# MASM22/FMSN30: Linear and Logistic Regression, 7.5 hp
# FMSN40: ... with Data Gathering, 9 hp

## Lecture 3a, spring 2023
## Multiple linear regression - general and interaction

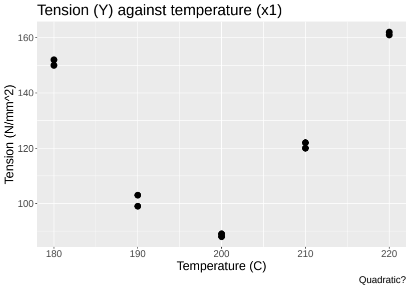
Mathematical Statistics / Centre for Mathematical Sciences
Lund University

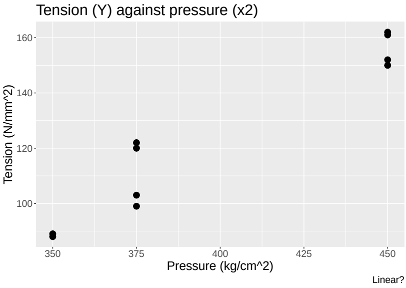
27/3-23

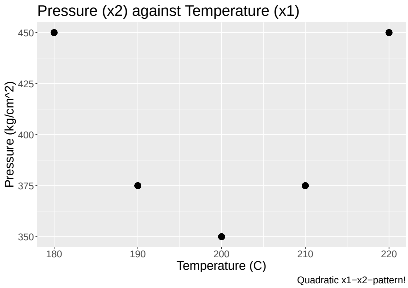
# Multiple regression: Example

Module of elasticity as a function of pressure and temperature:
Temperature and pressure and resulting tension in 10 plastic parts;

| Tension ($Y$) (N/mm$^2$) | Temperature ($x_1$) ($^\circ$C) | Pressure ($x_2$) (kg/cm$^2$) |
|---|---|---|
| 152 | 180 | 450 |
| 150 | 180 | 450 |
| 103 | 190 | 375 |
| 99 | 190 | 375 |
| 88 | 200 | 350 |
| 89 | 200 | 350 |
| 122 | 210 | 375 |
| 120 | 210 | 375 |
| 162 | 220 | 450 |
| 161 | 220 | 450 |




The quadratic relationship between $x_1$ and $x_2$ makes the (marginal) relationship between $Y$ and $x_1$ look quadratic as well!

See `p0` in `elasticity3d.rda` for a rotatable 3D-plot.

- plots of $Y$ vs individual covariates only unveil *partial relationships*. We do not know what happens when other covariates vary together.
- we can discover pairwise relationships between covariates by plotting $x_1$ vs $x_2$
- plotting $x_1$ vs $x_2$ does not say anything about the 3D joint relationship of $(x_1, x_2, Y)$
- if the plot $(x_1, Y)$ is nonlinear, you can perhaps transform $x_1$ and/or $Y$ but again, this is only going to linearize a partial relationship...
- our model (next slide) is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \qquad (*)$$

and in this case even if some **partial** relationships $(x_j, Y)$ are nonlinear, the **entire** surface (*) above is perfectly suitable for the joint relationship.

# Multiple linear regression model

See Lecture 1 for details.

- $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$ for $i = 1, \ldots, n$ where $\epsilon_i \sim N(0, \sigma^2)$ are independent.
- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$.

## Estimates

- Least squares: find the $\boldsymbol{\beta}$ that minimizes $Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$
- The solution satisfies the normal equztions: $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$
- $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \sim N_{p+1}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$
- Fitted values: $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$
- Residuals: $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} =$ observed $-$ predicted
- Residual variance: $\hat{\sigma}^2 = s^2 = \dfrac{\sum_{i=1}^n e_i^2}{n-(p+1)} = \dfrac{\mathbf{e}'\mathbf{e}}{n-(p+1)}$

- A given $\beta_j$ expresses the *effect* of a change in covariate $x_j$ on the expected value of $Y$, *given all other covariates in the model*;
- that is, $\beta_j$ gives the change in $E(Y)$ when $x_j$ increases by 1 units, when all other covariates are kept fixed.
- in other words, $\beta_j$ can only represent the partial (marginal) effect of $x_j$ on $Y$; the effect is *conditional* on what other variables we have in the model.
- The relevance of $x_j$ (hence the relevance of $\beta_j$) can be different if we introduce other covariates in the model.

The latter two concepts will be emphasized when we talk about hypothesis tests later.

## Collinearity (re-discussed later in the course)

- in order to determine $\hat{\boldsymbol{\beta}}$ the matrix $\mathbf{X}'\mathbf{X}$ must be invertible.
- The matrix $\mathbf{X}$ is singular and $(\mathbf{X}'\mathbf{X})^{-1}$ has no unique solution if a linear combination of some of the $x$-variables equals one of the other $x$-variables. Then $\boldsymbol{\beta}$ cannot be uniquely estimated.
- The matrix $\mathbf{X}$ is nearly singular and $(\mathbf{X}'\mathbf{X})^{-1}$ has an unstable solution if a linear combination of some of the $x$-variables almost equals one of the other $x$-variables. The same (almost) information included in several variables.
- $\beta$-estimates will have huge variance.
- Correlated $x$-variables "compete" (one variable might be necessary if the other is not in the model, but not if both are in the model, etc.)
- Found by: plotting all $x$-variables against each other.
- Solution: Use only one of the problematic variables

## Elasticity: estimates

$Y$ = tension (N/mm$^2$), $x_1$ = temperature (°C), $x_2$ = pressure (kg/cm$^2$).
Model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$, $i = 1, \ldots, 10$, $\epsilon_i \sim N(0, \sigma^2)$

| Variable | | est. | s.e. | 95 % C.I. | unit |
|---|---|---|---|---|---|
| intercept | $\beta_0$ | $-215.7$ | $31.9$ | $(-291.1, -140.2)$ | N/mm$^2$ |
| temperature | $\beta_1$ | $0.41$ | $0.13$ | $(0.10, 0.72)$ | $\frac{\text{N/mm}^2}{°\text{C}}$ |
| pressure | $\beta_2$ | $0.65$ | $0.04$ | $(0.54, 0.75)$ | $\frac{\text{N/mm}^2}{\text{kg/cm}^2}$ |
| resid.std.dev | $\sigma$ | $5.90$ | df $= 7$ | | N/mm$^2$ |

Fitted plane: $\hat{Y} = -215.7 + 0.41 x_1 + 0.65 x_2$.

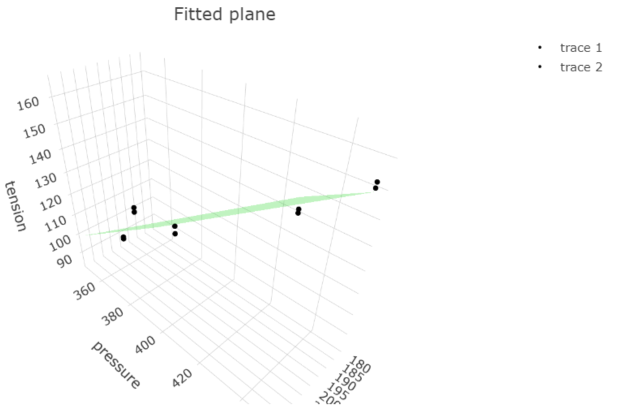
## Effect of temperature change

Increase the temperature by $\Delta_1$ °C from $x_{01}$ to $x_{01} + \Delta_1$
while keeping the pressure fixed at $x_{02}$:

$\hat{Y}_{\text{old}} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02}$

$\hat{Y}_{\text{new}} = \hat{\beta}_0 + \hat{\beta}_1 (x_{01} + \Delta_1) + \hat{\beta}_2 x_{02}$

$\hat{Y}_{\text{new}} - \hat{Y}_{\text{old}} = \hat{\beta}_1 \Delta_1 = 0.41 \Delta_1$ (N/mm$^2$) regardless of the pressure.


Fitted plane
trace 1
trace 2
tension
160
150
140
130
120
110
100
90
pressure
360
380
400
420
180

## Predictions

If temperature = 200 °C and pressure = 400 kg/cm$^2$,
$\mathbf{x}_0 = \begin{pmatrix} 1 & 200 & 400 \end{pmatrix}$

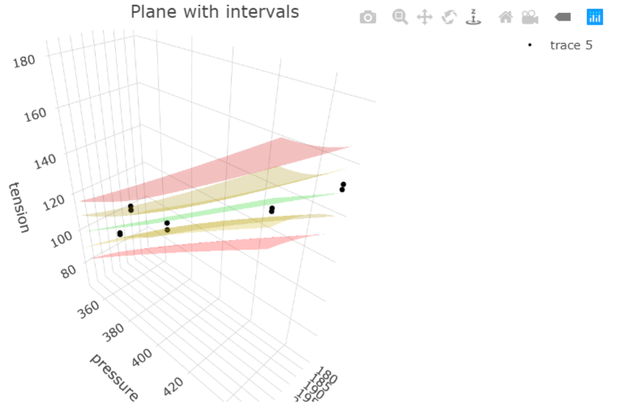
- what is the expected tension?
  $\hat{Y}_0 = \mathbf{x}_0\hat{\boldsymbol{\beta}} = -215.7 + 0.41 \cdot 200 + 0.65 \cdot 400 = 124.6$ (N/mm$^2$).
- What tension values might we observe?
  $\hat{Y}_{\text{pred}_0} = \mathbf{x}_0\hat{\boldsymbol{\beta}} + \epsilon_0 = 124.6 + \epsilon_0$ (N/mm$^2$).

| | estimate | s.e. | 95 % interval | |
|---|---|---|---|---|
| on average | $\hat{Y}_0 = 124.6$ | $1.87$ | $(120.2, 129.0)$ | conf. |
| single obs. | $\hat{Y}_{\text{pred}_0} = 124.6 + \epsilon_0$ | $6.19$ | $(110.0, 139.2)$ | pred. |

Note: $6.19 = \sqrt{5.90^2 + 1.87^2}$.

# Intervals for the plane

Estimated plane (green); confidence interval for the plane (brown) and prediction interval for observations (red).



# Interaction

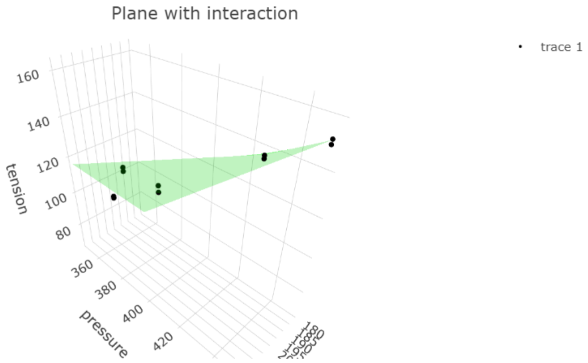
What happens if the effect of a change in temperature depends on the pressure? Add $x_3 = x_1 \cdot x_2 =$ temperature $\times$ pressure

## Model with interaction term

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$, $i = 1, \ldots, 10$, $\epsilon_i \sim N(0,\, \sigma^2)$

| Variable | | est. | s.e. | 95 % C.I. |
|---|---|---|---|---|
| intercept | $\beta_0$ | $-1071.2$ | $221.6$ | $(-1613.4, -529.0)$ |
| temperature | $\beta_1$ | $4.69$ | $1.11$ | $(1.98, 7.40)$ |
| pressure | $\beta_2$ | $2.61$ | $0.51$ | $(1.37, 3.86)$ |
| temp$\times$press | $\beta_3$ | $-0.0098$ | $0.0025$ | $(-0.016, -00036)$ |
| resid.std.dev | $\sigma$ | $3.40$ | df $= 6$ | |

Fitted plane: $\hat{Y} = -1071.2 + 4.69 x_1 + 2.61 x_2 - 0.0098 x_1 x_2$.


Plane with interaction
trace 1
tension
160
140
120
100
80
pressure
360
380
400
420

## Effect of temperature change: interaction

Increase the temperature by $\Delta_1$ °C from $x_{01}$ to $x_{01} + \Delta_1$ while keeping the pressure fixed at $x_{02}$:

$$\begin{aligned}
\hat{Y}_{\text{old}} &= \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \hat{\beta}_3 x_{01} x_{02} \\
\hat{Y}_{\text{new}} &= \hat{\beta}_0 + \hat{\beta}_1 (x_{01} + \Delta_1) + \hat{\beta}_2 x_{02} + \hat{\beta}_3 (x_{01} + \Delta_1) x_{02} \\
\hat{Y}_{\text{new}} - \hat{Y}_{\text{old}} &= (\hat{\beta}_1 + \hat{\beta}_3 x_{02}) \Delta_1 \text{ depends on the pressure!}
\end{aligned}$$

Temperature effect for some different pressures:

$$\begin{aligned}
x_{02} = 350 &: \hat{Y}_{\text{new}} - \hat{Y}_{\text{old}} = (\hat{\beta}_1 + \hat{\beta}_3 \cdot 350) \Delta_1 = 1.24 \Delta_1 \\
x_{02} = 400 &: \hat{Y}_{\text{new}} - \hat{Y}_{\text{old}} = (\hat{\beta}_1 + \hat{\beta}_3 \cdot 400) \Delta_1 = 0.75 \Delta_1 \\
x_{02} = 450 &: \hat{Y}_{\text{new}} - \hat{Y}_{\text{old}} = (\hat{\beta}_1 + \hat{\beta}_3 \cdot 450) \Delta_1 = 0.26 \Delta_1
\end{aligned}$$

Note: without the interaction we always had $0.41 \Delta_1$.

## Predictions: with interaction

If temperature = 200 °C and pressure = 400 kg/cm$^2$,

$\mathbf{x}_0 = \begin{pmatrix} 1 & 200 & 400 & 80000 \end{pmatrix}$

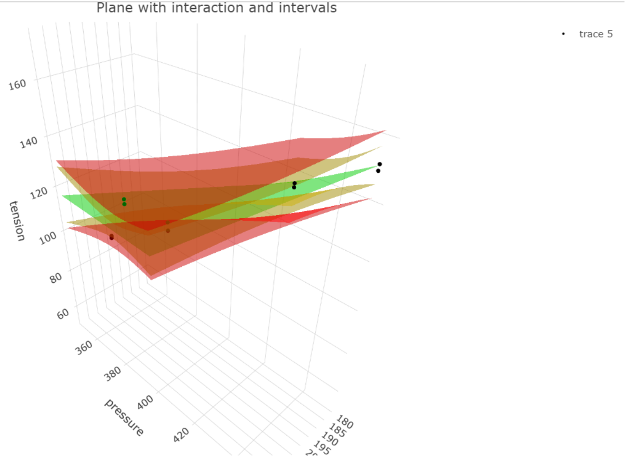
- what is the expected tension?
  $\hat{Y}_0 = \mathbf{x}_0\hat{\boldsymbol{\beta}} =$
  $= -1071.2 + 4.69 \cdot 200 + 2.61 \cdot 400 - 0.0098 \cdot 80000 = 124.6.$
- What tension values might we observe?
  $\hat{Y}_{\text{pred}_0} = \mathbf{x}_0\hat{\boldsymbol{\beta}} + \epsilon_0 = 124.6 + \epsilon_0.$

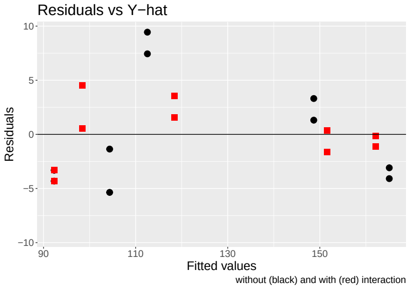
| | estimate | s.e. | 95 % interval | |
|---|---|---|---|---|
| on average | $\hat{Y}_0 = 124.6$ | $1.08$ | $(122.0, 127.2)$ | conf. |
| single obs. | $\hat{Y}_{\text{pred}_0} = 124.6 + \epsilon_0$ | $3.57$ | $(115.9, 133.3)$ | pred. |

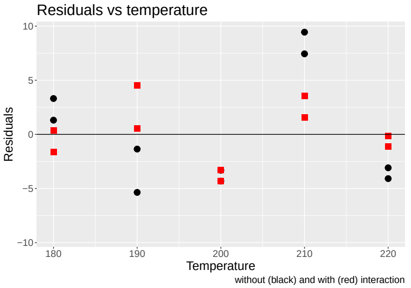
Note: $3.57 = \sqrt{3.41^2 + 1.08^2}$.

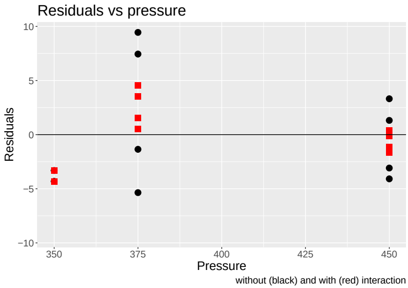
Note that both $\hat{\sigma}$ and all the intervals have become narrower. This model fits closer to the data!

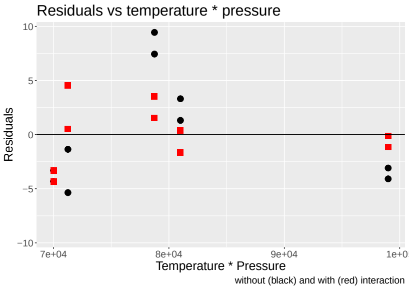
Too close? Overfitting? Next week...

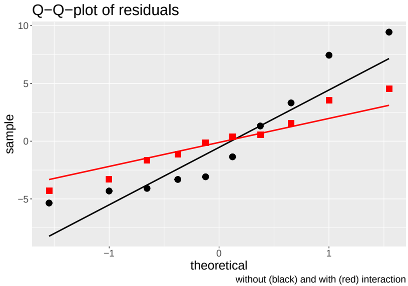

Plane with interaction and intervals
trace 5
tension
160
140
120
100
80
60
pressure
360
380
400
420
180
185
190
195


Residuals vs Y−hat
Residuals
10
5
0
−5
−10
90
110
130
150
Fitted values
without (black) and with (red) interaction


Residuals vs temperature
Residuals
10
5
0
−5
−10
180
190
200
210
220
Temperature
without (black) and with (red) interaction


Residuals vs pressure
Residuals
10
5
0
−5
−10
350
375
400
425
450
Pressure
without (black) and with (red) interaction


Residuals vs temperature * pressure
Residuals
10
5
0
−5
−10
7e+04
8e+04
9e+04
Temperature * Pressure
without (black) and with (red) interaction




## Conculsions from the basic residual analysis

The model with interaction has

- smaller residuals
- residuals closer to a normal distribution