Announcements

- Make sure to sign in on the google form
- Pset 9 due December 2 at 5 pm
- Final project due Wednesday December 14th at 5 pm

Many ways to ~~skin a cat~~ peel an orange

Recall the following linear model extensions:

- Heteroscedasticity-consistent standard errors: $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$, $\epsilon \sim MVN(\vec{0}, \vec{\sigma}^2\mathbf{I}_{n\times n})$ where $\vec{\sigma}^2$ has non-identical entries
- Weighted least squares: $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$, $\epsilon \sim MVN(\vec{0}, \mathbf{W}^{-1}\sigma^2\mathbf{I}_{n\times n})$ where $\mathbf{W}$ is a diagonal matrix of weights
- Huber's method: Minimize loss $= \sum_i^n g(\hat{\epsilon}_i)$ with

$$g(x) = \begin{cases} x^2/2 & |x| < c \\ c|x| - c^2/2 & |x| \geq c \end{cases}$$

- Block correlations: $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$, $\epsilon \sim MVN(\vec{0}, \mathbf{\Sigma})$ with a covariance matrix like

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho_1 & \rho_1 & 0 & 0 \\ \rho_1 & 1 & \rho_1 & 0 & 0 \\ \rho_1 & \rho_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho_2 \\ 0 & 0 & 0 & \rho_2 & 1 \end{bmatrix}$$

- Autoregressive: $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$, $\epsilon \sim MVN(\vec{0}, \mathbf{\Sigma})$ with a covariance matrix like

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

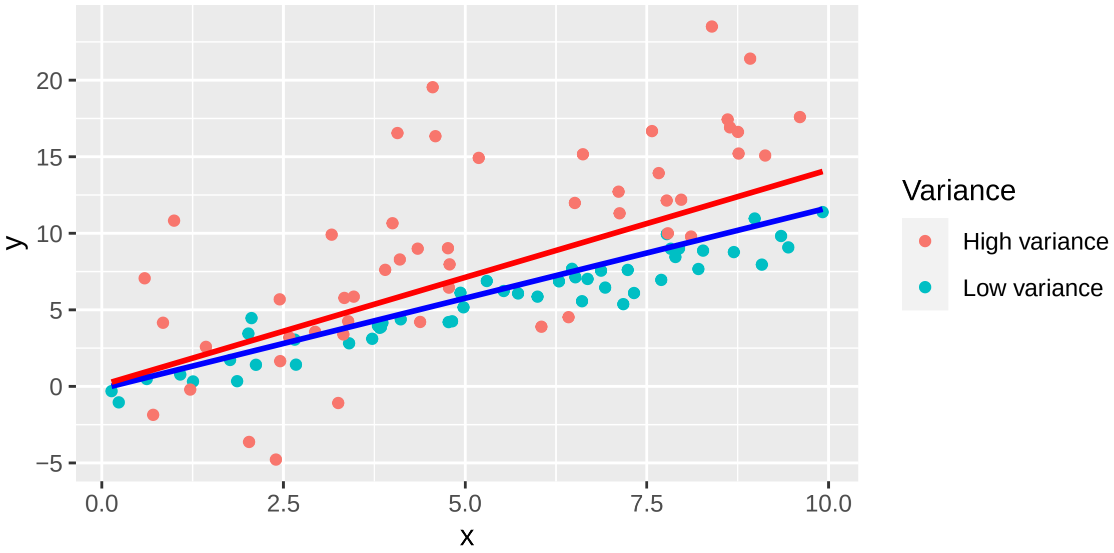
- Random intercepts: $Y_{ij} \sim \alpha + \alpha_i + \beta X_{ij} + \epsilon_{ij}$ with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_Y^2)$ and $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$. (Note that this version explicitly includes a fixed slope and a random slope with mean 0. Random effects as you'll see them elsewhere (and in `lmer`) are usually specified to have mean 0, so including the fixed effect separately is a good practice.)
- Random slopes: $Y_{ij} \sim \alpha + (\beta + \beta_i)X_{ij} + \epsilon_{ij}$ with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_Y^2)$ and $\beta_i \sim \mathcal{N}(0, \sigma_\beta^2)$.

1. For each of the scenarios below, which of the above linear model extensions would you use?

- Stock price over a week: Autoregressive
- Wheat yield vs phosphorus fertilizer use in various Kansas counties: Block correlation, random intercepts, random slopes
- Soil nitrate concentration vs fertilizer use where the nitrate measurements are taken with instruments of varying precision: Heteroscedasticity-consistent standard errors or weighted regression
- Spotify listens vs release year: Huber's method
- Day in year leaves start to fall vs latitude for various tree types: Block correlation, random intercepts, random slopes

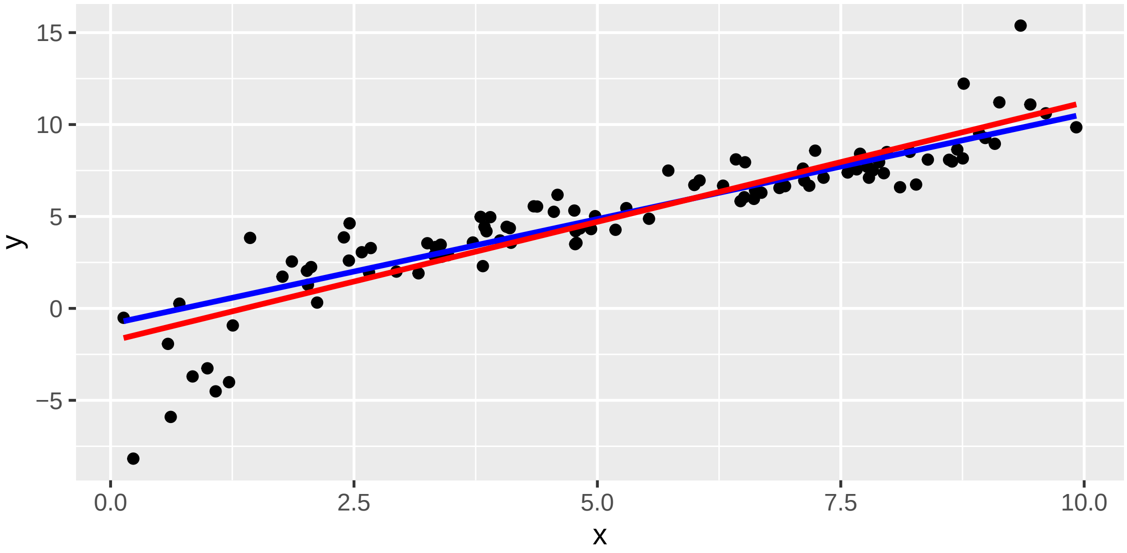
2. Which line is which in the following models?

- One line is a standard OLS; the other uses weighted regression.



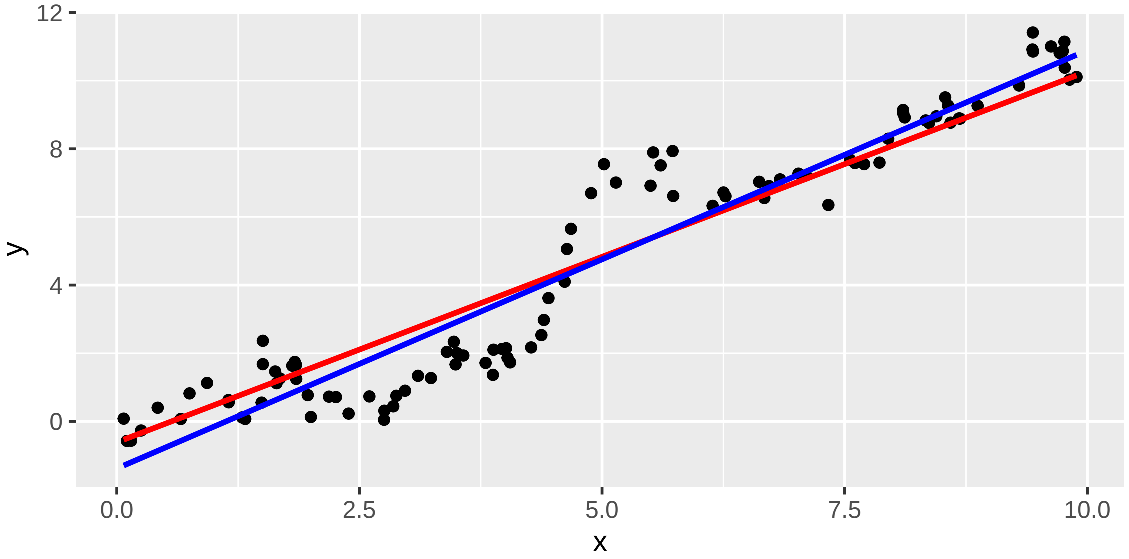
Blue is the weighted regression. It weights low variance observations more heavily as expected.

- One line is a standard OLS; the other uses Huber's method.



Blue is Huber's method. It is not as sensitive to outliers.

- One line is a standard OLS; the other uses autoregression.

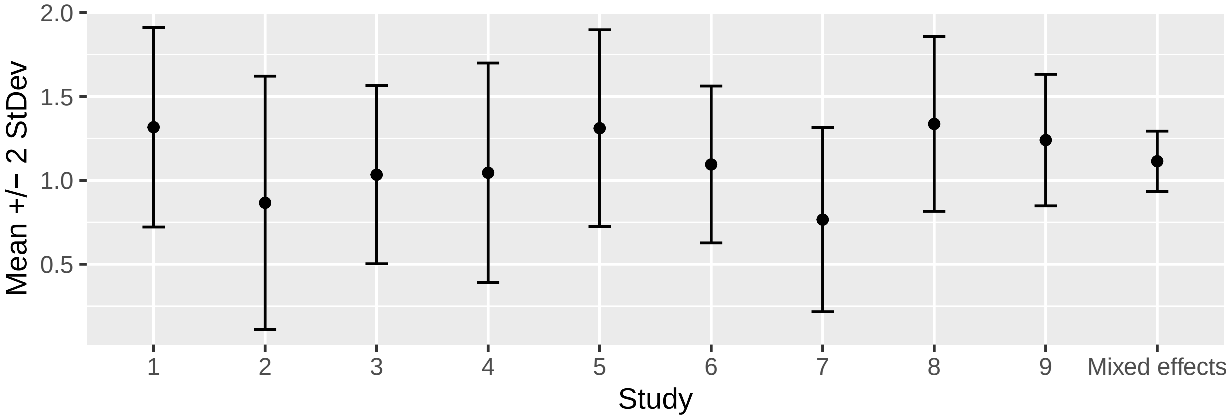


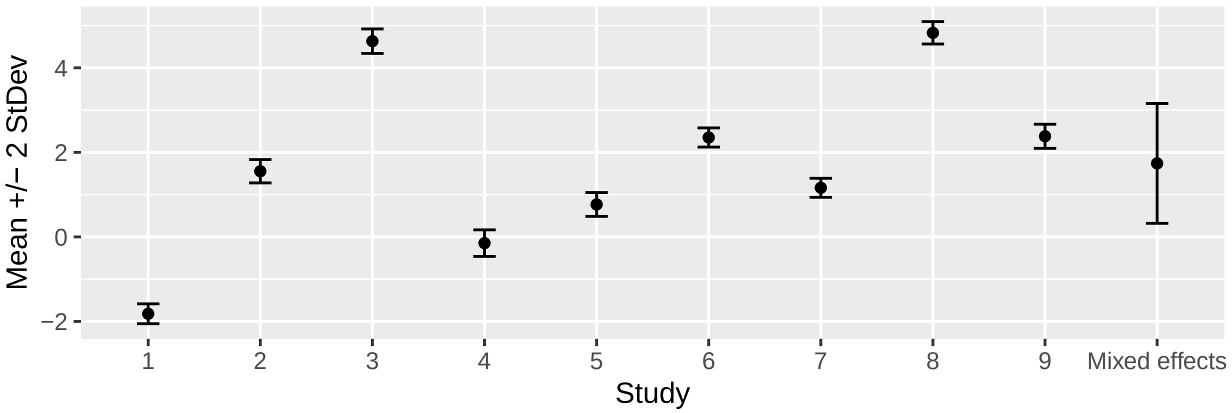
Red is the autoregression. It’s less influenced by tightly correlated nearby points.

Meta-analysis analysis

One common use of mixed effects models is in meta-analysis to aggregate the results of multiple studies. This question will explore various methods of meta-analysis aggregation.

1. A common plot in meta-analysis is a forest plot, showing the mean and standard deviation of some quantity of interest as determined by various studies as well as an aggregated mean and variance. Given the following forest plots (technically these are axis-flipped forest plots), determine what a linear mixed effects model combining the studies should look like.





2. Sometimes, studies will not make available all their individual data points. In these cases, we might only have a point estimate and standard error for a particular statistic of interest. Assume that for study i we have our statistic of interest

$$\hat{\beta}_i \sim \mathcal{N}(\beta, \sigma^2_{\hat{\beta}_i})$$

where we will approximate the true standard error $\sigma^2_{\hat{\beta}_i}$ with our estimated standard error $\hat{\sigma}^2_{\hat{\beta}_i}$. This model specifies that there is some underlying effect β and that each study will find some $\hat{\beta}_i$ with a standard error based on the study's sample size etc. Assuming we have n_{studies}, use maximum likelihood estimation to find a point estimate for β.

Keeping only the pieces that depend on β,

$$L \propto \prod_i^{n_{\text{studies}}} e^{\frac{1}{2}(\frac{\hat{\beta}_i - \beta}{\sigma_{\hat{\beta}_i}})^2} \implies \log(L) = \sum_i^{n_{\text{studies}}} \frac{1}{2}\left(\frac{\hat{\beta}_i - \beta}{\sigma_{\hat{\beta}_i}}\right)^2$$

Taking the derivative and setting it equal to 0 shows

$$0 = \sum_{i}^{n_{\text{studies}}} \left(\frac{\hat{\beta}_i - \beta}{\sigma_{\hat{\beta}_i}}\right) = \sum_{i}^{n_{\text{studies}}} \left(\frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}}\right) - \beta \sum_{i}^{n_{\text{studies}}} \left(\frac{1}{\sigma_{\hat{\beta}_i}}\right) \implies \beta = \frac{\sum_{i}^{n_{\text{studies}}} \left(\frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}}\right)}{\sum_{i}^{n_{\text{studies}}} \left(\frac{1}{\sigma_{\hat{\beta}_i}}\right)}$$

3. The code below simulates data from one of two scenarios. In the first scenario, there is a true effect β that holds in all the studies. The only thing that differs by study is the sample size. Therefore, for study i, the j^{th} observation is given by $Y_{ij} = \beta_0 + \beta X_{ij} + \epsilon_{ij}$ with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_Y^2)$. In the second scenario, the effect in study i is $\beta_i \sim \mathcal{N}(\beta, \sigma_\beta^2)$. Then, for study i, the j^{th} observation is given by $Y_{ij} = \beta_0 + \beta_i X_{ij} + \epsilon_{ij}$ with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_Y^2)$.

We will consider 5 methods:

- Use the likelihood method above
- Find $\hat{\beta}_i$ for each study and take the average to estimate β.
- Use a mixed effects model `ys ~ xs + (xs - 1 | study)` (random slopes, no random intercepts)
- Combine all the data points and run a linear model with `ys ~ xs`
- Combine all the data points and run a linear model with `ys ~ xs + xs:as.factor(study_id)`

Interpret the outputs: how do the methods perform on the two scenarios? (The code below should be cached and takes a while to run, so don't change it unnecessarily.)

```
##                           Method      Mean        SD       MSE
## 1                     Likelihood 1.0104931 0.5171895 0.2673276
## 2                      Raw means 1.0075483 0.6428431 0.4128909
## 3           Linear mixed effects 1.0096241 0.4829380 0.2330885
## 4 Fixed effects with interaction 0.9187366 1.8444425 3.4051700
## 5   Fixed effects no interaction 1.0142975 0.4690748 0.2200155
```

```
##                           Method      Mean        SD       MSE
## 1                     Likelihood 0.9575646 0.8740974 0.7650830
## 2                      Raw means 0.9620820 0.9234308 0.8533096
## 3           Linear mixed effects 0.9515375 0.8615242 0.7438303
## 4 Fixed effects with interaction 1.0185786 2.7047508 7.3087065
## 5   Fixed effects no interaction 0.9513935 0.9343959 0.8745852
```

The fixed effects with interactions is clearly the worst model: it actually just takes study 1 as the estimated effect and fits adjustments for the rest of the studies. The raw means also does rather poorly with a higher standard error than the other methods. Fixed effects with no interactions does the best when all the participants in all the studies are from the same distribution, but its standard error is higher when there are by-study differences. Linear mixed effects performs very well in both scenarios, and the likelihood method also does rather well considering that it doesn't have any of the original data points.

4. Under the random slopes model specified above, given a new X from a new study, find the probability that its associated Y value will be less than τ.

Conditioning on the group β_i and applying the Law of Total Probability,

$$\begin{aligned} P(Y < \tau) &= \int_{-\infty}^{\infty} P(Y < \tau | \beta_i) P(\beta_i) d\beta_i \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\tau} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{y-\beta_i}{\sigma_Y}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_\beta} e^{-\frac{1}{2}\left(\frac{\beta_i-\beta}{\sigma_\beta}\right)^2} dy d\beta_i \end{aligned}$$

Crops continued

This question will deal with a data set of country-level statistics from this source with an explanation of the data encoding found here.

A few useful columns:

- `mad_gdppc`: GDP per capita
- `wdi_araland`: Arable land (% of land area)
- `wdi_precip`: Average annual precipitation (mm per year)
- `ht_region`: Country's region of the world: Eastern Europe (1), Latin America (2), North Africa & the Middle East (3), Sub-Saharan Africa (4), Western Europe and North America (5), East Asia (6), South-East Asia (7), South Asia (8), Pacific (9), Caribbean (10)

1. Fit a quadratic regression model to predict `wdi_araland` from `wdi_precip`, using `log(mad_gdppc)` as a weight to account for the fact that wealthier countries might have more accurate crop tracking technology. Call this `lm_weight`.

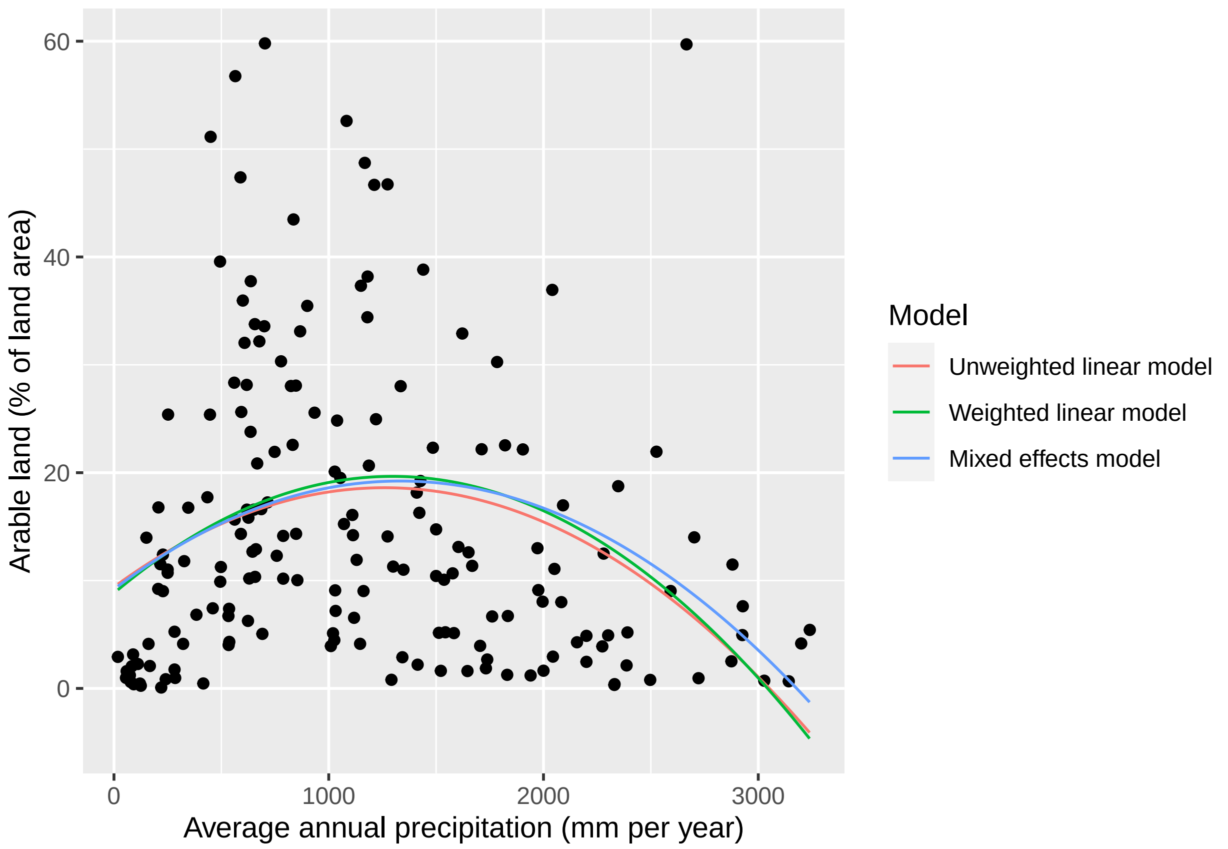
```
lm_weight <- lm(wdi_araland~poly(wdi_precip, 2, raw = TRUE), countries, weight=log(mad_gdppc))
```

2. Use the `lmer` function to fit a mixed effects model with a random intercept based on the country's `ht_region`, and use the argument `weight=log(mad_gdppc)` here as well. Call this `lmer1`.

```
countries$ht_region <- as.factor(countries$ht_region)
lmer1 <- lmer(wdi_araland ~ poly(wdi_precip, 2, raw = TRUE) + (1 | ht_region),
              countries, weight=log(mad_gdppc))
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

3. Use your models to plot the estimated relationship between precipitation and arable land.



There's not much to interpret here: both of the weighted models have fitted curves slightly higher than the unweighted model (plausibly suggesting that wealthier countries might just have more arable land given the same amount of precipitation). However, the curves are all quite similar and have maxima close to each other.