

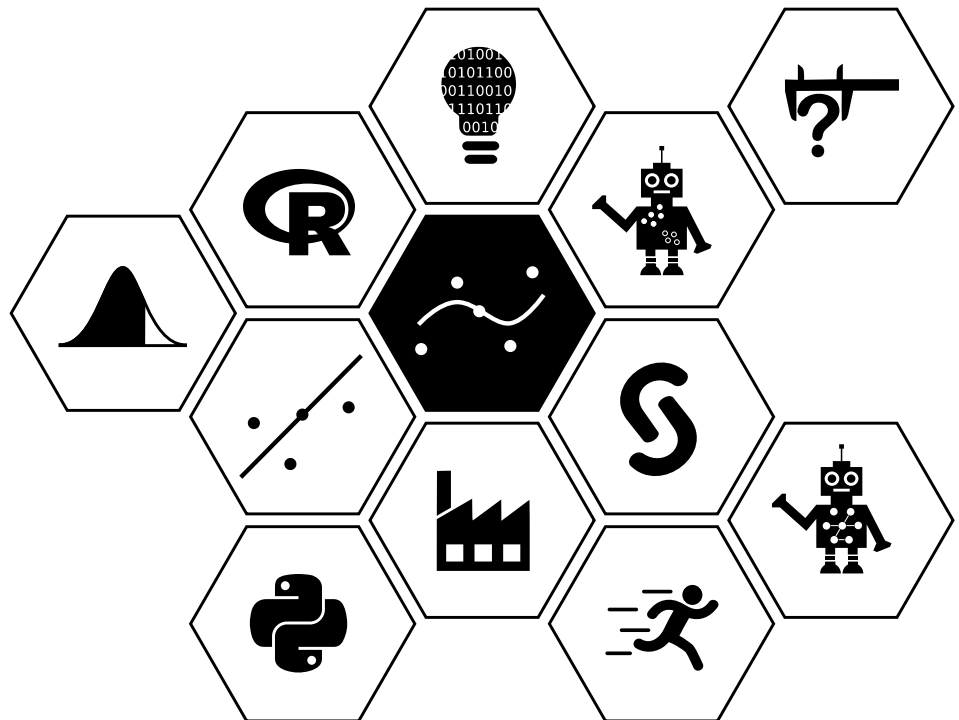
Advanced Predictive Models

Tereza Neocleous

Academic Year 2020-21

Week 9:

Linear Mixed Models



Models for correlated data

In the time series part of this course we saw how we can model correlated observations where the correlation occurs due to the ordering in time. For the last part of the course we will focus on correlated observations where the correlation structure can take different forms. We will begin this week with models for continuous responses, and generalise to models for discrete responses in Week 10.

The main class of models used for correlated continuous responses is the class of **linear mixed models**. We will introduce the main features of these models through an example. A nice visualisation of some of the same concepts can also be found [here](#).



Random coefficient models – College GPA example

<https://youtu.be/760rAfwuwwE>

Duration: 15m39s



Example 1 (College GPA over time).

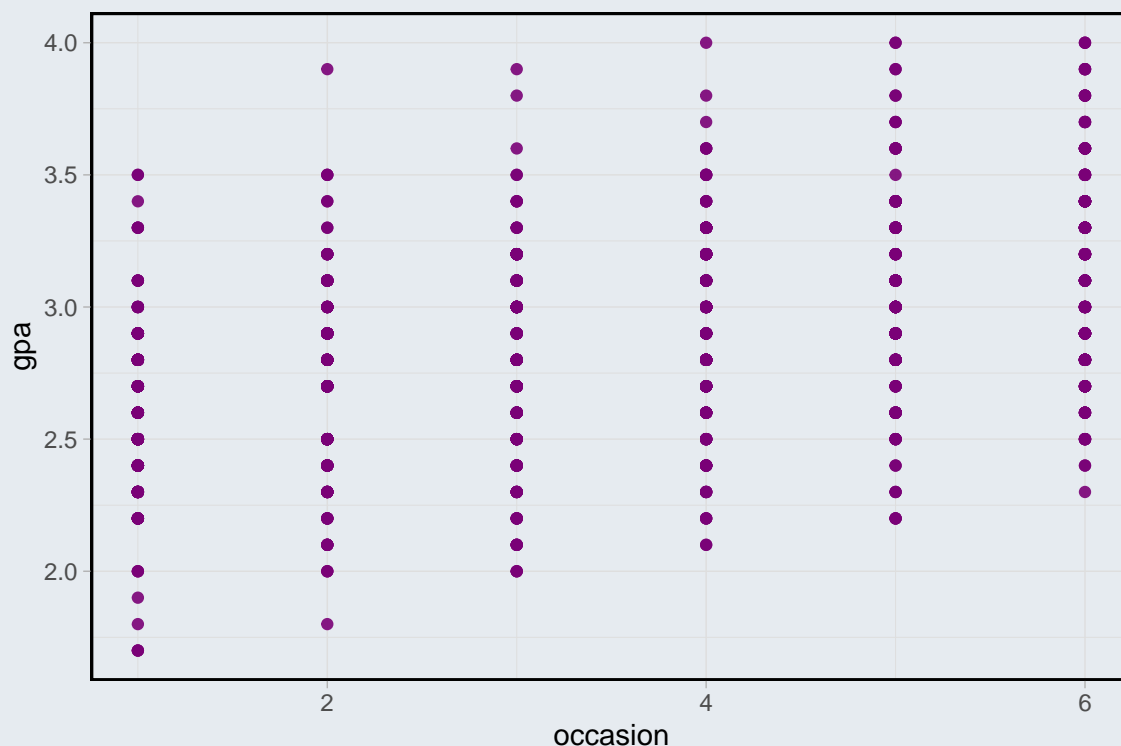
200 students' college GPA was recorded on six occasions: each semester for three years. The dataset `gpa.csv` also contains information on each student's gender, high school GPA, and hours of work for each semester.

```
gpa <- read.csv(url("http://www.stats.gla.ac.uk/~tereza/rp/gpa.csv"))
head(gpa)
```

student	occas	gpa	job	sex	highgpa	admitted	year	semester	occasion
1	year 1 semester 1	2.3	2 hours	female	2.8	yes	1	1	1
1	year 1 semester 2	2.1	2 hours	female	2.8	yes	1	2	2
1	year 2 semester 1	3.0	2 hours	female	2.8	yes	2	1	3
1	year 2 semester 2	3.0	2 hours	female	2.8	yes	2	2	4
1	year 3 semester 1	3.0	2 hours	female	2.8	yes	3	1	5
1	year 3 semester 2	3.3	2 hours	female	2.8	yes	3	2	6

We are interested in modelling the GPA as a function of time, so let us begin by looking at the scatterplot of the data.

```
ggplot(gpa, aes(x=occasion, y=gpa))+geom_point(alpha=0.9, color="#7a0177")
```



There is a positive association between GPA and occasion, suggesting that the GPA increases with time on average. We can model this relationship using a linear model as follows:

```
gpa_lm <- lm(gpa ~ occasion, data=gpa)
summary(gpa_lm)
```

Call:

```
lm(formula = gpa ~ occasion, data = gpa)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.90553	-0.22447	-0.01184	0.26921	1.19447

Coefficients:

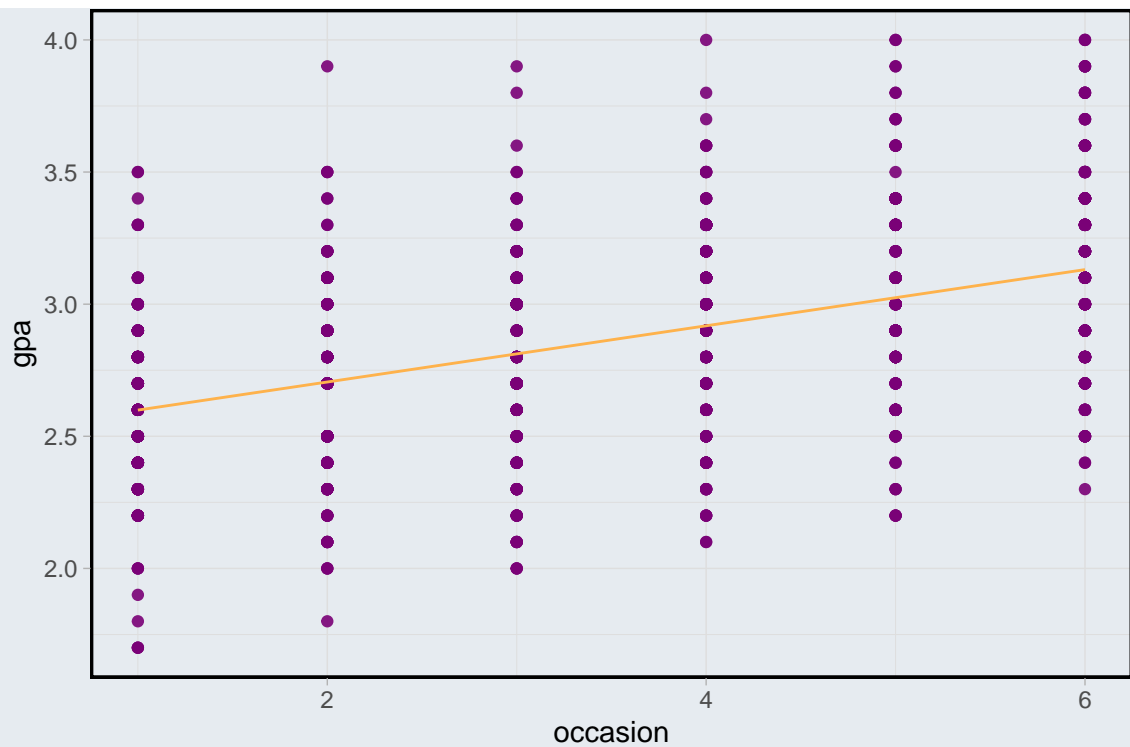
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.492900	0.022955	108.60	<2e-16 ***
occasion	0.106314	0.005894	18.04	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3487 on 1198 degrees of freedom
 Multiple R-squared: 0.2136, Adjusted R-squared: 0.2129
 F-statistic: 325.3 on 1 and 1198 DF, p-value: < 2.2e-16

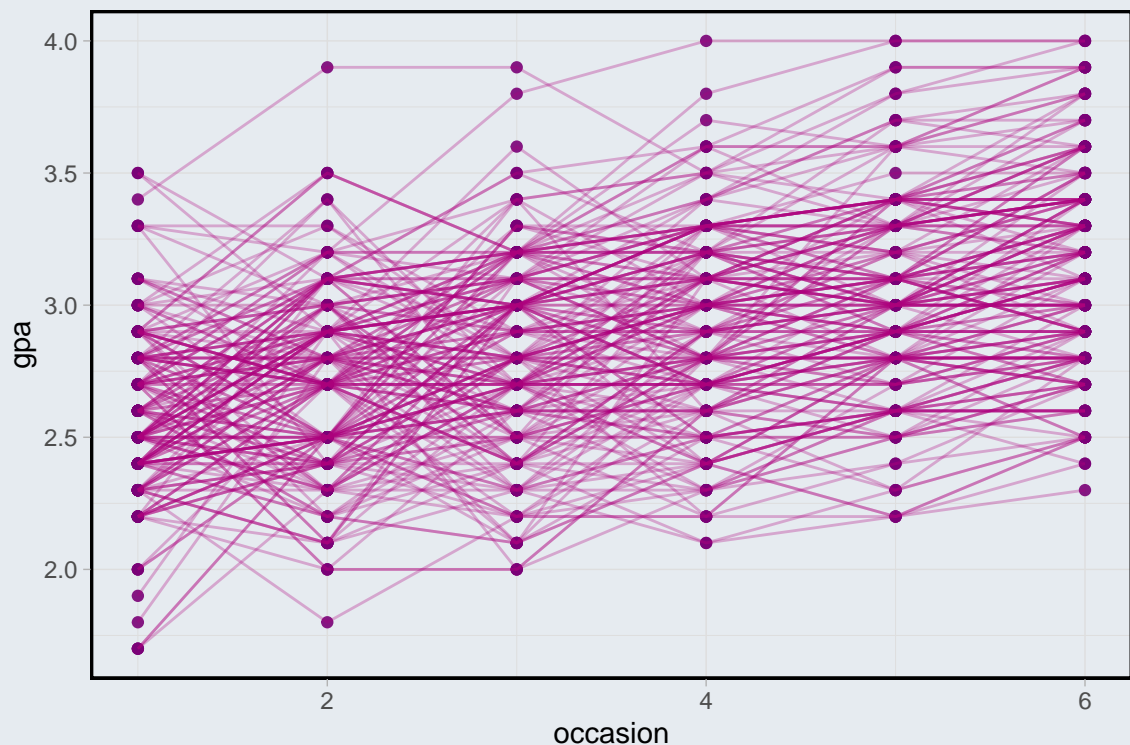
The fitted line can be superimposed on the scatterplot as shown below:

```
gpa_fit<- data.frame(x=c(1:6),
                     y=gpa_lm$coefficients[1]+ gpa_lm$coefficients[2]*c(1:6))
ggplot(gpa, aes(x=occasion, y=gpa)) + geom_point(alpha=0.9, color="#7a0177") +
  geom_line(data = gpa_fit, aes(x, y), col="#feb24c")
```



There are a few issues with this modelling approach. Firstly, it assumes that we have 1200 independent observations from a model of the form $y = \beta_0 + \beta_1 x + e$ where the errors e are independent $N(0, \sigma^2)$. This is not quite right, as we have multiple observations per student and it would be reasonable to expect that they are correlated. This is why the following plot might be a better depiction of the data than the scatterplot we saw earlier.

```
ggplot(gpa, aes(x=occasion, y=gpa)) + geom_point(alpha=0.9, color="#7a0177") +  
  geom_path(aes(group=student), alpha=0.3, col="#ae017e")
```



As it turns out, failing to take into account the correlation between measurements from the same student results in biased standard errors for the β estimates (the simulation study in the notes from Week 8 illustrates this). What's more, the linear regression model does not allow us to explore the student effect, which might be interesting in its own right.

To remedy this we can introduce some correlation for observations coming from the same student. Let us write this down more precisely. Let y_{ij} be the GPA of the i th student on the j th occasion and let $x_{ij} = 1, 2, 3, 4, 5, 6$ denote the occasion. We can consider a model of the form

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 x_{ij} + e_{ij}$$

where the b_{0i} are independent $N(0, \sigma_0^2)$ and the e_{ij} are independent $N(0, \sigma^2)$ and b_{0i} and e_{ij} are independent of each other.

The term b_{0i} is the **random effect** corresponding to the i th student. If we rewrite the model as

$$y_{ij} = b_i^* + \beta_1 x_{ij} + e_{ij}$$

we can see that the terms $b_i^* \sim N(\beta_0, \sigma_0^2)$ are **random intercepts** in the regression model, each corresponding to a different student. They are centered around β_0 , the intercept of the linear regression, with added variability represented by the variance parameter σ_0^2 .

The random intercept model above is an example of a **linear mixed model** which is a linear model that contains both **fixed** and **random** effects. Fixed effects are effects we wish to estimate explicitly, while random effects are effects that introduce variability. So for example here we wish to estimate the coefficient of occasion to understand how GPA changes with time, but we would not want to estimate separate regression coefficients for each student. Instead, we view the 200 students in this example as a random sample from a larger population of students, and we are hoping that the inference will hold for the general population.

Note also the convention to use greek letters for the fixed effects β_0 and β_1 and roman letters for the random effects b_{0i} and e_{ij} .

Fitting this random intercept model in R requires use of the function `lmer` from `library(lme4)`. Note the way in which we specify the random effect in the model by using brackets and declaring the grouping variable (here this is student).

```
gpa_ranint <- lmer(gpa ~ occasion + (1|student), data=gpa)
summary(gpa_ranint)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: gpa ~ occasion + (1 | student)
Data: gpa
```

```
REML criterion at convergence: 408.9
```

```
Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6169 -0.6373 -0.0004  0.6361  2.8310
```

```
Random effects:
 Groups   Name      Variance Std.Dev.
 student (Intercept) 0.06372  0.2524
 Residual                0.05809  0.2410
Number of obs: 1200, groups:  student, 200
```

```
Fixed effects:
              Estimate Std. Error t value
(Intercept)  2.492900   0.023881  104.4
occasion      0.106314   0.004074   26.1
```

```
Correlation of Fixed Effects:
      (Intr)
occasion -0.597
```

The output gives

- the fixed effects, i.e. the coefficient estimates for β_0 and β_1 , which are the same as the ones obtained from using `lm()`, and

- the random effects, for which we estimate the variance parameters.

There are two variances estimated here: σ_0^2 , corresponding to the student, and σ^2 (Residual), corresponding to the error. The standard deviation column in the output is simply the square root of the variance. So by modelling the data in this way we take into account the fact that there is correlation in the data due to multiple observations per student, and we also quantify the between-student variability.

One relevant quantity to compute is the **intraclass correlation coefficient** which can be estimated by

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}_0^2 + \hat{\sigma}^2}.$$

From the output this is:

```
0.06372/(0.06372+0.05809)
```

```
[1] 0.5231098
```

which is interpreted as "the correlation between any two observations coming from the same student is about 0.52".

Another thing to note about the output is that it does not contain p -values, as these can be based on different approximations depending on the design of the study. We can get some indication of the significance of the model parameters by obtaining approximate confidence intervals as shown below, although this approach too is an approximation. You can try different methods of computing the intervals, e.g. `method="boot"` for bootstrap (default is "profile"):

```
round(confint(gpa_ranint, oldNames=FALSE),3)
```

```
                2.5 % 97.5 %
sd_(Intercept)|student 0.225 0.282
sigma                  0.231 0.252
(Intercept)           2.446 2.540
occasion              0.098 0.114
```

```
round(confint(gpa_ranint, oldNames=FALSE, method="boot"),3)
```

```
                2.5 % 97.5 %
sd_(Intercept)|student 0.222 0.279
sigma                  0.231 0.251
(Intercept)           2.449 2.541
occasion              0.097 0.115
```

We can also predict the random effect for each student using the `ranef()` function from `lme4`.

```
round(head(ranef(gpa_ranint)$student),3)
```

```
(Intercept)
1      -0.071
2      -0.216
3       0.088
4      -0.187
5       0.030
6      -0.302
```

By adding these to the β estimate for the intercept, we get individual intercepts for each student.

```
gpa_ranint@beta[1]
```

```
[1] 2.4929
```

```
round(head(gpa_ranint@beta[1]+ranef(gpa_ranint)$student),3)
```

```
(Intercept)
1       2.422
2       2.277
3       2.581
4       2.306
5       2.523
6       2.191
```

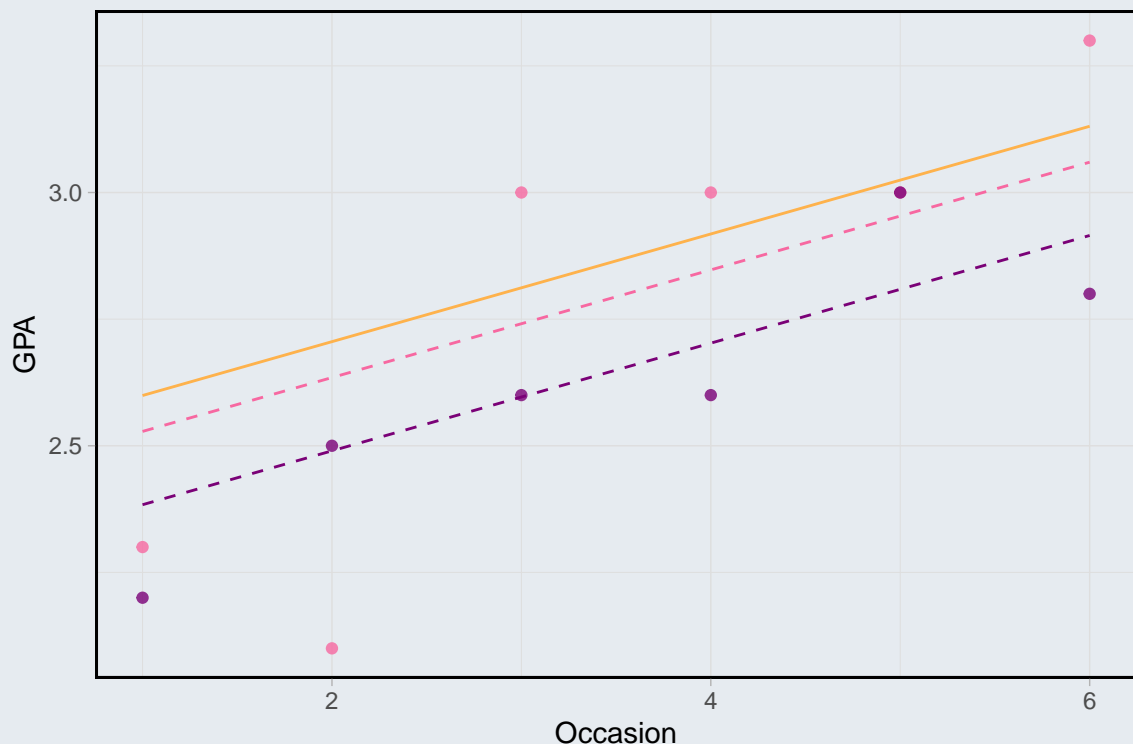
This is what the `coef()` function returns:

```
round(head(coef(gpa_ranint)$student),3)

      (Intercept) occasion
1          2.422      0.106
2          2.277      0.106
3          2.581      0.106
4          2.306      0.106
5          2.523      0.106
6          2.191      0.106
```

This allows us to plot the predicted regression line for each student. This is shown for two individual students in the plot below (pink and purple points and dashed lines), along with the overall regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ in yellow. Notice that the individual regression lines are parallel to the overall regression line.

```
p1<-predict(gpa_ranint)
p2<-predict(gpa_lm)
predictions <- data.frame(x=c(1:6), s1_m=p1[1:6], s2_m=p1[7:12], l=p2[7:12],
                           gpa_1=gpa$gpa[1:6], gpa_2=gpa$gpa[7:12])
ggplot(predictions, aes(x=x, y=gpa_1)) +
  geom_point(colour="#f768a1", alpha=0.8) +
  geom_line(aes(y=s1_m), colour="#f768a1", linetype=2)+
  geom_point(aes(x=x,y=gpa_2), colour="#7a0177",alpha=0.8) +
  geom_line(aes(y=s2_m), colour="#7a0177", linetype=2) +
  geom_line(aes(y=l), colour="#feb24c") + xlab("Occasion") + ylab("GPA")
```



We may wish to explore whether the coefficient of occasion should also be allowed to vary for each student in a more general **random coefficient model**. This model will take the form

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 x_{ij} + b_{1i} x_{ij} + e_{ij}$$

where, as before, the b_{0i} are independent $N(0, \sigma_0^2)$, the b_{1i} are independent $N(0, \sigma_1^2)$, and the e_{ij} are independent $N(0, \sigma^2)$. We further assume that each of b_{0i} and b_{1i} is independent of e_{ij} but that the two are allowed to be correlated with each other with $\text{Cov}(b_{0i}, b_{1i}) = \sigma_{01}$ and corresponding correlation $\text{Corr}(b_{0i}, b_{1i}) = \rho_{01}$.

In R, this model can be fitted using the following code:

```
gpa_rc<-lmer(gpa ~ occasion + (1+occasion|student), data=gpa)
summary(gpa_rc)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: gpa ~ occasion + (1 + occasion | student)
Data: gpa
```

REML criterion at convergence: 261

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.2695	-0.5377	-0.0128	0.5326	3.1939

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
student	(Intercept)	0.052503	0.22913	
	occasion	0.004504	0.06711	-0.38
Residual		0.042388	0.20588	

Number of obs: 1200, groups: student, 200

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.492900	0.021123	118.02
occasion	0.106314	0.005885	18.07

Correlation of Fixed Effects:

	(Intr)
occasion	-0.579

Notice that in the random effects part of the output we now have variance component estimates

- $\hat{\sigma}_0^2 = 0.0525$ for the random intercept term b_{0i} ,
- $\hat{\sigma}_1^2 = 0.0045$ for the random slope term b_{1i} ,
- $\hat{\sigma}^2 = 0.0424$ for the error term e_{ij} ,

as well as $\hat{\rho}_{01} = -0.38$, which is an estimate for $\text{Corr}(b_{0i}, b_{1i})$.

One question we may ask ourselves is which model we should prefer: random intercepts only, or random slopes and intercepts? And if we have both a random slope and intercept, should they be correlated or uncorrelated?

Let us fit the model with uncorrelated random effects for slope and intercept and compare it to the model in which these are correlated:

```
gpa_rc_uncor <- lmer(gpa ~ occasion + (1|student)+(0+occasion|student),
data=gpa)
summary(gpa_rc_uncor)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: gpa ~ occasion + (1 | student) + (0 + occasion | student)
Data: gpa
```

REML criterion at convergence: 271.9

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.5708	-0.5535	-0.0010	0.5437	3.1361

Random effects:

Groups	Name	Variance	Std.Dev.
student	(Intercept)	0.036496	0.1910
student.1	occasion	0.003283	0.0573
Residual		0.044480	0.2109

Number of obs: 1200, groups: student, 200

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.492900	0.019371	128.7
occasion	0.106314	0.005397	19.7

Correlation of Fixed Effects:

(Intr)
occasion -0.425

anova(gpa_rc_uncor,gpa_rc)

Data: gpa

Models:

gpa_rc_uncor: gpa ~ occasion + (1 | student) + (0 + occasion | student)

gpa_rc: gpa ~ occasion + (1 + occasion | student)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
gpa_rc_uncor	5	267.06	292.51	-128.53	257.06			
gpa_rc	6	258.23	288.77	-123.12	246.23	10.825	1	0.001001 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The p -value for the likelihood ratio test performed using `anova()` is small, suggesting that the model with correlated slope and intercept is a better fit to the data.

We can also look at the confidence intervals for model parameters to see whether the variance for the slope random effect is significantly different from zero. From the output below it seems the slope random effect is significantly different from zero since the confidence interval is entirely positive: (0.057,0.078).

round(confint(gpa_rc, oldNames=FALSE),3)

	2.5 %	97.5 %
sd_(Intercept) student	0.190	0.269
cor_occasion.(Intercept) student	-0.544	-0.174
sd_occasion student	0.057	0.078
sigma	0.196	0.216
(Intercept)	2.451	2.534
occasion	0.095	0.118

Finally we can predict from this model using the prediction of the random effect of each student:

round(head(ranef(gpa_rc)\$student),3) # random effects

	(Intercept)	occasion
1	-0.262	0.060
2	-0.209	-0.002
3	-0.049	0.043
4	-0.051	-0.042
5	0.113	-0.026
6	-0.162	-0.043

round(head(coef(gpa_rc)\$student),3) # random coefficients

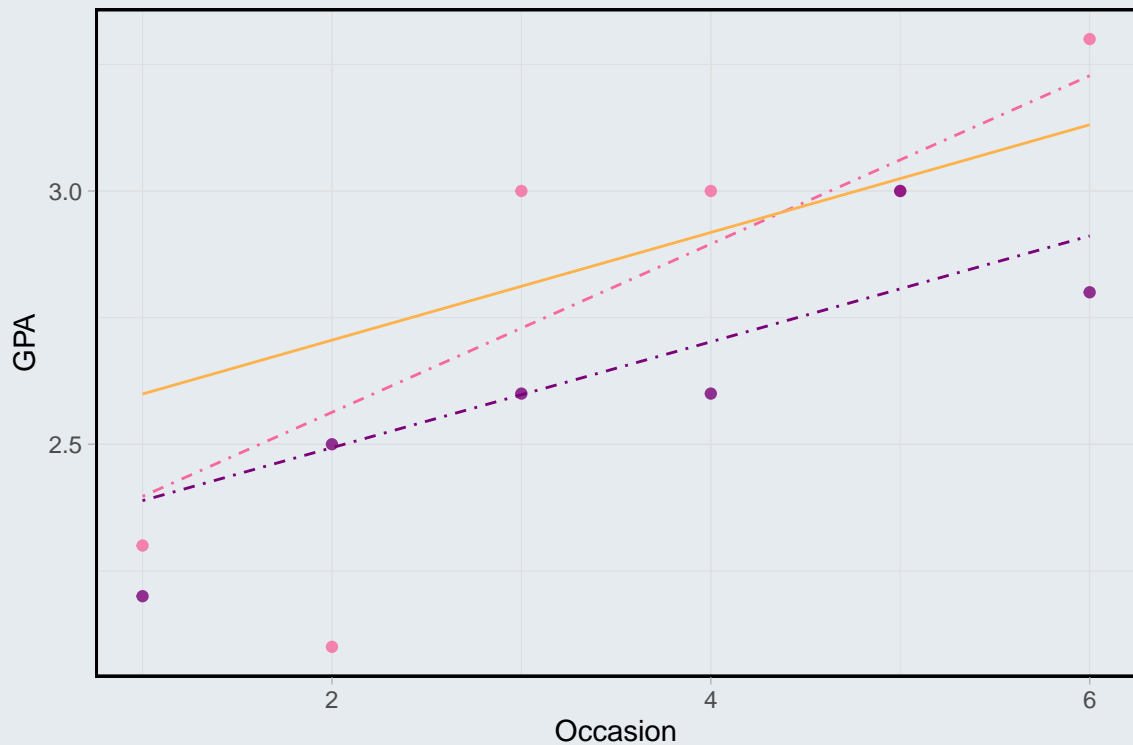
	(Intercept)	occasion
1	2.231	0.166
2	2.284	0.105
3	2.443	0.149
4	2.442	0.064
5	2.606	0.080
6	2.331	0.063

Let us now plot the predicted lines for the first two students along with the overall regression line as before:

```

p1<-predict(gpa_rc)
p2<-predict(gpa_lm)
predictions <- data.frame(x=c(1:6), s1_m=p1[1:6], s2_m=p1[7:12], s2_l=p2[7:12],
                           gpa_1=gpa$gpa[1:6], gpa_2=gpa$gpa[7:12] )
ggplot(predictions, aes(x=x, y=gpa_1)) +
  geom_point(color="#f768a1", alpha=0.8) +
  geom_line(aes(y=s1_m), colour="#f768a1", linetype=4)+
  geom_point(aes(x=x,y=gpa_2), colour="#7a0177",alpha=0.8) +
  geom_line(aes(y=s2_m), colour="#7a0177", linetype=4) +
  geom_line(aes(y=s2_l), colour="#feb24c") + xlab("Occasion") + ylab("GPA")

```



In this model each individual student's prediction (dashed lines) differs from the population line (solid yellow line) by a random amount both in the intercept and the slope.

In fact, the predicted line for an individual student can be viewed as a weighted average of

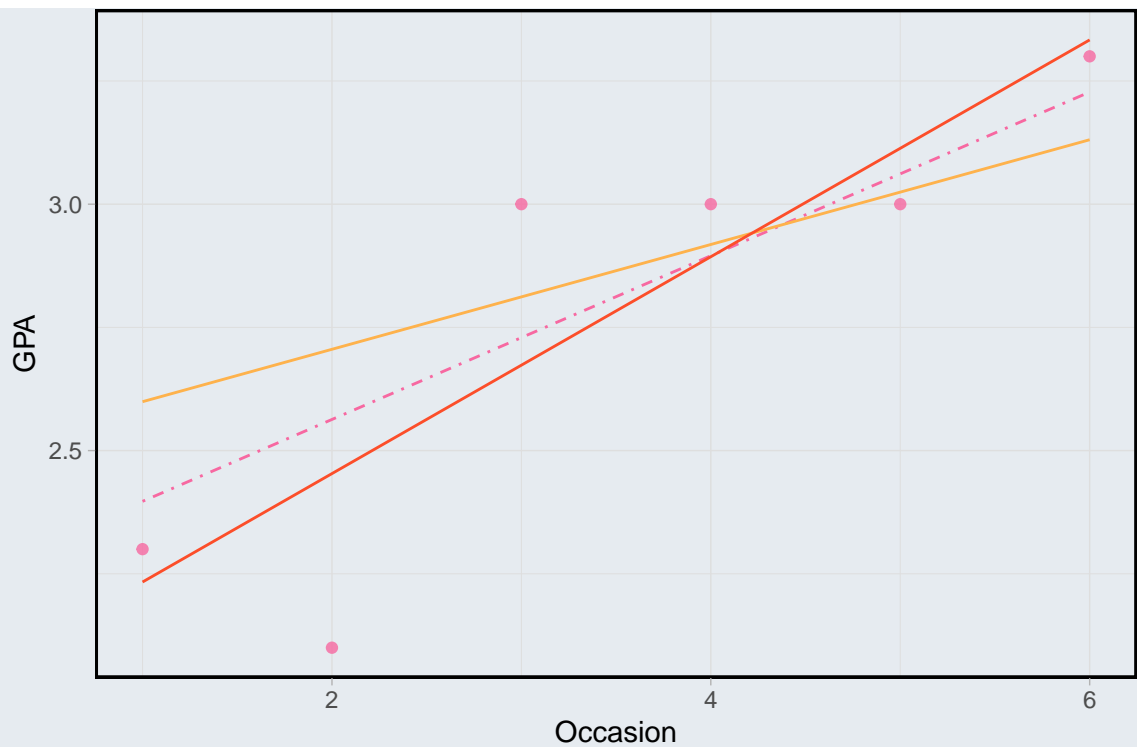
1. the regression line that we would have obtained by using just that student's data, and
2. the population line.

This is shown in the following plot for the first student:

```

p1<-predict(gpa_rc)
p2<-predict(gpa_lm)
p3 <- predict(lm(gpa~occasion, data=gpa[1:6,]))
predictions <- data.frame(x=c(1:6), s1_m=p1[1:6], s1_l=p2[1:6], s1_i=p3[1:6],
                           gpa_1=gpa$gpa[1:6])
ggplot(predictions, aes(x=x, y=gpa_1)) +
  geom_point(color="#f768a1", alpha=0.8) +
  geom_line(aes(y=s1_m), colour="#f768a1", linetype=4)+
  geom_line(aes(y=s1_l), colour="#feb24c") +
  geom_line(aes(y=s1_i), colour="#fc4e2a") + xlab("Occasion") + ylab("GPA")

```



The predicted line from the random coefficient model (pink dashed line) is somewhere between the population line (solid yellow line) and the individual student's regression line (solid orange line). This property is called **shrinkage**.



Task 1.

The file [pigweights.csv](#) contains weight measurements of 48 pigs over nine successive weeks. Let y_{ij} be the weight of the i th pig on week j . The following models are considered for these data.

- M_0 : no random subject effects
- M_1 : random intercept for each subject
- M_2 : independent random intercept and slope for each subject
- M_3 : correlated random intercept and slope for each subject

- (a) Write down each of the models in mathematical notation and then fit them in R using function `lmer()` from `library(lme4)`.
- (b) Which model would you choose for these data and why?
- (c) Using your selected model, predict the weight of:
 - (i) pig number 16
 - (ii) another pig, not included in this dataset

General form of a linear mixed model

We can write a linear model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$$

A **linear mixed model** is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where \mathbf{X} and \mathbf{Z} are given matrices and

$$E \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \text{Var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

We assume that $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$ independently of each other.

In general we can write

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{ZGZ}^\top + \mathbf{R}) = N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}^*$$

where $\mathbf{e}^* = \mathbf{Z}\mathbf{u} + \mathbf{e}$.

This is a linear model with correlated errors since

$$\text{Var}(\mathbf{e}^*) = \mathbf{V} = \mathbf{ZGZ}^\top + \mathbf{R}.$$

In Example 1, we had $\mathbf{R} = \sigma^2 \mathbf{I}$ where \mathbf{I} is the identity matrix. The correlation for observations from the same student comes from the matrix \mathbf{G} which contains terms for the variance components related to the student effect.



Supplementary material: Covariance structures

An alternative way to specify a linear mixed model so that the observations are correlated with each other is to omit the term $\mathbf{Z}\mathbf{u}$ from the model and introduce the correlation directly within matrix \mathbf{R} instead. In this case \mathbf{R} will be block-diagonal with each block corresponding to each subject/unit. For instance for the college GPA data, the blocks would have dimension 6×6 and they would represent the six GPA values at each occasion for each student.

Several covariance structures can be assumed, such as

- **exchangeable**, which assumes the same correlation between any two values from the same subject/unit, regardless of time distance between the observations,
- **AR(1)** which assumes that only adjacent observations are correlated with each other, and so on.

In R, it is possible to implement this type of covariance structure using `library(nlme)`. Note that this functionality is **not** available in `lme4`.



Supplementary material: Parameter estimation

Estimation of fixed effects:

One way to estimate the coefficients in a linear mixed model is maximum likelihood estimation. For a given \mathbf{V} , the **generalised least squares (GLS)** estimator of $\boldsymbol{\beta}$ is

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.$$

The estimator $\tilde{\boldsymbol{\beta}}$ is the **maximum likelihood estimator (MLE)** and the **uniformly minimum variance unbiased estimator (UMVUE)**. It maximises the log-likelihood

$$-\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \text{const}$$

for the normal model.

Variance component estimation: Maximum likelihood can be used to estimate $\boldsymbol{\beta}$ as well as variance components. The maximum likelihood estimate of

$$\mathbf{V} = \text{Var}(\mathbf{y}) = \mathbf{ZGZ}^\top + \mathbf{R}$$

is based on the model

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$

The log-likelihood of \mathbf{y} under this model is

$$l(\boldsymbol{\beta}, \mathbf{V}) = -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \text{const}$$

and the MLE of $(\boldsymbol{\beta}, \mathbf{V})$ is the one that maximises this expression. For any fixed \mathbf{V} , $l(\boldsymbol{\beta}, \mathbf{V})$ is maximised over $\boldsymbol{\beta}$ by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.$$

Substituting this back into the expression for the log-likelihood, we obtain the *profile log-likelihood* for \mathbf{V} :

$$\begin{aligned} l_P(\mathbf{V}) &= -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) + \text{const} \\ &= -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \mathbf{y}^\top \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}] \mathbf{y} + \text{const} \end{aligned}$$

This can be maximised for the parameters in \mathbf{V} .

However there is a problem with the maximum likelihood approach: it does not adjust for the degrees of freedom lost for estimation. The simplest illustration of this is the following: Consider X_1, \dots, X_n independent $N(\mu, \sigma^2)$. The sample variance is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

which takes into account the one degree of freedom lost in estimating the mean. In contrast, the variance estimate using the maximum likelihood approach is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

with no adjustment for the degree of freedom due to the estimation of the mean.

For this reason, **Restricted Maximum Likelihood (REML)** is the preferred method of estimation in linear mixed models. This maximises the likelihood of linear combinations of the elements of \mathbf{y} that do not depend on $\boldsymbol{\beta}$. The resulting criterion function is the **restricted log-likelihood**

$$l_R(\mathbf{V}) = l_P(\mathbf{V}) - \frac{1}{2} \log |\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}|.$$

REML works as follows. Starting with

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

we find all independent linear combinations of the response, \mathbf{k} such that $\mathbf{k}^\top \cdot \mathbf{X} = 0$. Then

$$\mathbf{k}^\top \cdot \mathbf{y} = \mathbf{k}^\top \cdot \mathbf{X}\boldsymbol{\beta} + \mathbf{k}^\top \cdot \mathbf{Z}\mathbf{u} + \mathbf{k}^\top \cdot \mathbf{e}$$

and taking \mathbf{K} to be the matrix with columns \mathbf{k} we have

$$\mathbf{K}\mathbf{y} = (\mathbf{KZ})\mathbf{u} + (\mathbf{K}\mathbf{e}).$$

Thus

$$\mathbf{K}\mathbf{y} \sim N(\mathbf{0}, \mathbf{K}^\top \mathbf{V} \mathbf{K}) = N(\mathbf{0}, \mathbf{K}^\top \mathbf{Z} \mathbf{G} \mathbf{Z}^\top \mathbf{K} + \mathbf{K}^\top \mathbf{R} \mathbf{K})$$

and the maximum likelihood approach can be used to estimate variance components based on the likelihood of $\mathbf{K}\mathbf{y}$. Since there are no longer fixed effects to estimate, we do not 'lose' degrees of freedom.

REML can also be viewed as **Residual** Maximum Likelihood as it is equivalent to:

- finding the least squares estimates of $\boldsymbol{\beta}$ from regressing \mathbf{y} on \mathbf{X} (ignoring random effects);
- taking the residuals;
- using maximum likelihood on the residuals.

Let us look at another example of fitting a linear mixed model in R.



Example 2.

In this example we will use the `pulp` data from `library(faraway)`, which contains data from an experiment to test the paper brightness depending on a shift operator. The data has 20 observations with `bright` as the response and `operator` as the explanatory categorical variable taking values `a-d`.

```
library(faraway)
data(pulp)
head(pulp)

  bright operator
1   59.8        a
2   60.0        a
3   60.8        a
4   60.8        a
```

5	59.8	a
6	59.8	b

We can fit the following model with operator as a random effect:

$$y_{ij} = \mu + a_i + e_{ij}$$

where

- y_{ij} is the paper brightness measured by the i th operator, $i = 1, \dots, 4$ and $j = 1, \dots, 5$ replicates per operator
- μ is the overall mean
- a_i is the random effect associated with the i th operator
- e_{ij} is the experimental error.

The code to fit the model in R is as follows:

```
library(lme4)
mmod <- lmer(bright ~ 1+(1|operator), data=pulp)
summary(mmod)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: bright ~ 1 + (1 | operator)
Data: pulp
```

```
REML criterion at convergence: 18.6
```

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-1.4666	-0.7595	-0.1244	0.6281	1.6012

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
operator	(Intercept)	0.06808	0.2609
Residual		0.10625	0.3260

Number of obs: 20, groups: operator, 4

```
Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	60.4000	0.1494	404.2

The intraclass correlation coefficient is given by

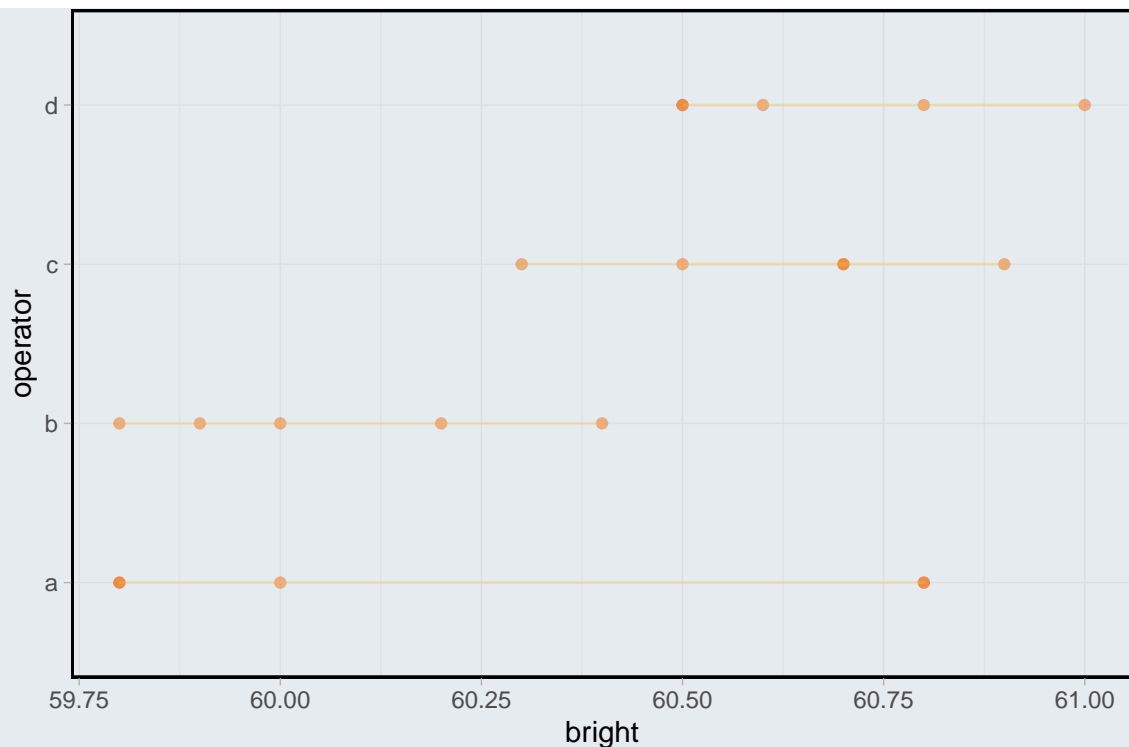
```
0.06808/(0.06808+0.10625)
```

```
[1] 0.3905237
```

so the correlation between brightness measurements from the same operator is estimated to be 0.39.

The main question of interest is whether there is significant variability due to the operator. We can visualize this in the following plot:

```
ggplot(pulp, aes(x=bright, y=operator))+ geom_point(alpha=0.5, color="#ec7014") +
  geom_path(aes(group=operator), alpha=0.3, col="#fec44f")
```



If we were to treat the operator effect as fixed, we would get the following linear model:

```
lmod <- lm(bright ~ operator, data=pulp)
summary(lmod)
```

Call:

```
lm(formula = bright ~ operator, data = pulp)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.440	-0.195	-0.070	0.175	0.560

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	60.2400	0.1458	413.243	<2e-16 ***
operatorb	-0.1800	0.2062	-0.873	0.3955
operatorc	0.3800	0.2062	1.843	0.0839 .
operatord	0.4400	0.2062	2.134	0.0486 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.326 on 16 degrees of freedom

Multiple R-squared: 0.4408, Adjusted R-squared: 0.3359

F-statistic: 4.204 on 3 and 16 DF, p-value: 0.02261

```
anova(lmod)
```

Analysis of Variance Table

Response: bright

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
operator	3	1.34	0.44667	4.2039	0.02261 *
Residuals	16	1.70	0.10625		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

This gives a significant operator effect with a *p*-value of 0.023.

Would we conclude the same from the mixed model? Let us look at the confidence interval for the oper-

ator standard deviation:

```
round(confint(mmod, method="boot", oldNames=FALSE), 3)
```

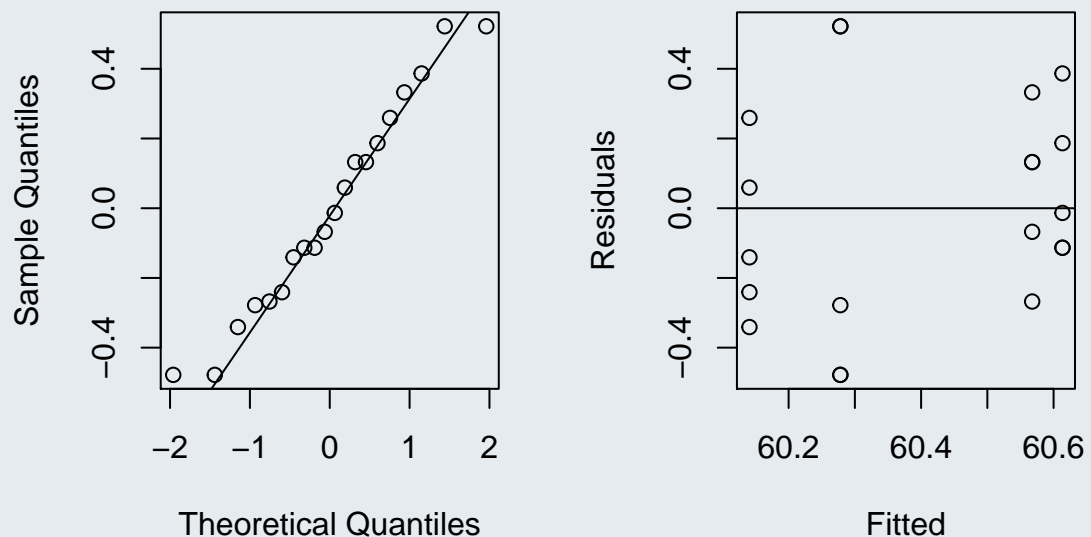
	2.5 %	97.5 %
sd_(Intercept) operator	0.000	0.505
sigma	0.212	0.424
(Intercept)	60.102	60.693

This includes zero, although we are not sure if that's because the variance component is truly zero or because of poor approximations. It is possible to use a parametric bootstrap approach to get a more accurate test of $H_0 : \sigma^2 = 0$, see Section 8.2 of [Faraway](#) for details.

Finally we can check the linearity and normality assumptions through the usual residual plots, e.g.

```
# Diagnostic plots:
```

```
par(mfrow=c(1,2))
qqnorm(resid(mmod), main="")
qqline(resid(mmod))
plot(fitted(mmod), resid(mmod), xlab="Fitted", ylab="Residuals")
abline(0,0)
```



The plots indicate no particular problems. If we had more than four operators, we could also look at the normality of the group level effects, but with so few groups this is not possible.

When should a factor be modelled as a random effect?

In Example 2, the operator variable was treated as a random effect. What is the justification for this?

A categorical variable in a regression model is considered a random effect if its levels can be thought of as samples from a larger population. A typical example is the subject effect in a study with repeated observations for each subject. In such studies we are not usually interested in estimating the mean of every subject, but we do want to take into account the variability introduced in the model by the fact that we do have different subjects. Other examples of effects that may be considered as random are the following:

- In studying the effectiveness of a new treatment for cancer with data collected from several hospitals, "hospital" can be considered as a random effect. The assumption here is that the hospitals participating in the study are only a sample of a larger population of hospitals, and that the results of the analysis will generalise to the wider population of hospitals.
- In a survey of student attitudes, if a few universities are randomly selected to be surveyed, then "university" can be considered as a random effect, as we would expect the results of the survey to generalise to all universities in the UK. On the other hand, if we had data from the Universities of Glasgow and Edinburgh and we were interested in how these two compare, we would treat "university" as a fixed effect and we would estimate means for each. In this case, any conclusions from the modelling would only apply to these two universities and would not generalise to a wider population.



Task 2.

A nutritionist is interested in estimating the average sugar content of breakfast cereal sold in the UK. She randomly selects ten brands of cereal from a list of all different brands sold in the largest supermarkets. Four boxes are bought of each brand and the sugar content (percent sugar) for each box is measured.

- Is the cereal brand here a fixed or a random effect?
- Write down a possible model for these data, stating your assumptions.



Task 3.

An experiment was conducted in order to compare the effects of two insect repellent products in warding off mosquitoes the chemical DEET and a skin product called “Skin so soft” (SSS). Twenty-four volunteers were recruited and they were randomly divided into three equal-sized groups. Three locations, in an area frequented by mosquitoes, were selected randomly and each of the groups was allocated to a location. At each location, one half of the volunteers had DEET spread on their right arm and the other half had “Skin so soft” spread on their right arm, and the volunteers were well spread out. The volunteers allowed mosquitoes to bite their exposed arms for 15 minutes and the number and severity of bites were recorded and converted into a severity score (higher meaning a worse biting experience). The data are stored in the file [bites.csv](#).

- Identify the factors in this experiment. Is each factor fixed or random? Are the factors crossed (do we have data for each combination of the factors?) or nested (does each level of Factor B only occur within one level of Factor A)?
- Write down a possible model for the biting severity data, stating clearly your assumptions.
- Fit the model in R and use it to assess whether the two mosquito repellents produce, on average, systematically different biting severity scores.

Hierarchical models



Example 3.

In Example 1, we studied the relationship between GPA and occasion but we also have other variables in the data. Gender and high school GPA are group-level variables (at the student level) and job status is at the observation (occasion) level. This is what we call a **hierarchical** or **multilevel** structure. Adding the two student-level variables to the model from Example 1 will result in the multilevel model

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 x_{ij} + b_{1i} x_{ij} + \beta_2 v_i + \beta_3 w_i + e_{ij}$$

where v_i is the gender and w_i is the high school GPA of the i th student. You can think of these as a modification to the intercept of the regression line we had computed earlier.

Fitting the model in R is as simple as adding these terms as fixed effects:

```
gpa_multil<-lmer(gpa ~ occasion + sex + highgpa + (1+occasion|student), data=gpa)
summary(gpa_multil)
```

```
Linear mixed model fit by REML ['lmerMod']
```

```
Formula: gpa ~ occasion + sex + highgpa + (1 + occasion | student)
```

```
Data: gpa
```

```
REML criterion at convergence: 250.1
```

```
Scaled residuals:
```

```
      Min       1Q   Median       3Q      Max
-3.0675 -0.5500 -0.0194  0.5376  3.2576
```

```

Random effects:
Groups   Name             Variance Std.Dev. Corr
student (Intercept) 0.050378 0.22445
         occasion    0.004504 0.06711 -0.44
Residual                0.042388 0.20588
Number of obs: 1200, groups: student, 200

Fixed effects:
              Estimate Std. Error t value
(Intercept)  2.271099   0.085635  26.521
occasion      0.106314   0.005885  18.066
sexmale      -0.117956   0.032963  -3.578
highgpa       0.092997   0.027682   3.359

Correlation of Fixed Effects:
          (Intr) occasn sexmal
occasion -0.150
sexmale  -0.113  0.000
highgpa  -0.953  0.000 -0.072

Computing the confidence intervals for the parameters, we see that both are significant:

round(confint(gpa_multil, oldNames=FALSE),3)

              2.5 % 97.5 %
sd_(Intercept)|student      0.184  0.263
cor_occasion.(Intercept)|student -0.596 -0.246
sd_occasion|student         0.057  0.078
sigma                      0.196  0.216
(Intercept)                 2.104  2.439
occasion                    0.095  0.118
sexmale                     -0.183 -0.053
highgpa                     0.039  0.147

```



Task 4.

Section 8.8 of [Faraway](#) gives a detailed analysis of a multilevel dataset. Reproduce the analysis making sure you can identify the different levels and variables associated with them.

Additional resources on linear mixed models



Chapter 8 of [Extending the Linear Model with R](#) by J. Faraway:

Section 8.1 on estimation has more details on how REML works. Section 8.2 on inference gives the basics on *testing the fixed effects* and *testing the random effects*. Section 8.8 goes over an example of a hierarchical model.

Chapter 9 of [Extending the Linear Model with R](#) by J. Faraway discusses longitudinal data (multiple observations from the same subject/unit over time).

Chapter 5 of [Mixed effects models and extensions in ecology with R](#) by Zuur et al.

Sections 5.1-5.3 describe the linear mixed model and show examples of models with random effects (random intercept and random slope).

Sections 5.4-5.9 discuss correlation structures, REML estimation and selection and validation for mixed effects models.

Chapter 7 of [Regression: models, methods and applications](#) by Fahrmeir et al describes the theory of linear mixed models and also provides an introduction to Bayesian linear mixed models.

Week 9 learning outcomes

By the end of this week, you should be able to:

- recognise when there is correlation in the responses of a linear model and why it is important to take it into account when fitting a model
- identify factors as fixed or random and crossed or nested according to the design of the study and the questions of interest
- fit linear mixed models in R including random coefficient models
- fit hierarchical (multilevel) models in R
- predict from linear mixed models, both for subjects/units from the dataset and also for new observations.

Answers to tasks

Answer to Task 1. (a) Suppose we have data (x_{ij}, Y_{ij}) where i labels subjects. Here Y_{ij} is the weight of the i th pig on the j th week and x_{ij} is the week, taking values $1, 2, 3 \dots, 9$.

The most general random coefficient model for Y_{ij} given x_{ij} is of the form

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i} x_{ij} + e_{ij},$$

where β_0, β_1 are unknown parameters; b_{0i}, b_{1i} are random effects; and e_{ij} is the error term.

- Assumptions under model M_3 : e_{ij} independent $N(0, \sigma^2)$, b_{0i} independent $N(0, \sigma_0^2)$, b_{1i} independent $N(0, \sigma_1^2)$ and $\text{Corr}(b_{0i}, b_{1i}) = \rho_{01} \neq 0$. Random variables b_{0i} are independent of e_{ij} and b_{1i} are independent of e_{ij} .
- Assumptions under model M_2 : Same as for M_3 except that now $\text{Corr}(b_{0i}, b_{1i}) = 0$.
- Assumptions under model M_1 , with equation

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + e_{ij}$$

Same as for M_3 except that now $b_{1i} = 0$ so there are no random slopes and no corresponding distributional assumptions involving b_{1i} .

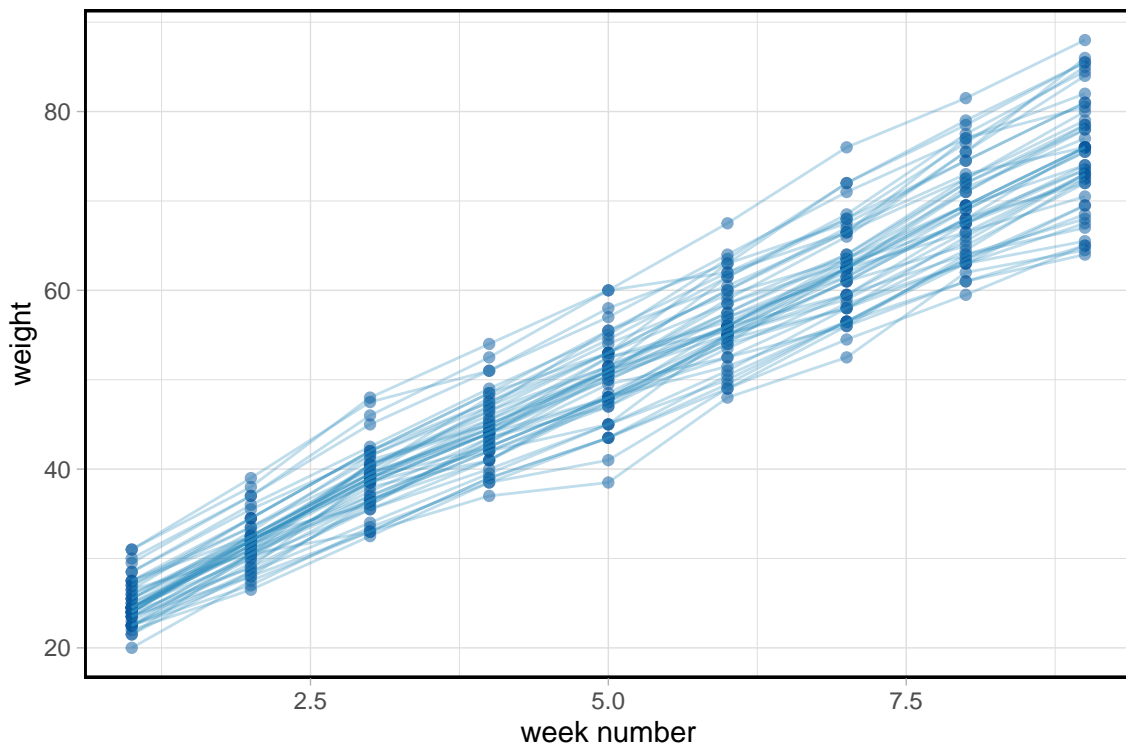
- Assumptions under model M_0 , with equation

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}$$

Same as for M_3 except that now $b_{0i} = b_{1i} = 0$. This is a simple linear regression model with the only random term being the error term e_{ij} .

To fit the models in R, let's start by reading in and plotting the data:

```
pig.weights <- read.csv(url("http://www.stats.gla.ac.uk/~tereza/rp/pigweights.csv"))
ggplot(pig.weights, aes(x=num.weeks, y=weight)) + geom_point(alpha=0.5, color="#08589e") +
  geom_path(aes(group=id.num), alpha=0.3, col="#2b8cbe") + xlab("week number")
```



The models are fitted in R using the following code. Note the model specification for Model M_3 , allowing for correlated random slopes and intercepts, and M_2 , assuming that they are uncorrelated.

```
wm0 <- lm(weight ~ num.weeks, data=pig.weights)
summary(wm0)
```

```

Call:
lm(formula = weight ~ num.weeks, data = pig.weights)

Residuals:
    Min       1Q   Median       3Q      Max
-11.9051  -2.5348  -0.1952   2.5949  13.1751

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.35561    0.46054   42.03  <2e-16 ***
num.weeks    6.20990    0.08184   75.88  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.392 on 430 degrees of freedom
Multiple R-squared:  0.9305,    Adjusted R-squared:  0.9303
F-statistic: 5757 on 1 and 430 DF,  p-value: < 2.2e-16

wm1 <- lmer(weight ~ 1 + num.weeks + (1|id.num), data=pig.weights)
summary(wm1)

Linear mixed model fit by REML ['lmerMod']
Formula: weight ~ 1 + num.weeks + (1 | id.num)
Data: pig.weights

REML criterion at convergence: 2033.8

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.7390 -0.5456  0.0184  0.5122  3.9313

Random effects:
 Groups   Name      Variance Std.Dev.
id.num    (Intercept) 15.142   3.891
Residual                4.395   2.096
Number of obs: 432, groups: id.num, 48

Fixed effects:
            Estimate Std. Error t value
(Intercept) 19.35561    0.60314   32.09
num.weeks    6.20990    0.03906  158.97

Correlation of Fixed Effects:
      (Intr)
num.weeks -0.324

wm2 <- lmer(weight ~ num.weeks + (1|id.num) + (0+num.weeks|id.num), data=pig.weights)
summary(wm2)

Linear mixed model fit by REML ['lmerMod']
Formula: weight ~ num.weeks + (1 | id.num) + (0 + num.weeks | id.num)
Data: pig.weights

REML criterion at convergence: 1741

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.6135 -0.5408  0.0195  0.5464  3.0117

Random effects:
 Groups   Name      Variance Std.Dev.
id.num    (Intercept) 6.9176   2.6301
id.num.1 num.weeks    0.3764   0.6135

```

```

Residual          1.5988    1.2644
Number of obs: 432, groups:  id.num, 48

```

Fixed effects:

```

          Estimate Std. Error t value
(Intercept) 19.35561    0.40211   48.13
num.weeks    6.20990    0.09164   67.77

```

Correlation of Fixed Effects:

```

(Intr)
num.weeks -0.075

```

```

wm3 <- lmer(weight ~ 1 + num.weeks +(1+num.weeks|id.num), data=pig.weights)
summary(wm3)

```

Linear mixed model fit by REML ['lmerMod']

Formula: weight ~ 1 + num.weeks + (1 + num.weeks | id.num)

Data: pig.weights

REML criterion at convergence: 1740.9

Scaled residuals:

```

      Min       1Q   Median       3Q      Max
-3.6202 -0.5474  0.0150  0.5486  2.9939

```

Random effects:

```

Groups   Name             Variance Std.Dev. Corr
id.num   (Intercept)  6.986      2.6432
          num.weeks    0.380      0.6164  -0.06
Residual                  1.597      1.2637
Number of obs: 432, groups:  id.num, 48

```

Fixed effects:

```

          Estimate Std. Error t value
(Intercept) 19.35561    0.40387   47.93
num.weeks    6.20990    0.09204   67.47

```

Correlation of Fixed Effects:

```

(Intr)
num.weeks -0.132

```

- (b) To compare models we can use the `anova()` function, although some caution is needed for testing hypotheses of the type $H_0 : \sigma_1^2 = 0$. If we get a small p -value suggesting that the random slope is significant, that's fine, but if we get a large p -value, we cannot be sure that this is because the random slope effect is not significant or because the test is underpowered.

```

anova(wm3,wm2)

```

Data: pig.weights

Models:

```

wm2: weight ~ num.weeks + (1 | id.num) + (0 + num.weeks | id.num)

```

```

wm3: weight ~ 1 + num.weeks + (1 + num.weeks | id.num)

```

```

      npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
wm2      5 1748.1 1768.4 -869.04  1738.1
wm3      6 1749.9 1774.3 -868.96  1737.9 0.1528  1    0.6959

```

```

anova(wm2,wm1)

```

Data: pig.weights

Models:

```

wm1: weight ~ 1 + num.weeks + (1 | id.num)

```

```

wm2: weight ~ num.weeks + (1 | id.num) + (0 + num.weeks | id.num)

```

```

      npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
wm1      4 2037.8 2054.1 -1014.93  2029.8

```

```
wm2      5 1748.1 1768.4 -869.04  1738.1 291.78  1 < 2.2e-16 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First we test $H_0 : \rho_{01} = 0$ by comparing Models M_2 and M_3 . The p -value is large, so we don't reject H_0 . We can drop the assumption of correlated random slopes and intercepts and go with Model M_2 .

To test $H_0 : \sigma_1^2 = 0$ we compare Models M_1 and M_2 . If H_0 is not rejected, we only need a random intercept, hence Model M_1 should be preferred. Here the p -value is very small suggesting that random slopes are needed and therefore we go with Model M_2 . We will use Model M_2 for predictions.

- (c) (i) The best prediction for the i th pig in the study will include the prediction for its random effects. For $i = 16$ and $x_{ij} = 1, \dots, 9$ we can obtain these as:

```
p1<-predict(wm2, newdata=data.frame(id.num=16, num.weeks=1:9))
p1
```

```
      1      2      3      4      5      6      7      8      9
21.65725 27.78136 33.90548 40.02960 46.15372 52.27783 58.40195 64.52607 70.65019
```

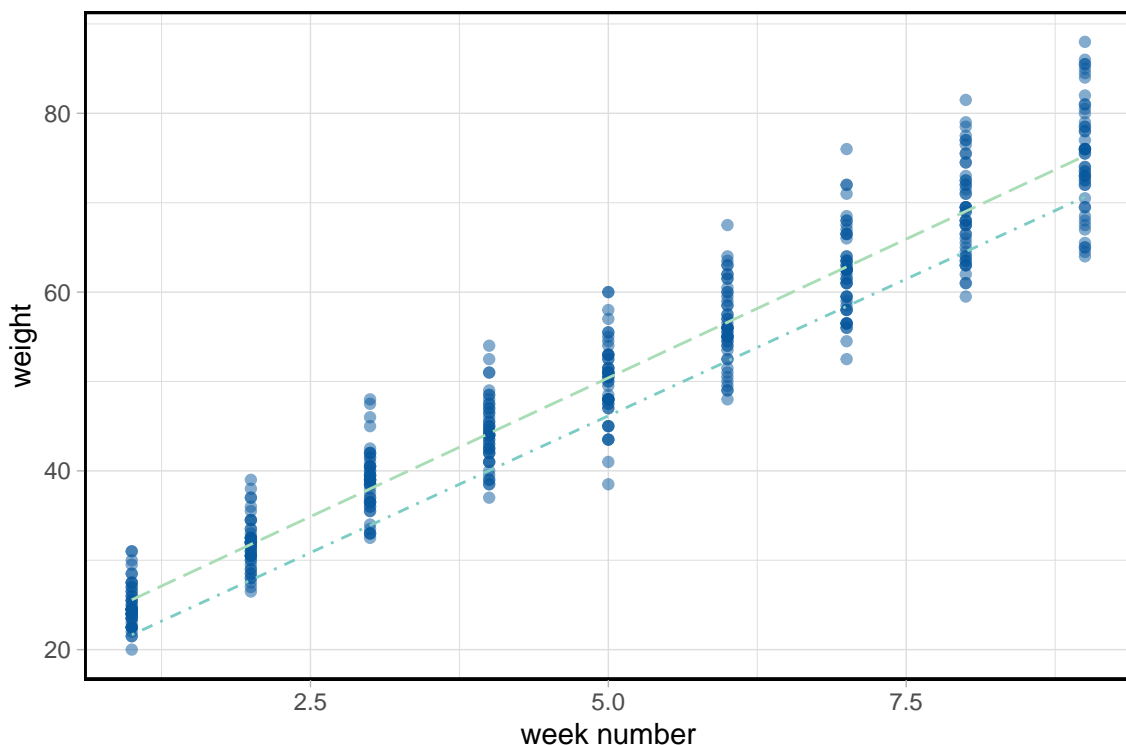
- (ii) The best prediction for a new pig is given by the fixed effect term in the model: $\hat{\beta}_0 + \hat{\beta}_1 x_{ij} = 19.35561 + 6.20990 x_{ij}$. This can also be obtained using the simple linear regression model output.

```
p2<-predict(wm0, newdata=data.frame(num.weeks=1:9))
p2
```

```
      1      2      3      4      5      6      7      8      9
25.56551 31.77541 37.98530 44.19520 50.40509 56.61499 62.82488 69.03478 75.24468
```

We can visualise these predictions as follows:

```
predictions <- data.frame(x=1:9, p1=p1, p2=p2)
ggplot(pig.weights, aes(x=num.weeks, y=weight)) + geom_point(alpha=0.5, color="#08589e") +
  xlab("week number") +
  geom_path(data=predictions, aes(x=x, y=p1), col="#7bccc4", linetype=4) +
  geom_path(data=predictions, aes(x=x, y=p2), col="#a8ddb5", linetype=5)
```



Answer to Task 2. (a) As the brands are a random selection from the population of all brands sold in the UK, we can treat “brand” as a random effect. This way, we can estimate how much variability there is in the sugar content from brand to brand.

- (b) Let y_{ij} be the percent sugar content for the j th box of the i th cereal brand, $i = 1, \dots, 10$ and $j = 1, 2, 3, 4$. A possible model for the sugar content is

$$y_{ij} = \mu + a_i + e_{ij}$$

where μ is the mean sugar content, the a_i are independent $N(0, \sigma_A^2)$ random variables, the e_{ij} are independent $N(0, \sigma^2)$ errors and the a_i and e_{ij} are independent of each other.

Answer to Task 3. (a) The factors in this experiment are product (fixed) and location (random). The factors are crossed (we have observations from all six combinations of product and location).

- (b) Let y_{ijk} be the biting severity score for the k th volunteer at the j th location using the i th product where $i = 1, 2$ for DEET and SSS respectively, $j = 1, 2, 3$ and $k = 1, 2, 3, 4$. A possible model for the severity score data is

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

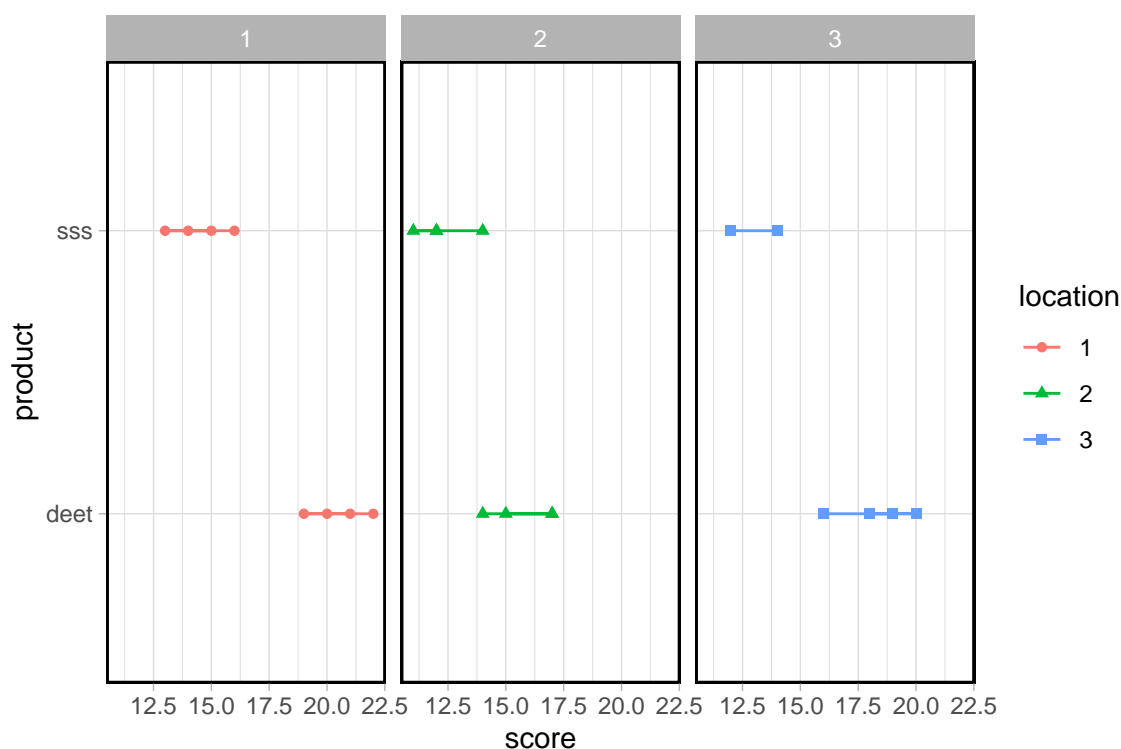
where $\sum_{i=1}^2 \alpha_i = 0$, b_j independent $N(0, \sigma_B^2)$, $(\alpha b)_{ij}$ independent $N(0, \sigma_{AB}^2)$ and e_{ijk} independent $N(0, \sigma^2)$ are all mutually independent.

- (c) Let us first read in and plot the data:

```
bites <- read.csv(url("http://www.stats.gla.ac.uk/~tereza/rp/bites.csv"))
head(bites)
```

```
  score product location
1    21    deet        1
2    19    deet        1
3    20    deet        1
4    22    deet        1
5    14     sss        1
6    15     sss        1
```

```
bites$location <- as.factor(bites$location)
ggplot(bites, aes(x=score, y=product)) +
  geom_point(aes(shape = location, color=location)) +
  geom_path(aes(group=product, color=location)) +
  facet_wrap(~location)
```



The severity score appears to be lower for SSS on average, with some variability by location.

We fit a model with a random effect for the location and a random effect for the interaction between product and location.

```
bm1 <- lmer(score ~ product+(1|location)+(1|product:location), data=bites)
summary(bm1)
```

Linear mixed model fit by REML ['lmerMod']

Formula: score ~ product + (1 | location) + (1 | product:location)

Data: bites

REML criterion at convergence: 87.1

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.65864	-0.73969	0.09461	0.55061	1.47008

Random effects:

Groups	Name	Variance	Std.Dev.
product:location	(Intercept)	0.3194	0.5652
location	(Intercept)	2.6771	1.6362
Residual		1.8472	1.3591

Number of obs: 24, groups: product:location, 6; location, 3

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	18.1667	1.0737	16.920
productsss	-4.7500	0.7217	-6.582

Correlation of Fixed Effects:

	(Intr)
productsss	-0.336

```
confint(bm1, oldNames=FALSE)
```

	2.5 %	97.5 %
sd_(Intercept) product:location	0.000000	2.013141
sd_(Intercept) location	0.000000	4.192534
sigma	1.012277	1.917831
(Intercept)	15.820055	20.513286
productsss	-6.394877	-3.105091

The product effect is significant with DEET (product 1) having a higher severity score than SSS by somewhere between 1.6 and 3.2 units on average.

It is not clear if the variability due to location and due to the interaction of product and location is significant, and one could experiment with dropping the interaction term and comparing models using an appropriate test. However, in this case we are interested in testing for a difference in the means of the fixed effect adjusting for any extra variability introduced by the random effect, so the emphasis is on the interpretation of the product coefficient.