Lab 8

Jada Zorn, Ricardo Javier Leiton Cubillo, Ceili DeMarais

# Rules

In groups of 2 or 3, complete the following.

# Modeling Snow Density

We will be using the data set from Wetlaufer, Hendrikx, and Marshall (2016) - study where they explored the relationship between snow density () or snow depth (snow, mm) or snow water equivalent (SWE, mm) with a suite of predictor variables. In this lab we will condition on there being snow present and then try to model snow density with models that will compete with what they discuss in Tables 3 and 4. We will be interested in using elev (Elevation, m), Land (forest cover with 0 = unforested and 10 = forested), rad (Potential Solar radiation, ), curvature (see <https://blogs.esri.com/esri/arcgis/2010/10/27/understanding-curvature-rasters/> for a description), aspect (orientation of slope in degrees (0 to 360 degrees)), and angle (angle of slope in degrees with 0 being flat) as predictors. Also pay attention to the strata variable (read its definition in the paper) and the role that played in the data collection and should in the analysis, as we will revisit this.

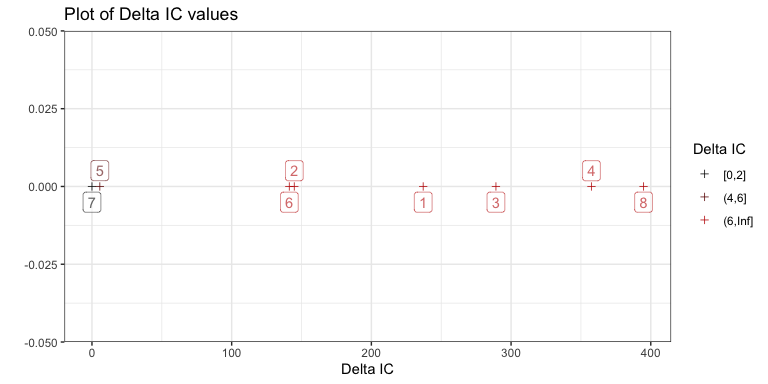
* Wetlaufer, K., Hendrikx, J., and L. Marshall (2016) Spatial Heterogeneity of Snow Density and Its Influence on Snow Water Equivalence Estimates in a Large Mountainous Basin. *Hydrology*, 3(1):3, <doi:10.3390/hydrology3010003>. Available at <http://www.mdpi.com/2306-5338/3/1/3/htm> and on D2L

Run the following code to get re-started with the data set.

data(snowdepths)  
snowdepths <- snowdepths %>%  
 mutate(AspectCat = factor(case\_when(  
 aspect %in% (0:45)~ "North",  
 aspect %in% (315:360)~ "North",  
 aspect %in% 45:(90+45) ~ "East",  
 aspect %in% (90+45):(180+45) ~ "South",  
 aspect %in% (180+45):315 ~ "West"  
 )),  
 SnowPresence = factor(case\_when(  
 snow == 0 ~ "None",  
 snow > 0 ~ "Some"  
 )),  
 Landf = factor(cover),  
 Landf = fct\_recode(Landf,  
 "Not Forested" = "0",  
 "Forested" = "10")  
 )  
  
snowdepthsR <- snowdepths %>% drop\_na(density)  
lmA <- lm(density ~ rad + snow + elev + AspectCat, data = snowdepthsR)  
lmDepth <- lm(density ~ snow, data = snowdepthsR)  
lmSWE <- lm(density ~ swe, data = snowdepthsR)  
lmC <- lm(density ~ snow + swe, data = snowdepthsR)  
lmElev <- lm(density ~ elev, data = snowdepthsR)  
lmRad <- lm(density ~ rad, data = snowdepthsR)  
lmSlope <- lm(density ~ angle, data = snowdepthsR)  
lmER <- lm(density ~ elev + rad, data = snowdepthsR)  
lmDR <- lm(density ~ snow + rad, data = snowdepthsR)  
lmmeanonly <- lm(density ~ 1, data = snowdepthsR)  
lmFull <- lm(density ~ elev + rad + snow + angle + AspectCat, data = snowdepthsR)  
d1 <- model.sel(list(lmDepth, lmElev, lmRad, lmSlope, lmER, lmDR, lmFull, lmmeanonly), rank = "AIC")  
d1

## Model selection table   
## (Intrc) snow elev rad angle AspcC df logLik AIC delta  
## 7 528.5 0.01325 -0.1284 0.0004326 -0.314 + 9 -4117.800 8253.6 0.00  
## 5 479.5 -0.1104 0.0004878 4 -4125.604 8259.2 5.61  
## 6 259.7 -0.04034 0.0004390 4 -4193.436 8394.9 141.27  
## 2 593.7 -0.1121 3 -4196.215 8398.4 144.83  
## 1 362.4 -0.04413 3 -4242.349 8490.7 237.10  
## 3 206.9 0.0005043 3 -4268.379 8542.8 289.16  
## 4 342.9 -1.699 3 -4302.568 8611.1 357.54  
## 8 320.7 2 -4322.212 8648.4 394.83  
## weight  
## 7 0.943  
## 5 0.057  
## 6 0.000  
## 2 0.000  
## 1 0.000  
## 3 0.000  
## 4 0.000  
## 8 0.000  
## Models ranked by AIC(x)

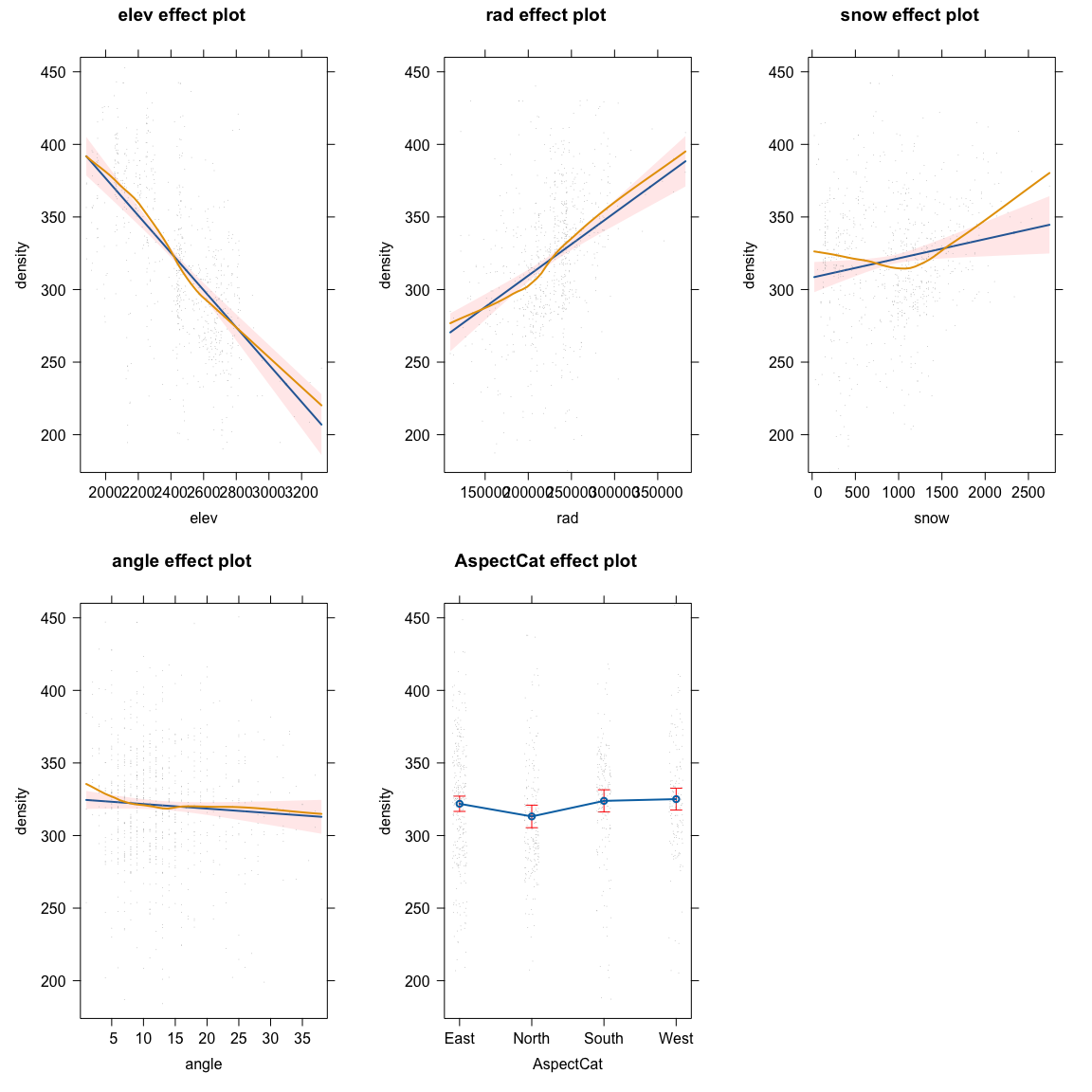
ggdredgeplot(d1)



**1) The following effects plot with partial residuals is modified so the plot options to work a bit better with large data sets. Use the provided plot to assess the linearity assumption as it relates to elevation and snow depth in the provided model.**

**There is a slight violation of linearity for elevation shown with the smoothing line being slightly above the linear fit for low elevations and below for middle and the maybe high after that where there are few points. It is not a clear violation since the partial residuals are fairly spread out and the pattern is now clearly showing in the points, but this could be explored more…**

plot(allEffects(lmFull, residuals = T), residuals.color = "darkgrey", lwd = 2, band.colors = "red", band.transparency = 0.1, residuals.pch = 16, residuals.cex = 0.1)



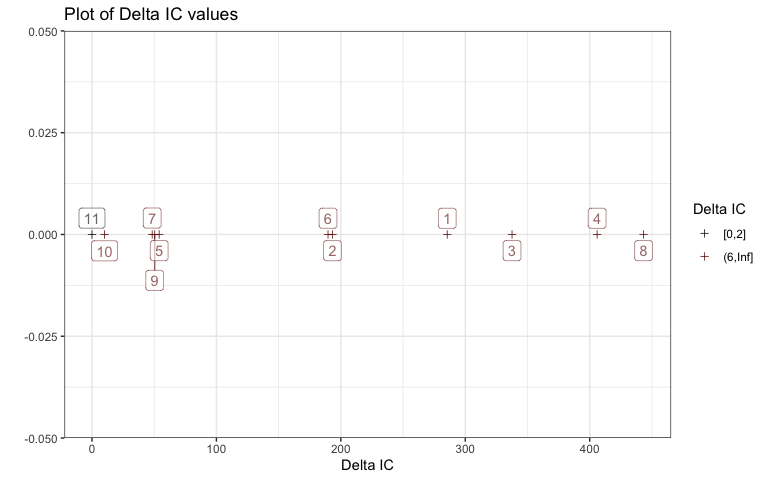
1. Regardless of your assessments of linearity in the previous model, consider the following three models to add to the previous model selection results. After adding those to the model.sel AIC comparisons provided above, what do you discover? Write up a couple of sentences to summarize the top model and its evidence versus other top models and what that suggests about the need for the two quadratic versions of the polynomial terms here.

According to model select and dredge plot, the top model is lmFull2. This model includes two polynomial terms for elevation and snow depth, along with the other variables of radiation, angle, and aspectcat. It has the lowest AIC value of 8205.2, and the second next model has a delta AIC of 10, which strongly supports this as the best model.

lmFulle2 <- lm(density ~ poly(elev,2) + rad + snow + angle + AspectCat, data = snowdepthsR)  
lmFulls2 <- lm(density ~ elev + rad + poly(snow,2) + angle + AspectCat, data = snowdepthsR)  
lmFull2 <- lm(density ~ poly(elev,2) + rad + poly(snow,2) + angle + AspectCat, data = snowdepthsR)  
  
d2 <- model.sel(list(lmDepth, lmElev, lmRad, lmSlope, lmER, lmDR, lmFull, lmmeanonly, lmFulle2, lmFulls2, lmFull2), rank = "AIC")  
d2

## Model selection table   
## (Int) snw elv rad ang AsC ply(elv,2) ply(snw,2) df  
## 11 226.9 0.0004471 -0.5280 + + + 11  
## 10 456.2 -0.09204 0.0004269 -0.6091 + + 10  
## 7 528.5 0.01325 -0.12840 0.0004326 -0.3140 + 9  
## 9 215.9 0.01330 0.0004321 -0.3182 + + 10  
## 5 479.5 -0.11040 0.0004878 4  
## 6 259.7 -0.04034 0.0004390 4  
## 2 593.7 -0.11210 3  
## 1 362.4 -0.04413 3  
## 3 206.9 0.0005043 3  
## 4 342.9 -1.6990 3  
## 8 320.7 2  
## logLik AIC delta weight  
## 11 -4091.592 8205.2 0.00 0.994  
## 10 -4097.635 8215.3 10.09 0.006  
## 7 -4117.800 8253.6 48.41 0.000  
## 9 -4117.795 8255.6 50.41 0.000  
## 5 -4125.604 8259.2 54.02 0.000  
## 6 -4193.436 8394.9 189.69 0.000  
## 2 -4196.215 8398.4 193.25 0.000  
## 1 -4242.349 8490.7 285.51 0.000  
## 3 -4268.379 8542.8 337.57 0.000  
## 4 -4302.568 8611.1 405.95 0.000  
## 8 -4322.212 8648.4 443.24 0.000  
## Models ranked by AIC(x)

ggdredgeplot(d2)



1. We started writing out the model with both elevation and snow depth as quadratic polynomials. Complete the redacted part of writing out the estimated model.

* where , , , and are orthogonal polynomials of the original elevation and snow depth predictors and where is 1 for a north facing slope location and 0 otherwise, is 1 for a south facing slope location and 0 otherwise, and is 1 for a west facing slope location and 0 otherwise,

lmFull2 %>% tbl\_regression(intercept = T)

| **Characteristic** | **Beta** | **95% CI***1* | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 227 | 200, 253 | <0.001 |
| elev |  |  |  |
| elev | -581 | -775, -386 | <0.001 |
| elev² | -180 | -282, -78 | <0.001 |
| rad | 0.00 | 0.00, 0.00 | <0.001 |
| snow |  |  |  |
| snow | -132 | -317, 54 | 0.2 |
| snow² | 413 | 302, 524 | <0.001 |
| angle | -0.53 | -0.98, -0.08 | 0.021 |
| AspectCat |  |  |  |
| East | — | — |  |
| North | -7.5 | -17, 1.8 | 0.12 |
| South | -0.69 | -9.5, 8.1 | 0.9 |
| West | 10 | 0.58, 20 | 0.038 |
| *1*CI = Confidence Interval | | | |

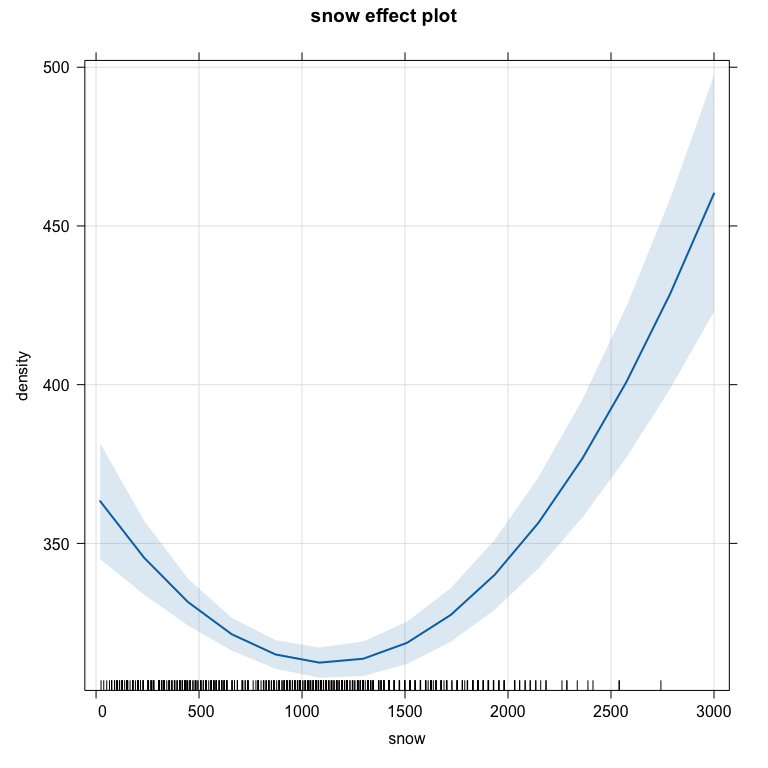
summary(lmFull2)

##   
## Call:  
## lm(formula = density ~ poly(elev, 2) + rad + poly(snow, 2) +   
## angle + AspectCat, data = snowdepthsR)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -153.679 -22.637 0.469 22.544 191.584   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 2.269e+02 1.348e+01 16.833 < 2e-16  
## poly(elev, 2)1 -5.806e+02 9.914e+01 -5.857 6.95e-09  
## poly(elev, 2)2 -1.800e+02 5.189e+01 -3.468 0.000554  
## rad 4.471e-04 5.440e-05 8.219 8.45e-16  
## poly(snow, 2)1 -1.318e+02 9.444e+01 -1.396 0.163154  
## poly(snow, 2)2 4.133e+02 5.651e+01 7.314 6.43e-13  
## angle -5.280e-01 2.289e-01 -2.307 0.021333  
## AspectCatNorth -7.514e+00 4.763e+00 -1.578 0.115076  
## AspectCatSouth -6.900e-01 4.494e+00 -0.154 0.878018  
## AspectCatWest 1.007e+01 4.837e+00 2.082 0.037639  
##   
## Residual standard error: 42.4 on 783 degrees of freedom  
## Multiple R-squared: 0.441, Adjusted R-squared: 0.4346   
## F-statistic: 68.64 on 9 and 783 DF, p-value: < 2.2e-16

1. The quadratic term for snow depth is extracted using the number of its panel from the suite of all the effects plots from the full model with two quadratic polynomial terms so you can focus on it. Write a size interpretation sentence for this term in this model. Use the grid lines to help with the discussion (you should discuss values of the estimated response and how they change across the range of the predictor). Make sure you use your knowledge of the units of the predictor and response in the interpretation. No CI reporting is done for polynomial size interpretations.

At a snow depth of 0, the estimated mean snow density is about 360. As snow depth increases to around 100, estimated mean density decreases to around 300. As depth continues to increase, snow density then begins to increase. At a snow depth of 3000, we then have an estimated mean snow density of around 460.

plot(allEffects(lmFull2), grid = T, 3)



1. A variable we have not considered much so far is whether the location had trees or not (forested/not), which is available in the cover variable in the original data set. Review the provided data wrangling code from the start of the lab to find a better version of that variable and note what changed about the variable besides its name.

The data wrangling code from the beginning of the lab created a new variable named “Landf”, which took the cover variable and changed it into a factor. These values were also recoded to show “0” as “Not Forested” and “10” and “Forested. The end result was a categorical variable named”Landf” with two categories of “Nor Forested” and “Forested”.

summary(snowdepthsR$Landf)

## Not Forested Forested   
## 499 294

#Landf = factor(cover),  
 #Landf = fct\_recode(Landf,  
 #"Not Forested" = "0",  
 #"Forested" = "10")

1. Using snowdepthsR, make a contingency table of counts of the improved version of cover and AspectCat. Is this a balanced design with respect to these two variables? What combinations had the smallest and largest counts?

This is not a balanced design with respect to these two variables as the count combinations were not equal throughout. West Forested areas had only 21 observations, but East Non-Forested areas had 176.

tally(AspectCat ~ Landf, data = snowdepthsR)

## Landf  
## AspectCat Not Forested Forested  
## East 176 104  
## North 114 81  
## South 93 88  
## West 116 21

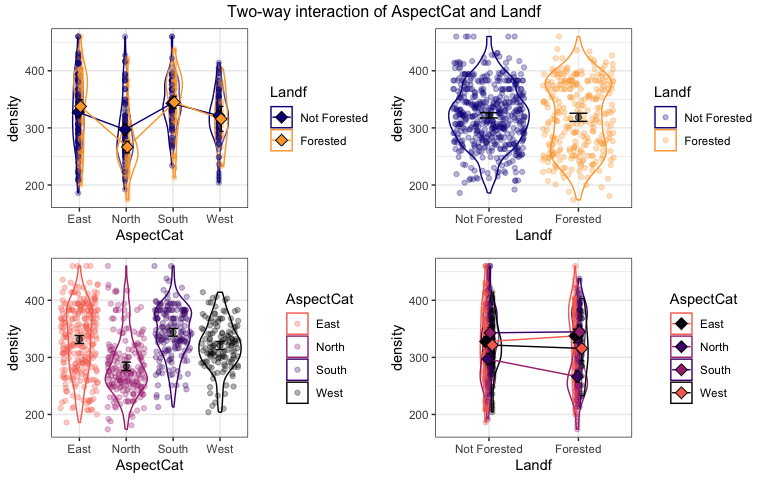
1. Make an interaction plot with ggintplot and discuss the pattern of a potential interaction, making sure you use the words “estimated mean snow density” at least once. Based on the plots, do you think you will find evidence that there is an interaction present? Why or why not?

**Pattern: In north facing slopes the estimated mean snow density for forest sites is 275 but the non-forested sites have an estimated mean of 300 (it decreases) but for every other level of aspect, the difference in land cover level is minimal even though the means for different aspects do vary with South being a bit higher than East or West.**

**Preliminary evidence: There is a large data set but there is also a lot of variability around the means in the observations, so it might be possible to find evidence related to an interaction, but it is not a very dramatic pattern of interaction, so we might fail to detect it. We need to get to modeling to find out...**

Looking at the two-way interaction plots, we think that we will find evidence of an interaction as the changes in estimated snow density do not seem to be the same for different groups. Looking at the bottom right plot, lines connecting estimated mean densities for aspect slope categories are not parallel to each other, suggesting that there may be an interaction between aspect and landF on density. We can also see this in the top left plot, with changes in estimated mean density differing for forested and non forested plots.

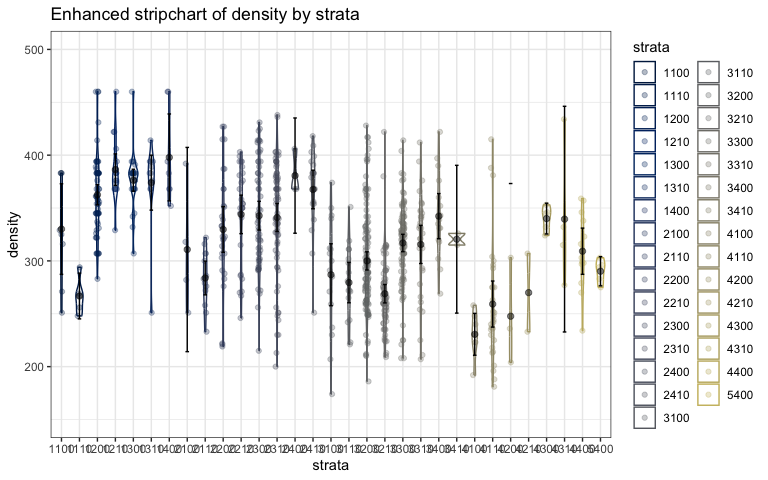
ggintplot(response = "density", groupvars = c("AspectCat", "Landf"), data = snowdepthsR)



1. Because the sampling design involved stratified random sampling, the strata (see paper for definition) constitute a blocking or grouping variable that should be included in the model to avoid violating the independence assumption. The following plot explores the density responses versus the strata and the tally result helps with counting. How many levels of strata are there? If you added strata to the linear model, how many indicator variables would it create? Is this a balanced design with respect to strata?

There are 31 levels of strata. If we added this into the linear model we would see 30 indicator variables, as one of them would be set at the baseline. This is not a balanced design with respect to strata as some strata only have 2-3 observations, while others have up to 124.

enhanced\_stripchart(density ~ strata, data = snowdepthsR) + ylim(c(150,500))



tally(~strata, data = snowdepthsR)

## strata  
## 1100 1110 1200 1210 1300 1310 1400 2100 2110 2200 2210 2300 2310 2400 2410 3100   
## 8 5 62 17 27 12 7 4 13 29 28 57 68 3 20 16   
## 3110 3200 3210 3300 3310 3400 3410 4100 4110 4200 4210 4300 4310 4400 5400   
## 16 124 68 91 32 16 2 7 29 3 2 5 4 13 5

1. Because we really do not care too much about differences in strata but need to account for it, we usually treat this sort of blocking variable with lots of levels as a random effect and fit a linear mixed model. The following code uses lmerTest to fit a lmer (pronounced el-mer or (my preference) lee-mer) with the previously explored aspect and forested variables and their interaction as fixed effects and strata as a random effect. The Anova(..., test.statistic = "F") code generates an F-test for the fixed effects with the previous rules based on “types” of F-tests that are also *conditional on the random effects in the model*. Complete the evidence sentence for the interaction in the provided model.

lmerA <- lmer(density ~ Landf\*AspectCat + (1|strata), data = snowdepthsR)  
Anova(lmerA, test.statistic = "F")

## Analysis of Deviance Table (Type II Wald F tests with Kenward-Roger df)  
##   
## Response: density  
## F Df Df.res Pr(>F)  
## Landf 0.1961 1 711.49 0.658028  
## AspectCat 5.3790 3 774.24 0.001144  
## Landf:AspectCat 2.7288 3 778.63 0.043031

* There is some evidence against the null hypothesis of no interaction between land cover and aspect on the snow density ($F\_{3,778.63} = 2.7288, p-value = 0.043), controlled for strata as a random effect, so we would conclude that there is evidence of an interaction between land cover and aspect on snow density and keep the interaction in the model.

1. The following results are from the linear model where we do not account for strata. Report the test statistic, distribution under the null, and p-value for the interaction in this model. How can you explain the difference in this result as compared to the previous mixed model result?

lmA <- lm(density ~ Landf\*AspectCat, data = snowdepthsR)  
Anova(lmA)

## Anova Table (Type II tests)  
##   
## Response: density  
## Sum Sq Df F value Pr(>F)  
## Landf 3875 1 1.4629 0.2268340  
## AspectCat 388163 3 48.8448 < 2.2e-16  
## Landf:AspectCat 47814 3 6.0167 0.0004707  
## Residuals 2079429 785

In the interaction model, there is strong evidence against the null hypothesis of no interaction between land cover and aspect on the snow density ($F\_{3,785} = 6.0167, p-value = 0.0004). This differs from the results of the mixed model because the linear model does not account for the variability that comes from strata. The mixed model treats strata as a random effect, which adjusts for the variability that comes from different strata. The linear model, which does not account for this variability, thus shows a much smaller p-value of 0.0004, inflating the term. By not accounting for strata, the linear model may mis-estimate the evidence for the interaction.

1. Report any resources outside your group and the course provided materials that you used and how they impacted your answers. Report NONE if there were none.

NONE