

どのように分析統計から標準規格が導き出されるのでしょうか

こちらは、英文記事「[How analytical statistics lead to standard specifications](#)」（2020年1月9日付）の和訳です。

分析科学の中心原則の1つは、実施された試験が検証済みであり、別の科学者が同じ手法・分析機器を用いて再現できることです。今回の Gard Insight では、Brookes Bell のモス博士とシェアド博士（Dr Tim Moss, Director of Science, Asia and Dr Daniel Sheard, Principal Scientist）に、貨物やバンカー燃料の規格における許容偏差の算出方法を説明いただきました。

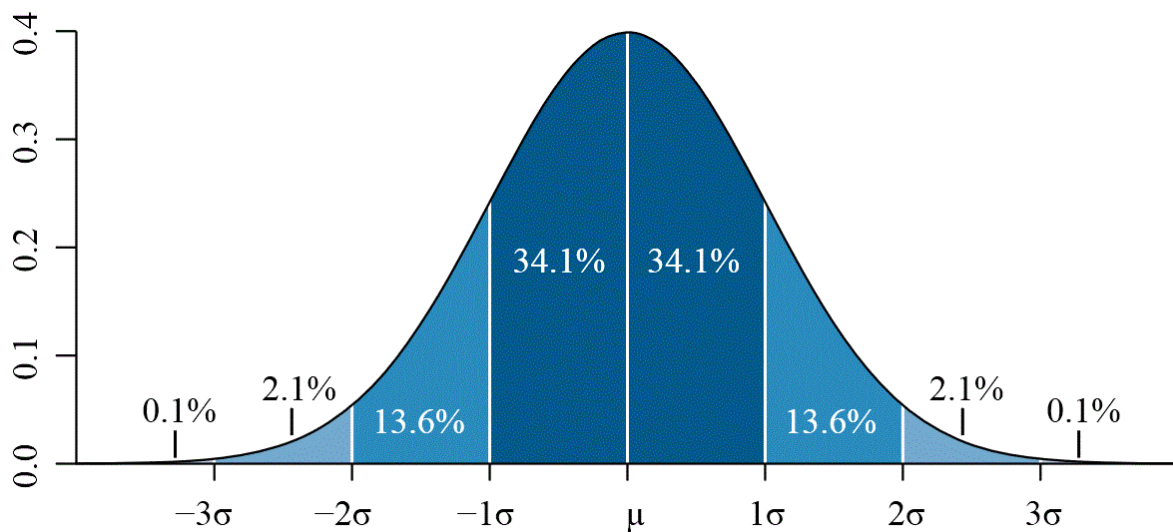


ASTM（米国試験材料協会）、ISO（国際標準化機構）、AOAC（公認分析化学者協会）などの国際標準化団体は、これまでに度々、所定のパラメーターを調査するための精密な方法を確立しようとしてきました。このことは、国際商品取引において極めて重要です。これによって、（別の国に所在することが多い）買い手と売り手が、商品の品質を確認することができ、ラボが違っていても採用されている試験は許容可能な公正なものであることの確信が持てるからです。

実際、ラボ間では物質分析において制御不能な変数が数多く存在しています。同じラボ内でさえも試験結果に多少の差が出ることは避けられません。

分析試験の設計者はこのような内在する差を十分認識しており、それを精度と呼びます。統計学者は、内在する差を、試験から得られる結果の母集団としてモデル化します。個々の分析結果は、モデル化された母集団の1要素となります。

これらの方法や統計上の標準偏差の確立に多くのラボが関与することになるでしょう。標準偏差は、「分散した」母集団のデータがどの程度平均値の周りにばらついているかを数値化したものです。標準偏差の値を求めるには、分散の平方根を取ります。分散の計算は簡単で、「平均からの差」の2乗の平均を計算して求めます。2乗を行うのは単に負の数を正にするためで、そうしなければ負の数（平均未満）と正の数（平均よりも大きい）が互いを相殺してしまいます。



上図（作成者不明）は、[CC BY](#)の許可を得て使用しています

試験結果の母集団の大半は、上図のような特徴的な釣鐘型の形状をしています。母集団の 68.2% が平均値の両側の 1 標準偏差内にあり、95.6% が 2 標準偏差内にあります。このようなデータが、多数のラボから数多く取得されると、当該試験方法の標準誤差が計算できるようになります。試験方法の精度を測定するための測定数が増えるほど、精度が上がり、結果として標準誤差が適切な値となります。標準誤差は、測定数が増えるにつれて減少します。標準誤差を求めるには、標準偏差を標本データ数の平方根で割ります。

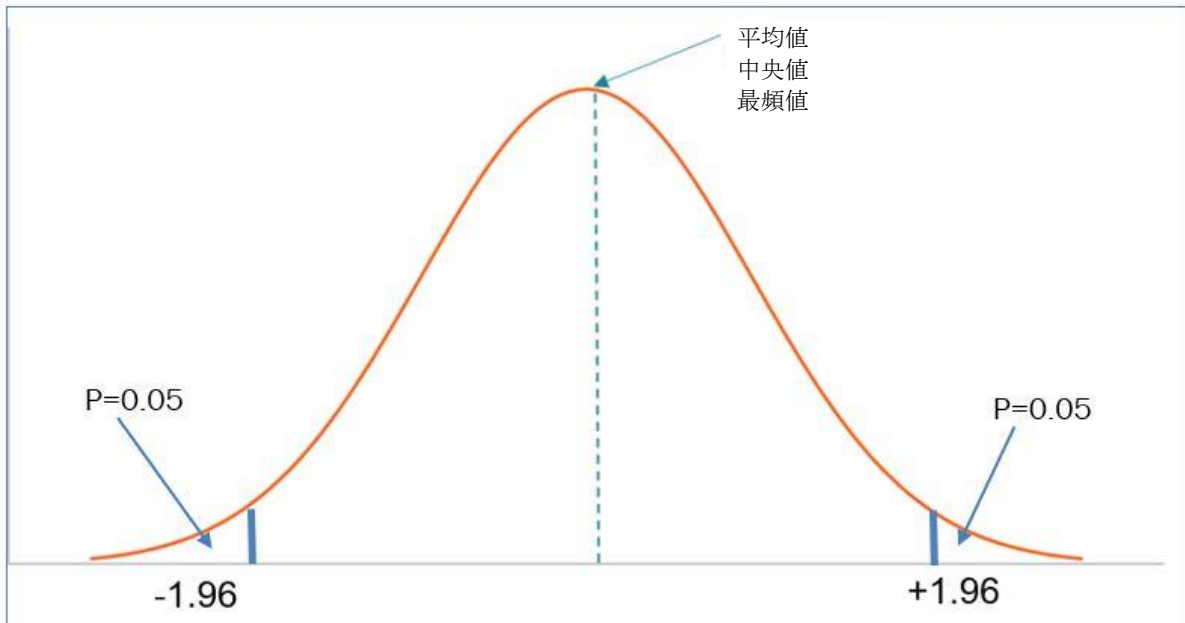
室間再現精度（reproducibility）と繰返し精度（repeatability）

国際機関が公表する分析方法の大半が、標準誤差の 2 種類の測定結果について規定することになるでしょう。これら 2 つは共に統計学的メソッドに準拠しています。繰返し精度（ r と略します）は、同じ技術者が同一の分析機器で試験した同一標本の 2 個の試験結果間の差を指します。室間再現精度（ R と略します）も、2 個の試験結果間の差を指しますが、別の技術者が別のラボで試験したときの、2 個の試験結果間の差を指します。ISO 5725 では、標準測定方法の繰返し精度と室間再現精度を測定するための基本的方法が紹介されています。

これら 2 つは共に「95% の確率」という観点から、2 個の試験結果間で推定される最大差を表します。たとえば、別々のラボで得られた 2 個の試験結果間の差が R 未満になる確率は 95% となります。予想されるとおり、別のラボ／オペレータによる試験と比べ、同一ラボ／オペレータによる同一試験方法で生じる差の範囲は大幅に小さく、実際、繰返し精度は室間再現精度よりもかなり小さいことがほとんどです。

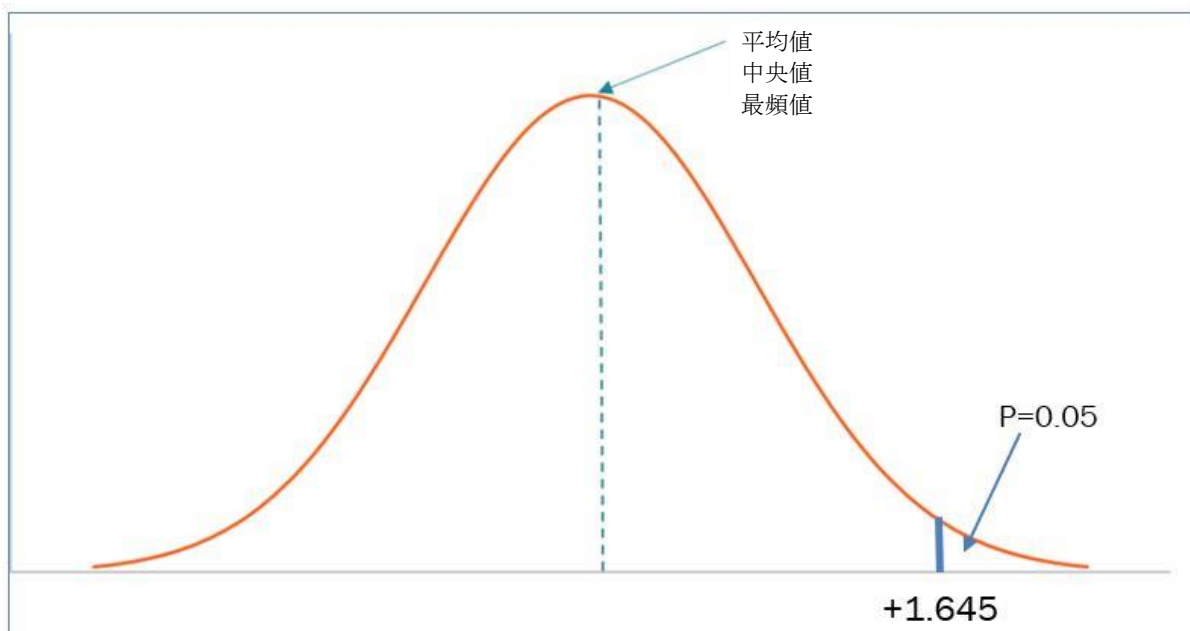
統計学者は、95% 信頼限界によく注目します。これは、当該統計データによって結論が間違っている可能性が 5% しかないことが示されていることを意味しています。分布曲線全体を考慮した場合の

95%信頼限界では、結果の95%が平均値の ± 1.96 以内に収まります。これは以下のとおり両側検定です。



上図（作成者不明）は [CC BY-SA](#) の許可を得て使用しています

大抵の場合（後述します）、分布曲線の最低部分または最高部分だけに着目することになります。この場合、試験結果の95%が平均値 + 標準偏差 1.645 内に収まっています。これは以下のとおり片側検定です。



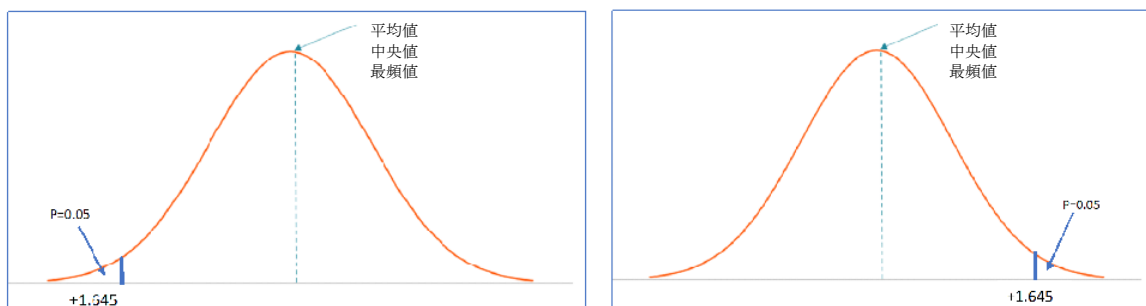
上図（作成者不明）は [CC BY-SA](#) の許可を得て使用しています

分析試験と規格値の適合

商品を出荷する場合、通常、商品のサンプリングが行われ、代表標本が作成されます。代表標本は、各パラメーターに関する標準方法に基づき、その地域のラボで試験されます。こうして得られた試験結果は貨物の品質証明書に盛り込まれ、多くの場合、売買契約の基礎となります。

荷揚げ時に、荷送人は再び代表標本を採取し、品質に問題がないことを確認します。荷揚げ時のサンプリングが重要であることは間違いありませんが、「貨物に変化がなく、各代表標本が品質証明書において同一である」ことが前提とされています。貨物が規格値内に十分収まっている場合は、上記方法に内在する精度は重要ではありませんが、貨物が合格品質のボーダーライン上にある場合は、各パラメーターの精度の値が重要になります。2つの港間は遠く離れている場合があり、異なるラボが使用される可能性が高いことから、**R**（室間再現精度）の値が関係してくることになるでしょう。

規格値は、ほとんどの場合、指定された値の「最大」または「最小」の一定のパラメーターとして表現されることから、片側検定とされる傾向があります。対象のパラメーターよりも上または下にあるデータだけが注目されます。



上図（作成者不明）は [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) の許可を得て使用しています

既定の試験の測定を 1 回行うと、規格値と比較するための値が取得できます。

様々なパラメーターに関する多くの特異的試験では、試験レベルは最大または最小範囲内に収まる必要があり、1 個の試験で僅かに規格値から外れただけでも、不適合と見なされます。例外もありますが、本稿の残りの部分では、燃料に関する事例に焦点を当てます。

燃料規格と分析結果の判定

2020 年に入り、特に関心を集める分野の 1 つは、燃料の品質と特に船舶燃料の硫黄分規制ですが、硫黄分については定期的な燃料試験の一環で測定されることは間違いありません。

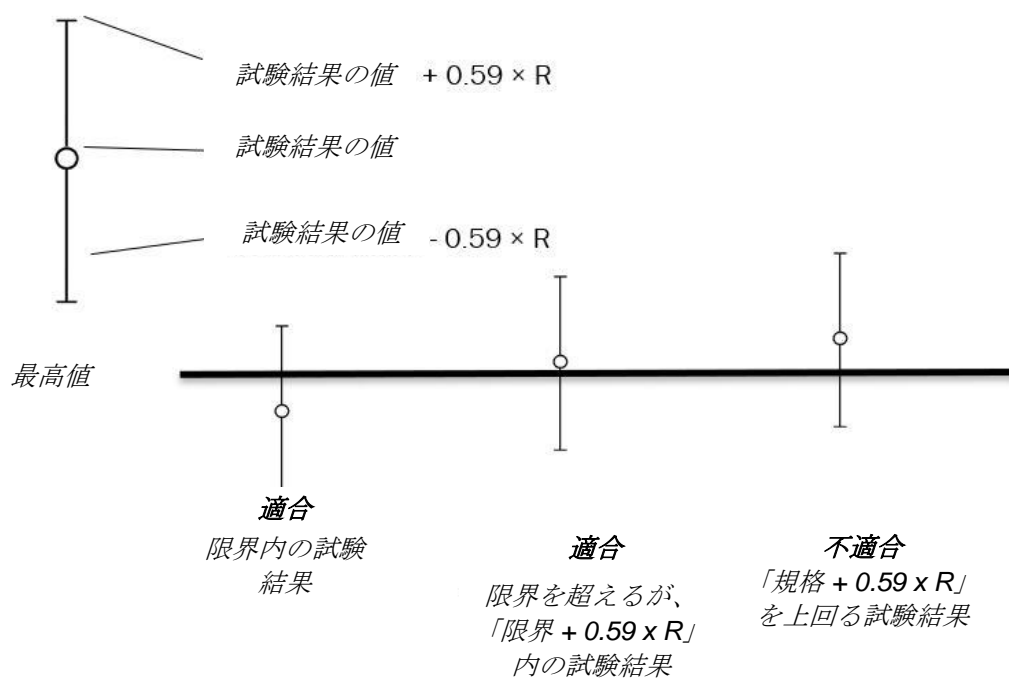
燃料分析、特に ISO 4259「石油および関連製品－測定方法の精度および結果」や ASTM D3244 などの関連規格には追加規定があり、試験結果がボーダーライン上にある場合に特定の試験の室間再現精度にある程度の許容差を与えています。ISO 4259 には次の 4 レベルの判定が規定されています。

第 1 レベルの判定

試験結果が規格値を上回る場合、ISO 4259 は係数 $0.59 \times R$ を使用することを認めています。パラメーターに上限がある場合、試験結果が規格限界を超えていても、規格限界 $+ 0.59 \times R$ を下回る場合は当該燃料の購入者／受取人は、当該燃料は規格に適合していると考えべきです。

したがって、たとえば、公称硫黄含有量が 0.5% の燃料は ISO 8754:2003 を基に試験され、許容限度は $0.5\% + 0.59 \times R (= 0.53\%)$ となります。

試験結果が規格限界 $+ 0.59 \times R$ を上回る場合（上記例では 0.53% より大きい場合）、規格に適合していないと見なされます。



第 2 レベルの判定

上記判定は必ずしも判定プロセスの最終段階ではありません。規格 ISO 4259（2017）第 2 部では、「試験方法に関する精密データの解釈と適用」が規定されています。

同規定では、受取人が、規格限界が満たされておらず、供給された当該燃料は規格不適合であるとして苦情を言うことができ、サプライヤーに、サプライヤーの保有サンプルを試験することを義務付けています。

「4.3.1 結果の許容性

2つのラボでそれぞれ1つの結果が得られ、その差がR以下である場合、その2つの結果は許容できると見なされ、平均値Xの計算に使用される。単一の結果を個別に使用するのではなく、平均値Xが、試験される特性の推定値として使用されるものとする。

…

もし、2つの試験結果の差がRを上回る場合、両方の結果が疑わしいと見なされるものとする。その場合、各ラボは、少なくとも3個以上の他の許容可能な結果を取得する義務がある」

第3レベルの判定

このように第3段階では、ISO 4259に従い、各ラボが繰返し精度の基準を満たす、パラメーターの3個の測定結果を提示する義務があり（各ラボでは既に数値測定のために測定結果を提示済みの可能性があります）、ISO 4259には以下のとおり、推定平均の測定に関する式が規定されています。

この場合、各ラボの基準を満たすすべての試験結果の平均値の差は、[Formula \(10\)](#)で示されたように、Rに代わる新しい値R₂を使用して、適合性が判定されるものとします。

$$R_2 = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2k_1} - \frac{1}{2k_2} \right)} \quad (10)$$

ここで、

Rは当該手法の空間再現精度。

rは当該手法の繰返し精度。

k₁は、1番目のラボの試験結果数。

k₂は、2番目のラボの試験結果数。

平均値の差がR₂以下である場合、これらの平均値は許容でき、それらの全平均値は試験される特性の推定値と見なされるものとします。平均値の差がR₂を上回り、試験される特性の仕様適合性に疑いがある場合は、[第7節](#)で指定された手順が適用されます。

第4レベルの判定

第7節の規定は細部にわたるため、現段階では直接言及しませんが、簡単に言いますと、第7節には、異なるラボ同士で明らかに異なる結果を取得した場合に、連絡を取り合って試験方法、分析機器、設定などについて協議を進める方法などが規定されています。第4レベルの精査に達することは非常にまれですが、メンバーはこれらの標準手順を通じて、自分たちにも適用の可能性があることに留意する必要があります。

バンカー燃料のサンプリングと試験の詳細については、Gard Insight [「使用中の VLSFO \(0.5%硫黄分の船舶燃料油\) が規格に合致し、MARPOL 条約の規制に準拠していることを 95%確信していますか？」](#) でご覧いただけます。

結論

通常の実験では、試験結果を判定するための詳細な統計計算は必要ありません。しかし、一部の業界、特に燃料業界では、ブレンドは非常に精密なレベルで行われており、試験の統計精度を満たし、確固たる判定を行わなければならないボーダーライン上の事例が必ず発生します。ISO 4259（および関連規格）には、統計計算を行うための体系的な方法が規定されています。

本情報は一般的な情報提供のみを目的としています。発行時において提供する情報の正確性および品質の保証には細心の注意を払っていますが、Gard は本情報に依拠することによって生じるいかなる種類の損失または損害に対して一切の責任を負いません。

本情報は日本のメンバー、加入者およびその他の利害関係者に対するサービスの一環として、ガードジャパン株式会社により英文から和文に翻訳されています。翻訳の正確性については十分な注意をしておりますが、翻訳された和文は参考上のものであり、すべての点において原文である英文の完全な翻訳であることを証するものではありません。したがって、ガードジャパン株式会社は、原文と内容の不一致については、一切責任を負いません。翻訳文についてご不明な点などありましたらガードジャパン株式会社までご連絡ください。