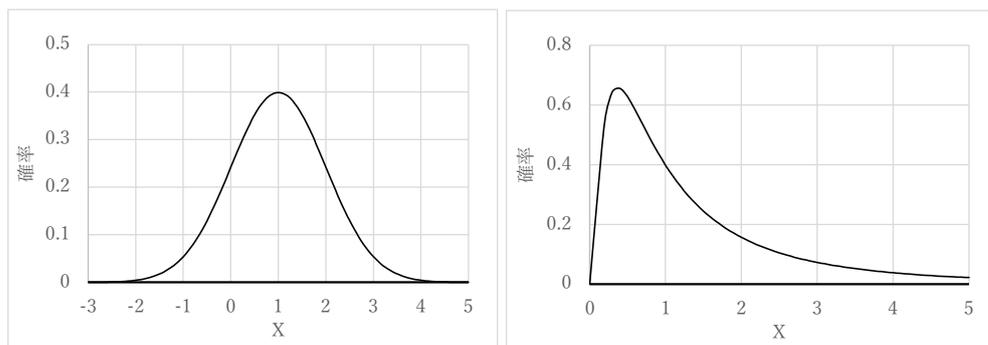


要約統計量～平均値、分散、標準偏差、標準誤差、信頼区間。

普段得られるデータは常に「ばらつき」を含み、特に臨床データは患者内や患者間での誤差が大きい場合が多く、ばらつきを無視して正確な解釈はできない。ばらつきの形状は「分布」として整理され「ヒストグラム（頻度分布）」でグラフ化できる。医薬領域で扱うデータの分布は「正規分布」あるいは「対数正規分布」の場合が多く、正規分布は図左に示すように平均値を中心とした左右対称の形で統計解析でも扱いやすい。対数正規分布は正規分布に比べて右に裾を引いた分布であり



裾を引いた分布であり

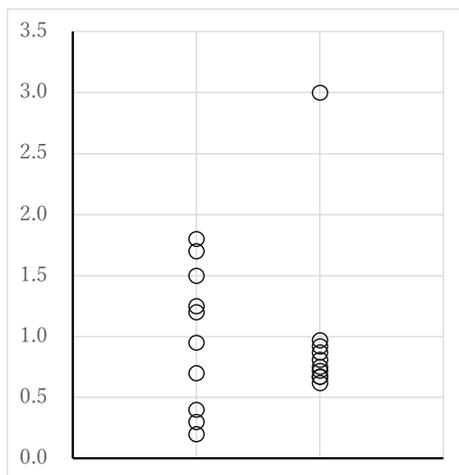
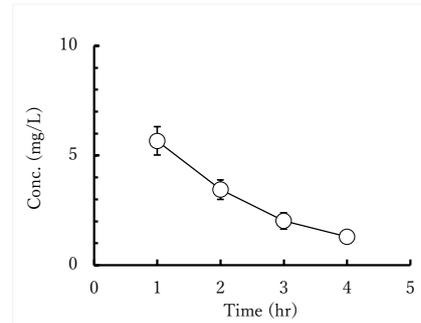
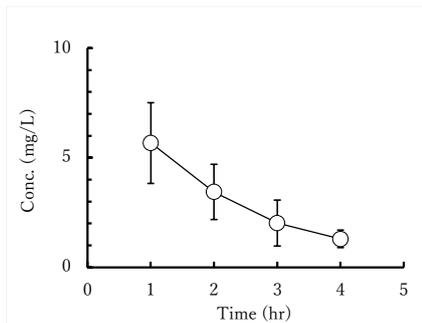
(図右)、血中濃度や臨床検査値が対数正規分布で説明されることが多い。

データを評価する場合、実測値の羅列やヒストグラムだけではデータの特徴をひとことで表すことが難しいため「要約統計量」が用いられる。要約統計量には、平均値や中央値といったデータの中心を示す「代表値」と、分散などデータのばらつきの程度を示す「散布度」がある。

- ・ 平均値：
 - ▶ 算術平均値（すべてのデータの和を個数で割った値で正規分布の平均値に相当する）
 - ▶ 幾何平均値（すべてのデータを掛け合わせてその数 N の N 乗根、あるいはすべてのデータの対数値を用いて得られる算術平均値）。
- ・ 中央値：データを大小順に並べたときの真ん中の値。偶数個ある場合には真ん中 2 個の算術平均値。
- ・ 最頻値：ヒストグラムで範囲ごとの分布を書いたとき、もっとも個数の多い「範囲」。
- ・ 分散：各データの平均値からのばらつきの二乗を加え、個数で割った値。このとき個数 N で割る場合と $N-1$ で割る場合があり、前者を標本分散、後者を不偏分散という。通常得られるデータでは後者の不偏分散を用いるほうが真の値に近いとされる。
- ・ 標準偏差（Standard Deviation, SD）：分散の平方根。通常 $N-1$ で割った分散から算出する。
- ・ 標準誤差（Standard Error, SE）：SD をさらに N の平方根で割った値。統計量のばらつきを示すが通常は平均値のばらつきの程度と理解する。
- ・ その他、分布から大きく離れた数値である、外れ値などがある。

実際のデータがどのような分布を示すかは、多くの量のデータを得てヒストグラムを描かないとわからないが、通常は十分多くのデータが得られるとは限らないので、従来の経験から例えばある検査値は「もしたくさんのデータがあれば正規分布になるであろう」と仮定して話を進めることも多い。このような場合「データは正規分布に従うと仮定する」といった表現が使われる。

データをグラフ化するとき、SDあるいはSEの幅を「バー」で示す場合がある。SD（左図）、あるいはSE（右図）どちらも用いている例があるが、グラフの説明にどちらの数値を用いたのか明記すべきである。



実際のデータには外れ値が含まれることがあり、外れ値を正しく解釈するためには実際のデータをそのままプロットして眺めてみるのが大切である。図はある2セットのデータの実測値をプロットしたもので（どちらも10個のデータ）、右側には外れ値が含まれている。実際これらのデータの平均値はどちらも1.0で、SDは左が0.58、右が0.72であり要約統計量だけでは外れ値の有無を見抜くことはできない。

要約統計量に加えて「信頼区間」を算出し、統計的な考察に用いることがある。特に正規分布を考えるとき（対数正規分布など正規分布以外の分布では定義が異なるので注意）、データ（実測）の95%信頼区間は平均値 $\pm 1.96 \times SD$ 、平均値の95%信頼区間は平均値 $\pm 1.96 \times SE$ 、でありSDとSEとを使い分ける。例えば上図データの場合、平均値の95%信頼区間は（定義に従いSEを算出して求めると）
 $1.0 \pm 1.96 \times 0.58 / \sqrt{10} = -0.64$ （下限値）, 1.36 （上限値）と得られる。

各種統計的検定では平均値の差を検定することが多く、それに対応して平均値の（すなわちSEを用いて計算した）95%信頼区間が用いられる。1.96という値は標準正規分布（平均値が0、分散が1の正規分布）の両側95%区間を示す値であり頻繁に用いられる。

統計的検定で2群間のそれぞれの平均値の差が有意に異なるかどうかを調べることもある。t検定などの検定法が存在するが、2群の平均値の差の95%信頼区間を算出して（計算式は省略、成書参照）、その区間が0を挟まない場合には有意な差があると判断することもできる。比率についてオッズ比やハザード比では比の95%信頼区間を算出して、その区間が1（0ではなく）を挟まない場合には有意な因果関係があると判断する。