

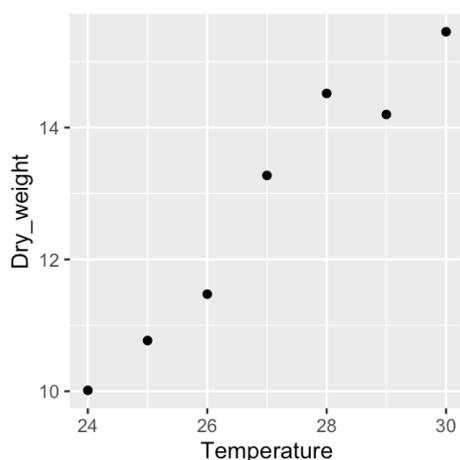
実習試験 2 (大問 2) 【模範解答】

問題 1

問 1 11.48886 10.68633 10.01386 13.30461

問 2 p -value = 0.003406

問 3



問 4 0.9752806

問 5 pH の p value が 0.060566 であるため、線形モデル $Y \sim aX + b$ の係数 a が 0 であることを否定できない。従って pH の変化に伴って培養後の乾燥重量が増減する、とは判断できない。

問 6 温度の p -value が $1.35e-07$ 、pH の p -value が 0.00233 と 0.05 を下回っておりどちらの変数も乾燥重量の増加を説明できるといえる。

問 7 問 5 では pH の上昇とともに乾燥重量が増加するとは言えなかったが、これは pH と温度が同時に変化することが考慮されていなかったためである。重回帰分析では 24 度的时候、25 度的时候と個別に pH の増加に対して乾燥重量が変化するか分析したため、pH の影響を見出すことができた。

問題 2 【解答例】

問 1 細胞の長軸

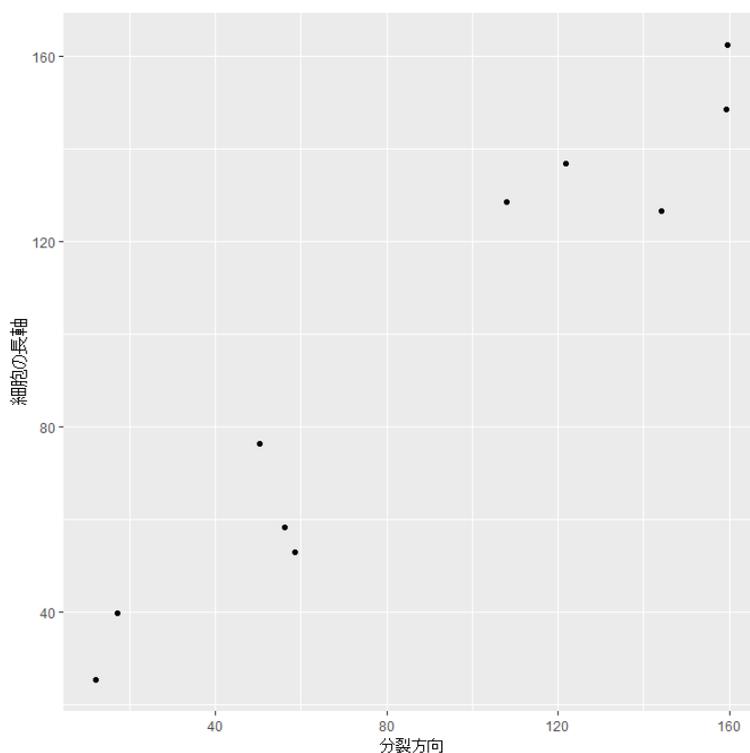
問 2 紡錘体の染色の画像 (X_red.tif) を二値化して紡錘体の領域を抽出する。このときの閾値は 60 とする。次に、Ellipsoid fitting を用いて紡錘体の長軸の角度を測定する。紡錘体の角度の分布をヒストグラムを用いて可視化する。

問 3 110.5°

問 4 細胞の輪郭の染色の画像 (X_green.tif) を二値化して細胞の輪郭を抽出する。このときの閾値は 60 とする。次に、Ellipsoid fitting を用いて細胞の長軸の角度を測定する。細胞の長軸の角度の分布をヒストグラムを用いて可視化する。

問 5 128.8°

問 6



細胞の長軸と分裂方向には相関がありそうだ。

実習試験 2 【出題意図と解説】

問題 1

問 1 表の読み解きができているか。統計データを扱う現代の生物学では基本的な能力。Rで表を読み込むことができるか。read.csv()を使う。df = read.csv()とした場合はxと実行して中身を表示させ、24°Cのデータを全て目視で探しても良いし、df[df\$Temperature == 24,]として抽出してもよい。自分で学習して subset(df, Temperature == 24)などとした場合も正解とする。

問 2 あるパラメタに注目して二つの条件の間に違いがあるのか調べる。data.frameに関する基本的な操作ができるか判定している。問1と同じ操作を30°Cでも繰り返す。次に24°Cの結果、30°Cの結果をベクトルに収められるか。x1 = c(11, 12, 12, 13, 14)などとして、ベクトルを作成する。基本的なt検定ができるかどうかの判定。予習で示したt.test()でx1、x2の二つのベクトル間でt検定をしていたのと同様にt.test(x1, x2)とすれば2群の間の検定ができる。t.test(c(11.48886, 10.68633, 10.01386, 13.30461), c(15.09553, 15.09553, 15.45261, 17.67019))としてもよい。welch検定をしていなくてもよい。もしt.test関数を勉強して片側検定をしていたとしても減点しない。

問 3 相関を調べるときの基本になるのが散布図の作成。偽相関を見抜くなども含めて必要な手順である。pH8のデータを抽出するところは問1、問2と同じ。温度と乾燥重量をベクトルとして抽出した場合は、data.frame(x1, x2)などとしてベクトルをデータフレームに収める必要がある。df2 = df[df\$pH == 8,]で抽出すると作業が早い。ggplotを使ってグラフを作成できるか判定。事前に示した例とほぼ同じ。ggplot(df2, aes(Temperature, Dry_weight)) + geom_point()で作成できる。plot(x = df2\$Temperature, y = df2\$Dry_weight)とすることも可能。

問 4 あるパラメタに注目して相関関係を評価できるか判定。cor.test()を使う。引数はdf2のTemperatureとDry_weightの列を用いる。相関係数の評価も含めて相関係数の理解を観ている。

問 5 回帰分析を行うことができるか判定している。lm関数は事前学習で用いたものと同じで、res = lm(data = df, formula = Dry_weight ~ pH)、summary(res)と実行すれば、Coefficientsについて

```
Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.3014 2.2543 4.126 0.000336 ***
```

pH 0.5832 0.2973 1.962 0.060566 .

という結果が得られる。ここでは $Y \sim aX + b$ の係数、 a と b がそれぞれ 0 であるか見ており、pH の p-value が設定した有意水準の 0.05 を下回らなかったため、 a が 0 であるか判断保留となる。

問 6、問 7 重回帰分析の結果が評価できるか。単回帰分析の結果となぜ違うのか理解できているか判定した。R の操作方法については、問 5 の `lm` 関数の線形モデルを $Y \sim aX + bY + c$ とする以外は同じ。

```
Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.53345 2.90738 -3.279 0.00306 **
Temperature 0.69759 0.09622 7.250 1.35e-07 ***
pH 0.58324 0.17212 3.388 0.00233 **
```

という結果が得られるので Temperature と pH の項の係数が、いずれも 0 ではない、と判断できる。問 7 に関しては、回帰分析においてモデルの違いが何を意味しているか理解しているか観た。t 検定などの検定などの 2 群の比較に偏重せず、回帰分析を適切に使うメリットを感じ取って欲しい。

問題 2

問 1 写真から染色されている細胞の構造を読み解くことができ、そこから細胞の分裂方向と細胞のかたちの関係について、自分なりの仮説を導くことができるかどうか。画像データの解釈は生物学において、非常に重要性が高い能力。

問 2 細胞の分裂方向を評価するために、どのような形態情報を利用するのがふさわしいか、またそれを定量するためには、どのようにして形態情報から分裂方向の情報を抽出することができるのか。形態の定量的な評価は画像データから客観的かつ定量的な情報を抽出するためには、現代の生物学において欠かすことができない能力。分裂方向を評価する方法は複数あるが、恣意的ではなく、ノイズに頑強であり、たくさんのサンプルを扱うことができる方法が優れた方法だということができる。

細胞が分裂する際には染色体が赤道面に整列し、紡錘体と結合する。その後紡錘体が染色体を引っ張ることによって染色体が分配される。このことに着目すると、たとえば紡錘体の配向を計測することにより細胞の分裂方向を評価することができる。

問 3 ImageJ を使って実際に形態データから定量的な情報を得ることができるか。たとえば、紡錘体の配向を計測する場合には、紡錘体を二値化して Ellipsoid fitting することにより紡錘体の長軸方向の傾きを測定することができる。

問 4 問 2 と同様に、細胞のかたちを形態情報から抽出することができるのか。細胞の長軸に注目する場合には、細胞の輪郭のデータを二値化して Ellipsoid fitting することにより細胞の長軸の配向を測定することができる。

問 5 問 3 と同様に、ImageJ を使って実際に形態データから定量的な情報を得ることができるか。

問 6 形態情報の定量データから、細胞の分裂方向と細胞のかたちの間の相関関係を評価することができるか。角度データを扱うことから、どのようにデータを整理して相関を評価できる形 (ex. 散布図など) に落とし込むことができるかがポイントとなる。