

データ解析基礎- Basic Data Analysis

問題集(例) - Sample Examination Questions

These questions may or may not appear on any given examination.

Problems

When you are asked to do a calculation, you do not need to compute the decimal equivalent of a fraction, or radical (square root). Fractions should be reduced to lowest terms for convenience in grading. Radicals do not need to be reduced. That is $\frac{7}{5}$ and $1\frac{2}{5}$ are OK, please try to avoid $\frac{28}{20}$. $\sqrt{5}$ is OK.

計算を行うときには分数または根数のままを書いてもよい。少数にする必要はない。ただし、分数の分母と分子は互いに素にすること。例： $\frac{7}{5}$ も $1\frac{2}{5}$ もよいが、 $\frac{28}{20}$ を避けてもらいたい。 $\sqrt{5}$ はよい。

observed. It also can be used to group explanatory variables that have similar effect into **factors**, to infer the meaning of previously unobserved data values from context, and to accurately **classify** data according to imprecise criteria using easily measureable variables.

3. [Problem ID #IDNO] sampling: random

Is a random sample perfectly representative? Explain your answer.

ランダムサンプルは完全に代表的サンプルになるか。説明せよ。

A random sample (properly designed and implemented) is representative, but not perfectly so. Because it is random, the frequency of each value will vary if samples are taken repeatedly. Not all of these frequencies can perfectly match the underlying distribution, so the sample cannot be perfectly representative. On the other hand, a large random sample will tend to have a relative frequency distribution very close to that of the underlying distribution, including correlations with unobserved variables. This makes it suitable (representative) for statistical inference, because it eliminates the possibility of selection bias in the sample.

4. [Problem ID #IDNO] independence: matrix game

Ann and Bob repeatedly play a simple game where Ann chooses Up or Down, and at the same time, Bob chooses Left or Right. The relative frequencies of each combination of choices observed is given in Table 1.

アンちゃんとボブ君がゲームで遊ぶ。アンは上(Up) か下(Down) を選択し、ボブは左(Left) と右(Right) を同時に選択する。観察した各選択組の頻繁さを表 1 でそろえてある。

		Bob	
		Left	Right
Ann	Up	$\frac{1}{12}$	$\frac{1}{6}$
	Down	$\frac{1}{4}$	$\frac{1}{2}$

Table 1: Game of Ann and Bob

(a) What is the relative frequency of Ann playing Up?

アンが Up を選択する割合を計算せよ。

$$f(Up) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$$

(b) What is the relative frequency of Bob playing Left?

ボブが Left を選択する割合を計算せよ。

$$f(\text{Left}) = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$$

(c) Interpreting the relative frequencies as probabilities, are the Up and Left independent events?

割合を確立として解釈した場合には Up と Left は独立した事象になるか について答えて説明せよ。

$$\text{Yes: } f(\text{Up}) \times f(\text{Left}) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12} = f(\text{Up and Left})$$

5. [Problem ID #IDNO] hypothesis test: P-value and type II error

Explain the relationship between the P -value of an hypothesis test and the Type II error of the test.

仮説検定における P 値とタイプIIエラーの関係を説明せよ。

The P -value of a test is the probability of a Type II error, calculated by assuming the values of the parameters of the distribution of the test statistic are given by the null hypothesis H_0 .

6. [Problem ID #IDNO] hypothesis test: type I, II error

Define the Type I error and Type II error of an hypothesis test.

仮説検定におけるタイプIエラーとタイプIIエラーの定義を書け。

A Type I error occurs when the null hypothesis H_0 of the test is false, but the test accepts H_0 . A Type II error occurs when the null hypothesis H_0 of the test is true, but the test rejects H_0 .

7. [Problem ID #IDNO] hypothesis test: power computation

Why is computation of the *power* of an hypothesis test (probability of accepting a false null hypothesis) more difficult than computation of the test's *significance* (probability of rejecting a true null hypothesis)? なぜ仮説検定のパワー（仮説が不実の場合に認められる確立）の計算が有意性（仮説が事実の場合に認められない確立）の計算より難しいことを説明せよ。

The null hypothesis H_0 is defined by some scientific theory characterized by a given value for the parameter of the distribution of the test statistic. Computation of the significance is therefore straightforward: simply compute the probability using that value of the parameter. Computation of the power is usually no more difficult as a numerical calculation, but theory doesn't give a value for the parameter, making choice of that parameter a challenge to the judgment of the statistician.