

データサイエンス実験 A

シリコン結晶精製工程における解析と改善の提案

中央大学理工学部ビジネスデータサイエンス学科

23D7104001I 高木悠人

1. 目的

本実験では半導体メーカー C 社におけるシリコン単結晶の直径を大きくするにはどうしたら良いかを解析結果とともに検討し、データの可視化とモデルの当てはめの役割の違いについて考察することを目的とする。当該メーカーではシリコンの多結晶を加熱炉で引き上げ単結晶を精製している。シリコンの単結晶は直径が大きいほど良品とされるものの現在、基準である 105mm を下回る製品が多く製造されてしまっている。そこで、生産工程における作業報告書を調査し改善策を模索することとする。

2. 調査方法

まずは解析に用いるデータについて示す。当該データは 4 月 1 日から 5 月 25 日までの全製品(計 160 本)についての作業報告書に掲載されているもの表形式のデータを利用した。行はそれぞれの製品を意味し、列は生産・提供条件を意味する。以下に生産・提供条件についての表を示す。

表 1. データの変数名とその説明

変数	生産・提供条件	備考
X1	日付	4月1日から5月25日
X2	午前・午後の別	午前:1/午後:2とする
X3	引き上げ回次	午前・午後の何回目か
X4	シード温度	シーディング中の炉内温度(単位°C)
X5	シード圧	シーディング中の炉圧力(単位PSI)
X6	シード後温度	肩ができあがるときの温度(単位°C)
X7	シード後圧	肩ができあがるときの圧力(単位PSI)
X8	原料重量	投入した多結晶の重量(単位10g)
X9	原料メーカー	A社製:1/ B社製:2/ 自社製:3 とする
X10	ヒーター使用回数	利用したヒーターで何回引き上げを行ったか
X11	シード回数	うまく種付けをできたまでに行ったシーディング回数
X12	シード時間	シーディングから引き上げ始めるまでの時間(分)
X13	シード濃度	混入させる不純物濃度(単位はppm)
X14	加熱炉の区分	1号機:1/ 2号機:2 とする
Y	シリコン単結晶の直径	単位はmm

※作業報告書をもとに作成した。

次に、シリコンの生産手順を以下に示す。

- ① 加熱炉の中の釜に原料であるシリコンの多結晶を入れ、単結晶の種(シード)をセットする。
- ② 加熱炉を密閉し、高圧のもとで多結晶を溶かす。溶解後シードを釜の中に入れシーディング(種付け)を行う。
- ③ シーディングが終了すると徐々に引き上げはじめ、単結晶を成長させる。
- ④ さらに引き上げ続け、原料がなくなると完了する。

上記の手順で得られたデータをもとに、ヒストグラム,時系列プロット,散布図によるデータの可視化,特性要因図による要因特定へのアプローチ,回帰分析によるモデルの当てはめを行い解析する。

3. 解析と結果

(1) 品質特性に関する全体のヒストグラム

本項では品質特性についてのヒストグラムを作成し、どのようなくせ,特徴があるのかをまとめる。以下に品質特性の全体のヒストグラムを示す。

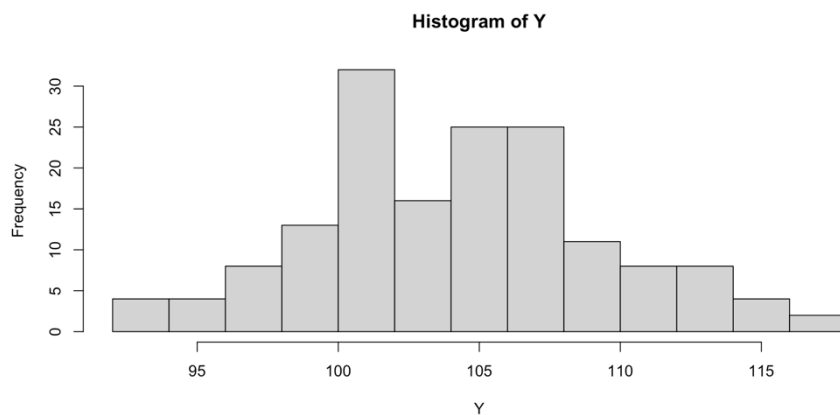


図 1. 品質特性の全体のヒストグラム

※上記の図は R を用いて作成した。単位は横軸,縦軸それぞれ mm,回とする。

上記より以下のことがわかる。

1. 102,107 を頂点とする二つ山型となっている

(2) 層別ヒストグラム

(ア) 午前・午後の層別ヒストグラム

まずは午前・午後による層別ヒストグラムを以下のように作成した。

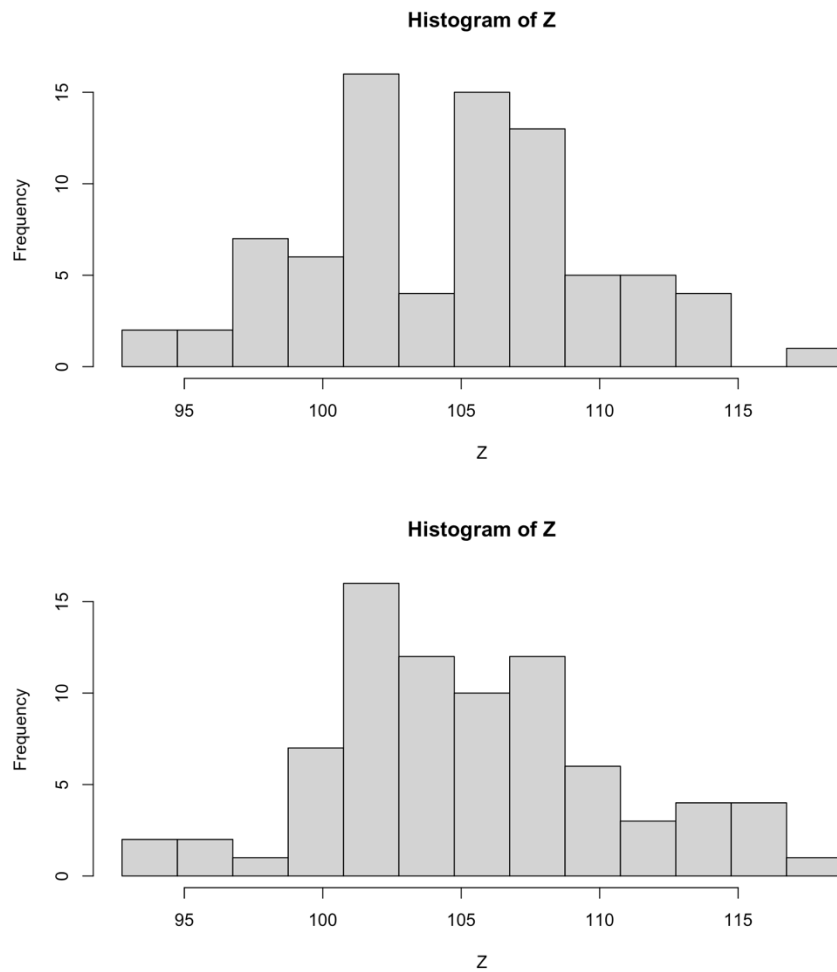


図 2. 午前・午後の層別ヒストグラム

※上を午前中のデータで下を午後のデータとした。

上記の図より以下のことがわかる。

1. 午前の層別ヒストグラムでは 102,107 付近に頂点を持つ二つ山型
2. 午後の層別ヒストグラムでは 102,107,116 を頂点とする三つ山型
102 から左右に行くと段々に減少するため一般型とも言える

(イ) 原料メーカー別層別ヒストグラム

次に原料メーカー(1:A 社製,2:B 社製,3:自社製)について層別ヒストグラムを作成した結果以下のようになった。

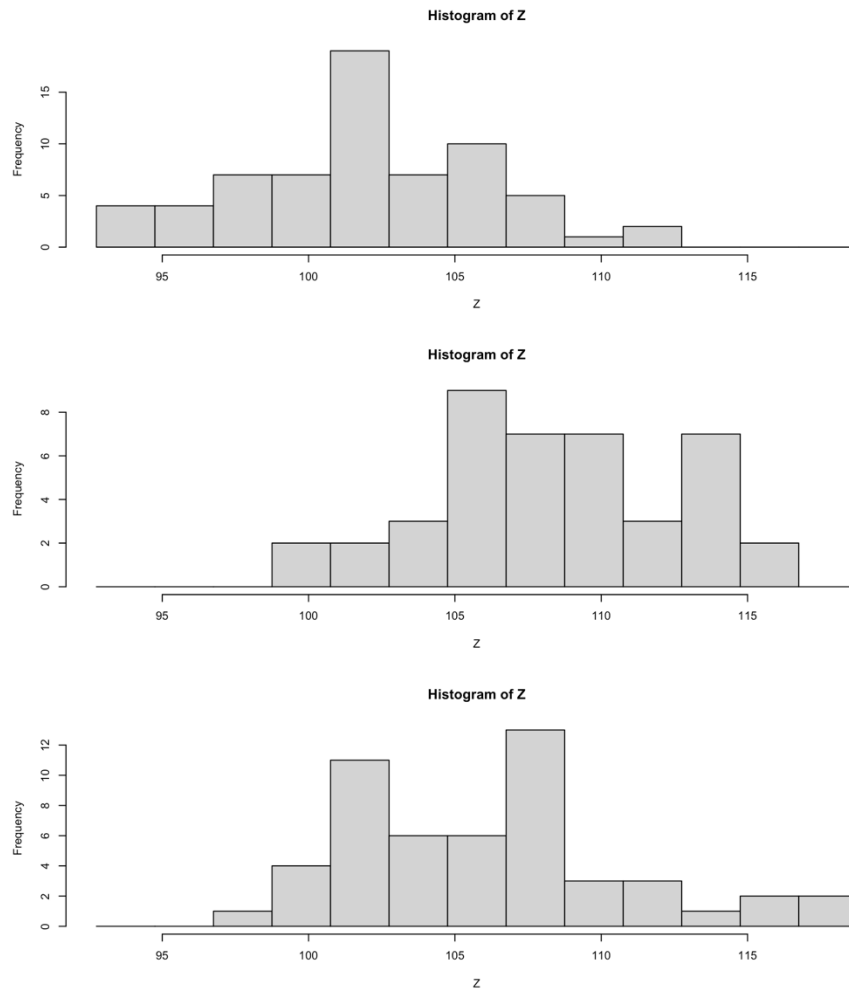


図 3. 原料メーカー別の層別ヒストグラム

※上から A 社製,B 社製,自社製として作成した。

上記の層別ヒストグラムより以下のことがわかる。

1. A 社製では 102,106 付近を頂点とする二つ山型となっている
2. B 社製では 106,114 付近を頂点とする二つ山型となっている
3. 自社製では 103,107,112 付近を頂点とする三つ山型となっている

(ウ) 設備の種類における層別ヒストグラム

次に設備である加熱炉の種類における層別ヒストグラムを作成した結果以下のようになった。

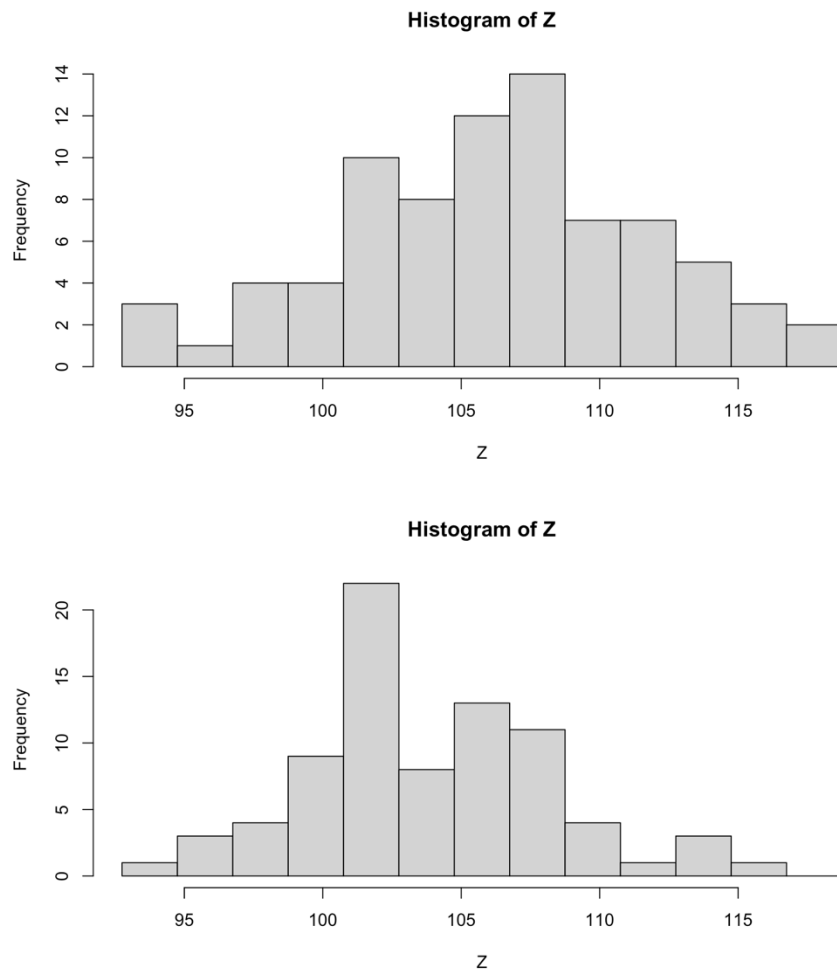


図 4. 加熱炉の区分による層別ヒストグラム

※上を 1 号機,下を 2 号機として作成した。

上記の層別ヒストグラムより以下のことがわかる。

1. 1 号機では 94,102,107 を頂点とする三つ山型となっている。
107 付近の度数が大きいので左右対称型ともみなすことができる。
2. 2 号機では 102,106,114 を頂点とする三つ山型となっている。

(3) 品質特性に関する時系列グラフ

(ア) 全体の時系列グラフ

以下に全体の時系列プロットを示す。

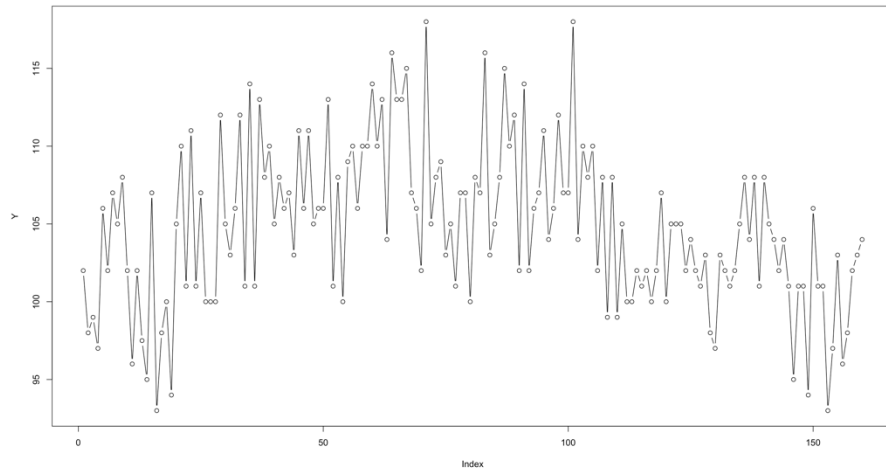


図 5. 全体の時系列グラフ

※横軸は製品製造の時系列,縦軸はシリコンの単結晶の直径を表す。

上記の時系列グラフからわかることを以下に示す。

1. 製造開始と製造終了時に比べ計測中間の単結晶の直径が大きいことがわかる。

(イ) 生産・提供条件で層別した時系列グラフ

次に、ヒストグラムと同様に層別し、時系列グラフを作成する。

① 午前・午後の層別時系列グラフ

以下に、午前・午後の層別時系列グラフを示す。

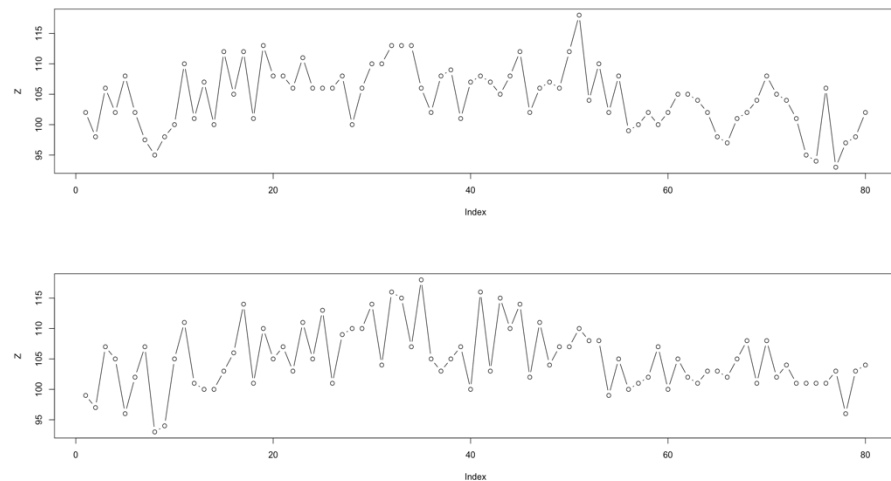


図 6. 午前・午後の層別時系列グラフ

※上が午前,下が午後とする

このデータから以下のことがわかる。

1. 午前・午後のグラフにおいて異なる目立った特徴は伺えない
2. 午前・午後のグラフにおいて測定初期は直径が小さく、測定中盤において大きくなり測定終期に再度小さくなるという同一の傾向が見られる。ただし、有意な傾向かは問わないものとする。
3. 午後の 60 以降において直線的な推移となっていることから、結晶の直径が安定していると言える。

② 原料メーカー 3 社での層別時系列グラフ

次に設備の種類での層別時系列グラフを作成する。

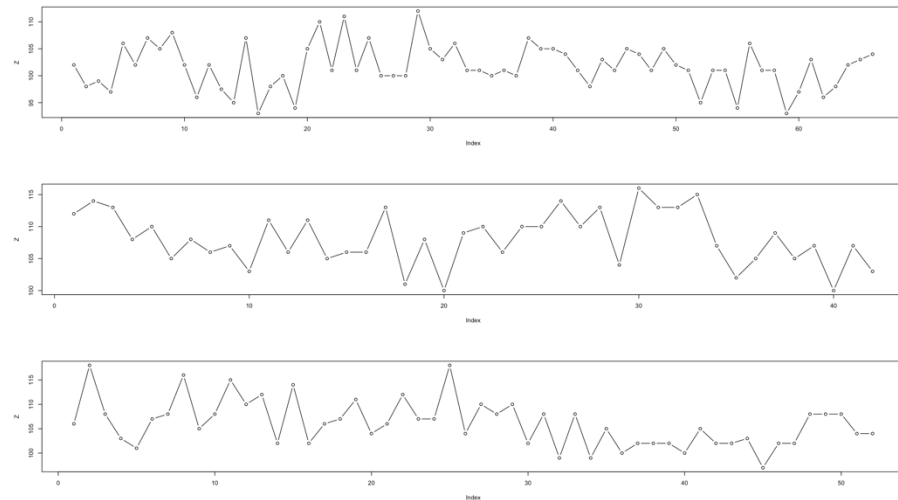


図 7. 原料メーカー 3 社における層別時系列グラフ

上から A 社, B 社, 自社を意味する

上記のデータの可視化により以下のことがわかる。

1. 可視的な共通する傾向は見受けられない。
2. A 社の時系列グラフは縦軸の最大値が 110 となっており他 2 社は最大値が 115 となっていることから、他 2 社よりも低く推移している。
3. 横軸に注目すると、B 社と自社は 55 以降利用されておらず、A 社でのみのデータとなっている。

③ 設備の種類での層別時系列グラフ

以下に加熱炉の種類において層別した時系列グラフを示す。

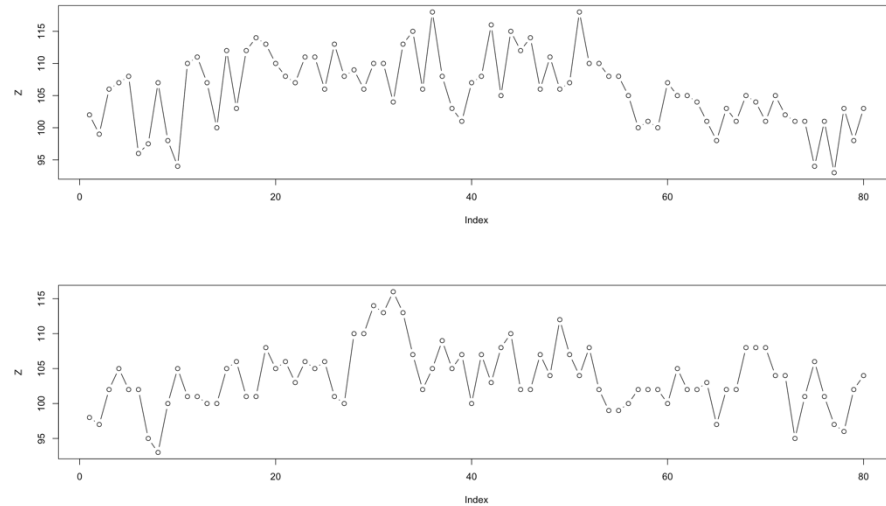


図 8. 設備の種類で層別した時系列グラフ

上から 1 号機,2号機を表す

上記の図から以下のことがわかる。

1. 1 号機,2号機ともに、山なりの推移となっていることから計測中盤(4 月後半からの時期)の直径が最も大きくなる傾向があると読み取れる。

(ウ) 計量値と品質特性の散布図

生産・提供条件のうち計量値である変数(シード温度,シード圧,シード後温度,シード後圧,原料重量,ヒータ使用回数,シード回数,シード時間,シード濃度)を横軸に,品質特性を縦軸に取った散布図を作成し、その形状からわかることを述べる。

① シード温度と品質特性の関係

はじめにシード温度と品質特性の関係を以下に示す。

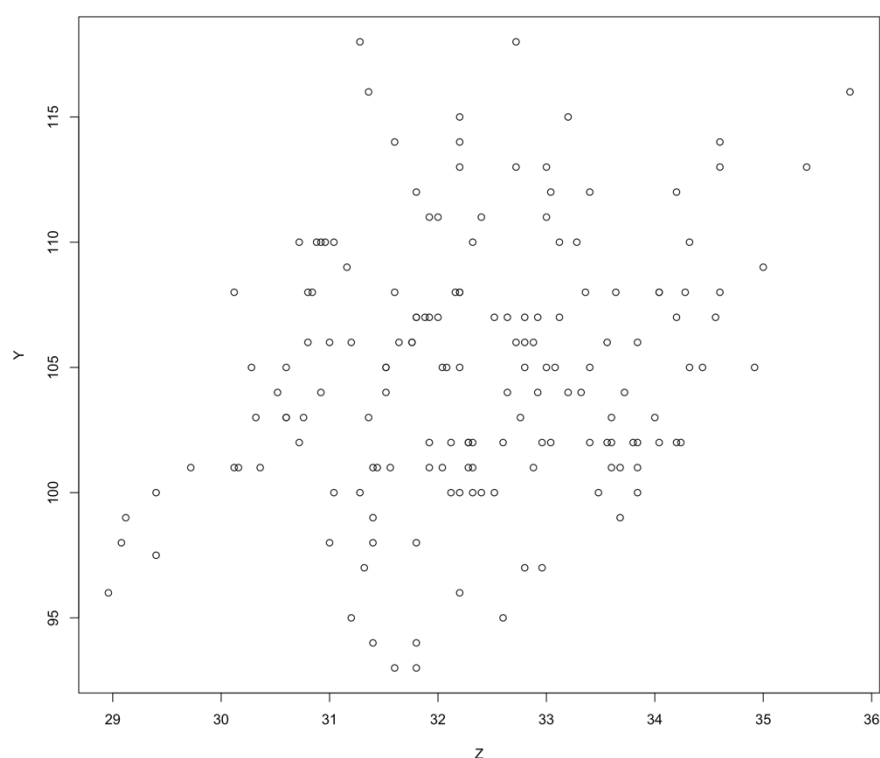


図 9. シード温度と品質特性の関係を示す散布図

横軸はシード温度,縦軸は直径を意味する。

上記の図から以下のことがわかる。

1. シード温度に対する品質特性の変化に強い傾向が見られないため、相関が弱いと言える。
2. ひろく見るとプロット群が全体的に右下から左上方向に伸びているため、正の相関があると言える。

② シード圧と品質特性の関係

次にシード圧と品質特性の関係を示す散布図を以下に表す。

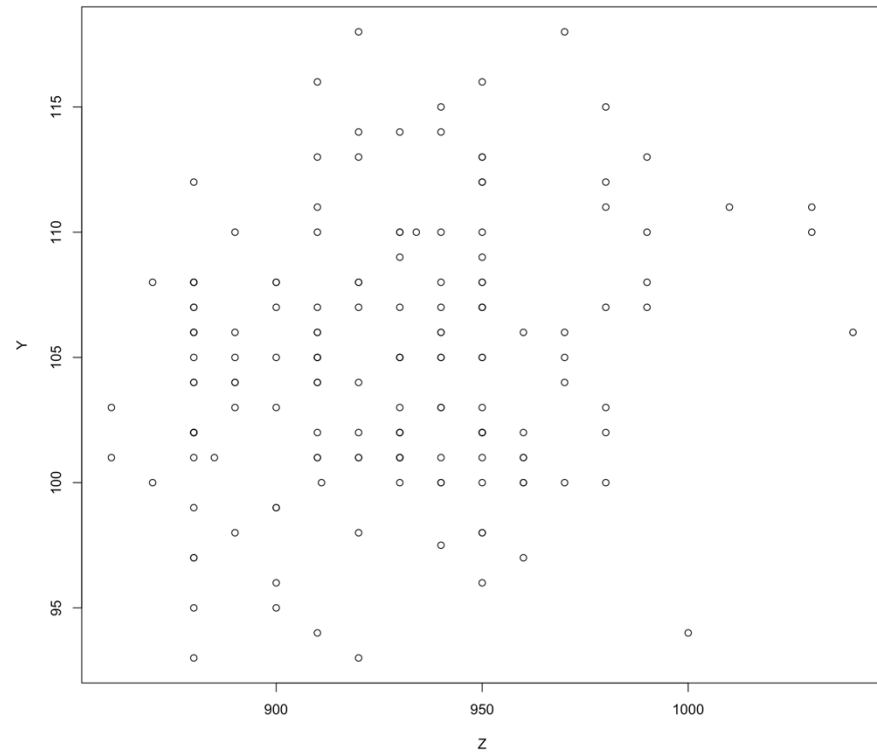


図 10. シード圧と品質特性の関係を示す散布図

横軸はシード圧,縦軸は品質特性を示す。

上記の図からわかることを以下に示す。

1. 散布図内にひろく分布していることから相関は見られないと言える。

③ シード後温度と品質特性の関係

次にシード後温度と品質特性の関係における散布図を以下に示す。

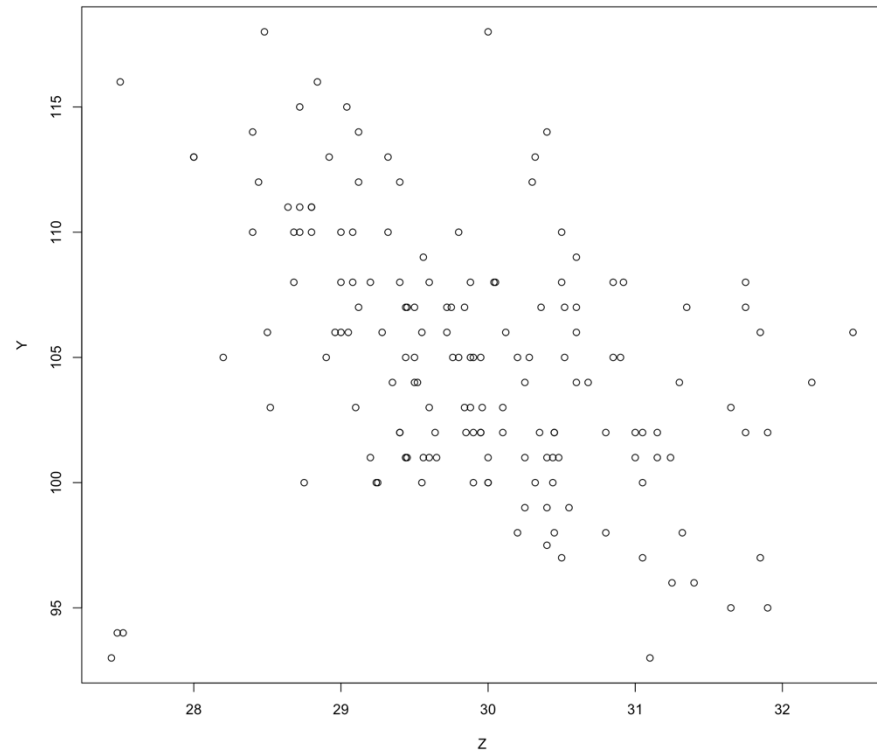


図 11. シード後温度と品質特性の関係を示す散布図

横軸がシード後温度,縦軸が品質特性を示す。

上記の図からわかることを以下に示す。

1. 図右上から左下に向けてプロットがひろがっているためシード後温度と品質特性の間に強い負の相関があると言える。

④ シード後圧と品質特性の関係

次にシード後圧と品質特性の関係を散布図で以下に示す。

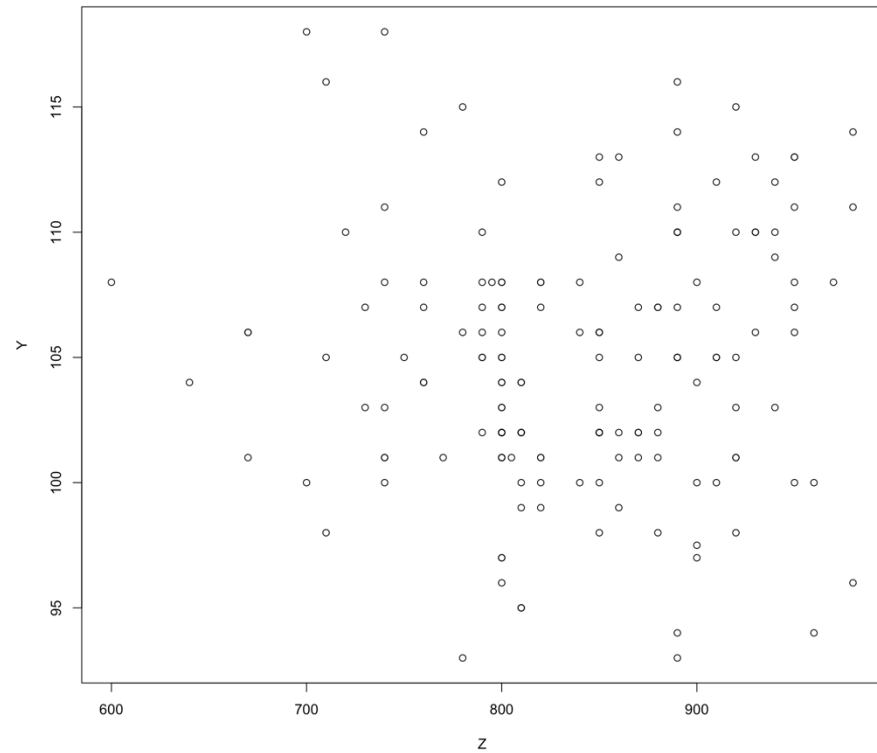


図 12. シード後圧と品質特性の関係

横軸がシード後圧,縦軸が品質特性を意味する。

上記の図からわかることを以下に示す。

1. プロット範囲内にひろく散らばっていることから相関はない,極めて弱いと言える。

⑤ 原料重量と品質特性の関係

次に原料重量と品質特性の関係を散布図で以下に示す。

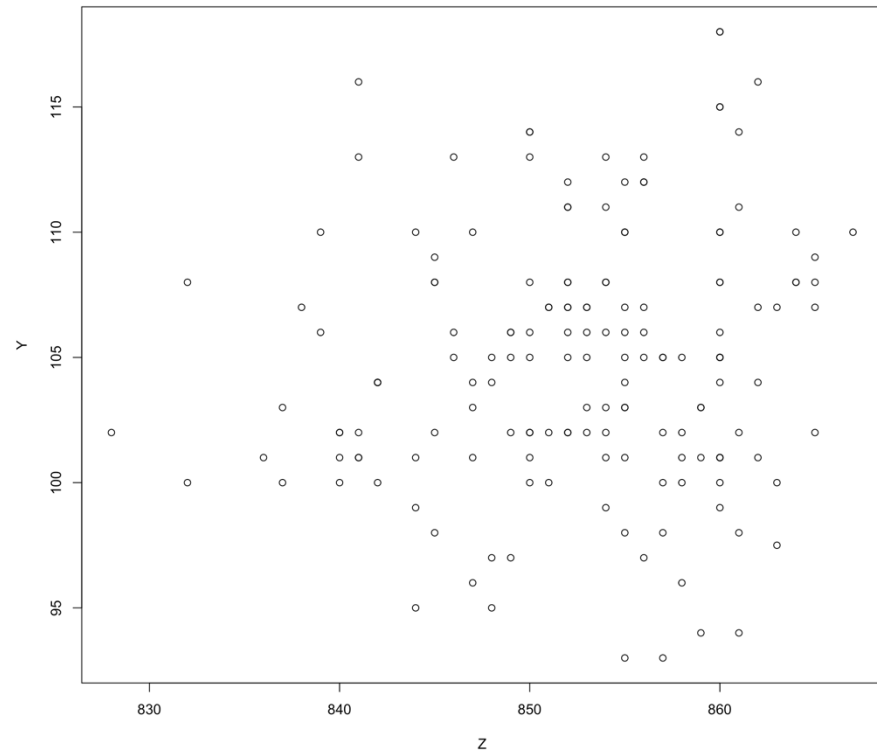


図 13. 原料重量と品質特性の関係

横軸が原料重量,縦軸が品質特性を意味する。

上記の図からわかることを以下に示す。

1. プロット範囲内にひろく散らばっていることから相関は極めて弱いと言える。
2. プロット範囲内左下に寄っているとも見られるため、上記の図のみでは原料重量と品質特性の間に関係がないとは言い切れない。

⑥ ヒータの使用回数と品質特性の関係

次にヒータの使用回数と品質特性の関係を散布図で以下に示す。

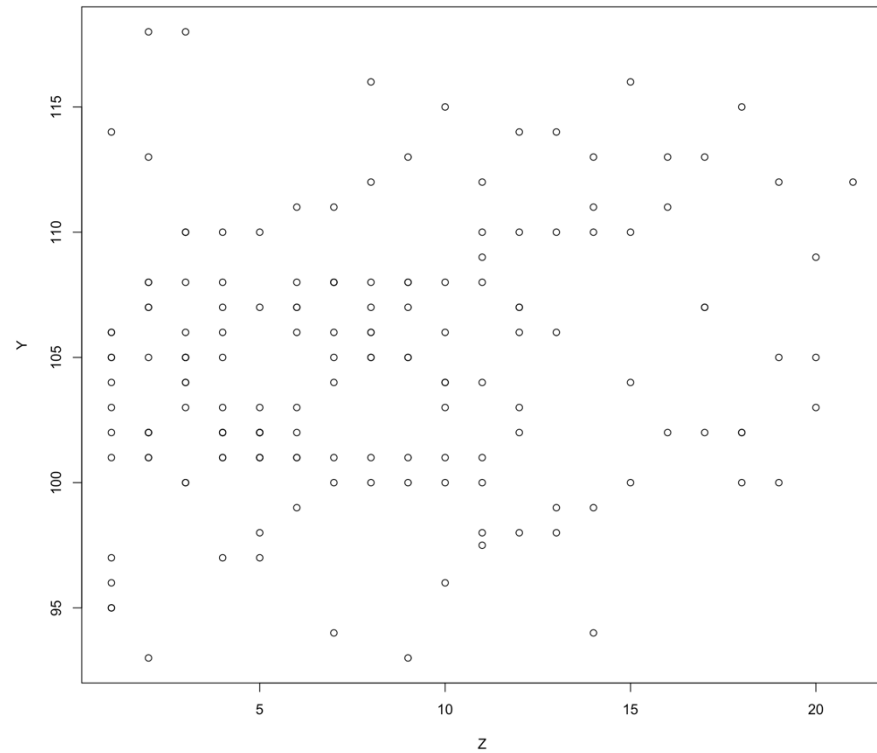


図 14. ヒータ使用回数と品質特性の関係

横軸がヒータ使用回数,縦軸が品質特性を意味する。

上記の図からわかることを以下に示す。

1. プロット範囲内にひろく散らばっていることから相関はない,極めて弱いと言える。

⑦ シード回数と品質特性の関係

次にシード回数と品質特性の関係を散布図で以下に示す。

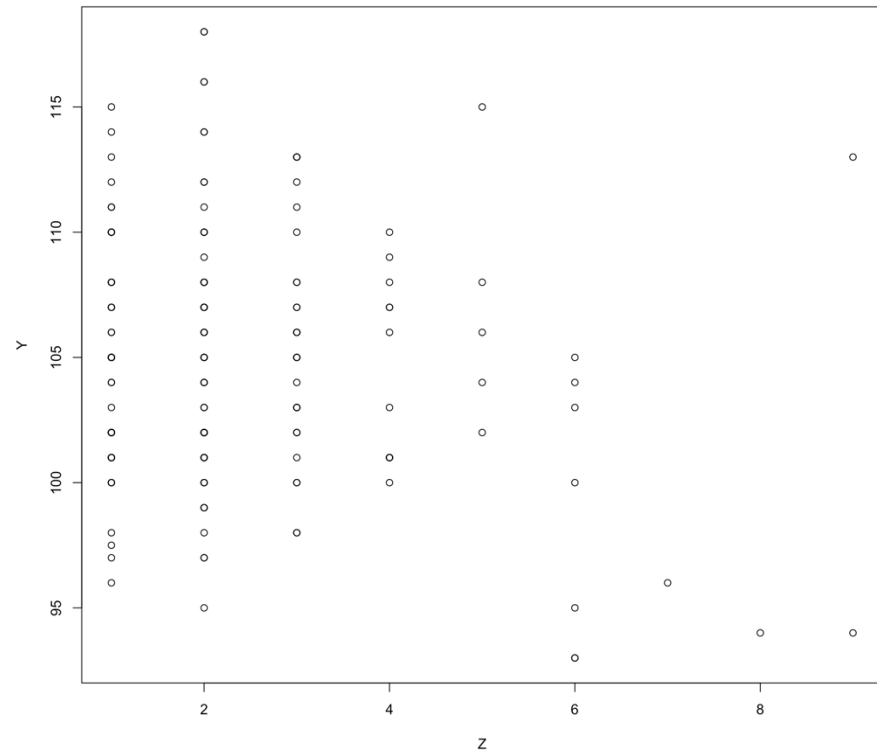


図 15. シード回数と品質特性の関係

横軸がシード回数,縦軸が品質特性を意味する。

上記の図からわかることを以下に示す。

1. プロットが線形的でなく範囲的に広がっているため、正,負で表現されるような直線的な相関はない,または弱いと言える。
2. プロットが左中央に寄っていることから、なんらかの関係があることがわかる。

⑧ シード時間と品質特性の関係

次にシード時間と品質特性の関係を散布図で以下に示す。

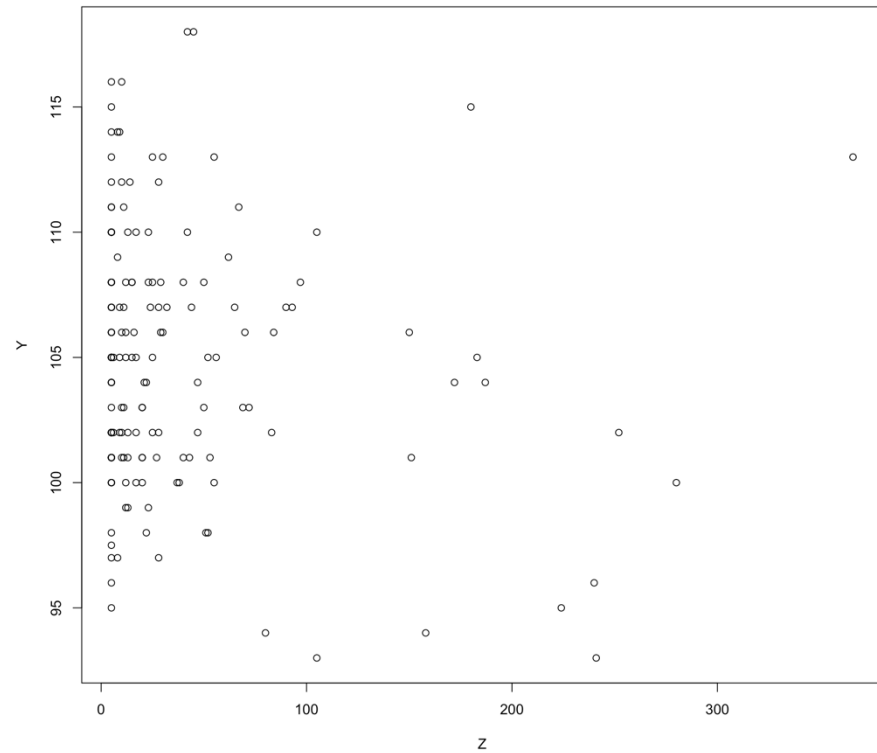


図 16. シード時間と品質特性の関係

横軸がシード時間,縦軸が品質特性を意味する。

上記の図からわかることを以下に示す。

1. プロットが縦軸付近に局所的に寄っていることから、この図のみでは直線的な関係性があると言い切れない。

⑨ シード濃度

次にシード濃度と品質特性の関係を散布図で以下に示す。

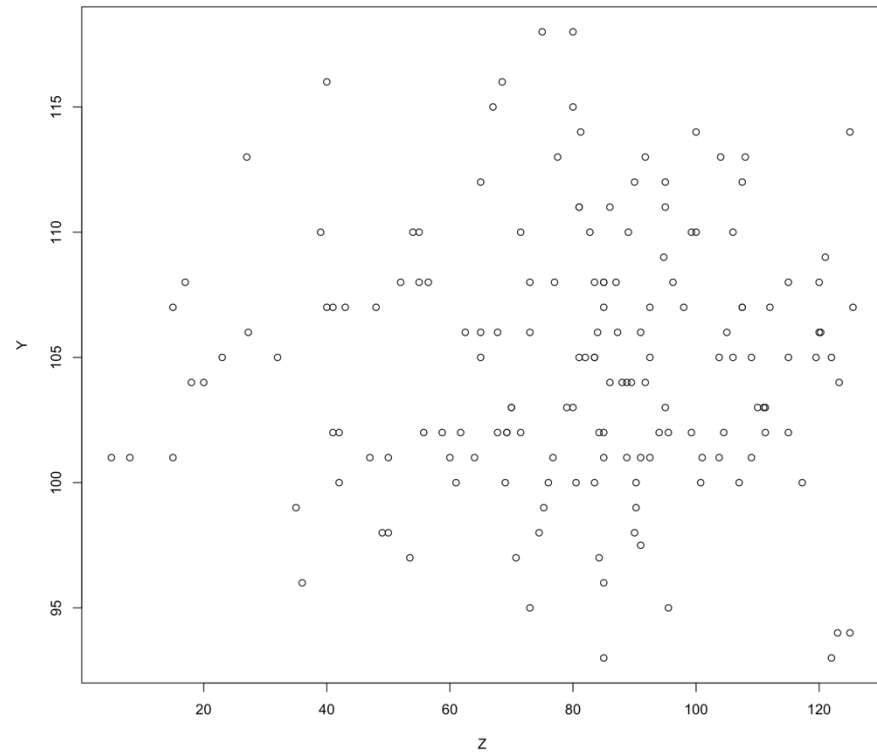


図 17. シード濃度と品質特性の関係

横軸がシード濃度,縦軸が品質特性を意味する。

上記の図からわかることを以下に示す。

1. 線形的なプロットではなく広範囲に広がっているため、正,負で表現できる相関関係があるとは言えない。
2. 図の左側(シード濃度が 80~120)にプロットが寄っていることがわかる。

(エ) 生産・提供条件で層別した散布図

この項では、層別散布図を作成する。本解析では、X2を除くX9,14に対してX4,5,7,8,10,11,12,13の8項目の層別散布図作成を行う。X2は時系列プロットにおいて、特徴的な相違点、変化がみられなかったため、層別した時に同様な分布になると推測したためである。そして、X4,5,7,8,10,11,12,13を選択した理由としては、先に挙げた散布図で強い相関があると言えなかったからである。相関が弱い、又はないと思われる原因としては、全く関係のないデータである場合以外にあらゆる要因が混ざっていることにより複雑化している場合もある。したがって、上記の組み合わせについて層別散布図を作成することとした。

① X4 と Y の散布図を X9 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9 は原料メーカーであり、X4 はシード温度であることから原料メーカーによるシードの適正温度が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

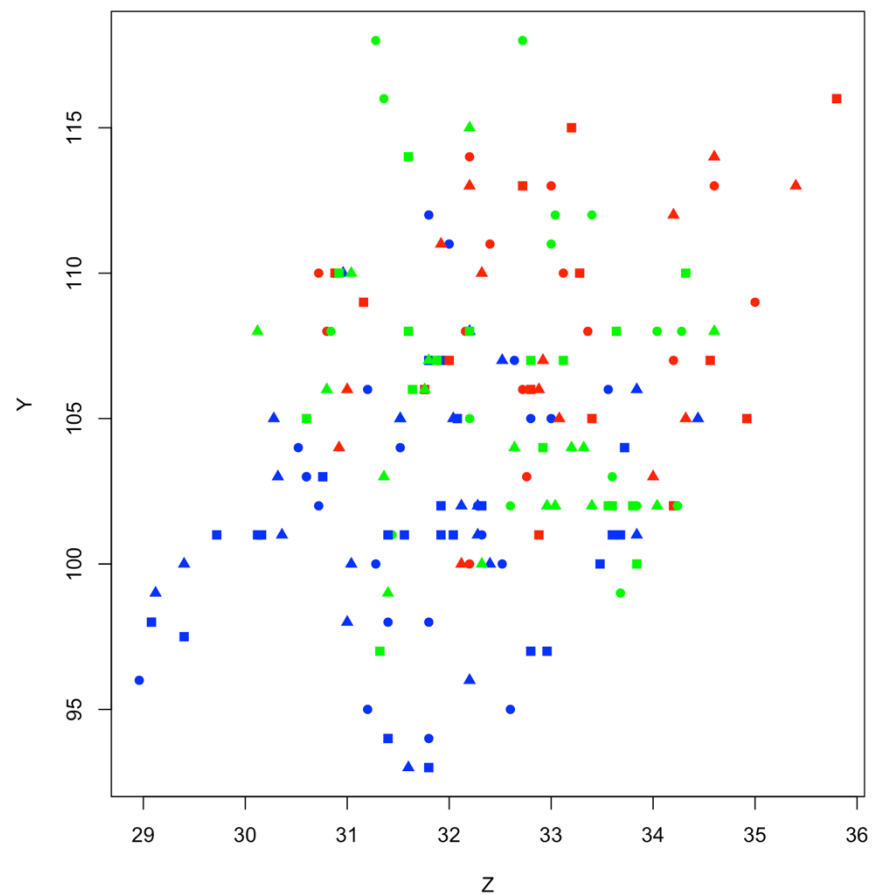


図 18. X4 と Y の散布図を X9 で層別した散布図

それぞれ、青:A 社/ 赤:B 社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青である A 社が低プロットのほとんどは 107 より低い範囲に分布しており、Y に対して正の弱い相関を持っているとも言える。
2. B 社と自社は A 社に比べて高い直径を示す範囲にプロットされている。
3. 1,2 より、原料メーカーの違いがシード温度に寄与していることがわかる。

② X5 と Y の散布図を X9 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9 は原料メーカーであり、X5 はシード圧であることから原料メーカーによるシードの適正圧が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

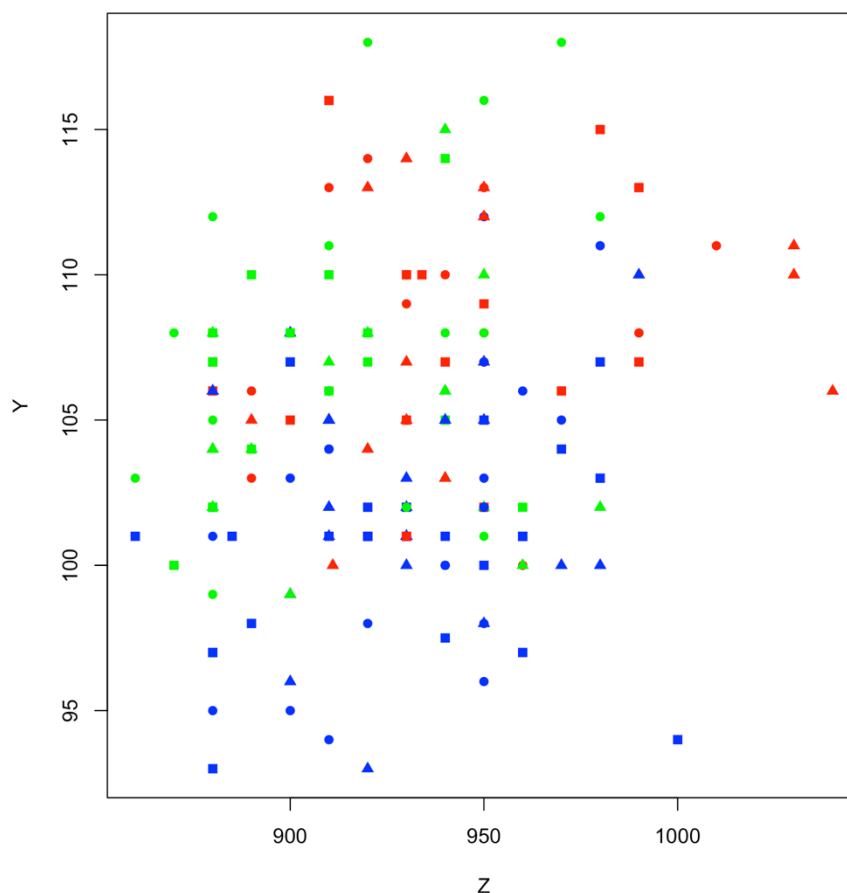


図 19. X5 と Y の散布図を X9 で層別した散布図

それぞれ、青:A 社/ 赤:B 社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青である A 社の Y が低プロットのほとんどは 107 より低い範囲に分布しており、Y に対して正の弱い相関を持っているとも言える。
2. B 社と自社は A 社に比べて高い直径を示す範囲にプロットされている。
3. 自社と Y は正の相関を持っていると言える。
4. 上記より原料メーカーの違いがシード圧に寄与していることがわかる。

③ X7とYの散布図をX9で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9は原料メーカーであり、X7はシート後圧であることから原料メーカーによるシート後適正圧が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

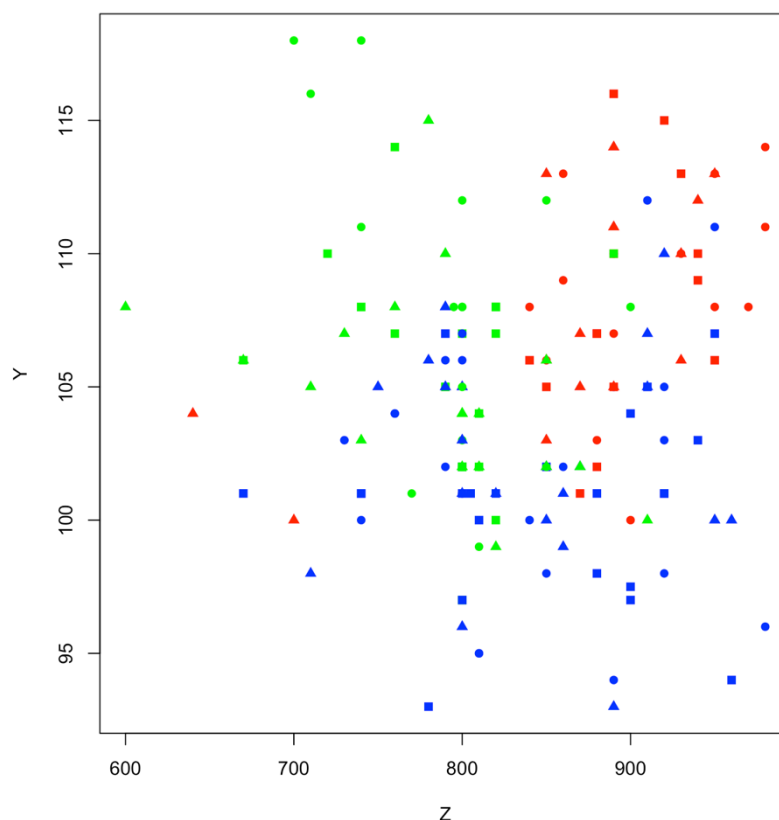


図 20. X7とYの散布図をX9で層別した散布図

それぞれ、青:A社/ 赤:B社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青であるA社は低い位置にプロットが多く、シート後圧が大きくなるとYも大きくなる傾向にあるため正の相関があると言える。
2. 赤であるB社は高い位置にプロットが多く、シート後圧 850～950 付近に局所的に固まっているため相関があるとはいえない。
3. 緑である自社はシート後圧 700～900 で縦方向に推移しており、傾きが急な負の相関と捉えることができる。
4. 1,2 に対して、3 で変化があるため原料メーカーの違いがシート後圧に寄与していることがわかる。

④ X8 と Y の散布図を X9 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9 は原料メーカーであり、X8 は原料重量であることから原料メーカーによる原料の適正量がまっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

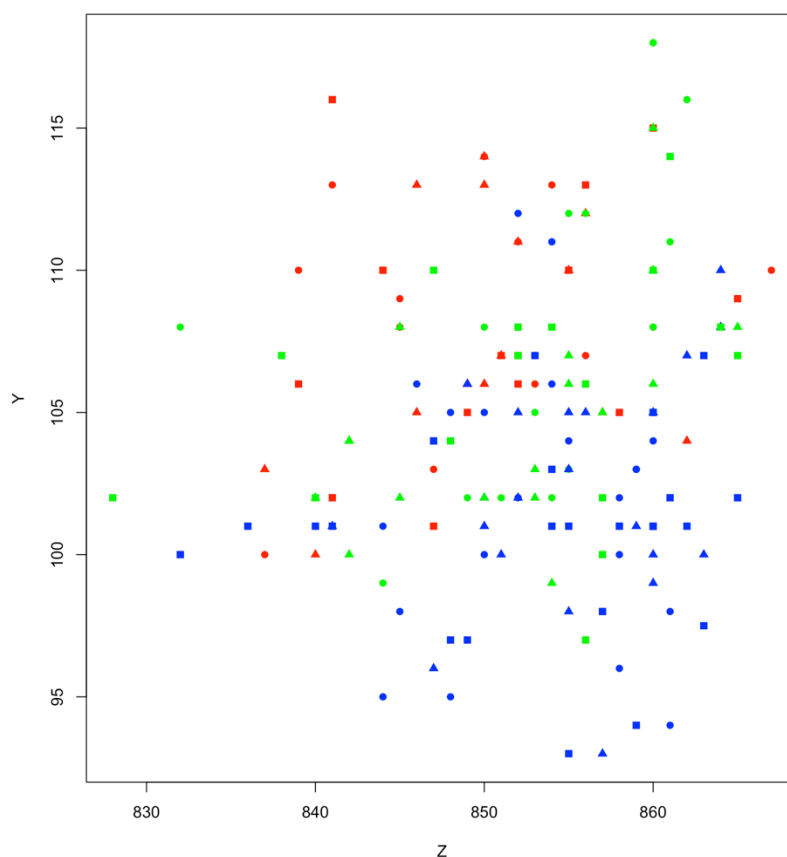


図 21. X8 と Y の散布図を X9 で層別した散布図

それぞれ、青:A 社/ 赤:B 社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青である A 社は低い位置にプロットが多く、原料重量が重くなっても大きな変化は見られないため相関が弱い、もしくはないと言える。
2. 赤である B 社は A より高い位置にプロットが多く、右肩上がりの傾向が強いため正の相関があると判断できる。
3. 緑である自社は A より高い位置にプロットが多く、右肩上がりの傾向が強いため正の相関があると判断できる。

⑤ X10 と Y の散布図を X9 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9 は原料メーカーであり、X10 はヒータ使用回数であることから原料メーカーによるヒータ使用の適正回数が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

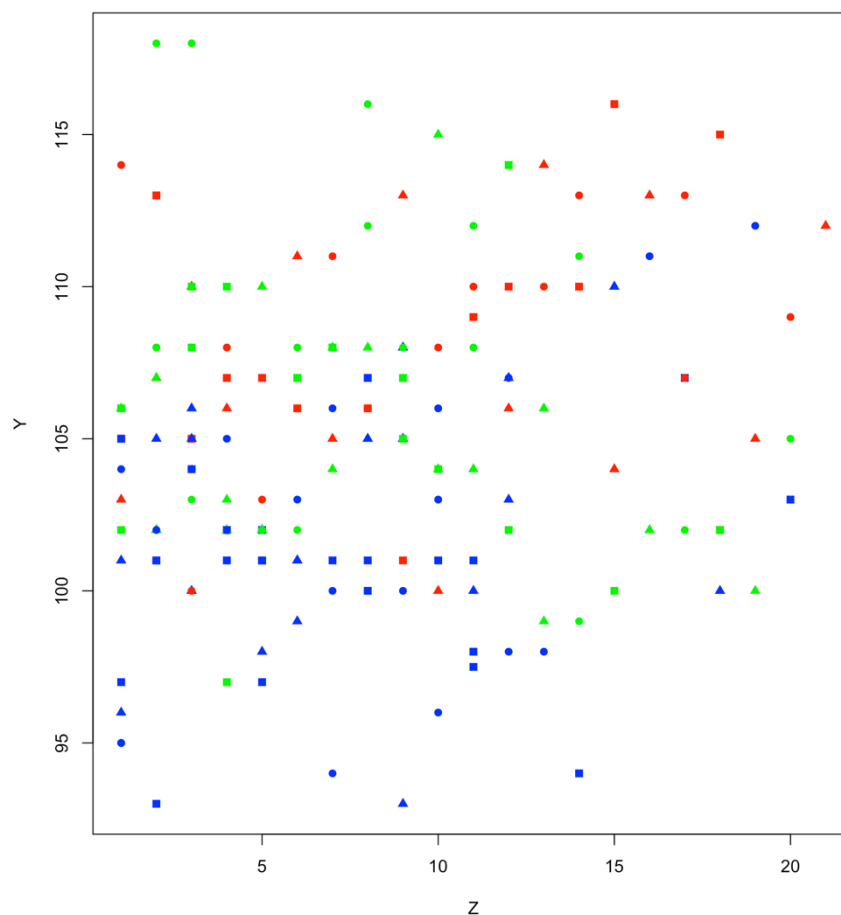


図 22. X10 と Y の散布図を X9 で層別した散布図

それぞれ、青:A 社/ 赤:B 社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青である A 社は右下に固まっているため相関は見受けられない。
2. 赤である B 社は直径 105 以上にプロットが多く、右肩上がりの傾向が強いいため正の相関があると判断できる。
3. 緑である自社はひろく分布しているものの右肩下がりとなる分布をしているため負の相関があると判断できる。

⑥ X11 と Y の散布図を X9 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9 は原料メーカーであり、X11 はシード回数であることから原料メーカーによるシードの適正回数が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

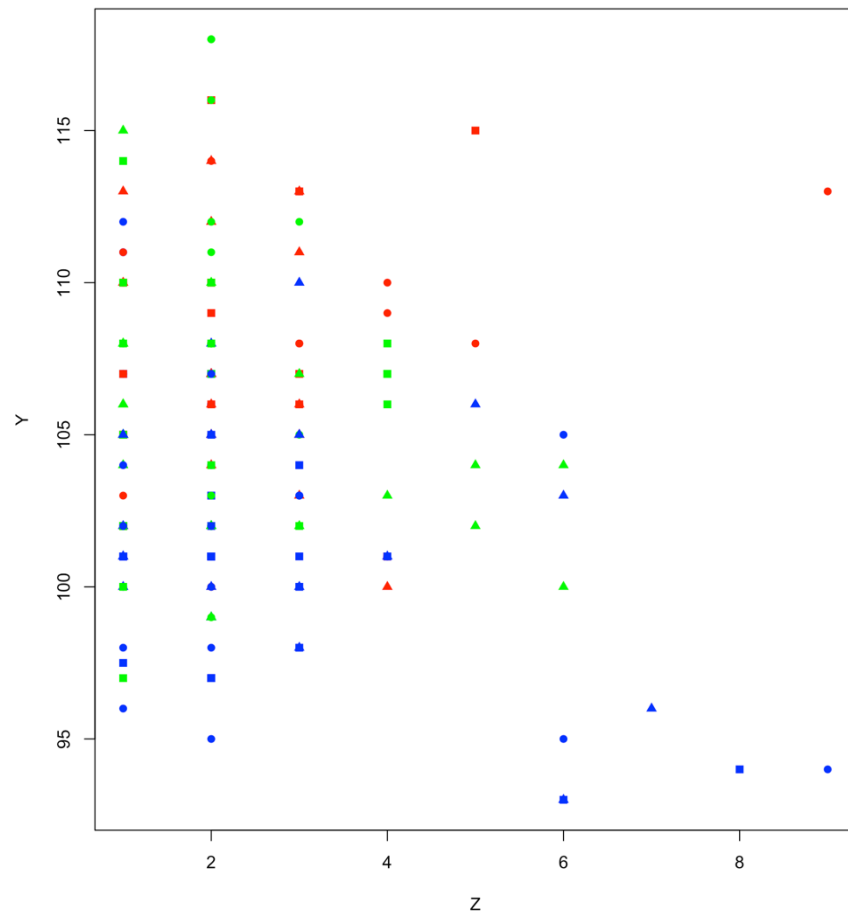


図 23. X11 と Y の散布図を X9 で層別した散布図

それぞれ、青:A 社/ 赤:B 社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青である A 社は右下に固まっているため相関は見受けられない。
2. 赤である B 社は 105 よりも大きい部分にひろく分布しているため相関は見られない。
3. 緑である自社は直径 97 から 120 までひろく分布しているため、相関は見られない。

⑦ X12とYの散布図をX9で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9は原料メーカーであり、X12はシード時間であることから原料メーカーによるシードの適正時間が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

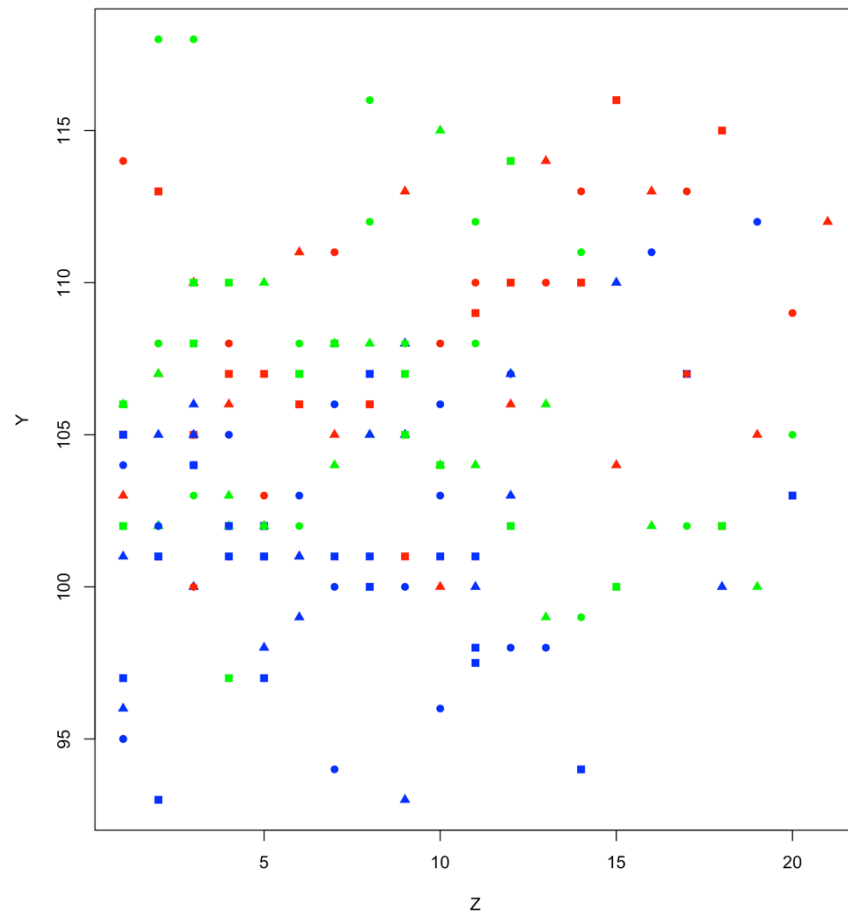


図 24. X12とYの散布図をX9で層別した散布図

それぞれ、青:A社/ 赤:B社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 図 23 と近い分布となっており、A社が右下に多くB社が上部に広範的に、自社が右部分にひろく分布している。

⑧ X13とYの散布図をX9で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X9は原料メーカーであり、X13はシード濃度であることから原料メーカーによるシードの適正濃度が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

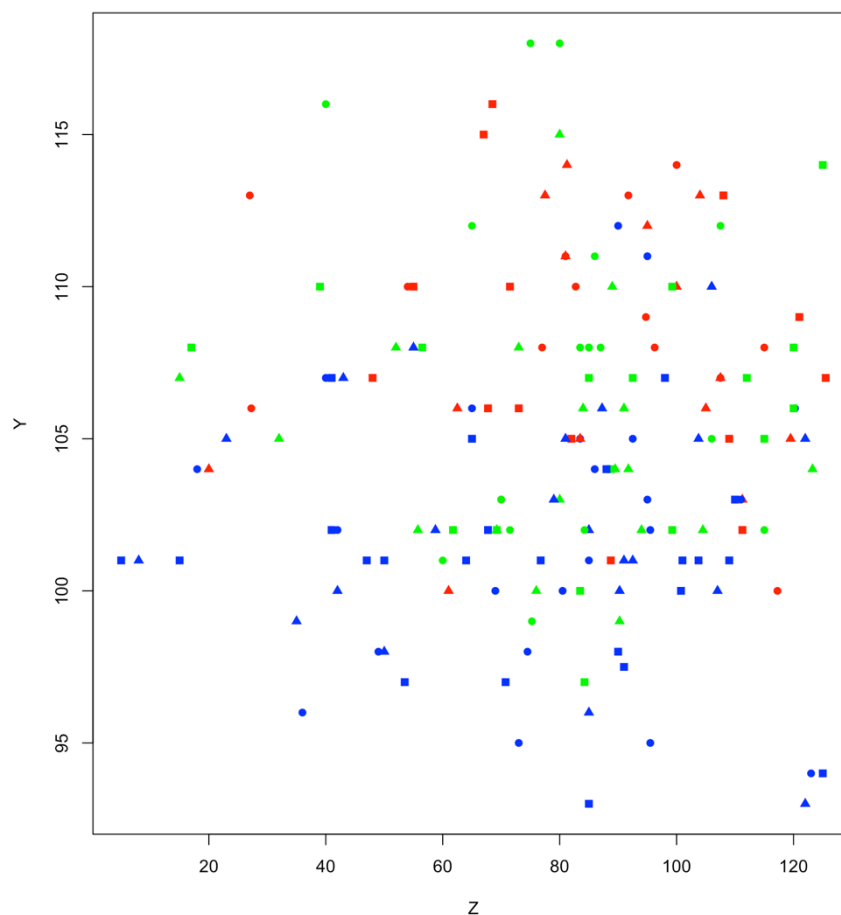


図 25. X13とYの散布図をX9で層別した散布図

それぞれ、青:A社/ 赤:B社/ 緑:自社としている。

上記からわかることを以下に示す。

1. 青であるA社は直径107以下にひろく分布しているため相関は見受けられない。
2. 赤であるB社は直径105以上にひろく分布しているため相関は見受けられない。
3. 緑である自社は全体にひろく分布しているため相関は見られない。

⑨ X4とYの散布図をX14で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14は加熱炉であり、X4はシード温度であることから加熱炉によるシードの適正温度が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

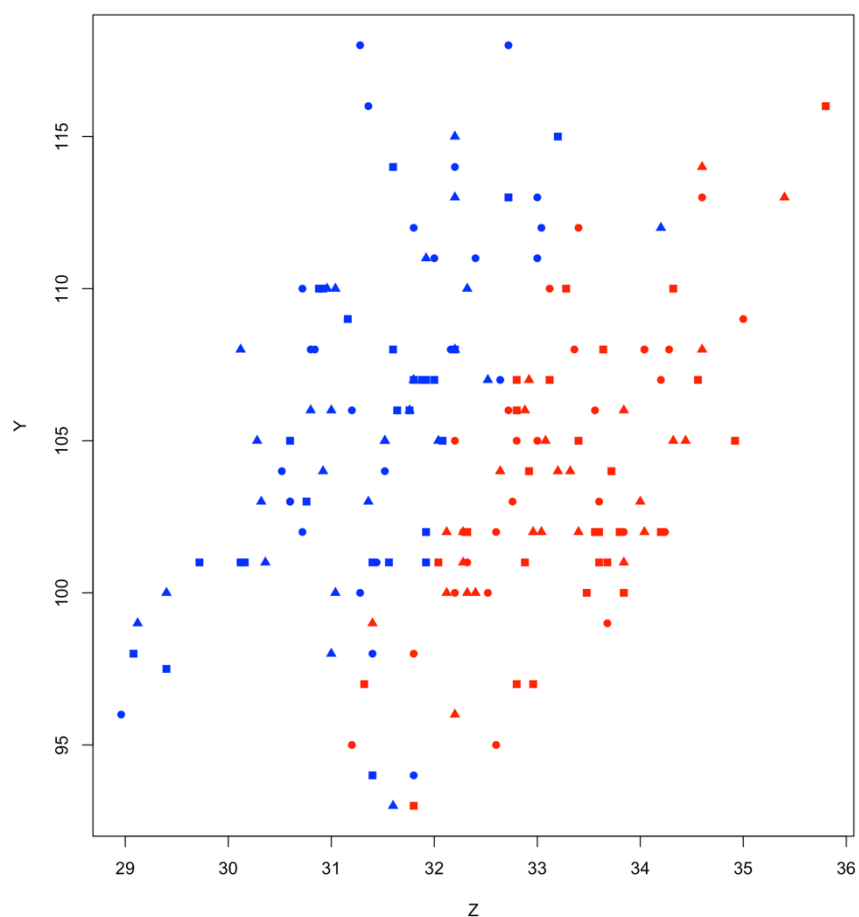


図 26. X4とYの散布図をX14で層別した散布図

それぞれ、青:1号機/赤:2号機としている。

上記からわかることを以下に示す。

1. 1,2号機両方とも強い正の相関があり、1号機の方が製造できる直径が大きく優位であるとわかる。

⑩ X5 と Y の散布図を X14 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14 は加熱炉であり、X5 はシード圧であることから加熱炉によるシードの適正圧が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

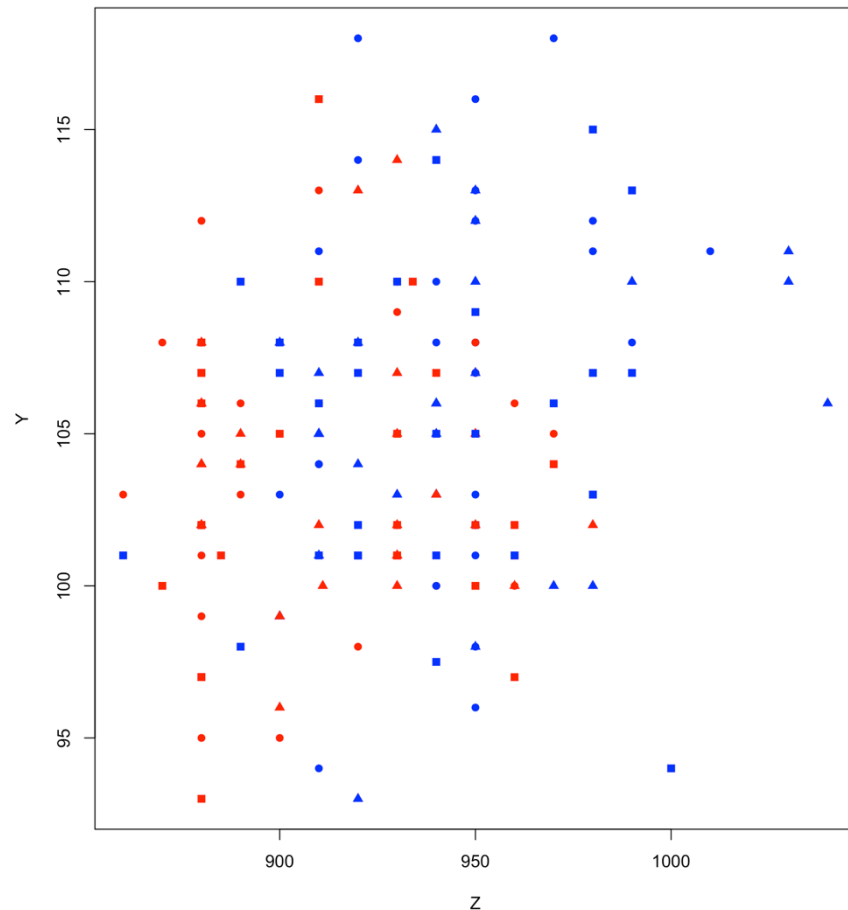


図 27. X5 と Y の散布図を X14 で層別した散布図

それぞれ、青:1 号機/赤:2 号機としている。

上記からわかることを以下に示す。

- 1,2 号機両方とも正の相関が見受けられないが、赤である 1 号機の方がシード圧の低い範囲にひろく分布していることがわかる。

⑪ X6とYの散布図をX14で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14は加熱炉であり、X6はシード後温度であることから加熱炉によるシード後の適正温度が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

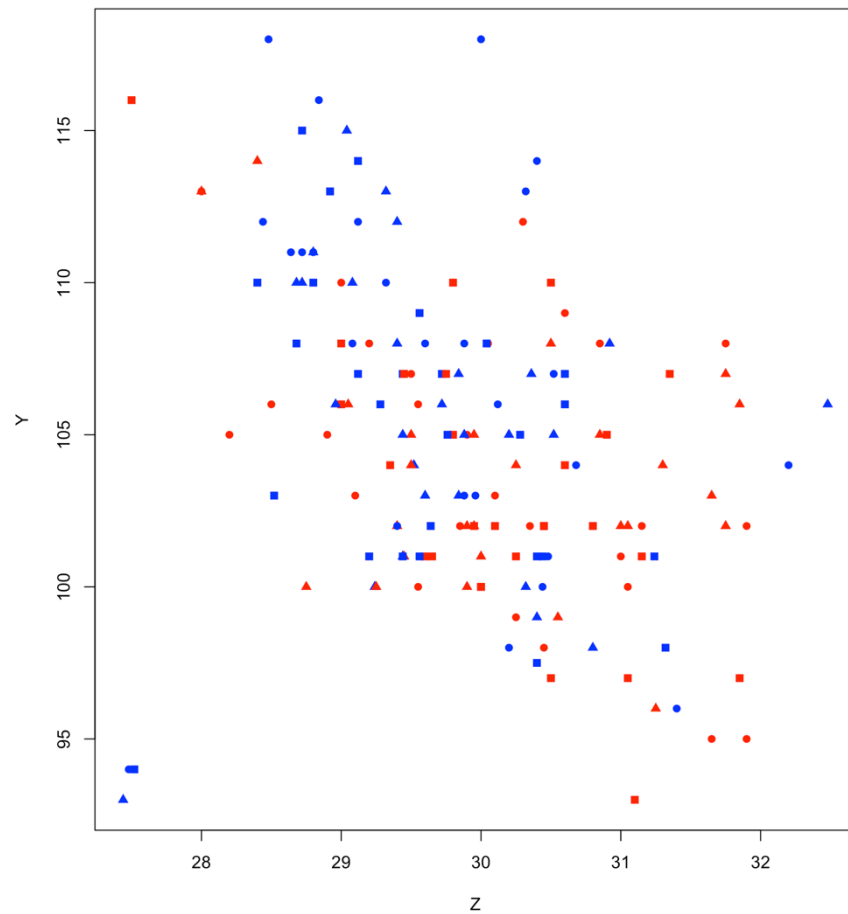


図 28. X6とYの散布図をX14で層別した散布図

それぞれ、青:1号機/赤:2号機としている。

上記からわかることを以下に示す。

1. 1,2号機両方とも同様な分布を示しており、強い負の相関があるとわかる。

⑫ X7 と Y の散布図を X14 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14 は加熱炉であり、X7 はシード後圧であることから加熱炉によるシード後の適正圧が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

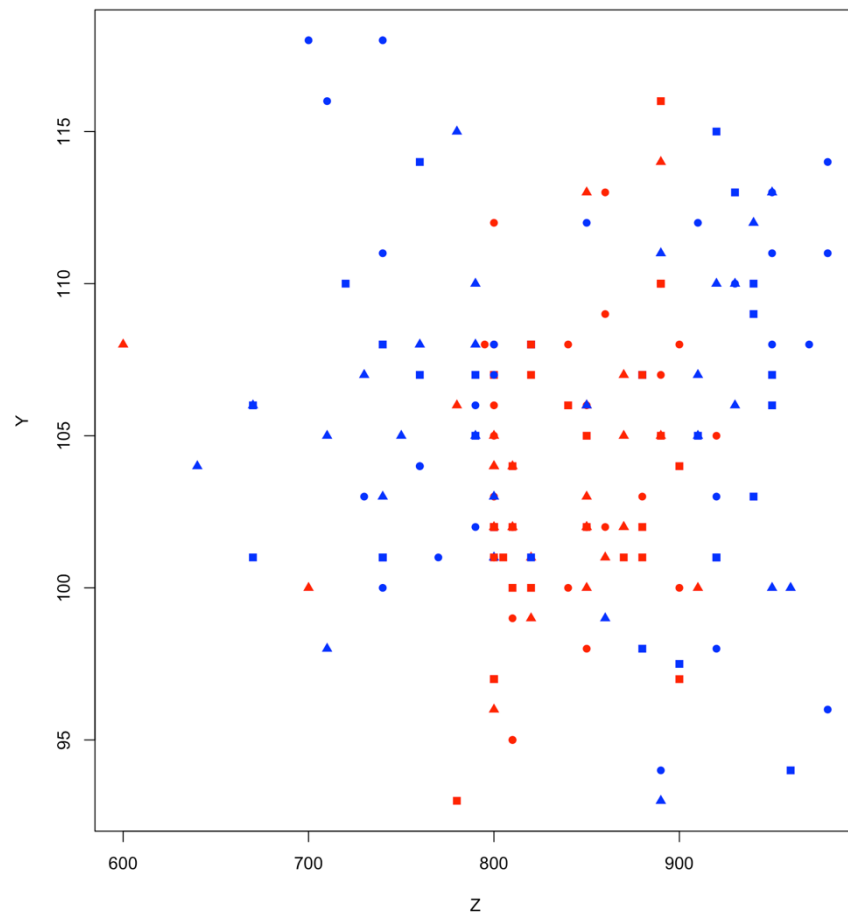


図 29. X7 と Y の散布図を X14 で層別した散布図

それぞれ、青:1 号機/赤:2 号機としている。

上記からわかることを以下に示す。

1. 1,2 号機両方とも正の相関が見受けられないが、赤である 2 号機の方が原料質量 800 から 900 の範囲で局所的に分布していることがわかる。

⑬ X8 と Y の散布図を X14 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14 は加熱炉であり、X8 は原料質量であることから加熱炉による原料の適正量が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

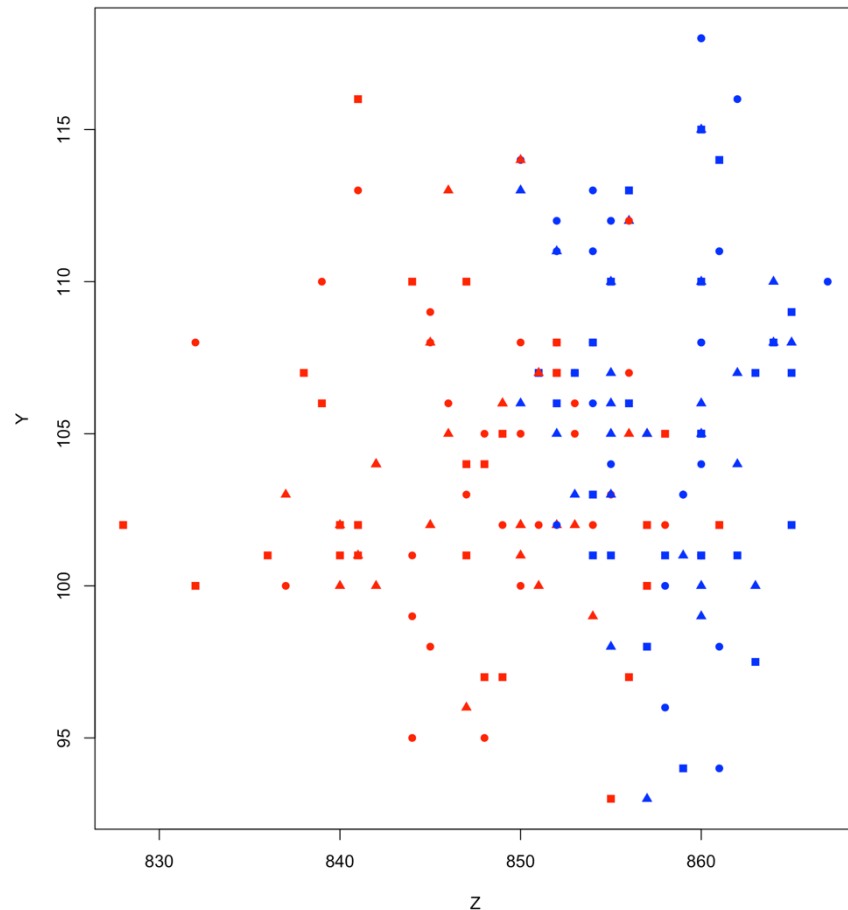


図 30. X8 と Y の散布図を X14 で層別した散布図

それぞれ、青:1 号機/赤:2 号機としている。

上記からわかることを以下に示す。

- 1,2 号機両方とも正の相関が見受けられないが、青である 1 号機の方が原料質量 851 以上の範囲で広範的に分布している一方で赤である 2 号機の方が質量 859 以下の範囲で広範的に分布しており、1,2 号機により異なる分布を示していることがわかる。

⑭ X10 と Y の散布図を X14 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14 は加熱炉であり、X10 はヒータ使用回数であることから加熱炉によるヒータ使用の適正回数が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

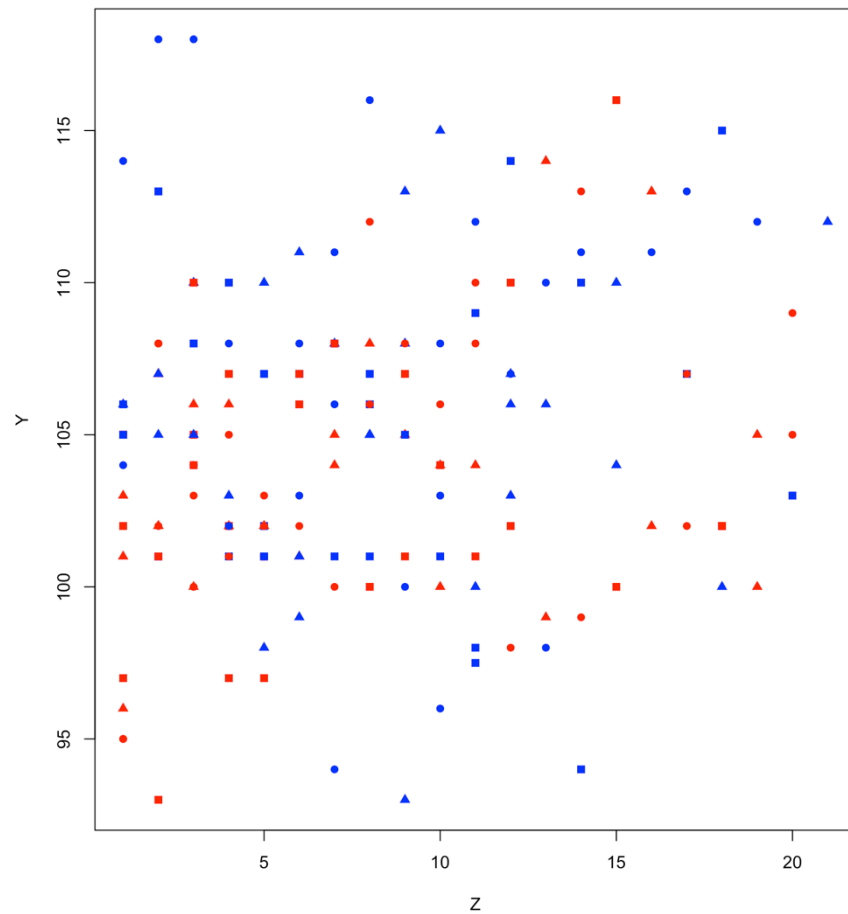


図 31. X10 と Y の散布図を X14 で層別した散布図

それぞれ、青:1 号機/赤:2 号機としている。

上記からわかることを以下に示す。

1. 1,2 号機両方ともどちらも同様に広範的に分布しており、相関がないように見受けられる。

⑮ X11 と Y の散布図を X14 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14 は加熱炉であり、X11 はシード回数であることから加熱炉によるシードの適正回数が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

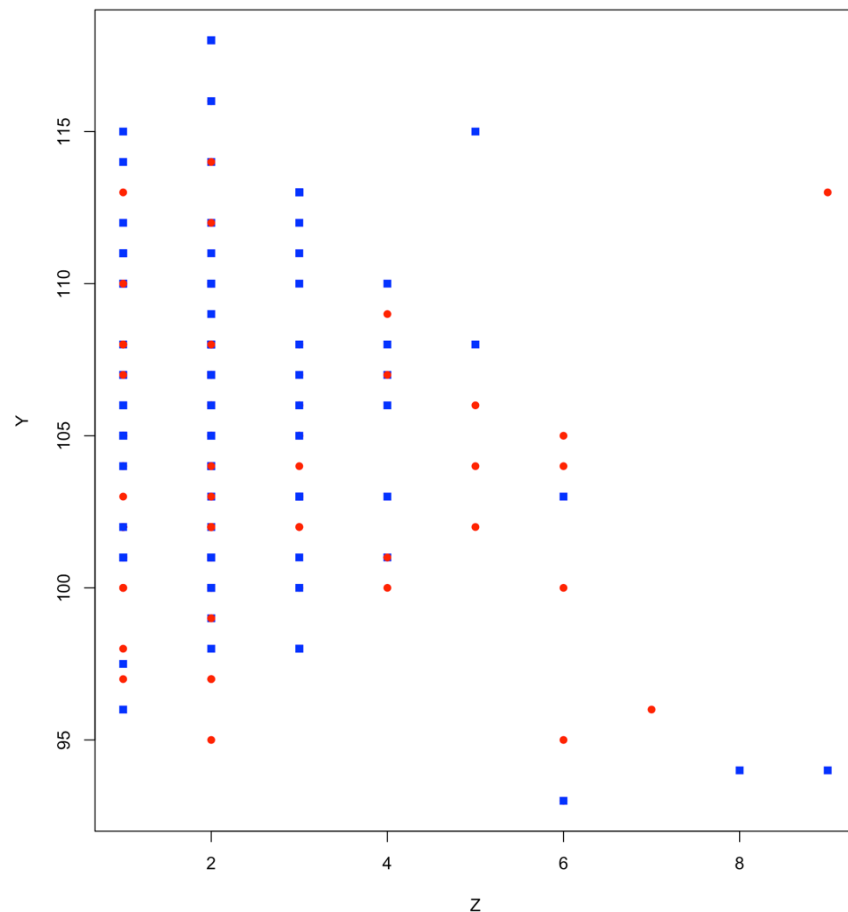


図 32. X11 と Y の散布図を X14 で層別した散布図

それぞれ、青:1 号機/赤:2 号機としている。

上記からわかることを以下に示す。

1. 1,2 号機両方ともどちらも同様に広範的に分布しており、相関がないように見受けられる。

⑩ X12 と Y の散布図を X14 で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14 は加熱炉であり、X12 はシード時間であることから加熱炉によるシードの適正時間が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

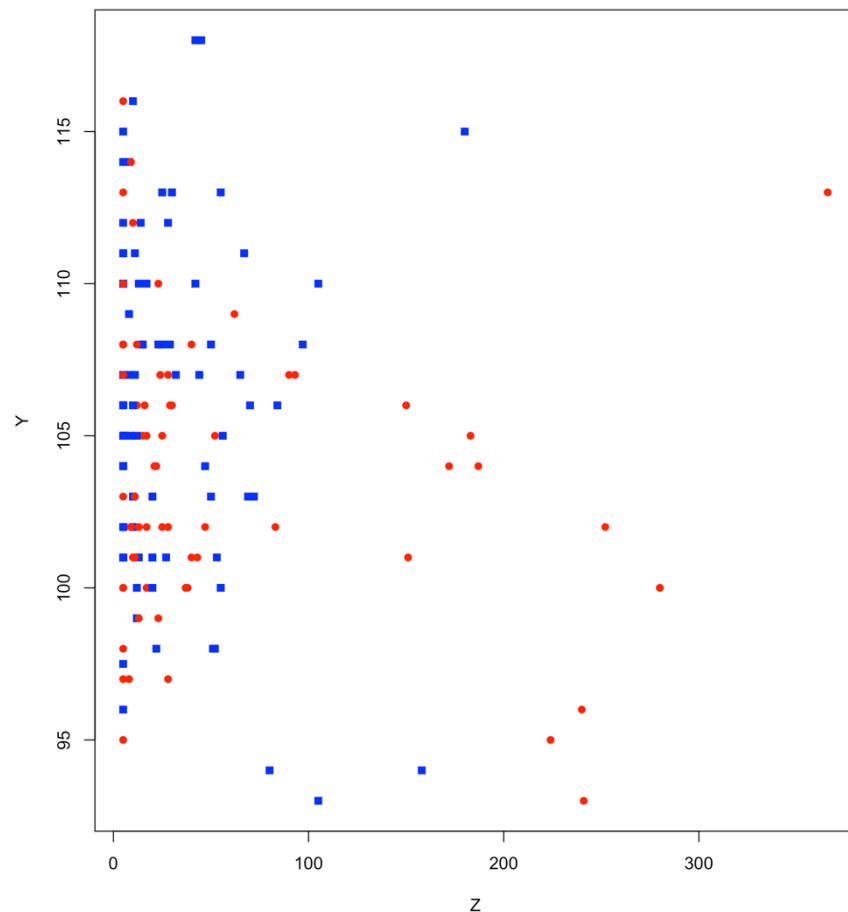


図 33. X12 と Y の散布図を X14 で層別した散布図

それぞれ、青:1 号機/赤:2 号機としている。

上記からわかることを以下に示す。

1. 1,2 号機両方ともシード時間 0 から 100 の範囲に局所的に分布している。赤である 2 号機は 100 から 300 にも点々と分布しているものの、数量で比較すると誤差とも考えられることから、加熱炉の違いがシード時間の分布に寄与しないと言える。

⑰ X13とYの散布図をX14で層別した散布図

はじめに上記の組み合わせの層別散布図を作成する。X14は加熱炉であり、X13はシード濃度であることから加熱炉によるシードの適正濃度が決まっている場合に相関や特徴が生じると言える。以下に層別した図を示す。

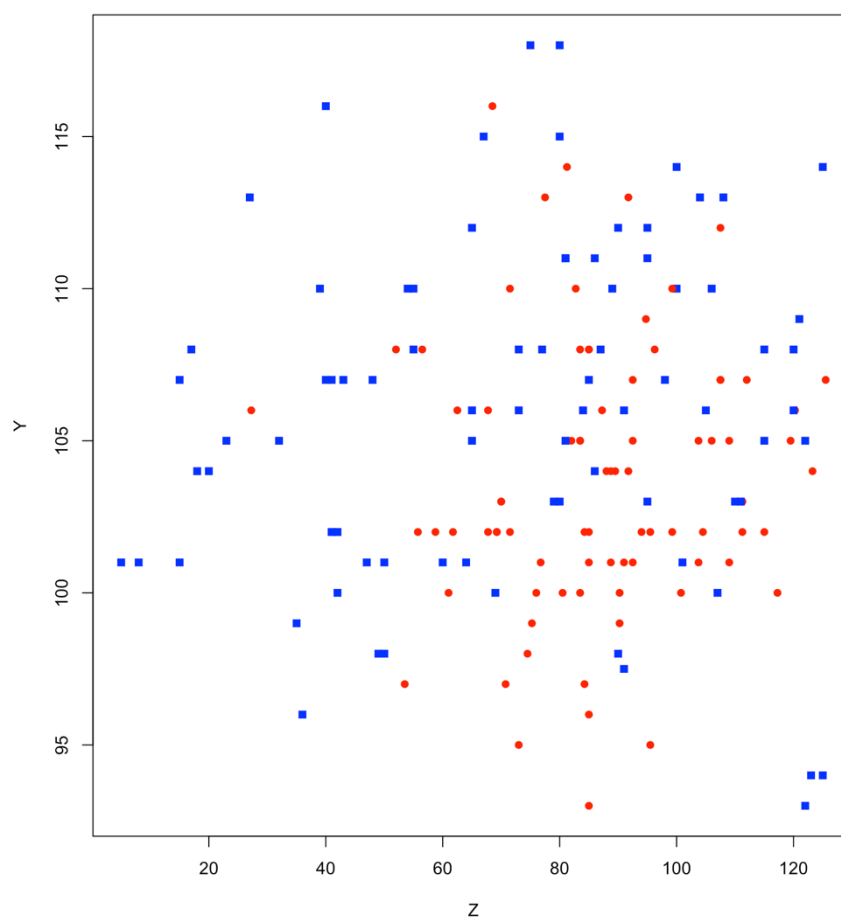


図 34. X13とYの散布図をX14で層別した散布図

それぞれ、青:1号機/赤:2号機としている。

上記からわかることを以下に示す。

1. 青で示される1号機では、プロット範囲に広範的に分布している。
2. 赤で示される2号機では、シード時間50以上の範囲内で広範的に分布している。
3. 上記より、分布に差があるため、設備の違いがシード濃度に寄与していることがわかる。

(オ) 特性要因図

上記の特性要因図を踏まえて特性要因図を作成する。今回は直径を主軸とし、設備,シード,日時,原料,シード後に特性を設定した。その中で、要因を分類し影響を与えると判断できたものを黒枠として示した。

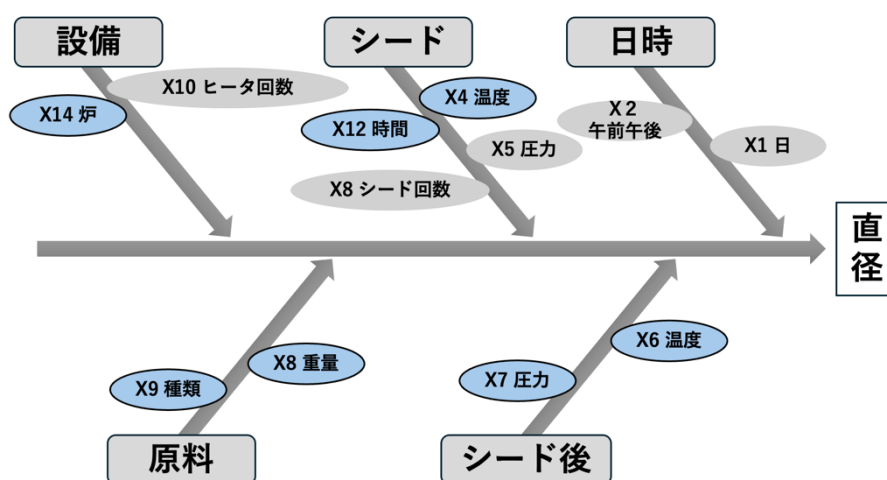


図 35. 特性要因図

※上記の特性要因図はヒストグラム,時系列グラフ,散布図から判断し作成した。

(カ) 回帰分析

上までの解析により調べた品質特性のうち X4,6,11,12 が影響すると判断できた。そして、複数の層別図より、X4 と X9, X5 と X9, X7 と X9, X8 と X9, X10 と X9, X4 と X14 の交互作用もあるように判断できた。

よって、それらを説明変数とする回帰分析を行うこととした。以下にその結果を示すこととした。

表 2. 分散分析表 1

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	49.1800678	21.7588434	2.26023355	0.02525597
X4	3.23517484	0.54898297	5.89303311	2.43E-08
X6	-1.37135553	0.31999168	-4.28559746	3.26E-05
X11	-1.11464519	0.33919961	-3.28610395	0.00126673
X12	0.01807485	0.00903426	2.00070044	0.04724123
X4X9	-0.37548457	0.23114543	-1.62445166	0.106393
X5X9	0.00044247	0.00473655	0.09341684	0.92569791
X7X9	0.0032982	0.00217977	1.51309528	0.13237368
X8X9	0.01307967	0.01026933	1.27366361	0.20476592
X10X9	-0.02524065	0.0273056	-0.92437629	0.35678504
X4X14	-0.19420482	0.02792166	-6.95534711	1.03E-10

※上記の回帰分析は上記を説明変数と仮定して実施した。

この表より、X4,X6,X11,X12,X4X14 が有意であるとわかった。このとき、有意にならなかった説明変数を除いてもう一度回帰分析を実施する。次は説明変数を、X4,X6,X11,X12,X4X14 とする。

表 3. 分散分析表 2

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	59.4451837	16.1264142	3.68619973	0.00031487
X4	3.06493764	0.35355911	8.66881253	5.68E-15
X6	-1.3567874	0.32740559	-4.1440571	5.61E-05
X11	-1.3983442	0.35940116	-3.8907616	0.00014835
X12	0.02505165	0.00954735	2.62393764	0.00956683
X4X14	-0.2184626	0.02778129	-7.8636608	6.08E-13

前ページの分散分析表では全て有意となったため、分散分析表により説明変数を上記の 5 つとする回帰式が妥当であると言える。次に、ANOVA を用いたモデルの評価を行う。

表 4. anova を用いた分散分析表

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	159	4236.64844				
2	154	2205.93202	5	2030.71642	28.3535782	2.65E-20

上記の 5 つを説明変数としたモデルを 2 行目に示す。

上記の分散分析表を見ると、F 値より有意であると言えるため上記のモデルが anova の分散分析より、有意と判断できる。

最後に、回帰診断を用いてモデルを可視化させて評価する。以下の図は R を用いて回帰診断をした結果である。

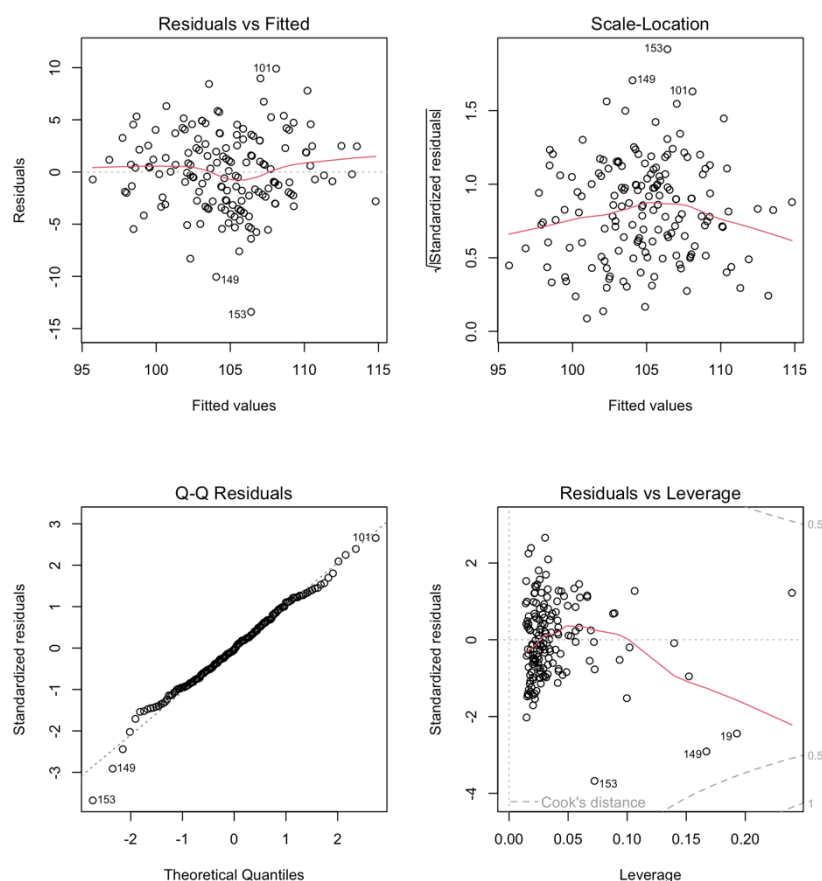


図 36. 回帰診断結果

※R による 5 つの説明変数を持つ回帰分析の回帰診断結果

前ページの回帰診断を見ると以下のことがわかる。

1. Residuals vs Fitted において分布に偏りがなく等分散であると判断できる。
2. Residuals vs Fitted において直線的であり説明変数と従属変数の関係が線形であると示唆している。
3. Scale-Location において、等分散性、線形性を満たしている。
4. Q-Q Residuals において、対角線に沿って直線的に伸びている。
5. Residuals vs Leverage においてプロットのほとんどが左側に多く分布しており、残差がモデルに対して大きな影響を与えていないとわかる。
6. 外れ値 19,149,153 の影響が大きく、直線が右肩下がりとなっている。

上記より、外れ値の影響が大きかった。そのため、外れ値を除いて再度実施したところ、直線系が水平に近づいたため、回帰診断により有意なモデルであるといえた。よって、回帰式は以下のように書くことができる。

$$y = 59.44 + 3.06 * X4 - 1.36 * X6 - 1.40 * X11 + 0.03 * X12 - 0.22 * X4X14$$

4. 考察

まず、品質特性の改善についての考察を行う。上記の分散分析、回帰診断等の解析により、シード温度、シード後温度、シード回数、シード時間、シード温度と加熱炉の交互作用が大きく影響を与えていることがわかった。本解析に用いたデータにはシード情報以外にもヒータや時間などの情報もあったものの、有意に影響を与えると判断できた説明変数がシードに関することが多かったことから、品質特性を効率的に改善させるためにはシードの改善が最も効果を与えると考察できる。特に説明変数の係数より、下記の優先順位で改善することが効率的であると考察する。

1. X4(シード温度)を従来よりも上昇させる。
2. X11(シード回数)を従来よりも少なくする。
3. X6(シード後温度)を従来よりも低くする。

次に、データの可視化とモデルの当てはめの役割について考察する。本実験では、データの可視化、モデルの当てはめそれぞれで複数の手法を用いた。具体的には、可視化ではヒストグラム、時系列グラフ、散布図などである。そして、モデルの当てはめでは回帰分析、診断を行った。データの可視化では視認できる大まかな特徴を掴むことや、手法の設定の前準備などを目的として実施する。本実験では解析手法がある程度決まっている状態だったが、従来の問題解決では目的がわかっており、手法などは解析者に依存する形となる。つまり、データを取得したときに、どのような形式となっているかを判断し解析手順を定める必要がある。例えば、今回のようにヒストグラムを用いてデータを可視化し、交互作用の有無を判定した。この時、二つ山型となっている場合は散布図や時系列グラフを用

いて変化,特徴を掴む手法を選択した。このように、モデル当てはめの前段階として、探索的データ分析のように理解を深めるために実施すると考察する。

一方、モデルの当てはめは可視化されたデータの判断を統計的に有意であると言右ことを目的に実施される。工場での問題解決の場合、改善策を施行するのに巨額の予算を必要とする場合が多い。そのため、手法の信頼性が担保されている必要がある。ただし、データの可視化のみでは数値を伴わない解析者の主観的評価にすぎない。したがってモデルの当てはめにより数値的に,客観的に評価し施策を統計的に評価することを目的に実施すると考察する。

このように、それぞれのセクションで目的が異なるため、どちらかのみを実施するのではなく双方を行い、効率的に問題解決を行うべきであると考ええる。

5. 結論

本研究ではシリコン結晶の直径が基準値を超えることを目的とする実データに近いデータを用いた解析を行った。本解析により、改善策を提示でき統計的に評価することができた。今後の展望としては、データの取得を逐次取得し続け PDCA サイクルを高速化することで品質向上を安定させることとする。

(10,938 文字)

6. 参考文献

1. データサイエンス実験 A,中央大学理工学部ビジネスデータサイエンス学科,pp.3-6,2024.
2. 白山晋,「可視化から何がわかるのか」,システム創成学第二回学術講演会,東京大学人工物工学研究センター,
(https://www.panda.sys.t.u-tokyo.ac.jp/SysInnov2009/A-3_Shirayama.pdf),
2024/10/29 取得.