

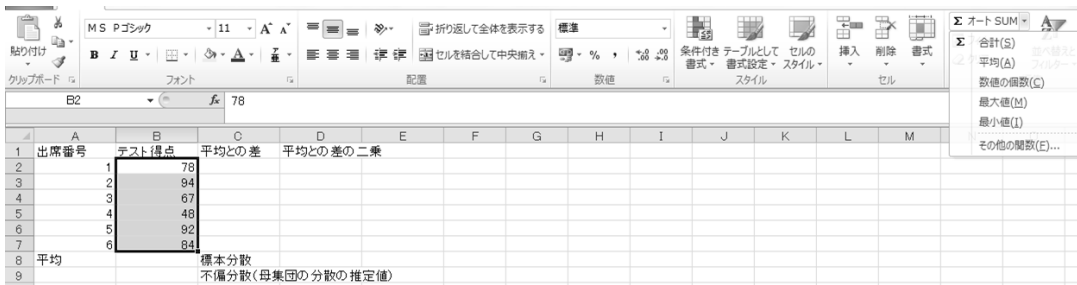
Excel を用いた基本統計量の算出

1. 平均値

■ 平均値 = $\frac{\text{(テスト得点の総和)}}{\text{(データの個数)}}$

- 「Σオート SUM」 → 「平均」
または、「=AVERAGE(セルの範囲)」

■ 結果を少数第何位まで表示するかは、ツールバー $\pm 0.00 \rightarrow 0.0$ で調節。



2. 分散

2.1 分散

■ 分散：データのバラツキの度合いを示す値

分散 = $\frac{\text{(各データー平均値)}^2 \text{の総和}}{\text{データの個数}}$

■ 分散の値が大きいほどバラツキが大きい

ex) C 列：「平均との差(データー平均値)」 (ex. =B2-B\$8)

D 列：「平均との差(データー平均値)」を二乗した値 (ex. =C2^2)

(二乗：「=二乗する数値が入力されているセル^2」)

■ 標本分散= $\frac{\text{平均との差の二乗}}{\text{データの個数(6)}}$

= $\frac{\text{SUM(D2:D7)}}{6}$

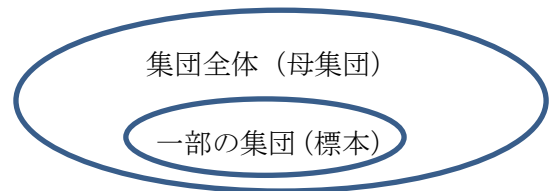
■ 標本分散と母分散

- 母分散：測定対象全体の分散
- 標本分散：母集団の中の一部の集団（標本）の分散

ex)1000 人の生徒の能力を測定したいが、全生徒にテストを実施するのが難しい場合。

偏りのない一部の生徒のテスト結果から集団全体の能力を測定する。

一部の生徒の分散 = 「標本分散」 集団全体の推定される分散 = 「母分散」



2.2 分散の自動算出

- 母分散 = VAR.S(セルの範囲)
- 標本分散 = VAR.P(セルの範囲)

■ 母分散 = $\frac{(\text{各データ} - \text{平均値})^2 \text{の総和}}{\text{データの個数} - 1}$

- 母分散は、標本から推定される値であるため、「データの個数-1」で「(データ-平均値)の二乗」を割ることによって分散を大きく見積もっている。

3. 標準偏差

- 標準偏差：データが平均からどの程度ずれているかを表す統計量。
数値が大きいほど、バラツキが大きいデータであることを示す。
- 標準偏差の計算式: 母分散に基づく場合: =STDEV.S
標本分散に基づく場合: =STDEV.P
- 標準偏差 = $\sqrt{\text{分散}}$

4. ヒストグラムの作成

- 「データ」 → 「分析」 の 「データ分析」 → 「ヒストグラム」 を選択、[OK]
- 「入力範囲」 、 「データ区間」 ボックスをそれぞれ選択。
- 「出力オプション」 の下の 「新規ワークシート」 と、「グラフ作成」 をチェック。 → [OK]



4. 別シートに、ヒストグラム テーブルと埋め込みのグラフが含まれる新しいブックが生成される。

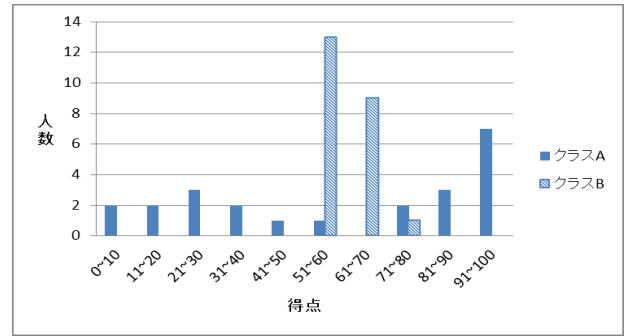
クラス A 平均 = (61.435), 標準偏差 = (34.360)

クラス B 平均 = (61.270), 標準偏差 = (6.214)

よってクラス(B)は、クラス(A)よりばらつきが少ない。

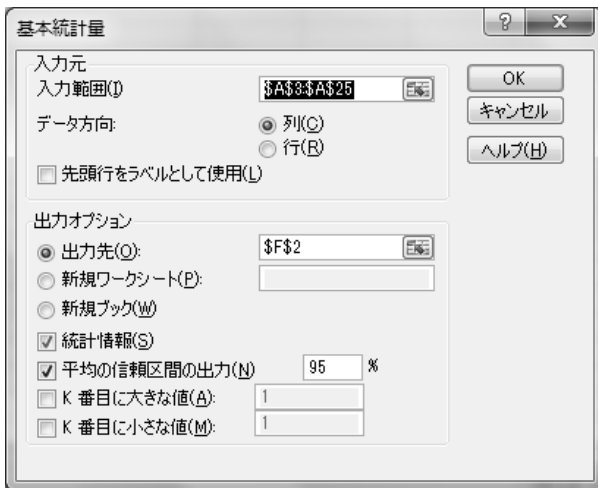
■ グラフの作成

「挿入」→「グラフ」の「縦棒」(「2-D 縦棒」)
 「グラフのレイアウト」でグラフの種類を選択。



5. 基本統計量の一括出力

- 「データ」→「分析」の「データ分析」→「基本統計量」を選択、[OK]
- 「入力範囲」: データの範囲を入力
 出力オプションの該当箇所(下記参照)をチェックし、[OK]。



基本統計量の結果

クラスA		クラスB	
平均	61.43478	平均	61
標準誤差	7.325527	標準誤差	1.324742
中央値(メジアン)	80	中央値(メジアン)	60
最頻値(モード)	100	最頻値(モード)	63
標準偏差	35.13199	標準偏差	6.353238
分散	1234.257	分散	40.36364
尖度	-1.58832	尖度	0.339749
歪度	-0.31296	歪度	0.590554
範囲	100	範囲	26
最小	0	最小	51
最大	100	最大	77
合計	1413	合計	1403
標本数	23	標本数	23
信頼区間(95.0%)	15.19221	信頼区間(95.0%)	2.747346

6. 区間推定と信頼区間

- 一部の標本から、母平均を推定しようとするとき、区間推定を行う。
- 区間推定: 母平均がとる値に一定の範囲を想定し、その範囲内に入る確率を統計的に求めること。
 ex) 一部の中学3年生(標本)の平均から推定される全国の中学3年生(母集団)の平均値は、「A値からB値の範囲内に~%の確率で入る」。
 → 区間推定では推定した範囲に入る確率を用いる。
- 信頼区間(confidence interval): あらかじめ決めた確率(信頼性係数)に基づいて算出された母平均がとる範囲。通常信頼係数は、95%か99%を用いる。
- 信頼区間の解釈: 母平均は95%の確率で、「平均±値」の範囲内にある。
 ex) クラスBの母平均は95%の確率で $61 \pm 2.75 (58.25 \sim 63.75)$ の範囲内にある。

<参考文献>

寺内正典(編著)(2012).『英語教育学の実践的研究法入門—Excelで学ぶ統計処理』研究社

<補足>

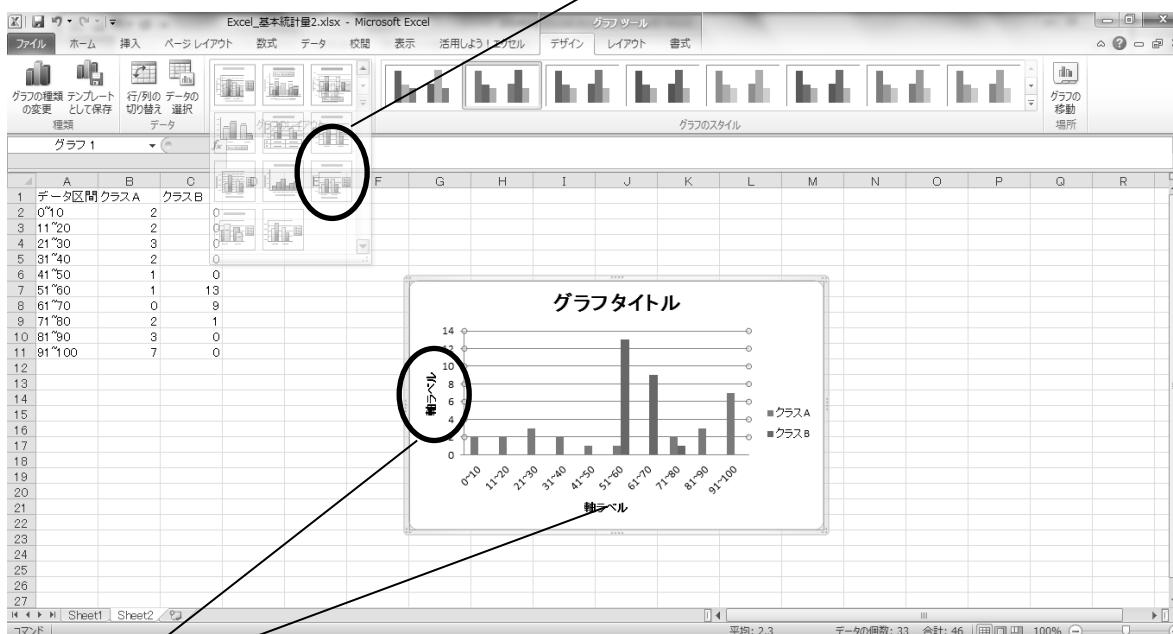
■ 「分析ツール」の表示方法

(Excel2010は、アドインを行わないと、「分析ツール」に表示されないため、最初に使用する場合は、設定が必要)

1. 「ファイル」→「オプション」→「Excelのオプション」
2. 「Excelのオプション」から「アドイン」をクリック。
3. 「管理(A)」欄で「Excelアドイン」を選択→「設定」
4. 「有効なアドイン」内の「分析ツール」をチェック。
5. 「データ」タブの右側にデータ分析が表示される。

■ グラフの作成

グラフの作成後、「グラフのレイアウト」で、縦軸、横軸両方の表示があるグラフを選ぶ。



軸ラベルは、何を示すか明記する。例えば、人数(%)など。