

データ解析基礎レポート課題（第1回）

出題 10月3日（金），提出期限 10月14日（火）

提出先：情報システム学研究科棟2階事務室脇のポスト

1. 日本の女性の金融アナリストにおいて経験年数が年収に与える影響を調査している。このとき、以下の量がそれぞれ変数であるか、定数であるかを答えよ。
 - a. 性別
 - b. 年収
 - c. 国籍
 - d. 職業
 - e. 経験年数
2. 次の量は、それぞれ、名義尺度、順序尺度、間隔尺度、比尺度のいずれであるか？ なぜ、そう考えるのか、その理由を付して答えること。
 - a. 花の品種
 - b. TOEFL の点数
 - c. 馬の走る速さを表す量としてのレース順位
 - d. 摂氏温度
 - e. 画面の輝度（明るさ）
3. ホームページ上にあるデータファイル data01.txt は講義の例題に用いたものである。このファイルをダウンロードして、試験点数(point)に関して以下の分析を行なえ（学科など他の項目は無視してよい）。
 - a. 階級幅 50 点として度数分布、および相対度数分布を求めよ。階級を決めるときは、各階級の下限がきりのよい数字になるようにせよ。
 - b. a. で求めた度数分布をもとに、ヒストグラムと累積度数のグラフを作成せよ。
 - c. a. で求めた度数分布をもとにして、ID=1 の学生の percentile rank を求めよ。
 - d. 同様に、50% の percentile point を求めよ。
 - e. この分布は正の歪度をもつか、負の歪度をもつか。
 - f. この分布は単峰型であるか。
 - g. もとのデータを用いて、平均値、分散、標準偏差を求めよ。
 - h. 度数分布に基づいて、平均値、分散、標準偏差を求めよ（階級値としては小数点以下の数字がない値を用いてよい）。
4. 第 1 回の講義の内容で理解できなかったこと、また、改善してほしいことなどについて、自由に記述せよ（なければ、特に記載しなくてよい）。