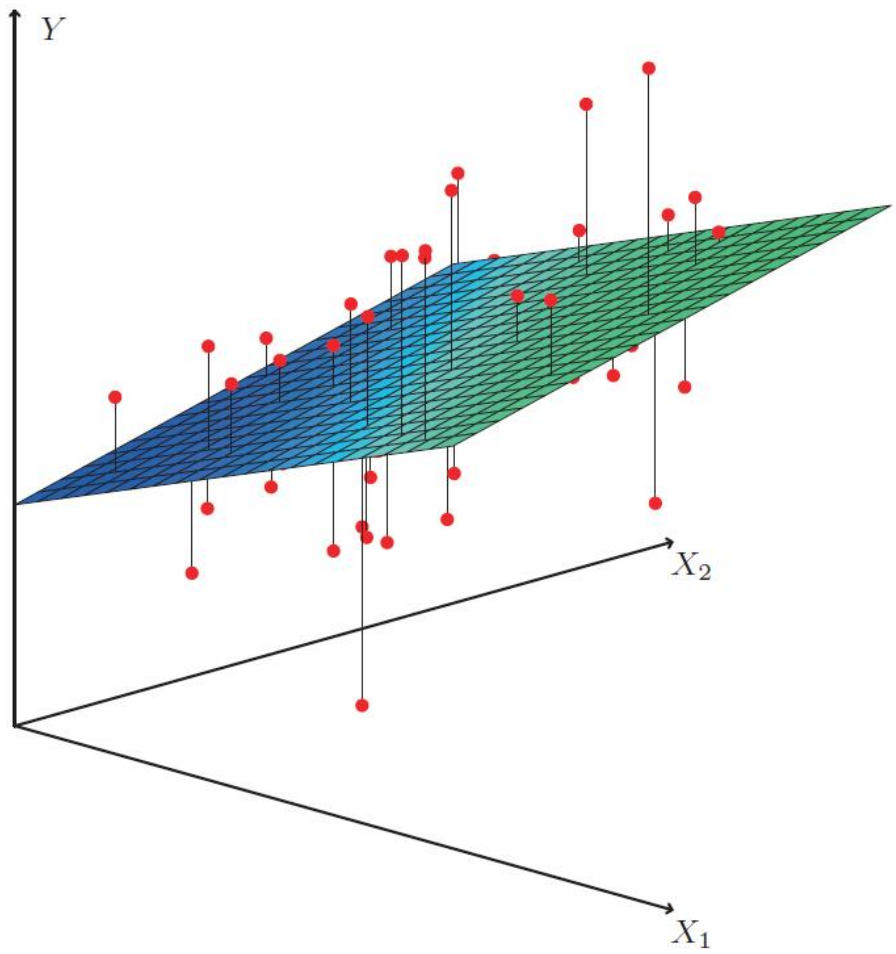
勾配情報を使って再急降下法、
準ニュートン法などを行う

例

- 予測変数が2個

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

- 回帰「直線」ではなく平面



単回帰と重回帰

- Advertising data

重回帰の結果

	Coefficient	Std. error	t-static	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	−0.001	0.0059	−0.18	0.8599

矛盾？？？
（newspaper の効果の有無が全く違う）

単回帰の結果

	Coefficient	Std. error	t-static	p-value
Intercept	12.351	0.621	19.88	< 0.0001
newspaper	0.055	0.017	3.30	< 0.0001

単回帰と重回帰

- Advertising data

相関行列

	TV	radio	newspaper	sales
TV	1.0000	0.0548	0.0567	0.7822
radio		1.0000	0.3541	0.5762
newspaper			1.0000	0.2283
sales				1.0000

newspaper と radio に相関あり

newspaper による宣伝費増加で売り上げが上がるように見えたのは、
実は radio の宣伝費が増加していたから

他の例） アイスクリームの売り上げとサメによる被害の数

推論

- 重回帰分析を行うときに知りたいこと
  - 応答変数 Y の値を予測するにあたり、予測変数 $X_1, X_2, \ldots, X_p$ のうち少なくとも一つが有用であるかどうか
  - すべての予測変数が Y を説明するのに有用なのか、あるいは一部だけが有用なのか
  - モデルはどれぐらいデータにフィットしているのか
  - 予測変数の値が与えられたとき、応答変数の値をどれだけ正確に予測できるのか

推論

- 応答変数と予測変数の間に関係があるかどうかの検定
  - 帰無仮説 H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
  - 対立仮説 H_a: 少なくとも一つの β_j が 0 ではない
  - 帰無仮説が正しい場合、以下のF統計量（f-static）はF分布（p, n-p-1）に従う

$$F = \frac{(\mathrm{TSS} - \mathrm{RSS})/p}{\mathrm{RSS}/(n-p-1)}$$

$$\mathrm{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$E[\mathrm{RSS}/(n-p-1)] = \sigma^2$$

H_0が正しい場合は $E[(\mathrm{TSS} - \mathrm{RSS})/p] = \sigma^2$

補足

E[TSS], E[RSS] について

- 帰無仮説が正しい場合

$$E\left[\frac{\mathrm{TSS}}{n}\right] = \frac{n-1}{n}\sigma^2$$

← 自由度が1あるぶん標本分散は真の分散より小さくなる

$$E[\mathrm{TSS}] = (n-1)\sigma^2$$

$$E\left[\frac{\mathrm{RSS}}{n}\right] = \frac{n-p-1}{n}\sigma^2$$

← パラメータが p+1 個あるので

$$E[\mathrm{RSS}] = (n-p-1)\sigma^2$$

参考

F分布

- 確率変数

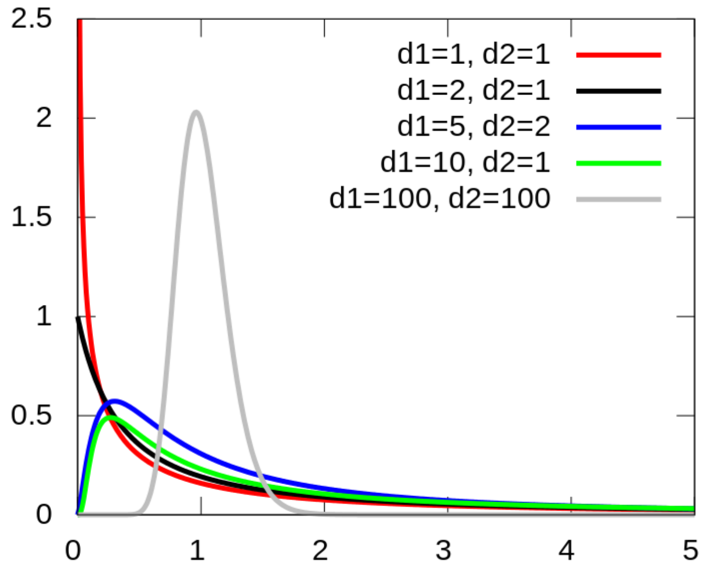
$$X = \frac{U_1/d_1}{U_2/d_2}$$

が従う確率分布

U_1: 自由度 d_1 のχ^2分布に従う確率変数

U_2: 自由度 d_2 のχ^2分布に従う確率変数

U_1とU_2は独立



推論

- Advertising data での例

Quantity	Value
Residual Standard Error (RSE)	1.69
R^2	0.897
F-static	570

  – F統計量が570
  – F分布(p, n-p-1) と F統計量からp値が計算できる
    - この場合p値はほとんどゼロ

→　帰無仮説は棄却される

推論

- 応答変数と予測変数（の特定の部分集合）の間に関係があるかどうかの検定
  - 帰無仮説 H_0: $\beta_{p-q+1} = \beta_{p-q+2} = \ldots = \beta_p = 0$
  - 対立仮説 H_a: 少なくとも一つの β_j $(j > p-q)$がゼロではない
  - 帰無仮説が正しい場合、以下のF統計量（f-static）はF分布（q, n-p-1）に従う

$$F = \frac{\left(\mathrm{RSS}_0 - \mathrm{RSS}\right)/q}{\mathrm{RSS}/\left(n - p - 1\right)}$$

RSS_0: 最後のq個の予測変数を使わないモデルで得られたRSS