# １．データの集計【第1回】

## 1.1　質的データの集計

基礎

単純集計　　１次元分割表　→　棒グラフ（値重視），円グラフ（割合重視）

クロス集計　２次元分割表　→　積み重ね棒グラフ

例

　２０人に以下のようなアンケートを取った。入力フォームをExcelで作成せよ。

質問１　あなたの性別は。

　　１．男性　　２．女性

質問２　あなたは学校改革案に賛成ですか。

　　１．はい　　２．いいえ　　３．どちらともいえない

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 性別 | 回答 | 性別 | 回答 | 性別 | 回答 | 性別 | 回答 |
| １ | １ | ２ | ３ | １ | ２ | ２ | １ |
| ２ | １ | １ | １ | １ | ２ | １ | ２ |
| １ | ２ | ２ | ２ | ２ | １ | １ | １ |
| １ | ２ | ２ | ３ | ２ | １ | ２ | ３ |
| ２ | １ | １ | １ | １ | ３ | １ | １ |

入力されたデータをCollege Analysisに移し、以下の問いに答えよ。

１）回答に関する１次元分割表を描け。

２）１）の分割表を用いて棒グラフと円グラフを描け。

３）性別と回答に関する２次元分割表を描け。

４）３）の分割表を用いて積み重ね棒グラフを描け。

問題

　Samples\テキスト9.txtを用いて以下の問いに答え、結果は文書にまとめよ。但し、地域について1：市街、2：郊外、意見１について1：賛成、2：反対、意見２について1：はい、2：いいえ、3：どちらとも（いえない）とする。

１）地域に関する１次元分割表を描け。

|  |  |  |
| --- | --- | --- |
| 市街 | 郊外 | 合計 |
|  |  |  |

２）意見１に関する１次元分割表を描け。

|  |  |  |
| --- | --- | --- |
| 賛成 | 反対 | 合計 |
|  |  |  |

３）意見２に関する１次元分割表を描け。

|  |  |  |  |
| --- | --- | --- | --- |
| はい | いいえ | どちらとも | 合計 |
|  |  |  |  |

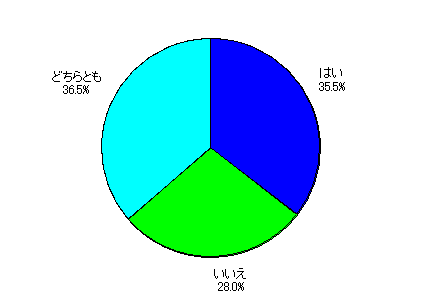
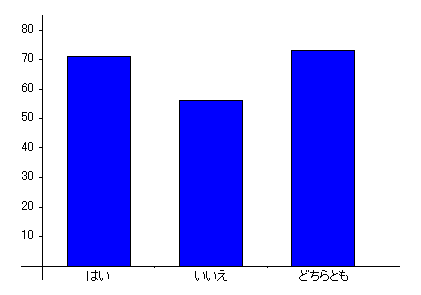
４）地域と意見１に関する２次元分割表を描け。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 賛成 | 反対 | 合計 |
| 市街 |  |  |  |
| 郊外 |  |  |  |
| 合計 |  |  |  |

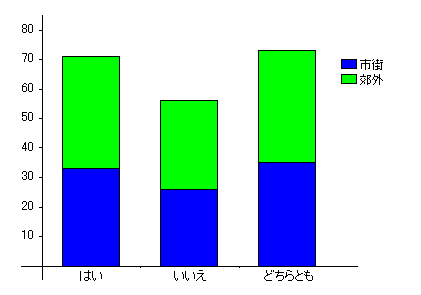
５）地域と意見２に関する２次元分割表を描け。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | はい | いいえ | どちらとも | 合計 |
| 市街 |  |  |  |  |
| 郊外 |  |  |  |  |
| 合計 |  |  |  |  |

６）意見２に関する棒グラフと円グラフを描け。



７）地域と意見２に関する積み重ね棒グラフを描け。



## 1.2　量的データの集計【第2回】

### 1.2.1　量的データの集計

度数分布表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 階級 | 度数 | 相対度数(%) | 累積度数 | 累積相対度数(%) |
| 50 <= x < 60 | 4 | 20 | 4 | 20 |
| 60 <= x < 70 | 8 | 40 | 12 | 60 |
| 70 <= x < 80 | 5 | 25 | 17 | 85 |
| 80 <= x < 90 | 3 | 15 | 20 | 100 |
| 計 | 20 | 100 |  |  |

　注）各階級の幅を階級幅、各階級の中央の値を階級値という。

ヒストグラム



　基本統計量（要約統計量）　【データ　3,3,4,2,8 】

分布の中心を表わす統計量（代表値）

注）基本統計量を代表値の意味で使う場合も多い

平均値（average, mean）



中央値（中間値，メジアンmedian）

データを小さい方から順番に並べて中間の値

2, 3, 3, 4, 8 →　3

2, 3, 3, 4, 6, 8 →　(3+4)/2=3.5

　　最頻値（モード mode）

　　　度数分布表やヒストグラムでまとめられている場合は、最大度数の階級値

分布の拡がりを表わす統計量（散布度）

レンジ（range）

*R*=最大値－最小値=6

分散（variance）



標準偏差（standard deviation）



不偏分散



標準偏差（standard deviation）



例

　Samples\テキスト1.txtのデータを用いて次の問いに答えよ。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学校 | 身長(cm) | 体重(kg) | 学校 | 身長(cm) | 体重(kg) |
| 2 | 169 | 71 | 1 | 170 | 62 |
| 1 | 175 | 68 | 1 | 182 | 75 |
| 2 | 170 | 67 | 2 | 177 | 70 |
| ： | ： | ： | ： | ： | ： |
| 1 | 181 | 65 | 2 | 173 | 58 |
| 1 | 179 | 74 | 2 | 169 | 59 |
| 2 | 178 | 71 | 2 | 170 | 73 |

問題

１）身長についての基本統計量を求めよ。

２）体重について平均、中央値、標準偏差を求めよ。

平均［　　　　　　］　中央値［　　　　　　］　標準偏差［　　　　　　］

３）身長について5cm毎の度数分布表を描け。

４）身長について5cm毎のヒストグラムを描け。

５）体重について10kg毎のヒストグラムを描け。

６）学校別に身長についての基本統計量を求めよ。

７）学校別に体重の平均と標準偏差を求めよ。

学校１　平均［　　　　　　］　標準偏差［　　　　　　］

学校２　平均［　　　　　　］　標準偏差［　　　　　　］

８）学校1について、身長のヒストグラムを描け。

### 1.2.2　相関係数と回帰式【第３回】

散布図（分布図，相関図），回帰直線，相関係数



例

　以下のデータ（Samples\テキスト1.txt）を用いて次の問いに答えよ。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学校 | 身長(cm) | 体重(kg) | 学校 | 身長(cm) | 体重(kg) |
| 2 | 169 | 71 | 1 | 170 | 62 |
| 1 | 175 | 68 | 1 | 182 | 75 |
| 2 | 170 | 67 | 2 | 177 | 70 |
| 1 | 179 | 72 | 1 | 175 | 70 |
| 1 | 176 | 69 | 1 | 172 | 62 |
| 2 | 174 | 81 | 2 | 166 | 58 |
| 2 | 173 | 75 | 2 | 168 | 60 |
| 1 | 181 | 65 | 2 | 173 | 58 |
| 1 | 179 | 74 | 2 | 169 | 59 |
| 2 | 178 | 71 | 2 | 170 | 73 |

１）身長と体重に関する散布図を描け（体重を縦軸）。

２）身長と体重の相関係数を求めよ。

３）身長で体重を予測する回帰式を求めよ。

問題

　Samples\テキスト9.txtを用いて以下の問いに答え、結果は文書にまとめよ。但し、地域について1：市街、2：郊外とする。

１）年収に関する基本統計量を求めよ。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| データ数 | 最小値 | 最大値 | 平均値 | 中央値 | 不偏分散 | 標準偏差 |
|  |  |  |  |  |  |  |

　　データの拡がりをみるには上のどの指標が適切か［　　　　　　　］

２）地域別の年収に関する基本統計量を求めよ。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | データ数 | 最小値 | 最大値 | 平均値 | 中央値 | 不偏分散 | 標準偏差 |
| 市街 |  |  |  |  |  |  |  |
| 郊外 |  |  |  |  |  |  |  |

　　市街と郊外ではどちらの年収が高いか ［市街・郊外］

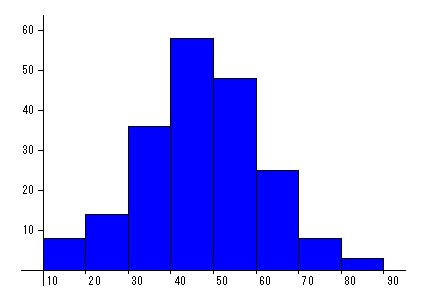
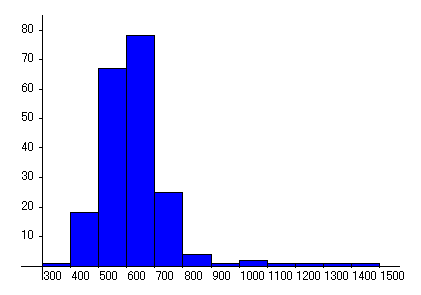
　　市街と郊外ではどちらの年収の拡がりが大きいか ［市街・郊外］

３）年収に関するヒストグラムを描け。（下図左）

　　このヒストグラムの階級幅はいくらか［　　　　　］

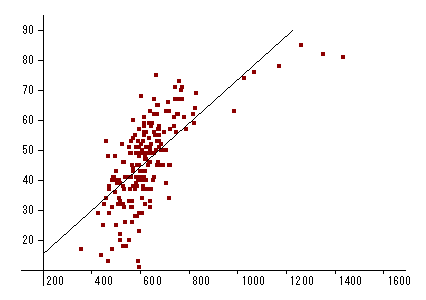
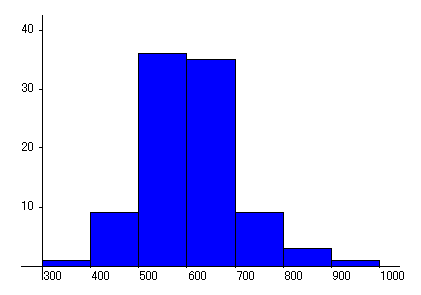
　　このヒストグラムの最頻値はいくらか［　　　　　］

４）支出に関するヒストグラムを描け。（下図右）



５）地域:1の年収に関するヒストグラムを描け。（下図左）

６）年収と支出に関する散布図を描け（支出を縦軸，下図右）。



７）年収と支出に関する相関係数を求めよ。 相関係数［　　　　　　］

８）支出を目的変数に年収を説明変数としたときの回帰式を求めよ。

　　　支出＝［　　　　　　］×年収＋［　　　　　　］

# ２．確率分布と検定【第４回】

## 2.1　確率密度関数

　データ数を十分多く取ったヒストグラムの上端をつなぎ、全体の面積が１になるように、目盛りを付けたものを確率密度関数と呼ぶ。この確率密度関数の形で分布の名前が付けられている。

：確率密度関数

## 2.2　正規分布 (normal distribution) と標準正規分布

正規分布（は平均 ** 分散 ** の正規分布：）

正規分布とは偶発的なデータのゆらぎによる分布（量的データの基本となる分布）

 両側　約32%

 両側　約5%

 両側　約0.3%

概数は覚えること

よく使う正規分布の性質

１）のとき　

の分布を標準正規分布といい、統計処理では非常によく利用される。

標準正規分布の詳しい確率の値は、例えばExcelでは、以下で求められる。

昔は表を使って求めていた。



２）を個のデータの平均とすると　

１つ１つのデータに対して、平均を取るとデータの精度が上がる。

標準偏差は、例えば100個だとになる。

これは*X*の分布によらない。中心極限定理

問題（１個のデータについて）

体重の平均20kg、標準偏差2kg（分散4kg2）の子供1000人の集団がある。データは正規分布するとして以下の問いに概数（大体の値）で答えよ。

１）22kgの子供は重い方から大体何％か ［　　　　］％

２）24kgの子供は重い方から大体何％か ［　　　　］％

３）24kgの子供は重い方から大体何番目か ［　　　　］番目

４）18kgの子供は重い方から大体何％か ［　　　　］％

５）18kgの子供は重い方から大体何番目か ［　　　　］番目

問題（１個のデータについて）

前の問題で、子供の体重から平均の20kgを引き、その結果を標準偏差の2kgで割るとする。以下の問いに答えよ。

１）20kgの子供の値はいくらになるか ［　　　　　］

２）22kgの子供の値はいくらになるか ［　　　　　］

３）17kgの子供の値はいくらになるか ［　　　　　］

４）この計算結果は平均［　　　］、分散［　　　］の正規分布になる。

５）３）について、17kg以下となる正確な確率を求めよ。

　　Excelの関数normsdist(x)を用いると ［　　　　　］

６）２）について、22kg以上となる正確な確率を求めよ。

　　Excelの関数normsdist(x)を用いると ［　　　　　］

７）17kg以上、22kg以下となる正確な確率を工夫して求めよ。

　　Excelの関数normsdist(x)を用いると ［　　　　　］

問題（データの平均について）

体重の平均20kg、標準偏差2kgの子供の大きな集団（母集団）がある。この中から100人の集団（標本）をランダムに取り出し、その平均を取るとする。以下の問いに答えよ。

１）の平均（標本平均の平均）はいくらか。 　［　　　　　］kg

２）の標準偏差（標本平均の標準偏差）はいくらか　［　　　　　］kg

３）の値が20.2kgの標本は重い方から大体何％か 　［　　　　　］％

## 2.3　標準正規分布から導かれる分布【第５回】

χ2分布

で独立なとき、

分布（自由度*n*の分布）

F分布

分布，分布で独立なとき、

分布

（自由度*n*1, *n*2のF分布）

t分布

分布，分布で独立なとき、

分布（自由度*n*のt分布）

注）分布

注）で分布

問題　College Analysisを使って以下の値を求めよ。

１）分布，*x*値1.5のときの上側確率 *p*/2 ［　　　　　　］

２）分布，*x*値1.5のときの両側確率 *p* ［　　　　　　］

３）分布，*x*値180のときの上側確率 *p*/2 ［　　　　　　］

４）分布，値10のときの上側確率 *p* ［　　　　　　］

５）分布，上側確率0.05のときの値 ［　　　　　　］

６）分布，*F*値10のときの上側確率 *p* ［　　　　　　］

７）分布，上側確率0.05のときの*F*値 ［　　　　　　］

８）分布，値2のときの上側確率 *p*/2 ［　　　　　　］

９）分布，値2のときの両側確率 *p* ［　　　　　　］

10）分布，両側確率0.05のときの値 ［　　　　　　］

## 2.4　検定の考え方

母集団と標本

母集団：調査の対象，日本人・日本の中小企業等（全数調査不可能な場合がある）

標本：　偏りがないように選抜（ランダムサンプリング）された実際に調査する対象

母集団の全数調査が不可能な場合、標本をとって母集団を推測する。



どんな検定があるか

１）指定値と母集団のある指標を比較する。

量的データの比較：

標本調査世帯と全国平均との所得の比較

質的データの比較：

標本調査の結果（割合）と期待される結果

（割合）との比較

２）いくつかの母集団のある指標を比較する。

量的データの比較：

２つの標本調査世帯の所得の比較

（対応がない場合）

標本店における宣伝前後の売り上げ比較

（対応がある場合）

質的データの比較：

男女間での意識調査の結果（割合）の比較（対応がない場合）

標本店における従業員教育前後の評判の変化（対応がある場合）

# ３．質的データの検定【第６回】

## 3.1　母集団の比率と指定比率との検定

例

　ある大学の学生50人を任意抽出し、大学改革のアンケートを行ったところ、賛成35反対15であった。学生の過半数が賛成している（賛成の比率が1/2と異なる）といえるか、有意水準5%で判定せよ。

理論　適合度検定

出現比率が指定比率と比べて差がないとすると

分布

分布

（Yatesの連続補正）

解答







判定　賛成は過半数といえる。

問題１

ある工場で１年間におきた事故の件数を曜日毎に調べたところ、以下の表が得られた。事故は曜日による差があるといえるか？有意水準5％で判定せよ。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 曜日 | 月 | 火 | 水 | 木 | 金 | 計 |
| 事故件数 | 23 | 14 | 16 | 11 | 16 | 80 |

解答

P =［　　　　　］

判定　曜日による差があると［いえる・いえない］

問題２

　前の問題で、月曜日は特に事故が起こっているといえるか。月曜日とその他の曜日に分けて有意水準5%で判定せよ。

解答

P =［　　　　　］

判定　月曜日に事故が多く起こっていると［いえる・いえない］

## 3.2　対応のない２群間の比率の検定

例

ある問題についての調査で、男女別に賛成か反対かを集計したところ以下の結果を得た。賛成（または反対）の比率に男女差はあるといえるか。有意水準5%で判定せよ。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 賛成 | 反対 | 計 |
| 男性 | 18 | 10 | 28 |
| 女性 | 12 | 14 | 26 |
| 計 | 30 | 24 | 54 |

解答

，

より、男女差があるとはいえない。

理論（２×２分割表）

|  |  |  |  |
| --- | --- | --- | --- |
|  | 事象１ | 事象２ | 計 |
| 要因１ | *a* | *b* | *a*+*b* |
| 要因２ | *c* | *d* | *c*+*d* |
| 計 | *a*+*c* | *b*+*d* | *a*+*b*+*c*+*d*=*n* |

要因間で、事象の出現比率に差がないとすると

分布

分布　　（Yatesの連続補正）

理論（ｍ×ｎ分割表）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 事象１ | 事象２ | … | 事象*s* | 計 |
| 要因１ | *x*11 | *x*12 | … | *x*1*s* | *x*1･ |
| 要因２ | *x*21 | *x*22 | … | *x*2*s* | *x*2･ |
| ： | ： | ： |  | ： | ： |
| 要因*r* | *xr*1 | *xr*2 | … | *xrs* | *xr*･ |
| 計 | *x*･1 | *x*･2 | … | *x*･*s* | *N* |

要因間で、事象の出現比率に差がないとすると

分布　　表の統計量の一般形

分布　　（Yatesの連続補正）

問題３

　ある案についてのアンケートで以下の結果を得た。男女間の回答（賛成の比率）に差があるといえるか。有意水準5%で判定せよ。

|  |  |  |
| --- | --- | --- |
|  | 賛成 | 反対 |
| 男性 | 128 | 86 |
| 女性 | 107 | 95 |

確率［　　　　　］　判定　男女間に差があると［いえる・いえない］

問題４

　女性を対象とした調査で、ある化粧品の所有の有無を職業別に分類してみると、以下の結果が得られた。職業間で商品所有の割合に差があるといえるか。有意水準5%で判定せよ。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 所有あり | 所有なし | 計 |
| 主婦 | 90 | 199 | 289 |
| 事務 | 32 | 47 | 79 |
| 販売・生産 | 53 | 71 | 124 |
| 計 | 175 | 317 | 492 |

確率［　　　　　］　判定　職業間に差があると［いえる・いえない］

問題５

　Samples\テキスト9.txtにおいて、以下の問いに答えよ。

１）意見１について１次元分割表を描け。（1：はい，2：いいえ）

|  |  |  |
| --- | --- | --- |
| はい | いいえ | 合計 |
|  |  |  |

２）意見１において、いいえは過半数といえるか。有意水準5%で判定せよ。

P =［　　　　　］

判定　過半数と［いえる・いえない］

３）意見２について１次元分割表を描け。（1：案１，2:案２，3:案３）

|  |  |  |  |
| --- | --- | --- | --- |
| 案１ | 案２ | 案３ | 合計 |
|  |  |  |  |

４）意見２において、回答間に差があるといえるか。有意水準5%で判定せよ。

P =［　　　　　］

判定　回答間に差があると［いえる・いえない］

５）意見１の回答に地域による差があるか。有意水準5%で判定せよ。

確率［　　　　　］　判定　地域による差があると［いえる・いえない］。

６）上の問題で有意水準を1%にすると結果はどう変わるか。

判定　地域による差があると［いえる・いえない］。

７）意見２の回答に地域による差があるか。有意水準5%で判定せよ。

確率［　　　　　］　判定　地域による差があると［いえる・いえない］。

８）意見２の回答に意見１による差があるか。有意水準5%で判定せよ。

確率［　　　　　］　判定　意見１による差があると［いえる・いえない］。

問題６

１）平均的な故障発生率が1%のとき、100個の製品で4個以上の故障発生が起こることは異常なことであろうか。2項分布の正確な確率検定を用いて、有意水準5%で判定せよ。

検定確率　［　　　　　　］　　異常なことと［いえる・いえない］

２）昨年の納品155個のうち、クレームのあったものが2個、今年の納品211個のうち、クレームのあったものが7個であった。昨年と今年のクレームの比率に差があると言えるか。Fisherの正確確率検定を用いて、有意水準5%で判定せよ。

検定確率　［　　　　　　］　　差があると［いえる・いえない］

演習５（適合度検定・χ2検定）【第７回】

　環境の変化と校舎の老朽化により、現在の小学校を少し離れた場所に移設しようとする案が出され、地域住民に以下のアンケートが実施された。

１．あなたの性別は。

１）男性 ２）女性

２．あなたの年齢はどれですか。

１）４０歳未満 ２）４０歳～５９歳 ３）６０歳以上

３．あなたには小学生以下のお子さんかお孫さんがいますか。

１）いる ２）いない

４．学校の移設に賛成ですか反対ですか。

１）賛成 ２）反対

５．学校移設に賛成の方、その理由を教えて下さい。（いくつでも選んで下さい。）

１）環境が良くなる　　２）設備が新しくなる　　３）その他

６．学校移設に反対の方、その理由を教えて下さい。（いくつでも選んで下さい。）

１）通学が不便になる　　２）現校舎に魅力がある　　３）その他

演習5.txtを用いて問題に答えよ。

問題

１）男性と女性の数を求める。男性［　　　］人　女性［　　　］人

２）年齢階級別の人数を求める。

　　４０歳未満［　　　］人　４０～５９歳［　　　］人　６０歳以上［　　　］人

３）小学生以下の子どもがいるかどうかの人数を求める。

　　いる［　　　］人　いない［　　　］人

４）賛成と反対の人数を求める。賛成［　　　］人　反対［　　　］人

５）男女別の賛成と反対の２次元分割表を作る。

|  |  |  |
| --- | --- | --- |
|  | 賛成 | 反対 |
| 男性 |  |  |
| 女性 |  |  |

６）子供がいるかどうか別の賛成と反対の２次元分割表を作る。

|  |  |  |
| --- | --- | --- |
|  | 賛成 | 反対 |
| 子供がいる |  |  |
| 子供がいない |  |  |

７）年齢層別の賛成と反対の２次元分割表を作る。

|  |  |  |
| --- | --- | --- |
|  | 賛成 | 反対 |
| 40歳未満 |  |  |
| 40～59歳 |  |  |
| 60歳以上 |  |  |

８）賛成は過半数といえるか、有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

結果　過半数と［いえる・いえない］

９）賛成の理由で「環境が良くなる」と答える人は４割より多いといえるか、有意水準5%で判定せよ。

ヒント：答えない（0），答えた（1）順に指定比率を入力すること（0.6,0.4）

検定名［　　　　　　　　　　］　確率［　　　　　］

結果　４割より多いと［いえる・いえない］

10）上の検定で有意水準を1%にすると結果はどうなるか。

結果　４割より多いと［いえる・いえない］

11）反対の理由で「通学が不便になる」と答える人は３割より多いといえるか、有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

結果　３割より多いと［いえる・いえない］

12）男性と女性で賛成・反対に差があるか、有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

結果　男女間に差があると［いえる・いえない］

13）小学生以下の子供がいるかどうかで賛成・反対に差があるか、有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

結果　子供がいるかどうかで差があると［いえる・いえない］

14）年齢層によって賛成・反対に差があるか、有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

結果　年齢層間で差があると［いえる・いえない］

15）賛成・反対に影響を与えている要素はどれか。

［性別・子供または孫の有無・年齢］

# ４．母集団と指定値との量的データの検定【第８回】

## 4.1　検定手順



## 4.2　正規性の検定

視覚的方法

データ数が多い場合 ヒストグラムによるグラフ化

データ数が少ない場合 正規確率紙（MS-Excelでも可能）

数値的方法

データ数が多い場合

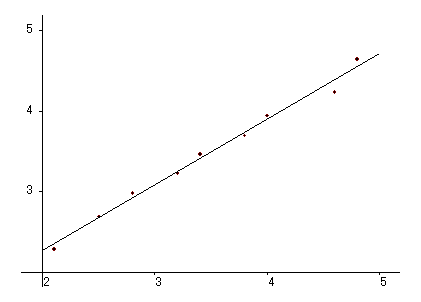
　コルモゴロフ－スミルノフ（Kolmogorov-Smirnov略してK-S）検定

データ数が少ない場合［今後主にこれを使用する］

シャピロ－ウィルク（Shapiro-Wilk略してS-W）検定　等

問題１

　以下のデータの正規性を調べよ。

2.5, 2.1, 3.4, 2.8, 4.6, 3.2, 3.8, 4.8, 4.0

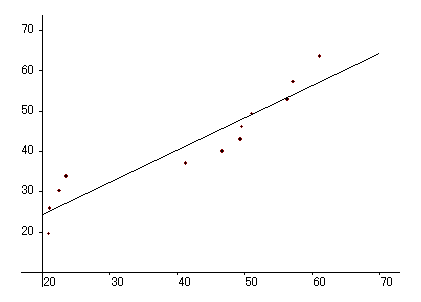
解答

　データの数が少ないので、ヒストグラムは使えない。正規確率紙の方法とS-W検定で調べる。

　S-W検定　　確率［　　　　　］

判定　正規分布と［みなす・いえない・判定困難］

問題２

　以下のデータの正規性を調べよ。

20.9, 61.1, 57.2, 51.0, 46.6, 41.2, 21.0, 56.3, 49.5, 49.3, 22.4, 23.5

解答

正規確率紙の方法とS-W検定で調べる。

　S-W検定　　確率［　　　　　］

判定　正規分布と［みなす・いえない・判定困難］

## 4.3　母集団の平均値と指定値との比較（正規性あり）

例

　ある地域のある規模の会社11社について１人当り売上高は以下の通りである。この地域の会社の１人当り売上高は同じ規模の会社の１人当り平均売上高2260（万円）に比べて差があるといえるか？検定を選んで有意水準5%で判定せよ。

2060, 2350, 1550, 1720, 1800, 1990, 1510, 1720, 2910, 1820, 2600

理論　母平均のt検定

　指定値と比べて平均に差がないとして、

分布

解答



 より、１人当り売上高に差があるといえない。

## 4.4　母集団の中央値と指定値との比較（正規性なし）

例

ある地域のある規模の会社の１人当り売上高（万円）は以下の通りである。これらの会社は同じ規模の会社の中央値2260（万円）に比べて売上高に差があるといえるか。検定を選んで有意水準5%で判定せよ。

2060, 2064, 2072, 2005, 2602, 1987, 1824, 1720, 2035, 1890, 2025,

概要　Wilcoxon（ウィルコクソン）の符号付き順位和検定

　データの順位により母集団の中央値が指定値と異なっているかどうか検定する。



図　検定概念図

　左右の順位和を求め、その小さい方をとする。

標本数が多いとき

 ～ 分布（正の部分）　（Yatesの連続補正）

解答

，　　より、中央値に差があるといえる。

問題３

　以下のデータの平均値（中央値）は5.5と比べて差があるといえるか。検定を選んで有意水準5%で判定せよ。

8.4, 4.6, 5.2, 6.3, 7.2, 5.8, 6.0, 5.4, 4.9, 6.9

　正規性の判定

　　ヒストグラムにはデータ不足。正規確率紙を描く。S-W検定　確率［　　　　　］

　　判定　正規分布と［みなす・いえない］。

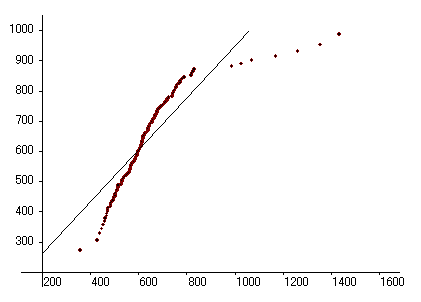
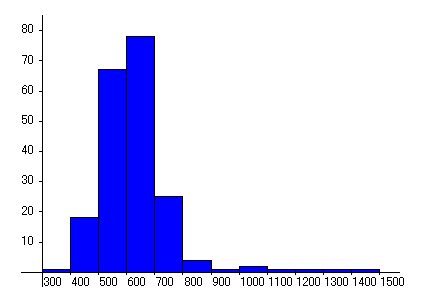
　検定名［　　　　　　　　　　］　確率［　　　　　］

　判定　5.5と比べて差があると［いえる・いえない］。

問題４

　Samples\テキスト9.txtのデータを用いて以下の問いに答えよ。

１）年収のデータの正規性をヒストグラム、正規確率紙、S-W検定で調べよ。



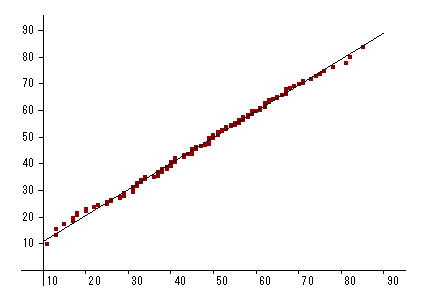
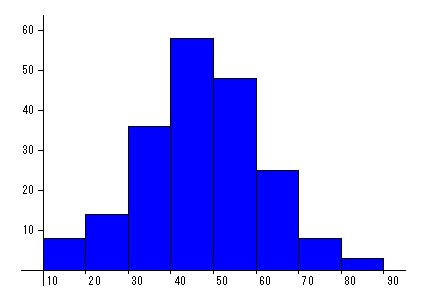
　　S-W検定　　確率［　　　　　］判定　正規分布と［みなす・いえない］。

２）年収の平均値（中央値）は610万円と比べて差があるとといえるか。分析を選んで有意水準5%で判定せよ。

検定名［　　　　　　　　　　　　］　確率［　　　　　］

判定　610万円より多いと［いえる・いえない］。

３）支出のデータの正規性をヒストグラム、正規確率紙、S-W検定で調べよ。



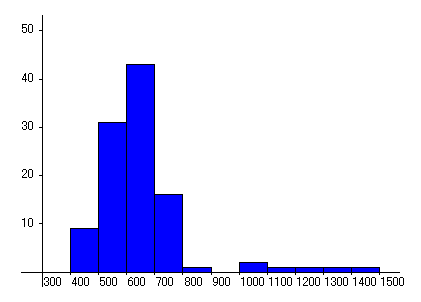
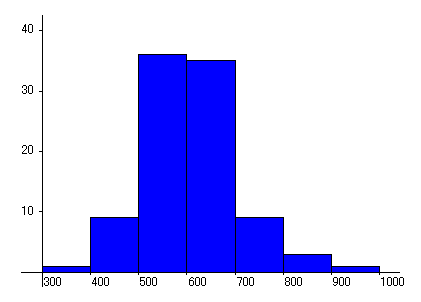
　S-W検定　　確率［　　　　　］判定　正規分布と［みなす・いえない］。

４）支出の平均値（中央値）は44万円と比べて差があるといえるか。分析を選んで有意水準5%で判定せよ。

　検定名［　　　　　　　　　　　　］　確率［　　　　　］

　判定　44万円と比べて差があると［いえる・いえない］。

５）地域別に年収のデータの正規性を調べよ。



　地域１　　確率［　　　　　］判定　正規分布と［みなす・いえない］。

　地域２　　確率［　　　　　］判定　正規分布と［みなす・いえない］。

# ５．２群間の量的データの検定【第９回】

## 5.1　対応のない検定手順



## 5.2　対応のない２群間の分散の検定（正規性あり）

例

　Ａ機を導入した会社18社（１群）とＢ機を導入した会社15社（２群）について、機械10台当り１年間の故障発生件数を調べ、不偏分散を求めたら以下の結果を得た。

|  |  |  |
| --- | --- | --- |
|  | 平均 | 不偏分散 |
| １群 | 10.56 | 10.68 |
| ２群 | 8.22 | 3.17 |

分布は正規分布であると仮定して、分散に差があるといえるか有意水準5%で判定せよ。

理論　F検定

母分散に差がないとすると

分布

解答

　　　より、分散に差があるといえる。

## 5.3　対応のない２群間の平均値の検定（正規性あり・等分散）

例

ある地域の同性・同年齢の児童について、ある要因の有無による２つの集団の体重を調べたところ以下のデータを得た。２つの集団の平均値に差はあるといえるか。正規性、等分散性を仮定して、有意水準5%で判定せよ。

|  |  |  |  |
| --- | --- | --- | --- |
|  | データ数 | 平均 | 不偏分散 |
| 要因なし | 20 | 40.2 | 25.5 |
| 要因あり | 20 | 36.4 | 16.0 |

理論（studentの）t検定

母平均に差がないとすると

分布

解答

　　　より、平均に差があるといえる。

## 5.4　対応のない２群間の平均値の検定（正規性あり・等分散性なし）

例

Ａ機を導入した会社18社（１群）とＢ機を導入した会社15社（２群）について、機械10台当り１年間の故障発生件数を調べ、平均と不偏分散を求めたところ以下の結果を得た。正規性があり、異分散であるとして、２群間の平均に差があるかどうか有意水準5%で検定せよ。

|  |  |  |
| --- | --- | --- |
|  | 平均 | 不偏分散 |
| １群 | 10.56 | 10.68 |
| ２群 | 8.22 | 3.17 |

理論　Welch(ウェルチ)のt検定

母平均に差がないとすると

　として、自由度を　　とし、

分布

解答

 （自由度）（小数点以下切り捨て）

 　より、平均に差があるといえる。

問題１

　ある1人当りの売上のデータについて、2つの地域の支店を比較したところ、以下の結果が得られた。２群間に差があるといえるか。有意水準5%の両側検定で判定せよ。

１群　2007, 2344, 2434, 2251, 2673, 1452, 2393, 2126, 2485, 1279, 2269

２群　2579, 2899, 2258, 3086, 2998, 2829, 2408, 2287, 3020, 1989, 2136

　検定名［　　　　　　　　　　］　確率［　　　　　］

　判定　母平均（母集団の中央値）に差があると［いえる・いえない］

## 5.5　対応のない２群間の中央値の検定（正規性なし）

例

　ある1人当りの売上のデータについて、2つの地域の支店を比較したところ、以下の結果が得られた。２群の売上は１群のそれに比べて大きいといえるか。有意水準5%の両側検定で判定せよ。

１群　2060, 2350, 1550, 1720, 1800, 1990, 1510, 1720, 2910, 1820, 2600

２群　1720, 2064, 2072, 2005, 2602, 1987, 1824, 2060, 2035, 1890, 2025

概要　Wilcoxon(ウィルコクソン)の順位和検定



両群のデータの小さい順に順位を付け、データ数の少ない群（）の順位和をとする。但し、同じ値にはそれらが異なると考えた場合の順位の平均値を付ける。

データ数が多い場合　両群の中央値が等しいとすると

分布（正の部分）　（Yatesの連続補正）

解答

 より、中央値に差があるといえない。

問題２

以下の標本データの母平均（母集団の中央値）には差があるといえるか。検定を選んで有意水準5%で判定せよ。

１群　112, 106, 101, 112, 102, 98, 108, 95, 101, 90, 110, 97, 95, 105, 101, 113, 114, 91

２群　98, 88, 105, 99, 96, 93, 109, 106, 103, 87, 107, 102, 97, 91

　検定名［　　　　　　　　　　］　確率［　　　　　］

　判定　母平均（母集団の中央値）に差があると［いえる・いえない］

問題３

以下の標本データの母平均（母集団の中央値）には差があるといえるか。検定を選んで有意水準5%で判定せよ。

１群　358, 469, 397, 350, 329, 446, 393, 379, 443, 348,  
455, 332, 311, 424, 420, 354, 353, 390, 434, 430

２群　335, 387, 385, 343, 394, 351, 404, 391, 330, 363,  
319, 334, 348, 396, 408, 403, 415, 353, 377, 399

　検定名［　　　　　　　　　　］　確率［　　　　　］

　判定　母平均（母集団の中央値）に差があると［いえる・いえない］

問題４

ラットの体重増加(g)を、条件を変えた２つのグループで測定したところ、以下の結果が得られた。２群の体重増加に差は認められるか、有意水準5%で判定せよ。

　１群：7.2, 8.3, 5.4, 6.0, 7.3, 11.7, 10.5, 8.0, 9.1

　２群：10.1, 13.2, 7.4, 9.1, 16.2, 14.5, 6.3, 11.2, 12.4, 7.4, 12.5, 9.1, 17.0

　検定名［　　　　　　　　　　］　確率［　　　　　］

　判定　体重増加に差があると［いえる・いえない］

問題５

　Samples\テキスト9.txtのデータを用いて以下の問いに答えよ。

１）地域別の年収に差があるか、検定を選んで有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　］

　　判定　地域別の年収に差があると［いえる・いえない］

２）地域別の支出に差があるか、検定を選んで有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　］

　　判定　地域別の支出に差があると［いえる・いえない］

３）意見１別の年収に差があるか、検定を選んで有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　］

　　判定　意見１で答え方が違う人で年収に差があると［いえる・いえない］

演習６－２【第10回】

　ある中学のクラスについて英語・数学・国語の試験結果を調べた。

１．性別

１）男子　　　２）女子

２．英語点数

３．数学点数

４．国語点数

演習6.txtのデータをCollege Analysisに読み込んで集計し、各科目の平均または中央値に差があるかどうか検定せよ。特に正規性が判定困難な場合、一般の分布の検定を行って判定を下すこと。

問題

１）男女の人数を求める。

　　男子［　　　　］人　女子［　　　　］人

２）英語、数学、国語の平均と中央値を求める。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 英語 | 数学 | 国語 |
| 平均 |  |  |  |
| 中央値 |  |  |  |

３）英語、数学、国語の標準偏差を求める。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 英語 | 数学 | 国語 |
| 標準偏差 |  |  |  |

４）最も成績の良かった科目は何か。 ［英語・数学・国語］

　　最も点数のばらつきの大きかった科目は何か。 ［英語・数学・国語］

５）それぞれ科目間の相関係数はいくらか。

　　　英語と数学　　［　　　　　］

　　　数学と国語　　［　　　　　］

　　　国語と英語　　［　　　　　］

　　　これらの科目間には関係があると思うか。　［ある・ない］と思う。

６）英語と数学の平均（中央値）に差があるか、（対応がないとして）有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

判定　科目間に差があると［いえる・いえない］。

７）数学と国語の平均（中央値）に差があるか、（対応がないとして）有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

判定　科目間に差があると［いえる・いえない］。

８）国語と英語の平均（中央値）に差があるか、（対応がないとして）有意水準5%で判定せよ。

検定名［　　　　　　　　　　］　確率［　　　　　］

判定　科目間に差があると［いえる・いえない］。

９）英語、数学、国語の平均について男女別に求める。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 英語 | 数学 | 国語 |
| 男子 |  |  |  |
| 女子 |  |  |  |

10）英語、数学、国語の中央値について男女別に求める。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 英語 | 数学 | 国語 |
| 男子 |  |  |  |
| 女子 |  |  |  |

11）男女別の英語の平均（中央値）に差があるか、有意水準5%で判定する。

検定名［　　　　　　　　　　］　確率［　　　　　］

判定　男女差があると［いえる・いえない］。

12）男女別の数学の平均（中央値）に差があるか、有意水準5%で判定する。

検定名［　　　　　　　　　　］　確率［　　　　　］

判定　男女差があると［いえる・いえない］。

13）男女別の国語の平均（中央値）に差があるか、有意水準5%で判定する。

検定名［　　　　　　　　　　］　確率［　　　　　］

判定　男女差があると［いえる・いえない］。

14）対応のない２群間の量的データの比較の検定手法で、最も一般的に使えるのは［　　　　　　　　］検定で、逆に最も制約が多いのは［　　　　　　　　］検定である。しかし［　　　　　］や等［　　　　　］などの制約が満たされるとき、後者は最も検出力の［高い・低い］検定となり、２群間の差は見つけ［やすく・にくく］なる。

## 5.6　対応がある検定手順【第11回】



## 5.7　対応がある２群間の平均値の検定（正規性あり）

例

ある商品の陳列位置を変える前と後とで売上高（千円）を規模の等しい８つの支店で比較したところ、以下の結果を得た。検定を選択して有意水準5%で差があるかどうか判定せよ。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 前 | 385 | 402 | 320 | 383 | 504 | 417 | 290 | 342 |
| 後 | 396 | 373 | 431 | 457 | 514 | 405 | 380 | 396 |

理論

対応する各標本の差（=(標本１－標本２)÷2）をとる。平均が等しいと仮定すると

分布

解答

　　より、平均に差があるとはいえない。

## 5.8　対応がある２群間の中央値の検定（正規性なし）

例

　ある商品の陳列位置を変える前と後とで売上高（千円）を規模の等しい８つの支店で比較したところ、以下の結果を得た。検定を選択して有意水準5%で売上高に差があるかどうか判定せよ。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 前 | 385 | 402 | 320 | 383 | 504 | 417 | 290 | 342 |
| 後 | 396 | 310 | 342 | 407 | 514 | 405 | 380 | 365 |

概要　Wilcoxonの符号付き順位和検定

　対応する各標本の差（=(標本１－標本２)÷2）について、の正負で2群に分けて順位和を求め、小さい方をとする。標本数が多いとき（少ない場合は数表を用いる）

　～　分布（正の部分）

解答

 より、２標本の中央値に差があるといえない。

注）2群のデータの分散は大きいが、各データ間の差が同じ符号の傾向がある場合、対応のある検定が非常に有効となる。（テキスト5.txt 7ページ）

問題

ある小学生の集団で国語・算数・社会・理科の学力を調べたところ以下のようなデータを得た。質問に答えよ。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 国語 | 68 | 58 | 60 | 63 | 55 | 69 | 63 | 79 | 62 | 74 | 53 | 75 | 64 | 77 | 66 |
| 算数 | 75 | 59 | 58 | 73 | 59 | 69 | 62 | 67 | 68 | 78 | 53 | 67 | 69 | 77 | 70 |
| 社会 | 66 | 58 | 50 | 55 | 57 | 66 | 54 | 91 | 57 | 56 | 65 | 55 | 80 | 90 | 63 |
| 理科 | 82 | 60 | 61 | 74 | 68 | 74 | 64 | 72 | 70 | 65 | 57 | 79 | 76 | 83 | 74 |

１）４科目の平均値と中央値を求める。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 国語 | 算数 | 社会 | 理科 |
| 平均値 |  |  |  |  |
| 中央値 |  |  |  |  |

２）各科目のデータの正規性を検討する。（下段にはみなす／いえないかを書き込む）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 国語 | 算数 | 社会 | 理科 |
| S-W検定確率 |  |  |  |  |
| 正規性ありと |  |  |  |  |

３）対応があるとして以下の科目間の点数の差の正規性を検討する。（同上）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 国語－算数 | 国語－社会 | 算数－理科 | 社会－理科 |
| S-W検定確率 |  |  |  |  |
| 正規性ありと |  |  |  |  |

　２群の比較ではデータ間に１対１の対応がある場合、通常対応がある検定手法を利用するが、対応がないとして検定しても間違いではない。以下の問題は両方の方法で検定を行い、結果を比較せよ。

４）国語と算数の平均値（中央値）に差があるといえるか、有意水準5%で判定する。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 検定名 | 確率 | 判定 |
| 対応なし |  |  | 差があると［いえる・いえない］ |
| 対応あり |  |  | 差があると［いえる・いえない］ |

演習７【Skip OK】

　都市圏と地方都市の小学生４年生からサンプルを選んで実施した調査で、演習7.txtのような結果が得られた。どんなことが読み取れるか考え、質問に答えよ。

１．住んでいる地域　　１）都市圏（東京・大阪）　２）地方都市（その他）

２．性別　　１）男子　　　２）女子

３．どんな遊びが好きか　　１）外で運動　　２）家の中でゲーム・読書他

４．体力測定（点数換算）

５．国語

６．算数

問題

１）外で遊ぶ人数と家の中で遊ぶ人数を求める。

|  |  |  |
| --- | --- | --- |
| 外で | 家の中で | 合計 |
|  |  |  |

２）家の中で遊ぶ割合は過半数といえるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　　確率［　　　　　　］

　　判定　家の中で遊ぶ割合は過半数であると［いえる・いえない］

３）地域による好きな遊びの傾向をみる。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 外で | 家の中で | 合計 |
| 都市圏 |  |  |  |
| 地方都市 |  |  |  |
| 合計 |  |  |  |

４）好きな遊びに地域差はあるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　　確率［　　　　　　］

　　判定　好きな遊びに地域差はあると［いえる・いえない］

５）男女による好きな遊びの傾向をみる。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 外で | 家の中で | 合計 |
| 男子 |  |  |  |
| 女子 |  |  |  |
| 合計 |  |  |  |

６）好きな遊びに男女差はあるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　　確率［　　　　　　］

　　判定　好きな遊びに男女差はあると［いえる・いえない］

７）地域別の体力測定の平均値と中央値を求める。

|  |  |  |
| --- | --- | --- |
|  | 都市圏 | 地方都市 |
| 平均値 |  |  |
| 中央値 |  |  |

８）体力測定の結果に地域差があるといえるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　　確率［　　　　　］

　　判定　地域による差があると［いえる・いえない］。

９）男女別の体力測定の平均値と中央値を求める。

|  |  |  |
| --- | --- | --- |
|  | 男子 | 女子 |
| 平均値 |  |  |
| 中央値 |  |  |

10）体力測定の結果に男女差があるといえるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　　確率［　　　　　］

　　判定　男女による差があると［いえる・いえない］。

11）地域別の算数の結果

|  |  |  |
| --- | --- | --- |
|  | 都市圏 | 地方都市 |
| 平均値 |  |  |
| 中央値 |  |  |

12）算数の結果に地域差があるといえるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　　確率［　　　　　］

　　判定　地域による差があると［いえる・いえない］。

13）国語と算数の結果に差があるといえるか、有意水準5%で判定する。

　　この問題は対応がないとしても、対応があるとしても答えを求めることができる。

　　対応がないとした場合

　　　検定名［　　　　　　　　　　］　　確率［　　　　　］

　　　判定　国語と算数に差があると［いえる・いえない］。

　　対応があるとした場合

　　　検定名［　　　　　　　　　　］　　確率［　　　　　］

　　　判定　国語と算数に差があると［いえる・いえない］。

# ６．相関係数の検定と回帰分析【第12回】

## 6.1　（Pearsonの）相関係数

例

　２つの商品A, Bの地域別使用率 (%) のデータは以下の通りである。それぞれの商品の使用率に線形の相関が認められるか。正規性を仮定して、有意水準5%で検定せよ。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A(%) | 33 | 24 | 30 | 50 | 42 | 15 | 15 | 56 | 13 | 45 | 44 | 21 | 18 | 31 | 27 | 40 |
| B(%) | 20 | 34 | 50 | 20 | 58 | 23 | 12 | 34 | 26 | 56 | 42 | 5 | 25 | 51 | 19 | 27 |

理論

　母相関係数を0と仮定して以下の性質を利用する。

分布

解答

，，

より、相関があるといえない。

　　（相関係数が0と異なるといえない。この検定は正規分布以外では使えない。）

## 6.2　（Spearmanの）順位相関係数

例

　前節の問題で、それぞれの商品の使用率に相関（非線形のものも含む）が認められるか。正規性を仮定せずに、有意水準5%で検定せよ。

理論

　順位相関係数  を求め、母相関係数を0と仮定して以下の性質を用いる。

分布

解答

，

　より、相関があるとはいえない。

## 6.3　回帰分析

例

　下の表のデータを用いて、身長により体重を推定する式を考える。ただし、式は１次式（体重 = *a*×身長 + *b*）と仮定し、その有効性を検討せよ。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 体重 | 71 | 68 | 67 | 72 | 69 | 80 | 75 | 65 | 74 | 71 |
| 身長 | 169 | 175 | 170 | 179 | 176 | 174 | 173 | 181 | 179 | 178 |
| 体重 | 62 | 75 | 70 | 70 | 62 | 58 | 60 | 58 | 59 | 73 |
| 身長 | 170 | 180 | 177 | 175 | 172 | 166 | 168 | 173 | 169 | 170 |

理論



回帰式の決定

2変数の関係を、の直線で表わすとすると、を説明変数、を目的変数と呼ぶ。データ点からこの直線へ垂直におろした線の長さの２乗が最小となるように係数を決める。

平均 ，標準偏差 ，相関係数 　とすると

， 

回帰式の有効性の検討

重相関係数  目的変数の実測値と回帰式による予測値の相関係数

（説明変数が１つの場合　）

寄与率（重決定係数） 　目的変数の変動のうち回帰式が説明する割合

回帰式の有効性の検定（残差が正規分布する場合のみ利用可能）

回帰式は無意味（傾きが0）と考えられる確率で検討する。

解答

，

，，

，

回帰式 

重相関係数 

寄与率 

回帰式の有効性の検定

　　確率　0.0207　　回帰式は有効であるといえる。

問題１

　以下の2変数のデータを用いて問いに答えよ。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 変数1 | 65 | 86 | 78 | 83 | 85 | 89 | 83 | 80 | 85 | 93 | 75 | 85 | 79 | 80 |
| 変数2 | 162 | 210 | 224 | 179 | 217 | 230 | 223 | 204 | 224 | 197 | 186 | 189 | 172 | 185 |

１）２変数のPearsonの相関係数とSpearmanの順位相関係数を両方を求めよ。

|  |  |
| --- | --- |
| 相関係数 | 順位相関係数 |
|  |  |

２）相関の検定にはどちらの相関係数を利用するか。

　　［相関係数・順位相関係数］

３）上で選んだ相関係数を用いて、相関の有無を有意水準5%で判定せよ。

　　検定確率［　　　　　　］　相関があると［いえる・いえない］

４）変数２を目的変数、変数１を説明変数として回帰分析を行う。

　　回帰式　変数２＝［　　　　　］×変数１＋［　　　　　］

　　重相関係数　［　　　　　］

　　寄与率　　　［　　　　　］

５）回帰分析の有効性の検定は［行える・行えない］。

　　検定確率［　　　　　］

　　回帰式は有効であると［いえる・いえない］

問題２

　Samples\テキスト9.txtのデータを用いて以下の問題に答えよ。

１）年収と支出についての相関係数と順位相関係数を求める。

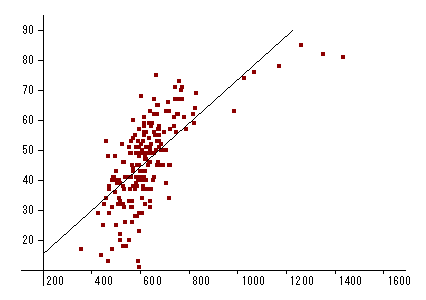
　　相関係数［　　　　　］　　順位相関係数［　　　　　］

２）年収と支出に相関があるといえるか、相関係数を選んで有意水準5%で判定する。

　　［相関係数・順位相関係数］で見る。

　　判定　確率［　　　　　］　相関があると［いえる・いえない］

３）年収（横軸）と支出（縦軸）について以下のような散布図を描く。



４）支出を目的変数、年収を説明変数として回帰分析を行う。

　　回帰式　支出＝［　　　　　］×年収＋［　　　　　］

　　重相関係数　［　　　　　］

　　寄与率　　　［　　　　　］

５）回帰分析の有効性の検定は［行える・行えない］。

　　検定確率［　　　　　］

　　回帰式は有効であると［いえる・いえない］

演習８【Skip OK】

　学生（男子）の地域別の身長・体重，試験の平均点と勉強時間を比べるために、調査を行い演習8.txtのデータを得た。以下の問題に答えよ。

１）地域（1：都市部 2：郊外）

２）身長

３）体重

４）実施した学力試験の点数（５科目の平均点）

５）１日平均の勉強時間

問題

１．都市部と郊外の調査数

　　都市部［　　　　］人　郊外［　　　　］人

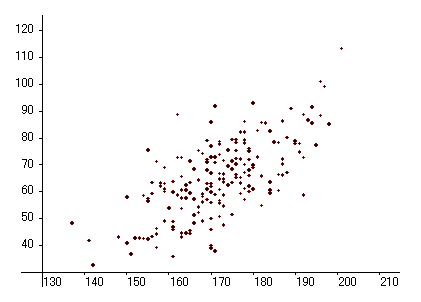
２．地域別にみた各変数の平均値と中央値を求める。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 身長 | 体重 | 点数 | 勉強時間 |
| 地域１平均値 |  |  |  |  |
| 地域１中央値 |  |  |  |  |
| 地域２平均値 |  |  |  |  |
| 地域２中央値 |  |  |  |  |

３．各項目に地域差があるといえるか、有意水準5%で判定する。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 検定名 | 検定確率 | 判定　差があると |
| 身長 |  |  | いえる・いえない |
| 体重 |  |  | いえる・いえない |
| 点数 |  |  | いえる・いえない |
| 勉強時間 |  |  | いえる・いえない |

４．身長と体重の散布図と相関係数



相関係数［　　　　　］，順位相関係数［　　　　　］

身長と体重には相関があるといえるか相関係数を選んで判定する。

［相関係数・順位相関係数］でみる。

検定確率［　　　　　］　相関があると［いえる・いえない］

５．体重を身長で予測する回帰分析結果

　　体重＝［　　　　　　］×身長＋［　　　　　　］

　　体重がどの程度身長で説明されるか示す。　寄与率　［　　　　　　］

　　回帰式の有効性の検定は可能か？可能な場合は有意水準5%で判定する。

　　［可能・不可能］

　　　検定確率［　　　　　］　この回帰式は有効であると［いえる・いえない］

６．勉強時間と点数の相関係数

相関係数［　　　　　］，順位相関係数［　　　　　］

勉強時間と点数には相関があるといえるか相関係数を選んで判定する。

［相関係数・順位相関係数］でみる。

検定確率［　　　　　］　相関があると［いえる・いえない］

７．点数を勉強時間で予測する回帰分析結果

　　点数＝［　　　　　　］×勉強時間＋［　　　　　　］

　　点数がどの程度勉強時間で説明されるか示す、寄与率　　［　　　　　　］

　　回帰式の有効性の検定は可能か？可能な場合は有意水準5%で判定する。

　　　［可能・不可能］

　　　検定確率［　　　　　］　この回帰式は有効であると［いえる・いえない］

８．勉強時間を２時間未満と２時間以上に分けて、点数について調べる。

|  |  |  |
| --- | --- | --- |
|  | ２時間未満 | ２時間以上 |
| 平均値 |  |  |
| 中央値 |  |  |

９．２時間未満と２時間以上で点数に差があるといえるか、有意水準5%で判定する。

　　検定名［　　　　　　　　　　］　検定確率［　　　　　］

　　点数に差があると［いえる・いえない］。

# ７．区間推定【第13回】

区間推定

標本から推測される母比率や母平均などがどの位の値の範囲に入るかを推定し、区間で表す方法。

信頼係数

推定した区間に母比率や母平均などが入る確率（%で表されることが多く、通常95%か99%）。１－信頼係数の値は検定での有意水準に相当する。

## 7.1　母比率の区間推定

例

　ある制度についてのアンケート調査をランダムに抽出された100人に対して行ったところ、賛成65人、反対35人であった。母集団の賛成の比率を、信頼係数95%（有意水準5% に相当）で推定せよ。また、調査数1000人で同じ比率ではどうか。

理論

データ数、標本比率の標本から、母比率を信頼係数で推定する。

として、信頼区間は以下で与えられる。



解答

、、



1000人では、以下のように精度が上がる。



## 7.2　正規母集団の母平均と母分散の区間推定

例

　ある標本データから所得について集計したところ以下の結果を得た。母集団は正規分布するとして母平均と母分散を信頼係数95%で推定せよ。

データ数 30，平均 620，標準偏差 90

　また、データ数を100にすると結果はどう変わるか？

理論

　正規分布する母集団から得られた標本より、母平均と母分散を信頼係数で推定する。データ数を，標本平均を，不偏分散を，，，として、各信頼区間は以下で与えられる。

母平均：

母分散：

解答

，，





データ数を100にすると、以下のように精度が向上する。





問題１

　ある500人に対する調査で支持205人、不支持295人という結果を得た。母集団における支持の比率を信頼係数95%で推定せよ。

信頼区間は［　　　　　　　　］≦母比率≦［　　　　　　　　］

問題２

　正規分布を仮定して、以下の身長データ(cm)から母平均と母分散を信頼係数95%で推定せよ。

184, 170, 164, 176, 177, 170, 171, 159, 174, 170,

165, 170, 171, 183, 175, 169, 181, 172, 171, 164

［　　　　　　　　］≦母平均≦［　　　　　　　　］

［　　　　　　　　］≦母分散≦［　　　　　　　　］

問題３

　Samples\テキスト9.txtのデータを用いて以下の問いに答えよ。

１）ここで用いた区間推定の手法は、支出に利用できるか。［可能・不可能］

　　可能な場合は支出の母平均と母分散を信頼係数95%で推定せよ。

［　　　　　　　　］≦母平均≦［　　　　　　　　］

［　　　　　　　　］≦母分散≦［　　　　　　　　］

２）上の結果を用いて、支出の平均は50（万円）と差があるかどうか有意水準5%で判定したい。　信頼区間［内・外］なので、差があると［いえる・いえない］。

演習９【できればやってみよう】

　都市部と郊外に分けて検討事項に関する意見、月収、ある品目に対する支出を調べて演習9.txtのデータを得た。

１．地域（1：都市部 2：郊外）

２．意見（1：支持 2：不支持）

３．月収（万円）

４．支出（万円）

　以下の問いに答えよ。但し、正規性の検定は簡単のためS-W検定の結果を元にすること。

１）地域についての回答者数を求めよ。

　　都市部［　　　］人　　　郊外［　　　］人

２）意見について回答者数を求めよ。

　　支持［　　　］人　　　不支持［　　　］人

３）意見で支持は過半数といえるか、有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　　］

　　判定　過半数と［いえる・いえない］

４）地域と意見についての２次元分割表を書け。

|  |  |  |
| --- | --- | --- |
|  | 支持 | 不支持 |
| 都市部 |  |  |
| 郊外 |  |  |

５）都市部と郊外とで支持率に差があるか、有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　　］

　　判定　差があると［いえる・いえない］

６）月収と支出の平均を求めよ。

　　月収［　　　　　　］万円　　支出［　　　　　　］万円

７）月収の平均は３０万円と比べて差があるか、検定を選んで有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　　］

　　判定　差があると［いえる・いえない］

８）地域別の月収と支出の平均を求めよ。

|  |  |  |
| --- | --- | --- |
|  | 月収 | 支出 |
| 都市部 |  |  |
| 郊外 |  |  |

９）地域別の月収について都市部と郊外とで差があるか、検定を選んで有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　　］

　　判定　差があると［いえる・いえない］

10）地域別の支出について都市部と郊外とで差があるか、検定を選んで有意水準5%で判定せよ。

　　検定名［　　　　　　　　　　］　確率［　　　　　　］

　　判定　差があると［いえる・いえない］

11）月収と支出の相関係数を求めよ。また相関の有無を調べるにはどちらを使うか。

　　相関係数［　　　　　］　順位相関係数［　　　　　］

　　［相関係数・順位相関係数］で調べる。

12）支出を月収で予測する回帰式を求め、支出の変動を回帰式が説明する割合である寄与率を示せ。また、この場合は回帰式の有効性の検定は行えるか。

　　支出＝［　　　　　　］×月収＋［　　　　　　］

　　寄与率［　　　　　　］　回帰式の有効性の検定は［行える・行えない］

13）データの先頭から３人の回帰式の予測値はいくらか。

　　［　　　　　　］［　　　　　　］［　　　　　　］

14）回帰式を使って月収で支出を予測した場合、月収30万円の人の支出はいくらか。

　　予測値　支出［　　　　　］万円

15）月収30万円未満とそれ以上に分けた場合、意見の支持率に差があるといえるか、有意水準5%で判定せよ。メニュー［分析－基本統計－量から質変換］を利用する。

　　検定名［　　　　　　　　　　］　確率［　　　　　　］

　　判定　差があると［いえる・いえない］

16）意見の支持率（１の比率）の値を信頼区間95%で推定する。メニュー［分析－基本統計－区間推定－比率の推定］を利用する。

［　　　　　］≦支持率≦［　　　　　］

17）支出（正規性が必要）の母平均の値を信頼区間95%で推定する。メニュー［分析－基本統計－区間推定－平均と分散の推定］を利用する。

［　　　　　］万円≦母平均≦［　　　　　］万円

# ８．アンケート調査【第14回】

## アンケート注意事項

１）アンケートは次の順序で作る。

タイトル，あいさつ文，調査団体または代表者名，

アンケート本文，謝辞

２）何を知りたいか十分検討し、アンケート対象者や項目を選ぶ。

アンケートの対象は、全数調査か、調べたい対象の中から無作為に抽出した標本とする。但し、年齢構成などで層別に抽出する場合もある。

質問に漏れがないか十分注意する。

　　例えば意見の男女差を知りたければ、当然性別を聞いておく必要がある。

最初に区分けのための質問、続いて具体的な意見などを聞く方が答え易い。

集計のことを頭に置いて質問項目を考える。

不必要なことはできるだけ聞かずに、アンケートをコンパクトにまとめる。

３）質問は答え易い形で書く。

数字を書かせる場合と自由記述を除いては、番号を選ぶのが無難。

　　例　あなたの性別は　　　１）男　　　２）女

集計と統計処理の簡単化のため、番号選択は１つか、いくつでもかが無難。

　　例　あなたの最も大切にしていることはなんですか。以下から１つだけ選んで下さい。  
あなたの大切にしているものはなんですか。以下の該当するものすべてを選んで下さい。

明らかな場合を除いて、選択肢の中には「その他」の項目を設け、具体的な内容を書く欄を添える。

　　例　１）製造業　　２）流通業　　３）サービス業

　　　　４）その他［　　　　　　　　　　］

具体的な数字を書かせる場合は、単位を明確に。（千円はやめておくべき）

　　例　あなたの年収は　　　　　　　　　　万円

質問項目の右側に回答欄を設けると集計に便利であるが、利用しない人もいるので注意する。

回答者を絞って答えてもらう場合は、分かり易さを心掛ける。

　　例　前問で「１）はい」と答えた人のみ回答して下さい。その他の人は設問５へ進んで下さい。

４）その他

予め集計用のフォームを考えておく。（大規模でなければExcelは有力）

あらかじめ少数の人で試し、集計までをシミュレーションしておく。

回収後、回答用紙には必ず整理番号を振っておく。

## 学生生活アンケート調査

この度統計学の授業において、アンケートの作成法とその集計方法を学ぶために仮想的なアンケート調査を実施することになりました。個人のプライバシー等につきましては十分な注意を払うことはもちろんですが、このアンケートをその他の目的に使用することはありません。どうかご協力をお願い致します。

福山××大学　福山太郎

質問１　あなたの性別は？

１）男性 ２）女性

質問２　あなたは自宅通学ですか？

１）自宅通学 ２）自宅通学でない

質問３　あなたの自由に使えるお金（生活費を除く）は１ヶ月におよそいくらですか？

［　　　　　　円］

質問４　あなたはアルバイトをしていますか？

１）している ２）していない

前問で**１）している**と答えた人だけ回答して下さい。その他の人は質問７へ進んで下さい。

質問５　どれ位の頻度でアルバイトをしていますか？１つ選んで下さい。

１）週５日以上　　２）週３，４日　　３）週１，２日

４）長期休業時のみ　　５）その他［　　　　　　　　　　　　　　　］

　　質問６　あなたのアルバイトの収入は１ヶ月におよそいくらですか？不定期にやっている人は、１ヶ月にならしてお答え下さい。

［　　　　　　円］

質問７　あなたの現在の悩みに当てはまるものがあればいくつでも選択して下さい。

１）特にない　　２）勉学上の問題　　３）金銭問題　　４）異性問題

５）健康上の問題　　６）就職・進路の問題

７）その他［　　　　　　　　　　　　　　　　　　　　　　　　　　　］

ご協力有難うございました。

## 学生生活アンケート調査報告書

福山××大学　福山太郎

　福山××大学では20XX年7月28日に、本学統計学の授業で受講生53名を対象に「学生生活アンケート調査」を対面して記述させる方式で実施した。調査結果の回収数は42で回収率は79.2%であった。

男女別にみると男34名、女8名であり、自宅通学かどうかをみると自宅通学26名、自宅通学以外は16名であった。アルバイトをしている学生は31名、していない学生は11名で、アルバイトをしている割合は、73.8%であった。アルバイトをしているかどうか通学区分別に見ると、表1のようになった。

表1 通学区分によるアルバイト状況

|  |  |  |
| --- | --- | --- |
|  | している | していない |
| 自宅 | 20 | 6 |
| 自宅外 | 11 | 5 |

これから通学区分によるアルバイト状況の有意差は見られなかった。また、アルバイトの頻度は、週5回以上11名、3～4回18名1～2回2名であった。

　自由に使える1ヶ月の金額は、平均3.96万円、標準偏差2.76万円であり、そのヒストグラムを描くと、図1のようになった。

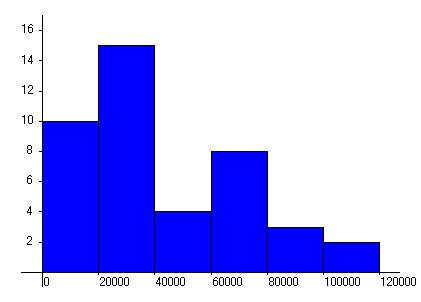


図1　自由に使える金額

性別、通学別、アルバイト状況別の自由に使える金額の平均は表2のようになった。

表2 各分類別平均（万円）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 性別 | | 通学 | | アルバイト | |
| 男 | 女 | 自宅 | 自宅外 | している | していない |
| 4.25 | 2.71 | 3.68 | 4.41 | 4.70 | 1.87 |

これらについて差を調べたところ（［　　　　　　］検定）、アルバイトをしているかどうかで有意な差が見られたが（*p<*［0.05・0.01・0.001］）、その他については有意な差は見られなかった。

　アルバイト収入の平均は6.65万円、標準偏差は2.35万円であった。また、自由に使える金額とアルバイト収入の関係は、図2で与えられ、アルバイト収入がないものを除いた順位相関係数は0.617であった。このことからアルバイト収入と自由に使える金額には相関関係があると思われる。

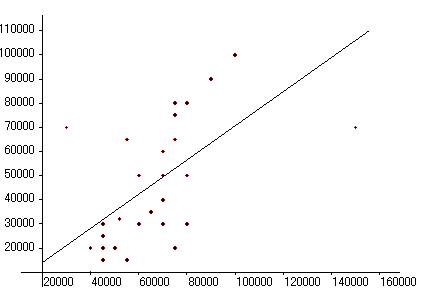


図2　アルバイト収入（横軸）と使える金額（縦軸）の相関

## アンケート報告書注意事項

１）タイトル、調査団体名または代表者名及び住所等（ここまで表紙でもよい）を最初に示す。

２）アンケートの実施時期と実施方法、対象数と回収数・回収率を明記する。

３）アンケート集計結果は以下の点に注意する。

単純集計から始めて、次にクロス集計をする。

図表には番号とタイトルを付け（通し番号または章ごと）、文中で指定して説明を加える。 例　　図1に設問３のヒストグラムを示す。

図表番号とタイトルを付ける位置として、表は上側、図は下側が多い。

必要があれば、その他を選んだ場合の内容を紹介してもよい。

質問用紙を最後に掲載するのもよい。

４）集計・検定結果の表示

集計値の桁数は、平均・標準偏差等でデータ桁数より1桁か2桁程度多く表示する。

　　例：171, 173, 174, ･･･ → 平均172.7

検定の際、検定統計量の値や自由度などはあまり書かれることがないが、検定手法の名前は書く場合もある（書かない場合もある）。

検定確率値についてはあまり具体的な数値を書くことはなくp<0.001, p<0.01, p<0.05, n.s.（\*\*\*, \*\*, \*, n.s. にしてもよい）のどれかの書き方にする。