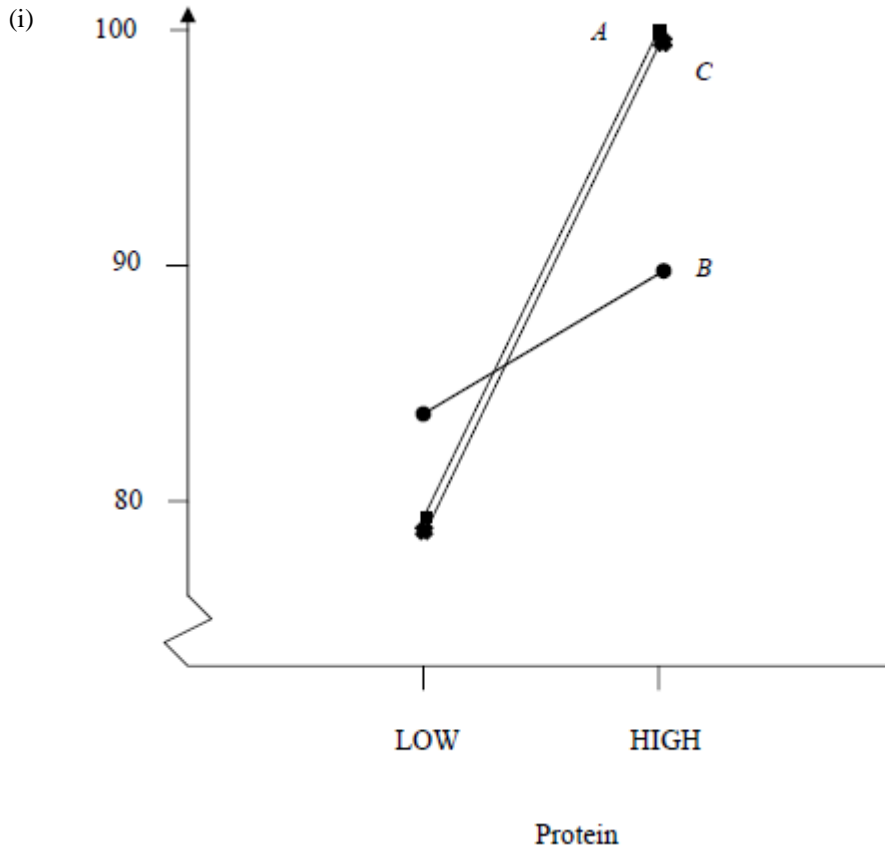


Question 1



交絡とは因子（ここではたんぱく質の量と種類）が独立に機能するように思われないときをいう。ここでは全ての種類 A, B, C で、量を多くすると平均の体重増加量も増加する（ B のふるまいは A や C とかなり違うが）。交互作用はあっても大きくはなさそうである

(ii) 分散分析表は以下のようなものである。「イタリック体」で書いてあるのは問題文で与えられたもので、他は計算する必要がある。

SOURCE	DF	SS	MS	F value
Level	1	<i>3776.3</i>	3776.30	17.60 Compare $F_{1,54}$
Type	2	<i>82.5</i>	41.25	0.19 Compare $F_{2,54}$
Level * Type (Interaction)	2	<i>730.1</i>	365.05	1.70 Compare $F_{2,54}$
Error (Residual)	54	11586.0	214.56	$= \hat{\sigma}^2$
TOTAL	59	<i>16174.9</i>		

$F_{1,50}$ および $F_{2,50}$ の上側棄却点は王立統計協会の(試験中に使用できる)数表から得られる。

(1, 54) や (2, 54) における棄却点はこれらに非常に近い値である.

	5%	1%	0.1%
$F_{1,50}$	4.03	7.17	12.22
$F_{2,50}$	3.18	5.06	7.96

たんぱく質の量 (level) に関する F 値は非常に高有意である. たんぱく質の 2 種類の量 (LOW, HIGH) では, 体重の増加量の平均は等しくないという非常に強い根拠が認められる. たんぱく質の種類に関する F 値は有意ではない. したがって 3 つの種類 (A, B, C) において, 体重の増加量の平均は異なっていると示す根拠がない.

同様に, 交互作用はある (つまり (たとえ B の応答が A や C の応答と多少異なっても) たんぱく質の量が増えるにつれ各たんぱく質の種類によって異なるふるまいをする) という根拠はない. (i) のグラフではパターンが見られたが, ここでの分析により, 変動の重要な原因が何であるのかが確認される.

(iii) 因子 (level と type) と交互作用 (合わせると自由度は 5 である) は全変動 (総平方和) のわずか 28.4% しか説明していない. これはかなり小さい.

実験誤差の推定値は $\hat{\sigma}^2 = 214.56$ ($\hat{\sigma} = 14.65$) であり, 観測値と比べかなり大きいことにも注意せよ.

おそらく単に, 体重の増加量に大きなばらつきが自然にあるためか, あるいは他の共変量 (たとえば始めの体重) から影響を受けているかもしれない.

体重の増加量はたんぱく質の量に強く依存していそうだと直感的にわかる. また, 応答変数についてもっと調べる必要があるのかもしれない.

Question 2

- (i) 機械が正常に稼働しているとき $X \sim N(22.5, 0.3^2)$ である. したがって $\bar{X} \sim N(22.5, 0.3^2/3)$, すなわち, \bar{X} は $N(22.5, 0.03)$ に従う.
- (ii) 一般に警戒限界 (warning limits) と処置限界 (action limits) には2つの工程が用いられる.

1つ目 (時々イギリス方式と呼ばれる) は, 警戒限界を「(目標値)±1.96 (標準誤差)」, 処置限界を「(目標値)±3.29 (標準誤差)」と定める. これらは正規分布の正確な両側5%点及び両側0.1%点を使うので, 機械が正常に稼働しているときの標本平均のそれぞれ5%, 0.1%を除外する限界になっている. 2つ目 (時々 (現代) アメリカ方式と時々呼ばれる) は, 前述の1.96と3.29をそれぞれ2と3に置き換えたものである. (このように, ここでの値3に対応する「シックス・シグマ」は, およそ0.25%を表している.)

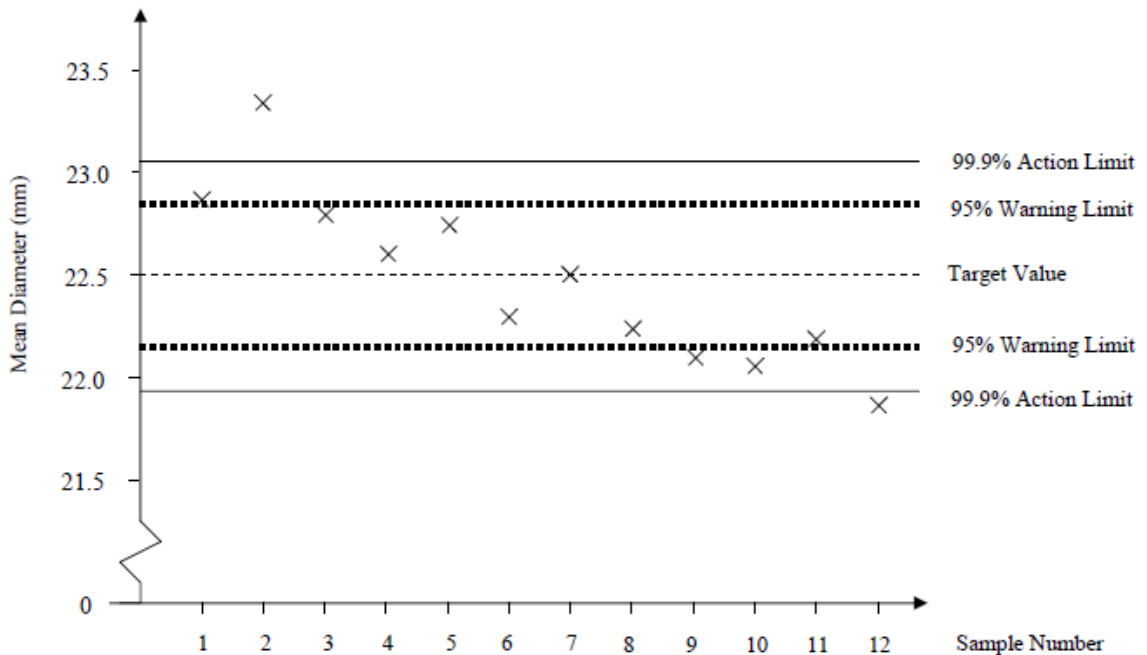
この工程では, 標本平均が警戒限界線の外側だが処置限界線の内側に得られたら, 通常の手順は標本抽出を続ける. そして次の標本で, 平均が警戒限界線の内側になるならば行動はとらないが, 警戒限界線の外側ならば工程を止めて点検する. 標本平均が処置限界線の外側にあるならば, 直ちにその工程を止める.

1.96と3.29を用いると, 各限界は以下のように得られる.

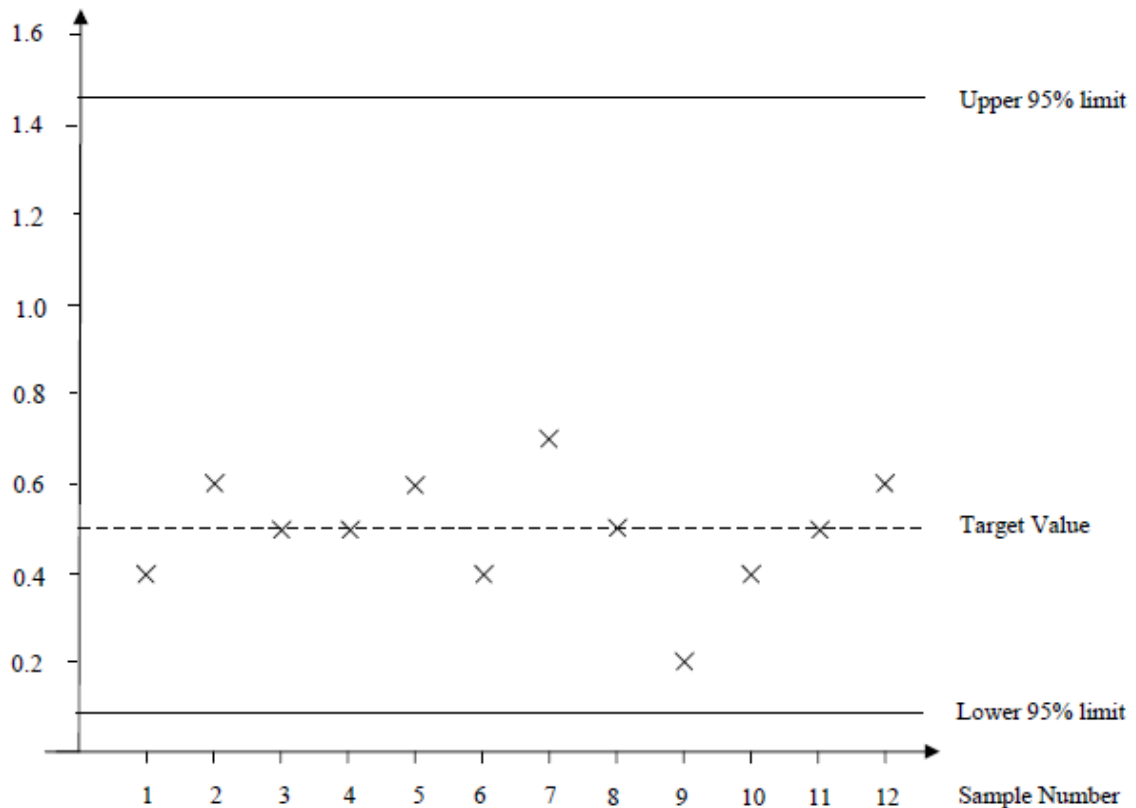
警戒限界 : $22.5 \pm 1.96\sqrt{0.03} = 22.5 \pm 0.34 = 22.16, 22.84$

処置限界 : $22.5 \pm 3.29\sqrt{0.03} = 22.5 \pm 0.57 = 21.93, 23.07$

ドリル加工における3つの値の平均値に関する管理図



(iii) ドリル加工における3つの値の範囲に関する管理図



(iv) 機械の問題は平均のふるまいによるものであり、ばらつきによるものではない。

平均値に関する管理図では、1番目のサンプルが（ほぼ正確に）警戒限界線上にあり、それだけでは工程を停止するのに十分でないが、2番目のサンプルが処置限界をはるかに超えているのでこの時点で工程が停止されたはずであり、機械は調整される。さらに図を見ると、9番目と10番目両方のサンプルの平均値が警戒限界を下回って工程が停止されるまで、平均に関してかなり安定してドリフトが下降している。12番目以降のサンプルの平均値は処置限界を下回る。

全体的にばらつきが管理されているにもかかわらず、明らかに平均値は管理されていない。（これらの単純な基本図と検査には、より複雑な（パターンをもつ点列に基づいた）停止ルールが追加されることがある。）

Question 3

「標準的な」 5×5 のラテン方格の表（たとえば Fisher-Yates の表）が 56 個ある。これらの方格の第 1 行と第 1 列にはアルファベット順に文字が並んでいる。まず、これらのからランダムに表を 1 つ選び、（ランダム順列の表を用いるか他の方法で）行の順序と列の順序をランダムに変え、処理を文字に割り当てればよい。

「修正項」は $1670^2/25 = 111556.0$ である。よって各平方和は以下のようになり、それぞれ自由度は 4 である。

全観測値： $112754 - 111556.0 = 1198.0$
 被験者： $558370/5 - 111556.0 = 118.0$
 実施日： $558092/5 - 111556.0 = 62.4$
 検査： $562750/5 - 111556.0 = 994.0$

残差平方和はこれらの引き算で得られる。したがって、分散分析表は以下のようになる。

SOURCE	DF	SS	MS	F value
Subjects	4	118.0	29.500	15.00 Compare $F_{4,12}$
Days	4	62.4	15.600	7.93 Compare $F_{4,12}$
Tests	4	994.0	248.500	126.36 Compare $F_{4,12}$
Residual	12	23.6	1.967	$= \hat{\sigma}^2$
TOTAL	24	1198.0		

$F_{4,12}$ の上側棄却点は 5% のとき 3.26, 1% のとき 5.41, 0.1% のとき 9.63 である。

適性検査の真の平均値の間に違いがあるという極めて強い根拠があり、同じ能力を測定していないことを示している。適性検査ごとの観測値の平均は、昇順に並べると

E 60.8 D 62.0 A 66.2 B 66.4 C 78.6

このそれぞれが 5 つの観測値の平均値である。2 つの平均値間の最小有意差は

$$t \sqrt{\frac{2 \times 1.967}{5}} = 0.887t$$

で得られる。ここで 1.967 は分散分析から得られる残差平均平方、 t は t_{12} 分布から得られる棄却点（5% のとき 2.179, 1% のとき 3.055, 0.1% のとき 4.318）を表す。これにより、最小有意差は 5% のとき 1.93, 1% のとき 2.71, 0.1% のとき 3.83 と求まる。以上から、適性検査 E と D は似た結果を与え、A と B も似た結果だが E や D よりも高く、C は更に大分高いスコアを得ている、と思われる。

被験者間の差について非常に強い証拠もあるし、実施日間の差についての強い証拠もある。これらは最小有意差の分析を用いて詳しく調べることもできる。慣れによる効果が存在する可能性をふまえ、実施日の平均値がかなり安定して増加しているのが見られそうだ。これは、実施日に関する合計の直線回帰を用いて調べることができる。

Question 4

- (i) 始めに Y の x_1, x_2, \dots, x_r への重回帰分析を行う。これより得られる残差平均平方はモデルの誤差分散 σ^2 の最良な推定値である。

その後、 $r - 1$ 個の説明変数 (x_i のうち 1 つだけ除外する) の可能な回帰を調べる。残差平均平方の増加が最小となる回帰が、この段階では“最適”として選択され、変数がモデルからふるい落とされることにより変数が残る。

同じ方法で手順は続けられる。説明変数が $r - 1$ 個のこの“最適”モデルで始まり、それぞれ順々に変数が取り除かれる。 $r - 2$ 個の説明変数を用いた回帰結果を調べて、同じように他の 1 つの変数が取り除かれる。

同じ方法で手順は更に続けられる。実質的に残差平均平方が増えずに、もはや変数を除外できないときに停止する。停止時の評価については様々な基準が使われうる。

ある変数をいったん取り除いたら、再び使われることはないことに注意せよ。これは、この手順の潜在的欠点である。早い段階でふるい落とされた変数は、他の少数の変数との組み合わせによっては、後進変数除去法で続けて調べられるモデルよりも実際にはよりよいモデルとなるかもしれない。

- (ii) この問題に後進変数除去法を適用すると、表の R^2 や F 列を明示的に使用することはない。これらの列は一般的に標準的なコンピュータの出力の一部として与えられる。残差平均平方の列に焦点を当てる。

4 つの説明変数の完全モデルでの残差平均平方は 4.983。各説明変数を順に省略していくとき、残差平均平方の増加が最小となるのは x_3 を省略したモデルにおいてである (これは x_4 か x_2 を省略したモデルよりも若干よい)。よってこの段階で x_3 を除外し、 x_3 を含まずに次に他の 1 変数が省略される 2 変数モデル (つまり、2 つの説明変数の組 (x_1, x_2) と (x_1, x_4) と (x_2, x_4)) を調べる。この中では明らかに (x_1, x_2) が最適なので、 x_3 と同じように x_4 を取り除く。今は順に 1 つずつを削除することを検討すべきだが、得られたただ 1 つの変数モデルが残差平均平方の非常に大きな増加を導くので好ましくないだろう。従って適切な回帰モデルは x_1 と x_2 の 2 変数モデルである。

- (iii) x_4 のみのモデルは、他のどの説明変数 1 つを用いたモデルと比べても残差平均平方が小さい。また R^2 や F 値も最大である。よって最適の 1 変数モデルといえる。しかし、 x_4 は後進変数除去法の第 2 段階で除去されているため、後進除去で得られるそれ以降のモデルでは現れないだろう。(実際には、「最適な」モデルを選択するため、主に R^2 や残差平均平方に基づいたいくつかの基準がある。)