STA302 week 5 continued: 4–9 October

Mark Ebden 2018. Section 3.3

With grateful acknowledgment to Alison Gibbs and Becky Lin

This week's lecture content will include:

- How to succeed in Assignment 1
- Transformations; reference: §3.3

Assignment 1

You should have received an email from Crowdmark last week introducing the assignment. If not, check your spam-filter settings and whitelist *crowdmark*. A few hints were added to the questions in recent days.

The deadline is Fri 12 October at 1 pm, but as explained in the syllabus you may wish to aim for earlier in case the website slows on the final day.

Auditing the course? Feel free to email me for a copy of the assignment to try for fun. Don't submit work.

Draft of marking rubrics to be sent to TA graders

Q1

- 1 mark for a graph of all points and the line
- 1 mark for showing R code with comments explaining what the lines do (it's ok to skip comments for lines that are very very obvious)

Q2

- 1 mark for stating the equation correctly
- 1 mark for showing how it was arrived at (e.g. commented R code, or a scan of something handwritten, or typed work in RMarkdown, etc)
- 2 marks for stating the hypotheses, (see e.g. step 1 of page 3 here), running the hypothesis testing, and clearly interpreting the result
- -1 mark for an incorrect number of significant figures in the final answer

Q3

- 1 mark for the final answer
- 5 marks for setting up the problem correctly and executing clearly
- Partial marks are available for showing accurately calculated, useful intermediate results — provided they are easy for TAs to follow

Draft of marking rubrics to be sent to TA graders

Q4

- 1 mark for arriving at approximately the correct probability (the range used by TAs will be published when solutions are posted)
- 4 marks for setting up the problem correctly, explaining steps clearly, and arriving at the exact answer
- This precludes simply listing some R commands without explanation of what the calculations are
- -1 mark for an incorrect number of significant figures in the final answer

Q5

- 1 mark for stating the correct response
- 2 marks for a clear, accurate explanation

Q6

- No marks can be deducted by TAs or instructors for integrity offences
- TAs and I regularly assist the OSAI office to handle offences. For my students in the past two years the office has decided various outcomes, including reduced marks, zero marks, an F in the course, and/or suspension

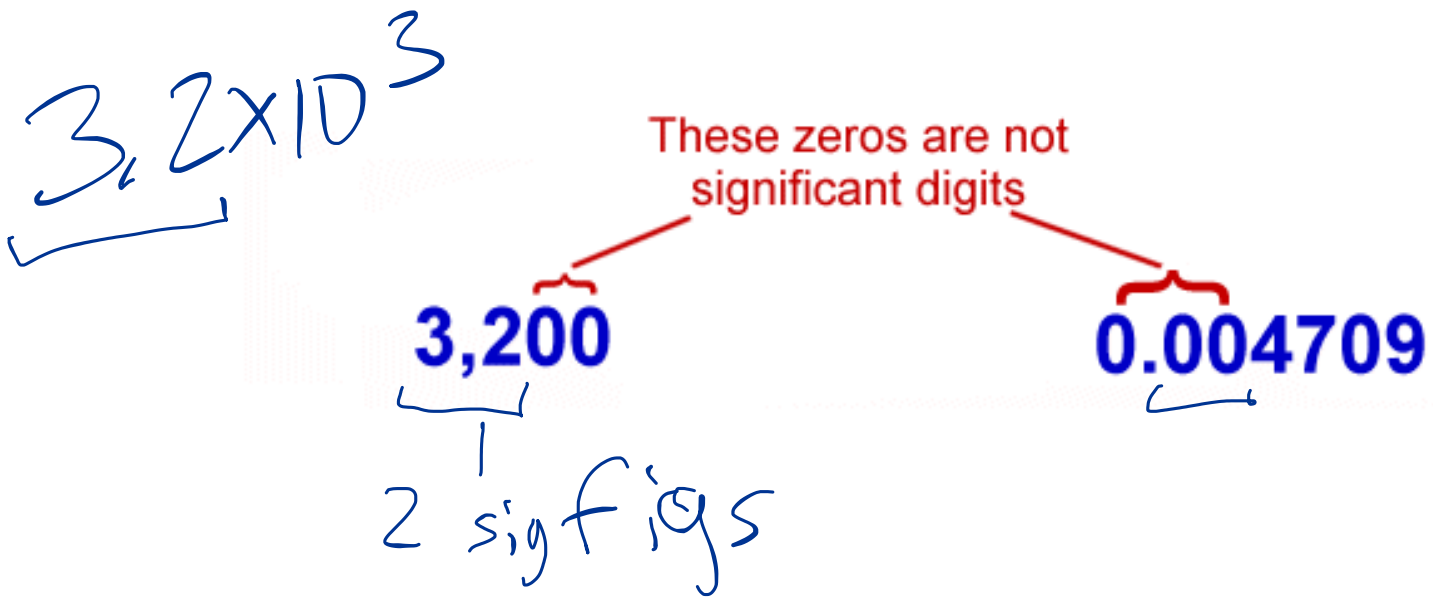
Review of Significant Figures

"Sig figs" (a.k.a. significant digits) include all **nonzero digits**: 1, 2, . . . 9

- e.g. the number 3153 has four sig figs

Zeros are also sig figs when:

- Between nonzero digits. e.g. 102 has three sig figs
- After a nonzero digit which is after a decimal point. e.g. 0.010 has two significant figures



Review of Significant Figures

Ambiguities about zeros exist, e.g. How many sig figs are here:

Scientific notation avoids confusion, but, often you can infer precision from context. This is why we don't write:

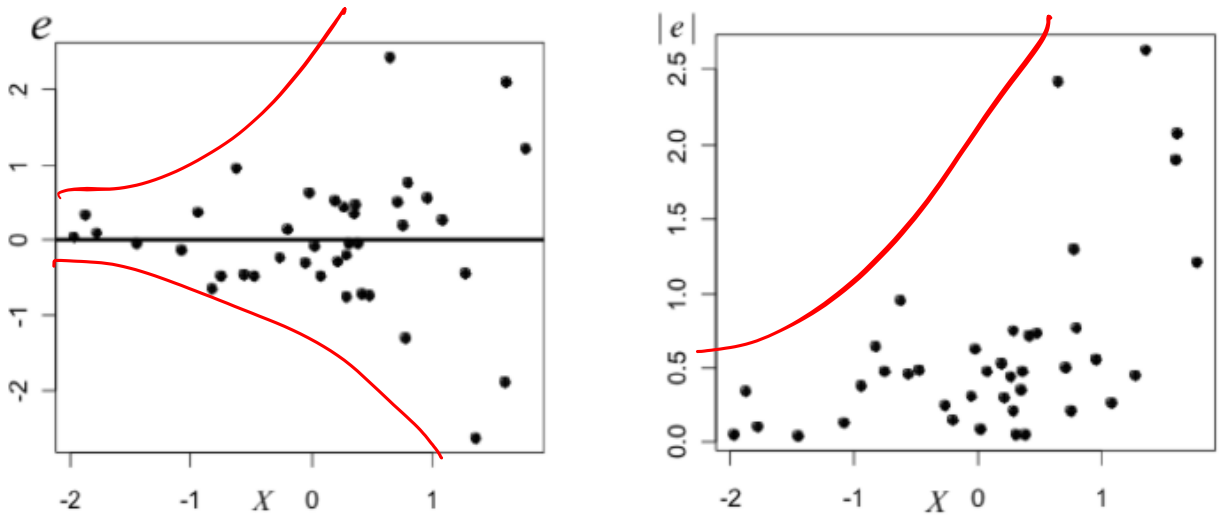
$$4.0 \times 10^{1}\text{th PArTy INVitE}$$

This week's lecture content will include:

- How to succeed in Assignment 1
- **Transformations; reference: §3.3**

Transformations

Recall Check 5, *error homoscedasticity*: sometimes the variance is found to be nonconstant.



There are a few things we can do in this case.

The Delta Method

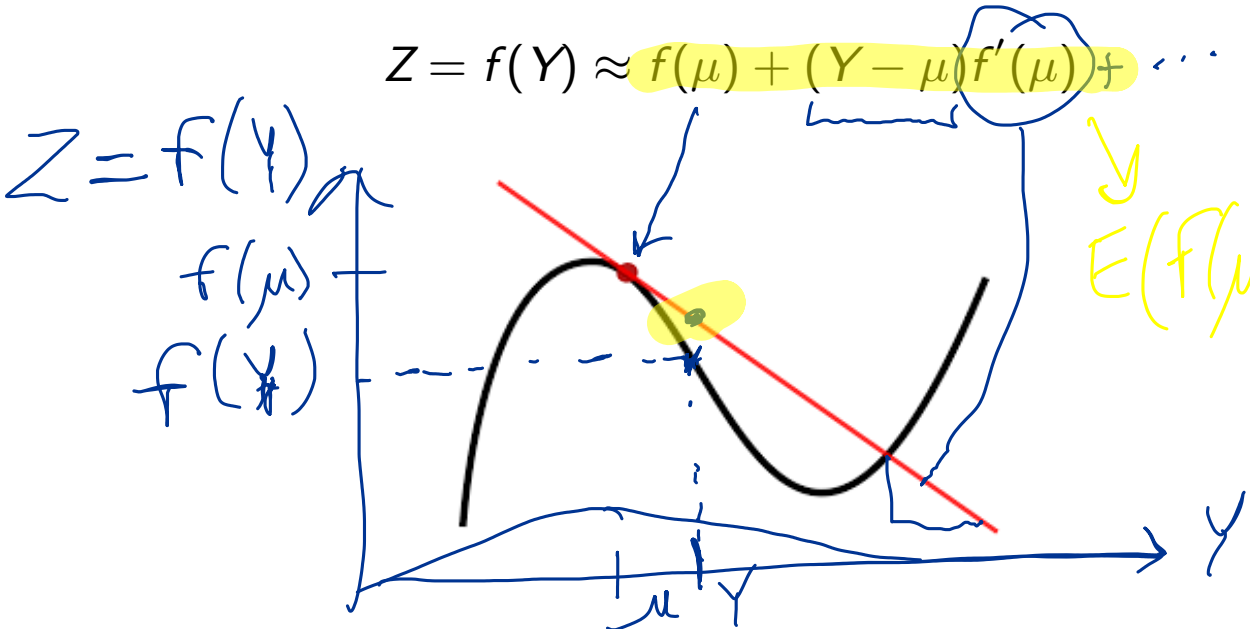
not just SLR

any function

Let Y be a distribution with mean μ and variance σ_Y^2, and let $Z = f(Y)$.

A first-order linear approximation to Z is:

$$Z = f(Y) \approx f(\mu) + (Y - \mu) f'(\mu) + \cdots$$



$E(f(\mu)) + E(Y-\mu) \cdot \text{const.}$ → 0

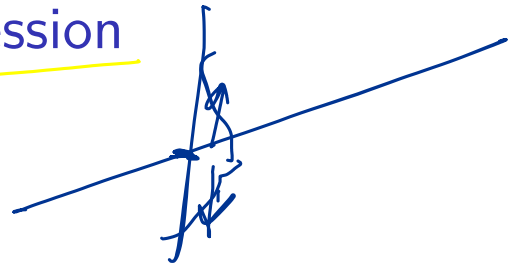
It's fairly easy to show that $\mathbb{E}(Z) \approx f(\mu)$ and that $\text{var}(Z) \approx \sigma_Y^2 [f'(\mu)]^2$.

This completes the Delta Method (in the univariate case). It estimates the mean and variance of a function of a random variable.

$\text{var}[f(\mu)] + \text{var}[(Y-\mu)] f'(\mu)^2 = \sigma_Y^2 f'(\mu)^2$

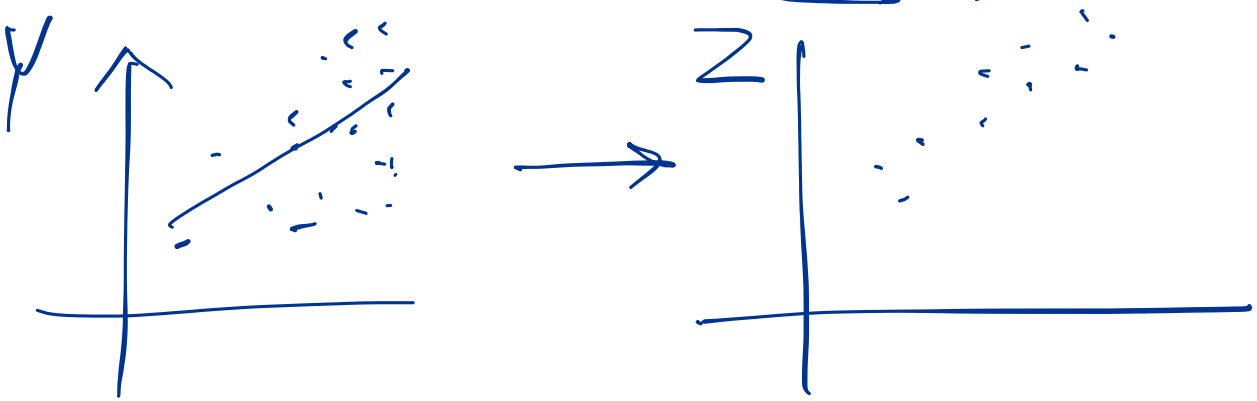
0

The Delta Method in Linear Regression



In SLR, we assume $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i = \mu_i$, and $\text{var}(Y_i) = \sigma^2$ which doesn't depend on i.

However, suppose that Y_i has a variance proportional to a function of μ_i. In other words, say $\text{var}(Y_i) \propto V(\mu_i)$. For our SLR model to continue to work, we want to find a transformation $Z = f(Y)$ so $\text{var}(Z) \approx \text{const}$.



Continued

Using our finding earlier that $\text{var}(Z) \approx \sigma_Y^2 [f'(\mu)]^2$, we want f such that $\text{var}(Z) \propto V(\mu)[f'(\mu)]^2 \approx$ const. So for some constant c, we need

$$[f'(\mu)]^2 = \frac{c}{V(\mu)}$$

So:

$$f'(\mu) \propto \frac{1}{\sqrt{V(\mu)}}$$

$$f(y) \propto \int \frac{d\mu}{\sqrt{V(\mu)}}$$

Key: If you suspect $\text{var}(Y_i) \propto V(\mu_i)$, then try transforming your y_i's as $f(y)$

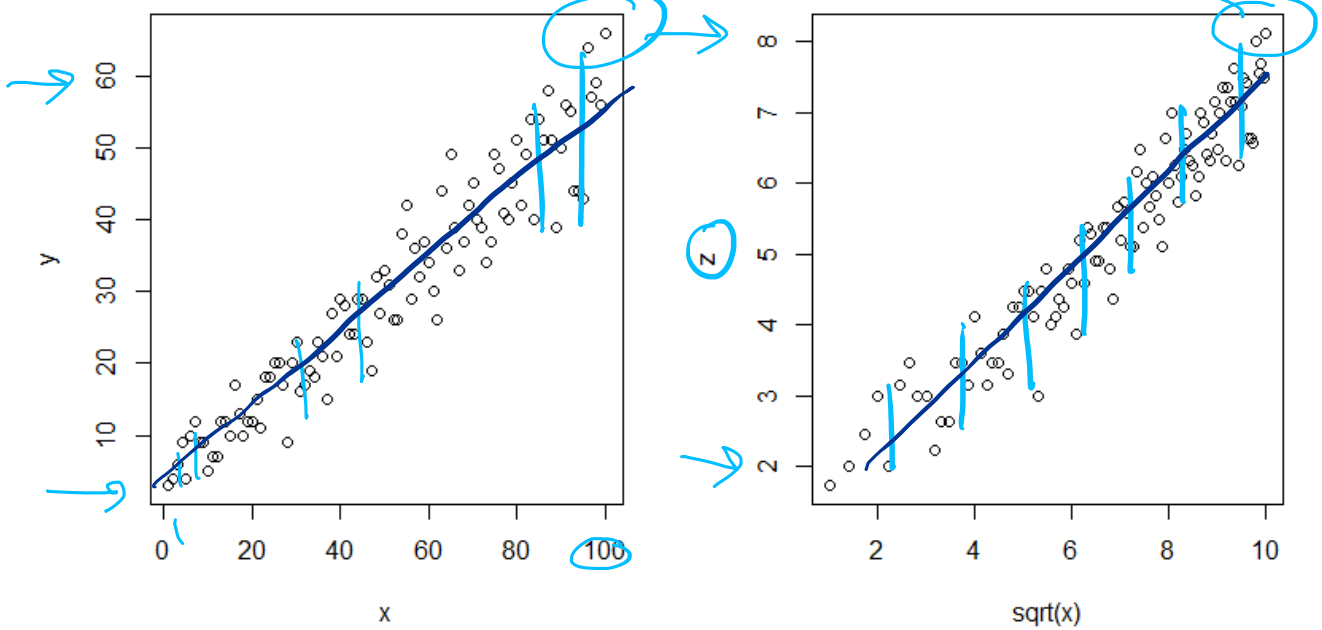
Example 1

Suppose $Y_i \sim \text{Pois}(\mu_i)$. Then $\mathbb{E}(Y_i) = \mu_i$ and $\text{var}(Y_i) = \mu_i$.

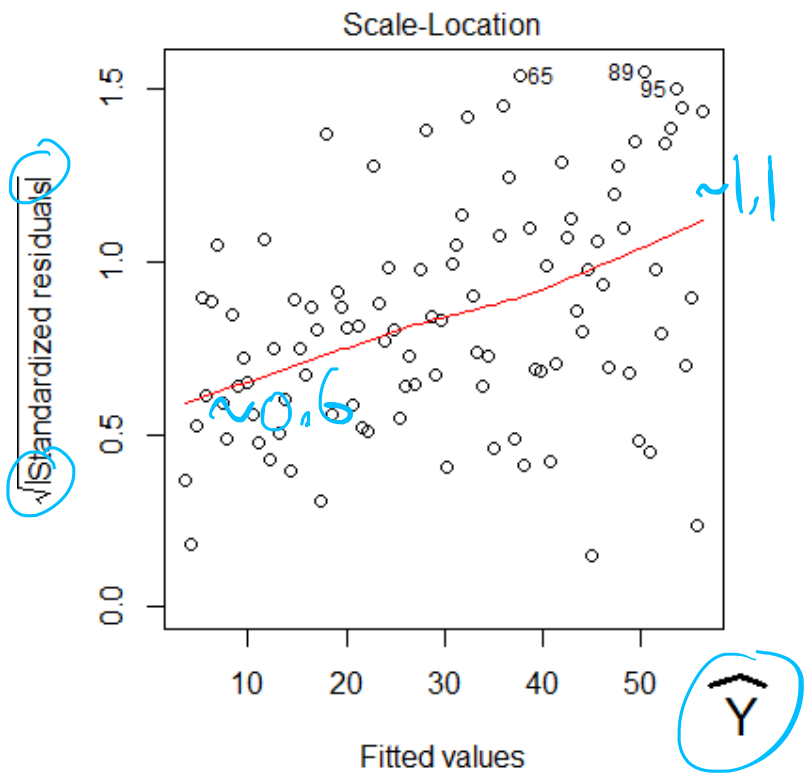
$$V(\mu) = \mu$$

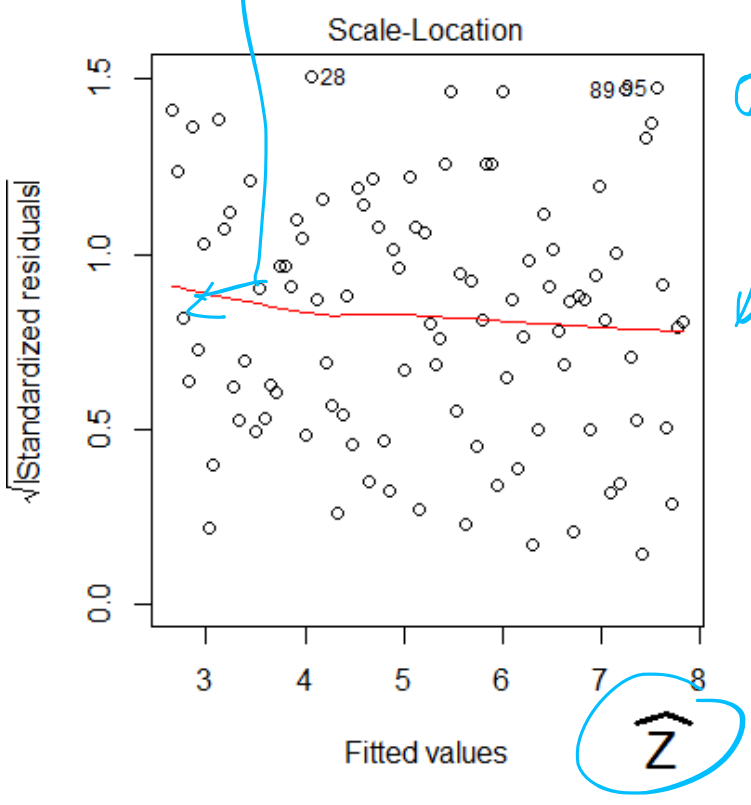
$$f'(\mu) \propto \frac{1}{\sqrt{V(\mu)}} \propto \frac{1}{\sqrt{\mu}} \qquad f(\mu) \propto \int \frac{d\mu}{\sqrt{\mu}} \propto \sqrt{\mu}$$

If $\text{var}(Y)$ is linearly proportional to $\mathbb{E}(Y)$, then $Z = \sqrt{Y}$ has a variance that's approximately constant.



Check 5: From fail to pass





0.8-something

0.7-something

nonconst var.

R code for Example 1

```
N <- 100 # Number of data points to create
beta0 <- 4; beta1 <- .5 # True population parameters
x <- 1:N # Given X values
mu <- beta0 + beta1*x # True line
#y <- mu + rnorm(N,0,sigma) # meets our model assumptions
y <- rpois(N,mu) # lambda can be a vector
z <- sqrt(y) # Transformation
par(mfrow=c(1,2))
plot(x,y); plot(sqrt(x),z)
```

left right

Notice that we plotted `sqrt(x)`. Often, a transformation in Y can help to correct nonconstant variance while a transformation in X can help to correct **nonlinearity**. After transforming Y to correct for nonconstant variance, you can check (again) for nonlinearity. You'll find that when both X and Y are measured in the same units, then it's often natural to consider the same tranformation for both X and Y. More on this later.

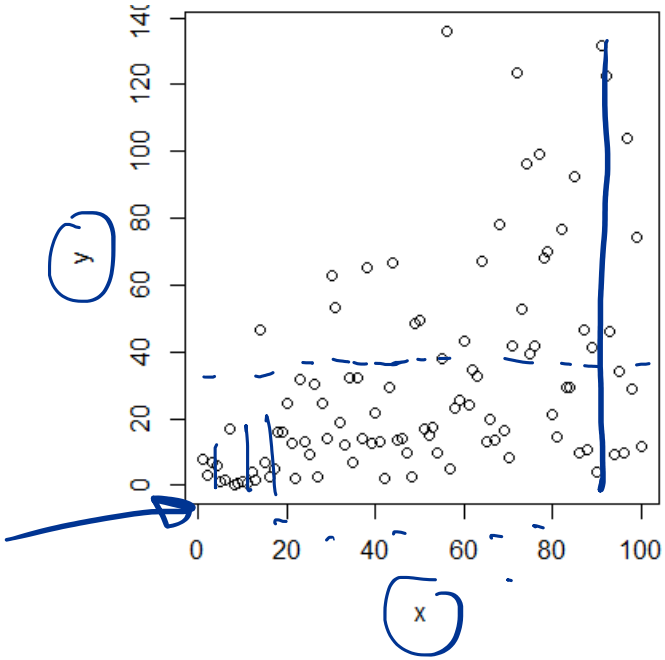
Example 2

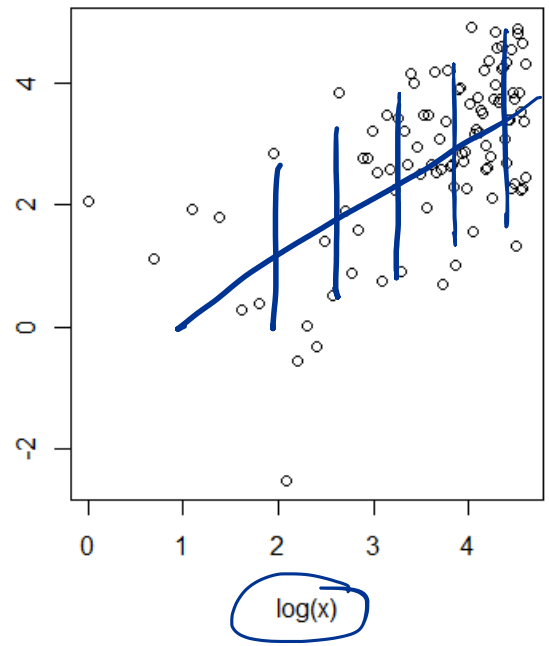
Suppose $Y_i \sim \text{Expon}(\lambda)$. Then $\mathbb{E}(Y) = \mu = \lambda^{-1}$ and $\text{var}(Y) \propto \mu^2$.

$$V(\mu) = \mu^2$$

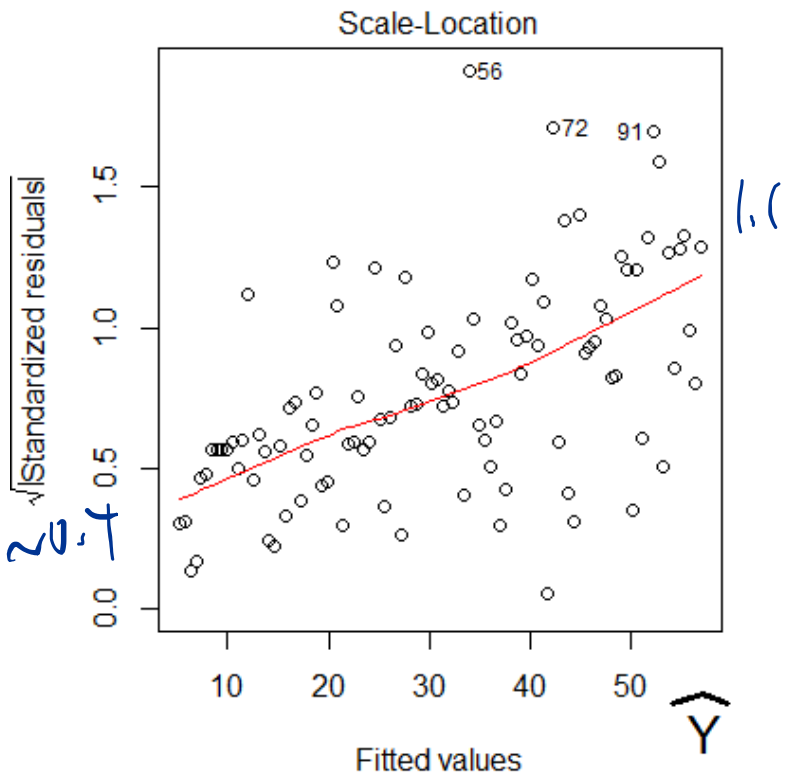
$$f'(\mu) \propto \frac{1}{\sqrt{V(\mu)}} \propto \frac{1}{\mu} \qquad f(\mu) \propto \log \mu$$

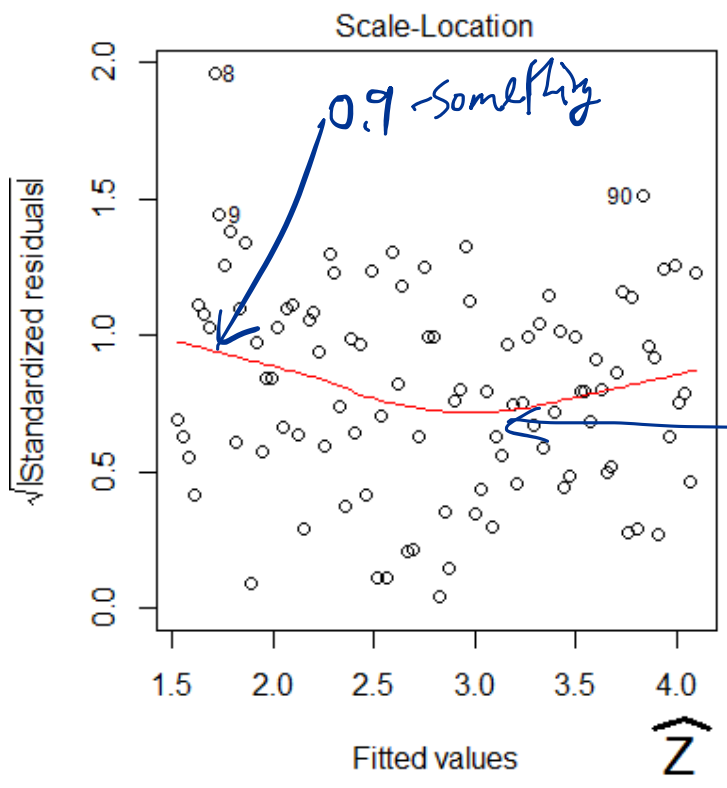
Note: "log" = "ln", natural logarithm, here and in R.





Example 2, Check 5: From fail to pass



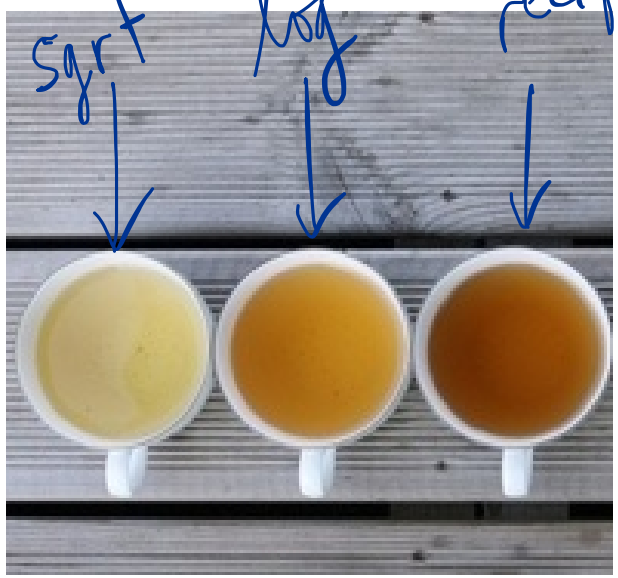


Logarithmic Transformations

In Example 2, var(Y) increased *more quickly than linearly* in $\mathbb{E}(Y)$. Whenever this happens, a log transformation often stabilizes the variance.

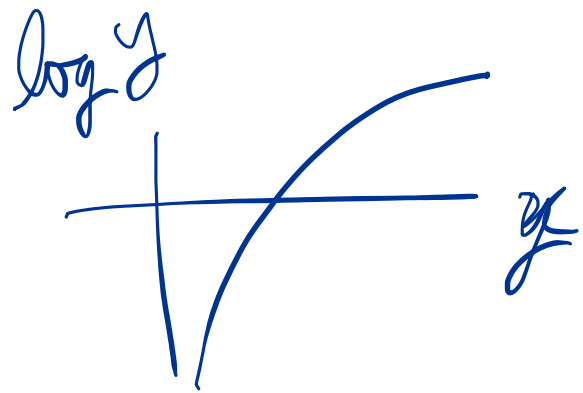
In regression, the most useful transformations include:

- Taking the square root
- `log` is stronger than $\sqrt{}$ (data tend to be affected more)
- Taking the reciprocal, $f(Y) = 1/Y$, is even stronger. Be careful when values are negative or close to zero



Logarithmic transformations

Exercise: How is var(Y) related to $\mathbb{E}(Y)$ if the reciprocal transformation is appropriate?



What if the data include 0's or negative numbers, but a logarithmic transformation seems otherwise appropriate?

Use $\log(Y+k)$, where k is a constant of your choice

Interpreting log-transformed data

If we only transform Y, our new model is

$$\log Y = \beta_0 + \beta_1 x + e$$

model error

$$Y = e^{\beta_0} e^{\beta_1 x} e^{e} \leftarrow \text{Euler}$$

An increase in x of 1 unit is associated with a multiplicative change in Y by a factor of e^{β_1}.

$$\frac{Y_{new}}{Y} = \frac{e^{\beta_0} e^{\beta_1 (x+1)} e^{e}}{e^{\beta_0} e^{\beta_1 x} e^{e}} = e^{\beta_1}$$

Log-transformed data: Electrical example

Suppose we were plotting time-to-breakdown (Y) versus voltage (x, in kiloVolts), for some equipment.

We fit

$$\widehat{\log Y} = 19 - 0.51X$$

So a 1-kV increase in voltage changes the estimated mean of Y by $e^{-0.51} = 0.6$. So if the voltage increases from 27 kV to 28 kV, the time to breakdown estimate is 60% of what it was.

Ensure that the transformation leads to reasonable interpretations for your problem under study.

What if we'd transformed x instead?

$\log(AB) = \log A + \log B$

Our new model would be: $Y = \beta_0 + \beta_1 \log(x) + e$

Our interpretation is now in terms of multiplicative changes in x. For each k-fold change in x, the estimated change in the mean of Y is $\beta_1 \log k$.

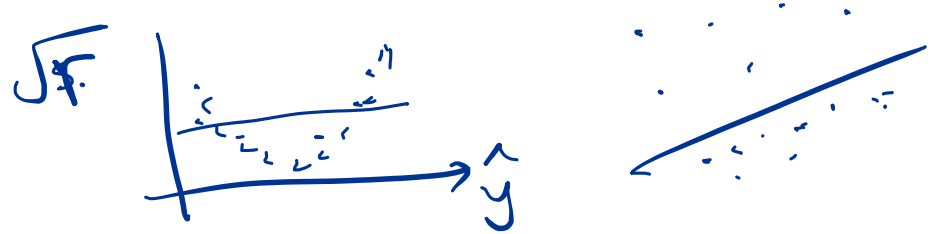
$$\mathbb{E}(Y_{\text{original}}) = \beta_0 + \beta_1 \log(x)$$

$$\mathbb{E}(Y_{\text{new}}) = \beta_0 + \beta_1 \log(kx)$$

$$\mathbb{E}(Y_{\text{new}}) - \mathbb{E}(Y_{\text{original}}) = \beta_1 (\log(kx) - \log(x))$$

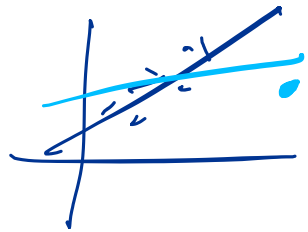
$$= \beta_1 \log k$$

Handling violated assumptions

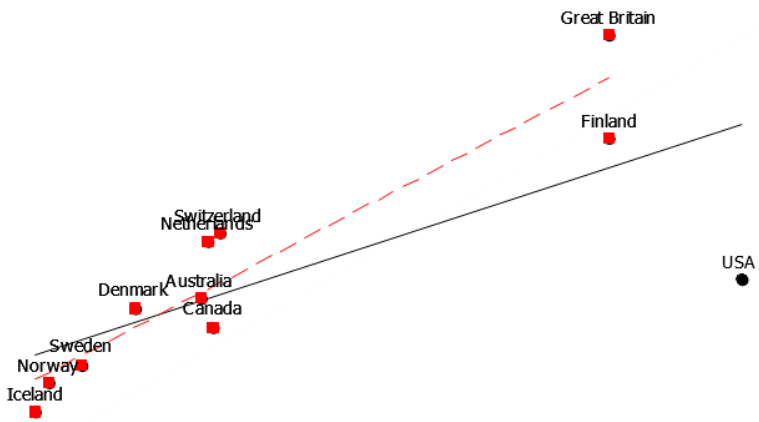


Last week, considering §3.2, we discussed briefly how we could change our underlying model, possibly swapping SLR for something more complicated, e.g.:

- Models that allow non-normal errors
- Nonlinear models to capture trends or unusual points
- Robust methods — reduce effects of outliers

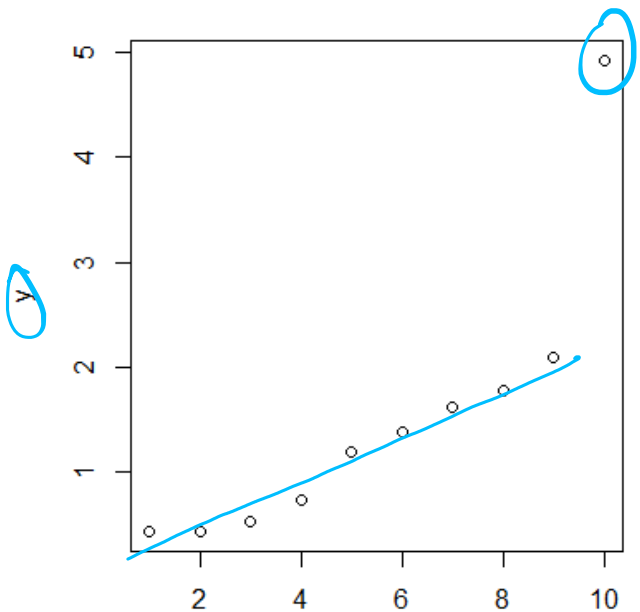


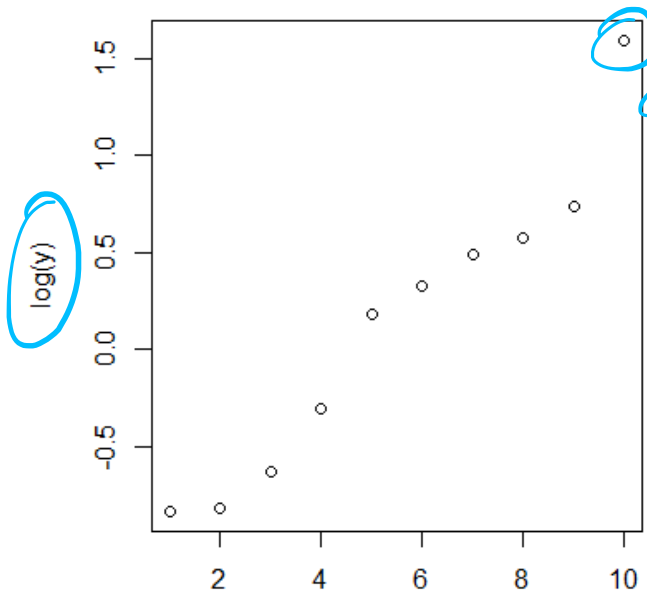
There are usually several options; remember you can report results with- and without outliers as well. For example in the cigarette dataset:



Handling violated assumptions

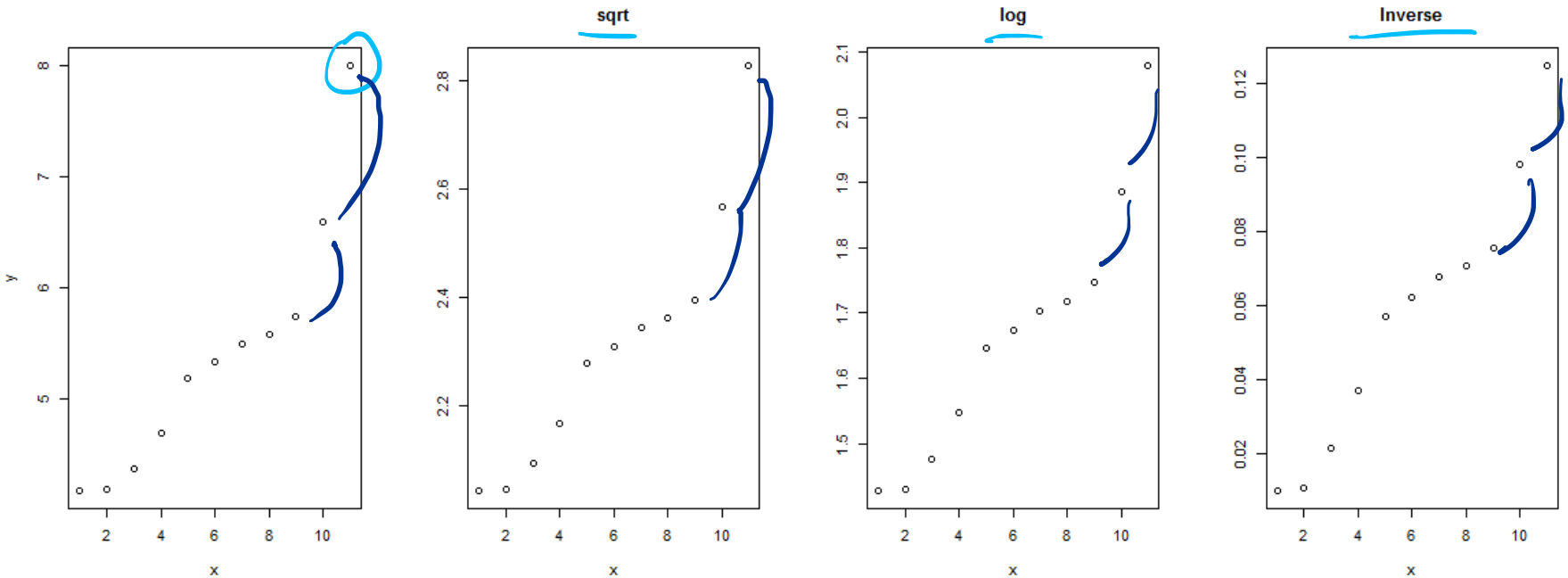
Whereas, our current investigations in §3.3 suggest that if **outliers** are occurring at the tails of a skewed distribution, we might mitigate their effects through a transformation. For example, notice the effect of a logarithmic transformation:





Handling violated assumptions

Here we see how the inverse transform can powerfully bring in outliers.

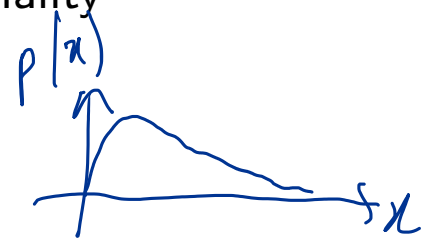


Handling violated assumptions

Besides addressing outliers, we've seen how transforming Y can help* with nonconstant variance and nonlinearity. It can also help with error **non-normality**: recall Check 7.

In case the errors are not normal:

- CLT says that linear combinations of r.v.s are normally distributed, even if original r.v.s aren't
- Our estimators of β_0 and β_1 are linear combinations of r.v.s, so tests and CIs for them are robust against non-normality, as long as they are not too skewed and there aren't extreme outliers
- Prediction intervals aren't robust against non-normality



* Transforming X can be useful as well. For example, if X is very right-skewed, perform a log, $\sqrt{\ }$, or $1/x$ transformation.

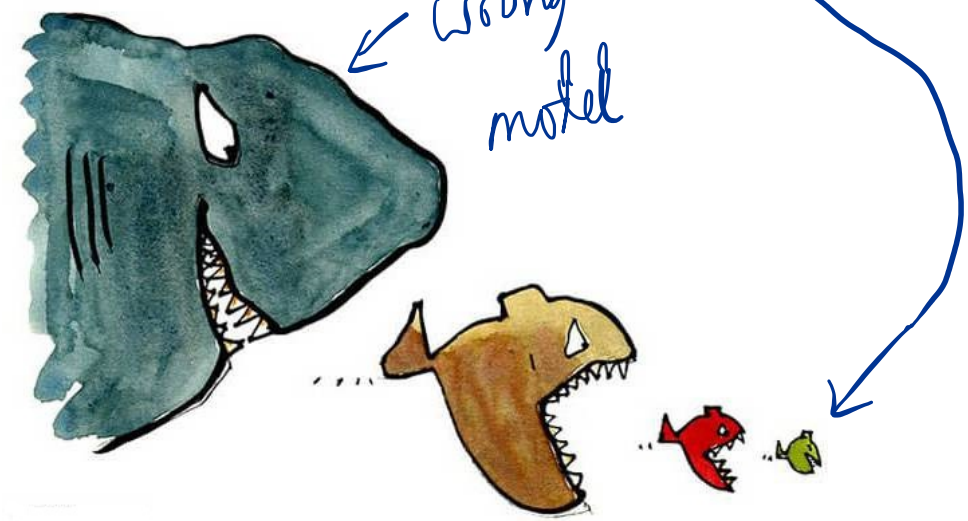
Relative importance of the assumptions for inference

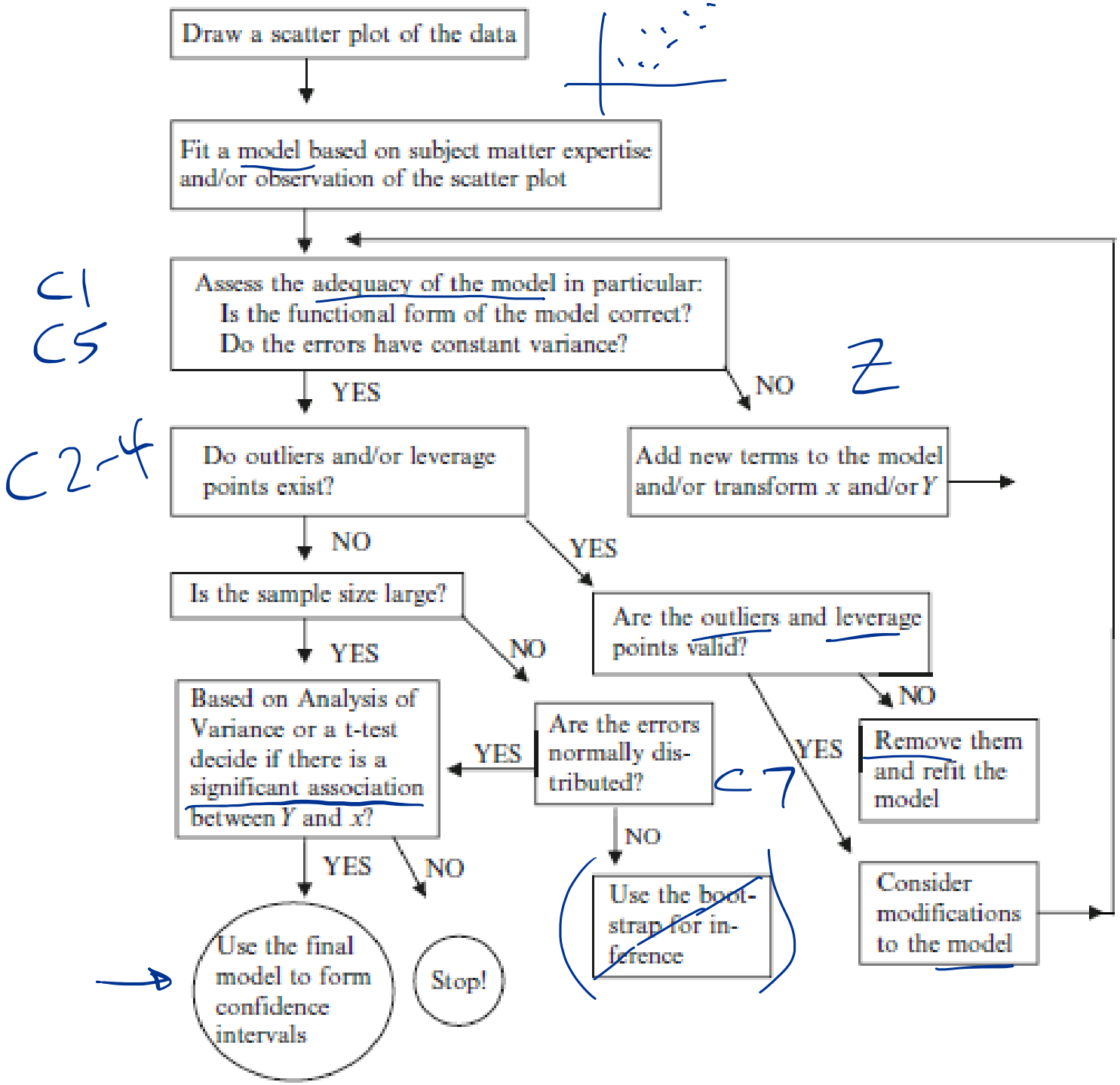
Most important is to have the right form of model: $\mathbb{E}(e) = 0$, etc

Then independence of errors.

Then constant variance — this is lower down the list because regression is reasonably robust to nonconstant variance provided we have a similar number of observations for each x.

Normality is less important (although it is necessary for PIs).




Draw a scatter plot of the data
Fit a model based on subject matter expertise and/or observation of the scatter plot
Assess the adequacy of the model in particular:
Is the functional form of the model correct?
Do the errors have constant variance?
YES
NO
Add new terms to the model and/or transform x and/or Y
Do outliers and/or leverage points exist?
NO
YES
Are the outliers and leverage points valid?
NO
YES
Remove them and refit the model
Is the sample size large?
YES
NO
Are the errors normally distributed?
YES
Based on Analysis of Variance or a t-test decide if there is a significant association between Y and x?
YES
NO
NO
Use the bootstrap for inference
Consider modifications to the model
Use the final model to form confidence intervals
Stop!

Next steps

pp 45-85, 102-113

Recalling from the syllabus what we'll miss in **Chapter 3**, we're ~~nearly~~ finished our expected coverage. Homework:

- Try the remaining questions in the back of the chapter
- We may discuss one next week, with a focus on §3.2

skip 3B, 3C
skip 6

When we leave Chapter 3, we'll discuss **Chapter 5** first (not Chapter 4). We'll eventually cover all of Chapter 5 (no skipped material).