**Homework 2: Mixed-Models course 2016**

**The homework consists of 2 parts , each of which involves doing the steps of mixed-model analysis that we discussed today in class.**

**What you have to hand in:**

Hand in your answers, the R code, and the requested figures: Provide the answers and figures first, and copy/paste the whole R code and relevant output below into a word document, print it out, and bring it to the class/lab session on **Feb 22, 2016**.

**(1) Analyze the *N* = 30 Cherry Pit Spitting data file.**

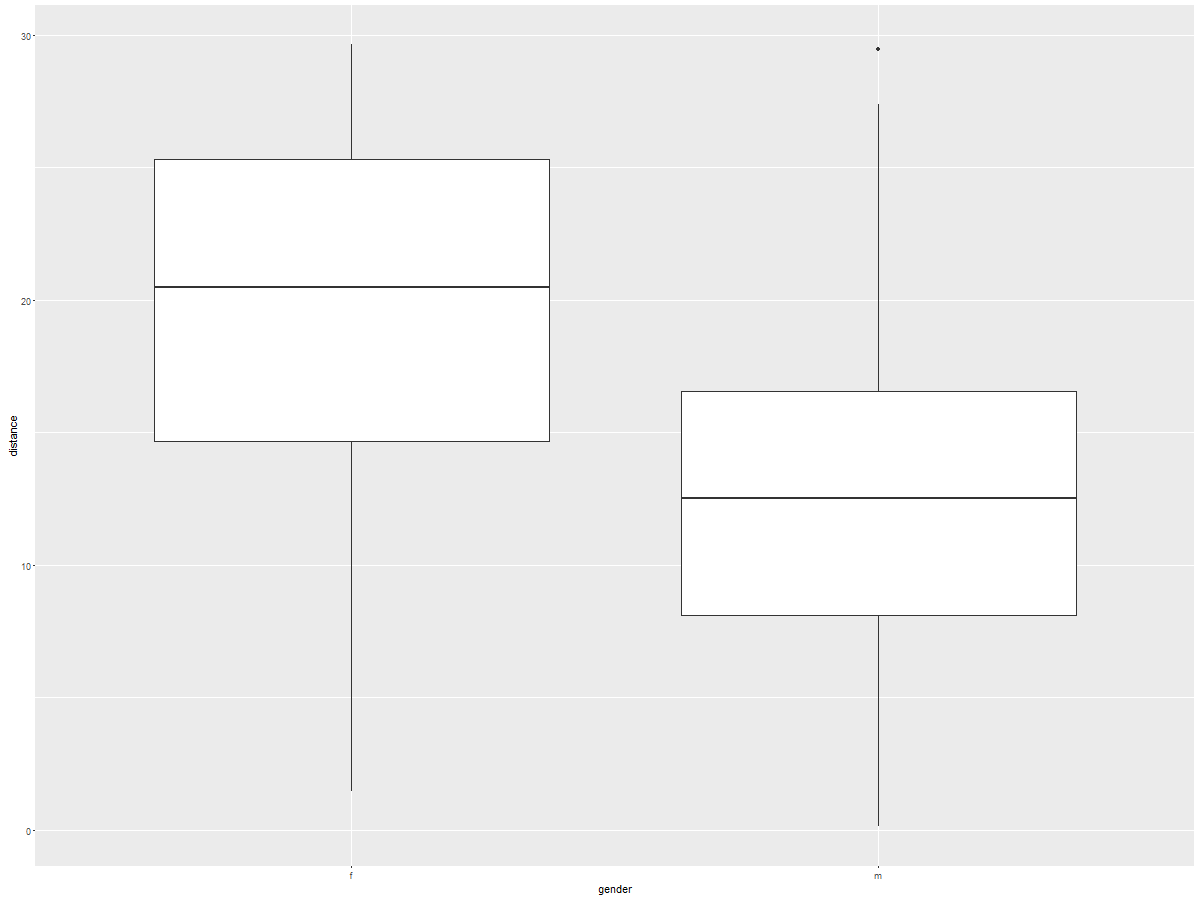
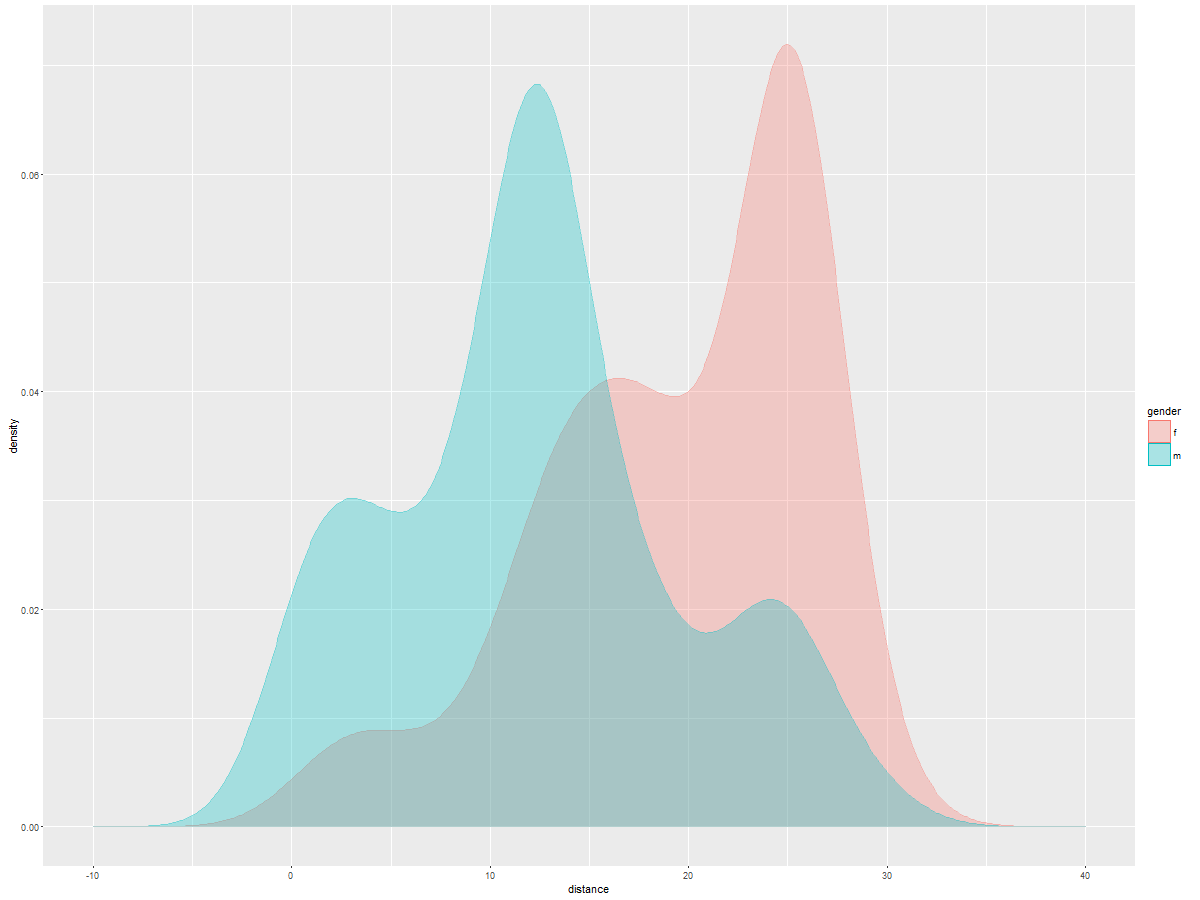
**The file is called** CherryPit\_large\_N30Gender.csv **and can be found on BlackBoard --> Week 2 --> Homework**

*Make sure you have set your contrast setting to sum-to-zero contrasts.*

The hypothesis you want to test is whether there is a gender effect in how far participants can spit their cherry pits.

**(a) Write down the lmer syntax for the model you want to use to test your hypothesis. Do not yet run the model.**

**(b) Create some plots to understand the data, before you are going to test this hypothesis statistically:**

* A boxplot showing distance as a function of gender
* 
* A densityplot showing a separate curve for females and males. Use ggplot2 to create a plot with a single panel that shows somewhat transparent curves for females (pink) and males (blue).
* 

**(c) Run your lmer model and do the following things:**

* Thoroughly inspect your summary() for the model:
* What is the variance associated with the participant code? **25.33**
* What is the residual variance? **25.29**
* Compute the ICC as discussed during class = **ICC= 25.33/(25.33+25.29) = 0.5003951**
* Use this ICC estimate to compute the "effective sample size" using the formula they give in the Aarts et al. (2014) paper (hint: they give it in Figure 2d) **neff = 150/(1+(5-1)\*0.5003951) = 49.97367**
* What is the estimate for the intercept? What does that number mean? **the estimate is 15.981,** **this means that the overall spitting distance of a person is 15.98**
* What is the estimate for the gender effect? **What does that number mean? 3.621, females spit 3.621 distance more than the average of 15.981, and males spit 3.621 less than the average of 15.981**
* So, based on these, what do you think is the average distance for females and the average distance for males? **For females is 19.602, for males is 12.36**
* Verify this by using describeBy from the library psych to get the raw means for males and **females.**

**cherry$gender: f**

**[1] 19.60231**

**---------------------------------------------------------------------------------------------**

**cherry$gender: m**

**[1] 12.35934**

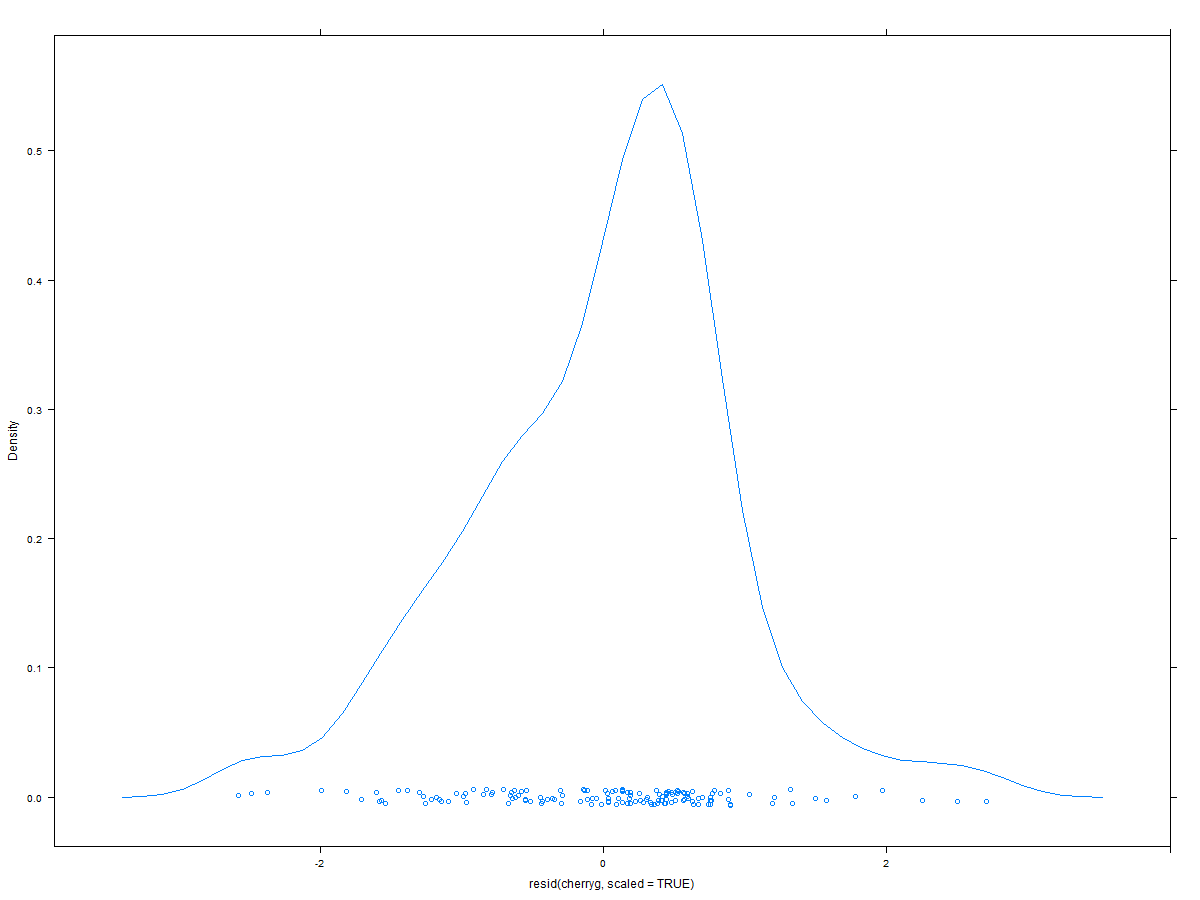
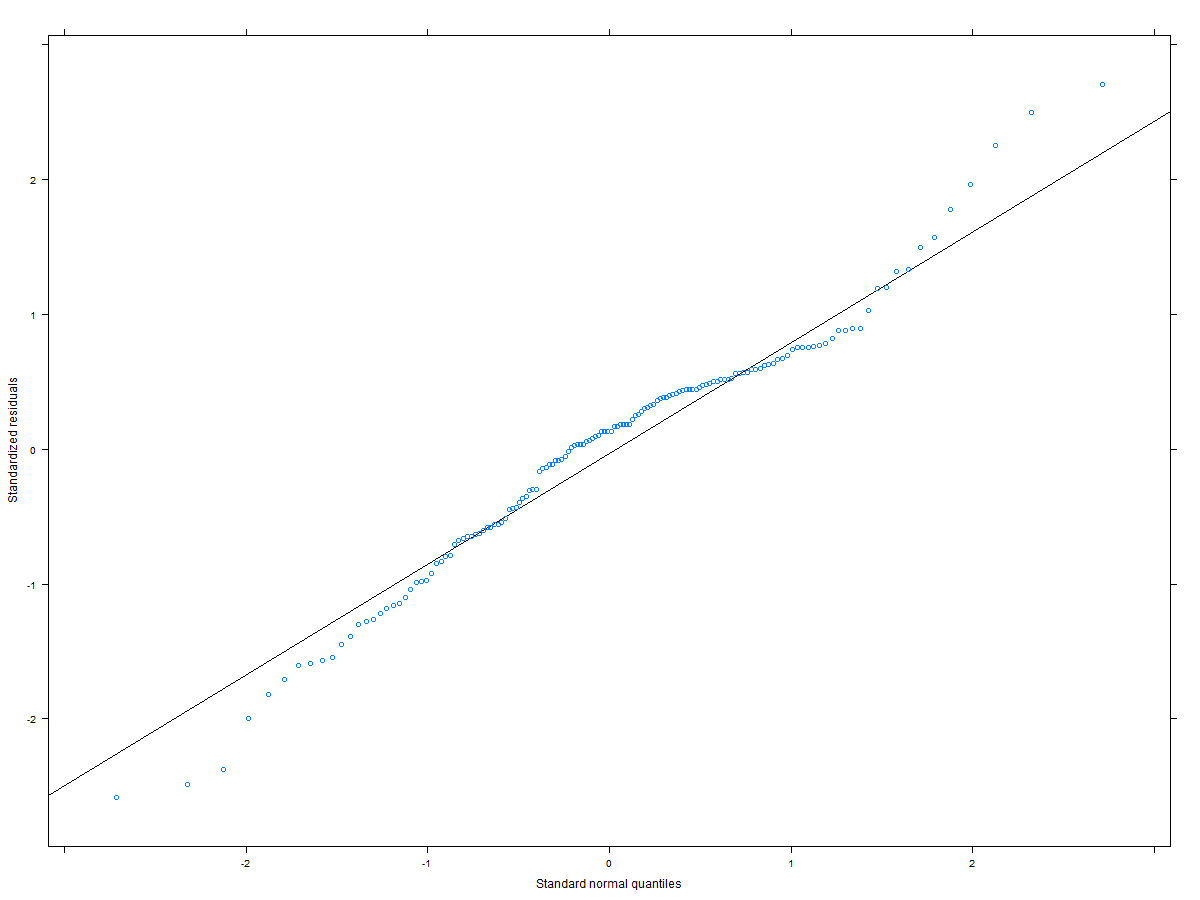
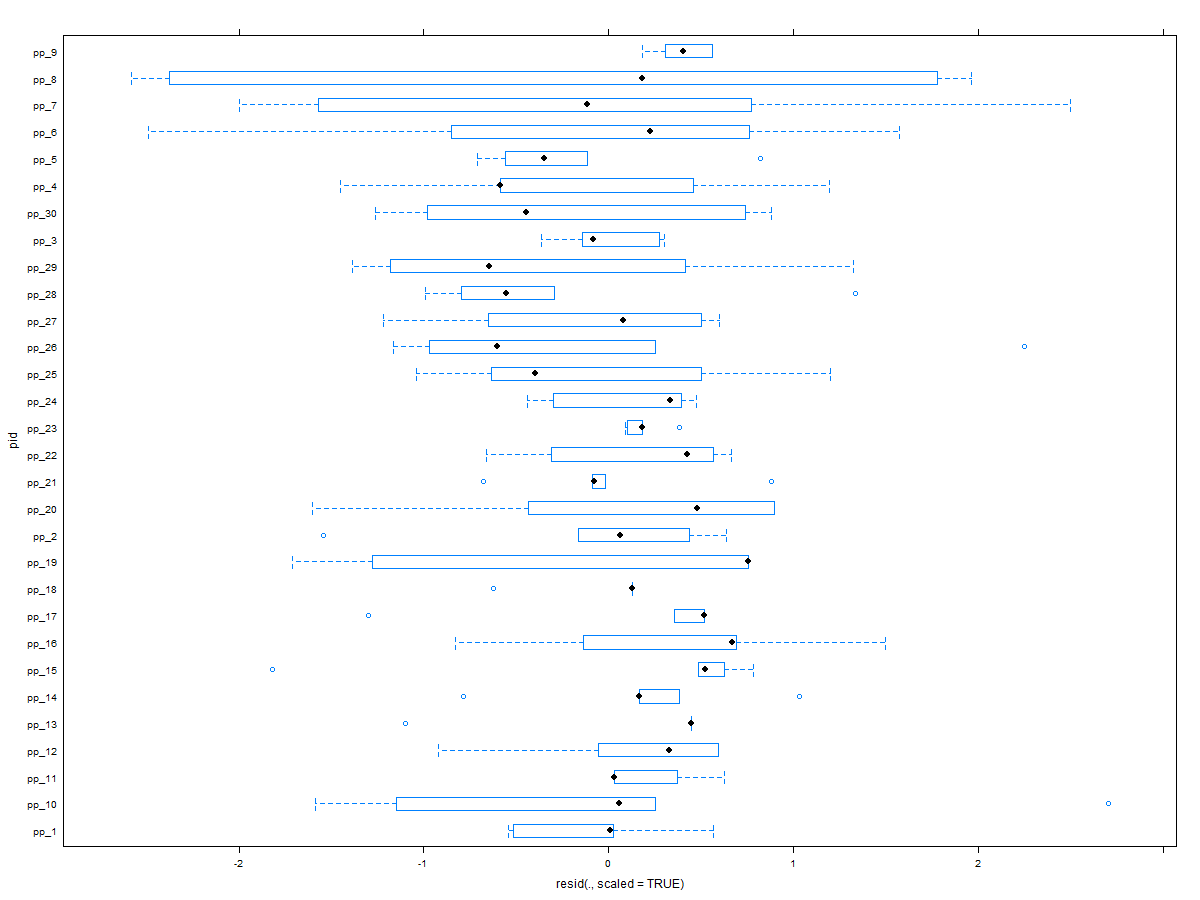
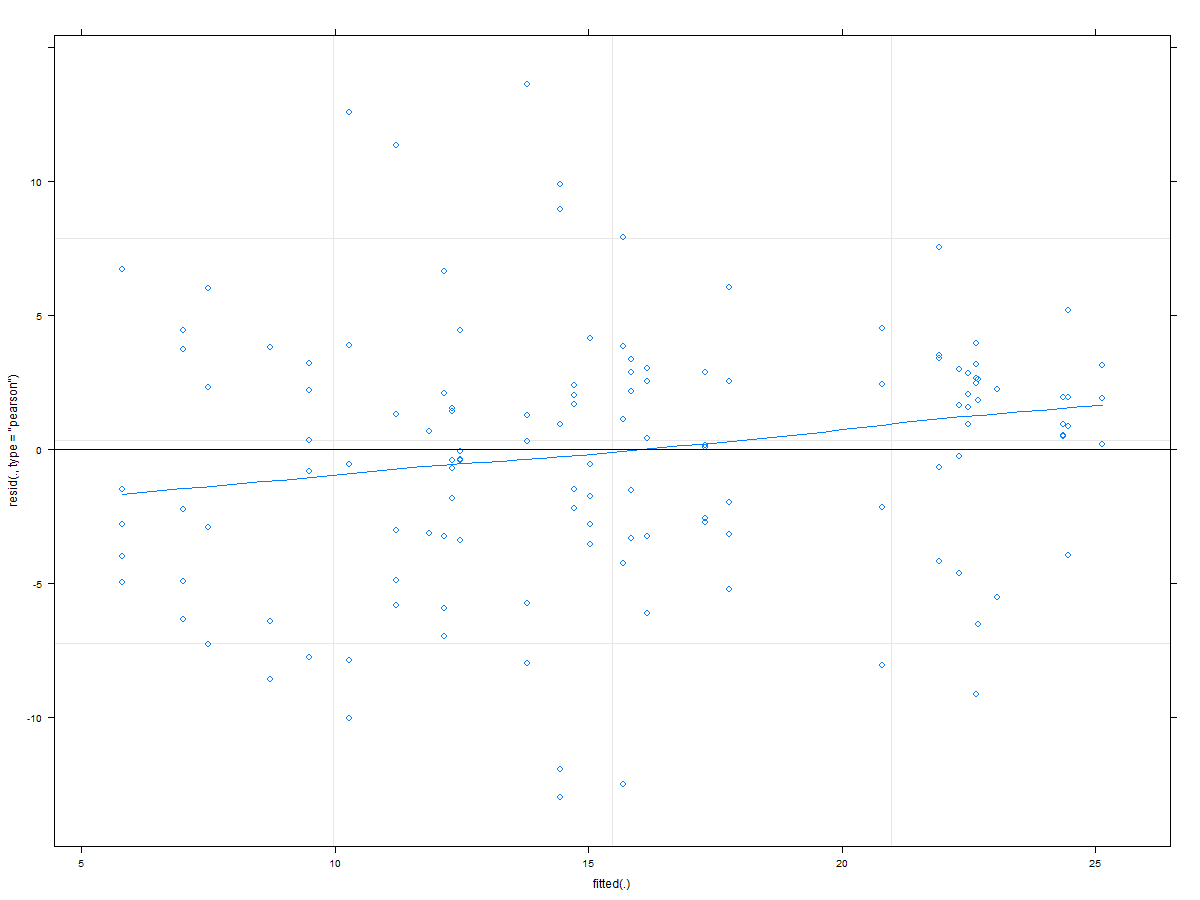
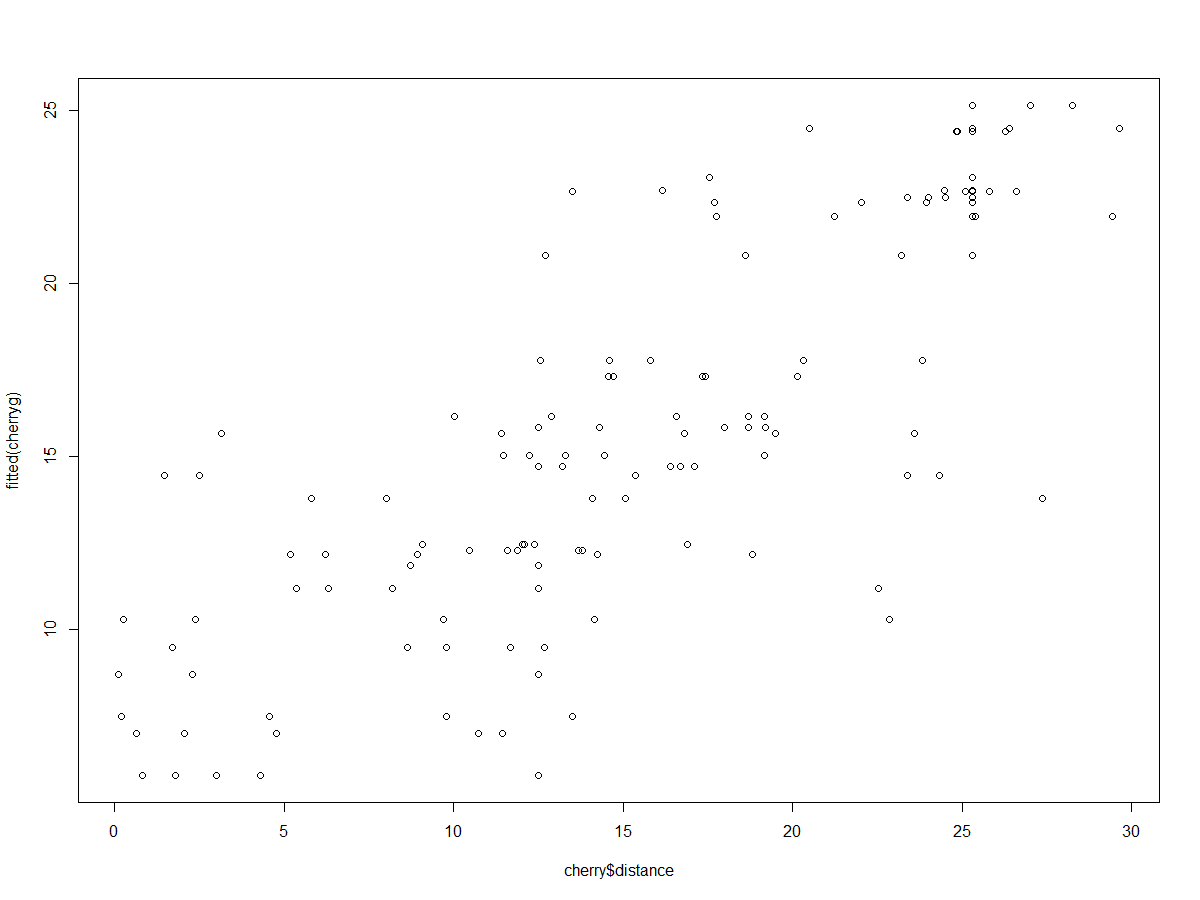
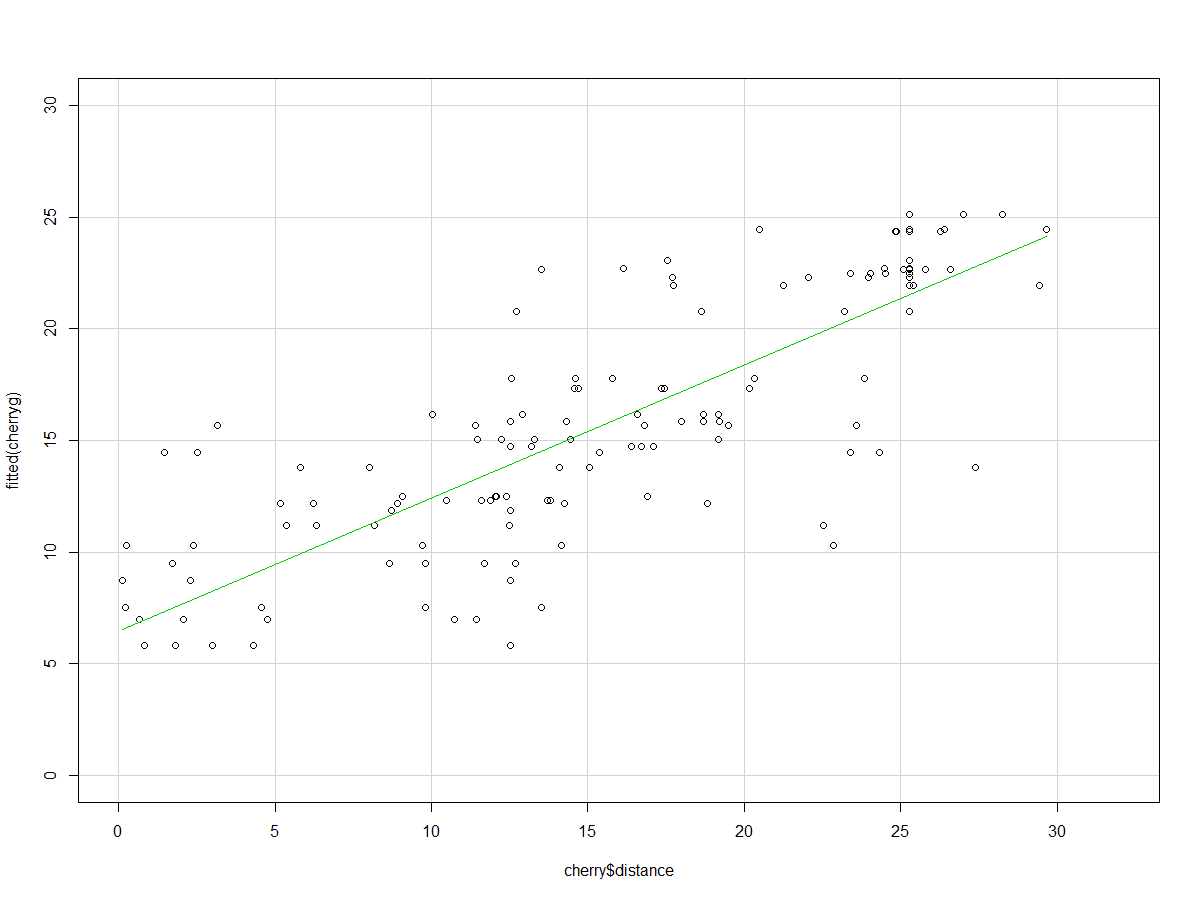
**> stat.desc(cherry$distance)**

**mean**

**15.9808273**

* Was your guess based on the intercept and slope correct or not? If not, why not? **Yes it was correct.**

**(d) Model diagnostics: Create the following plots**

* A densityplot of the scaled model residuals
* 
* A q-q plot of the residuals; you can simply use qqmath(MyModel) or you can use qqPlot() from the package car
* 
* Compute the following numbers: How many standardized residuals (i) > +/- 3, (ii) > +/- 2.5, (iii) > +/- 2 are there? **For i) 0 , for ii) 2, for iii) 6**
* A boxplot showing for each participant the distribution of the scaled residuals (hint: have a look at the slides: you do NOT use the command boxplot for this).
* 
* A scatterplot showing the fitted vs. the residuals to check for homo/heteroskedasticity
* 
* A scatter plot showing the observed vs. the fitted values
* 
* Add a regression line to the previous scatter plot (there are different ways how to do this; I like to use the function scatterplot() from the library car).
* 
* A simple (but not unproblematic) way to compute something like a Pseudo-R2 is to compute the correlation between the fitted and the observed values and then square it to get the R2; thus, compute the correlation and this Pseudo-R2. **Pseudo-R2= 0.6670434**
* What do you conclude in terms of the model based on these graphs and numbers? Is it ok to continue and make inferences (such as computing the CIs in the next step) or are there problems, and if so, what are they? **I don’t see a there are significant problems given the graphs, maybe normality of residuals could become a problem, a further influential points analysis is convinient just to double check.**

**(e) Confidence intervals**

To get the lme4 helpfile, you can type ?lme4::confint.merMod

Compute confidence intervals using all 3 methods available via lme4's confint() command, namely "Wald," "profile," and "boot."

* The Wald method is fast, but considered unreliable and not recommended.
* The profiled Likelihood is often quite fast and seems to have at least some implicit preferred status by the lme4 developers, as it is the default option (i.e., if you don't specify the method explicitly, this profile method will be used). If you add oldNames = FALSE to your command, you get names in the output that are easier to understand.
* The bootstrap method is typically seen as the most robust (i.e., even if conditions like normally distributed residuals etc are not met, you should still get trustworthy CIs), but they typically take longest to compute them. Use at least 500 simulations! Typically, I would use at least 1000 simulations; in particular, this is important if you use more extreme CIs like 99% or 99.9% or 99.99% etc. If you add the statement .progress="txt", PBargs=list(style=3)) to your command, you get a simple progress bar that gives you an idea how much longer it's going to take (at least on a Mac that works nicely). Note: In principle, you have the choice between norm, basic, or perc to compute the bootstrapped CIs; feel free to use just one (whichever you prefer) or more than one of those.

For each of the 3 methods (Wald, profile, boot), start with the 95% CI. If a CI does not include 0 at this 95% level (i.e., is considered significant with p < .05), compute the 99% CI. If it still doesn't include 0, compute the 99.9% CI. That way, you can check the traditional *p* levels of .05, .01, and .001.

#WALD

> confint(cherryg, level = 0.95, method = 'Wald')

2.5 % 97.5 %

.sig01 NA NA

.sigma NA NA

(Intercept) 14.008057 17.953598

gender1 1.648715 5.594255

> confint(cherryg, level = 0.99, method = 'Wald')

0.5 % 99.5 %

.sig01 NA NA

.sigma NA NA

(Intercept) 13.388168 18.573487

gender1 1.028825 6.214145

> confint(cherryg, level = 0.999, method = 'Wald')

0.05 % 99.95 %

.sig01 NA NA

.sigma NA NA

(Intercept) 12.6688002 19.292854

gender1 0.3094578 6.933512

#profile CI

2.5 % 97.5 %

sd\_(Intercept)|pid 3.551964 6.653075

sigma 4.454378 5.739479

(Intercept) 14.012276 17.949379

gender1 1.652934 5.590036

> confint(cherryprofci, level = 0.99)

0.5 % 99.5 %

sd\_(Intercept)|pid 3.2165446 7.408446

sigma 4.2962678 5.997255

(Intercept) 13.3310159 18.630639

gender1 0.9716735 6.271297

> confint(cherryprofci, level = 0.999)

0.05 % 99.95 %

sd\_(Intercept)|pid 2.8494040 8.439060

sigma 4.1243394 6.320787

(Intercept) 12.4699030 19.491750

gender1 0.1105606 7.132408

#bootstrap 1000 at 95%

2.5 % 97.5 %

.sig01 3.383625 6.700812

.sigma 4.373719 5.663667

(Intercept) 13.975660 17.954772

gender1 1.747485 5.696129

#bootstrap 1000 at 99%

0.5 % 99.5 %

.sig01 3.069967 7.249435

.sigma 4.153110 5.797279

(Intercept) 13.448466 18.451138

gender1 1.135488 6.117477

#bootstrap 10000 at 99.9%

0.05 % 99.95 %

.sig01 2.4291369 7.870026

.sigma 4.0273645 6.115391

(Intercept) 12.6711250 19.378680

gender1 0.3061441 6.916216

**(2) Analyze the Feather Propelling Contest Data.**

**The file is called** FeatherContest\_13Feb2016.csv **and can be found on BlackBoard --> Week 2 --> Homework**

Here are some explanations about the data file:

The data come from a similar contest to the Cherry Pit Spitting Contest:

Each of 100 participants got 3 feathers and attempted to propel them as far from the start line as possible.

Explanations of the variables in the data frame:

* pid: participant ID (i.e., participant code)
* trial: indexes the 3 attempts per participant: f1 (first feather the participant propelled, i.e., first trial); f2 (second trial), f3 (third trial)
* distance: how far the feather flew in meters --> note, there are negative values: the competition was held outside and it wasn't necessarily completely windless...
* smoking\_status = whether the participant is a smoker or a nonsmoker

Your task is to answer the following research question: Do smokers and non-smokers differ in how far they can propel a feather in this contest?

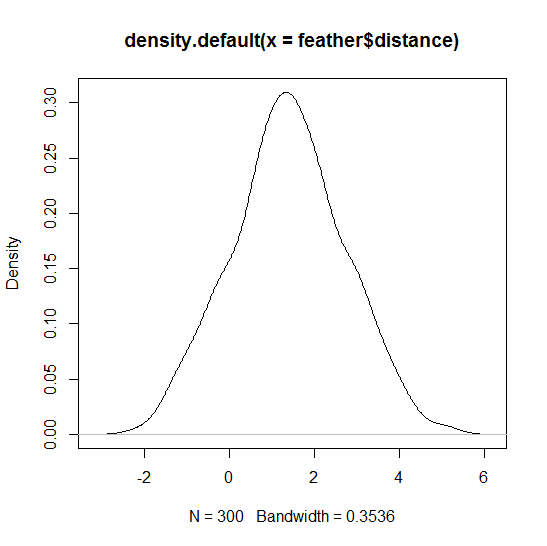
*Make sure you have set your contrast setting to sum-to-zero contrasts.*

Thus, do the following things in the following order:

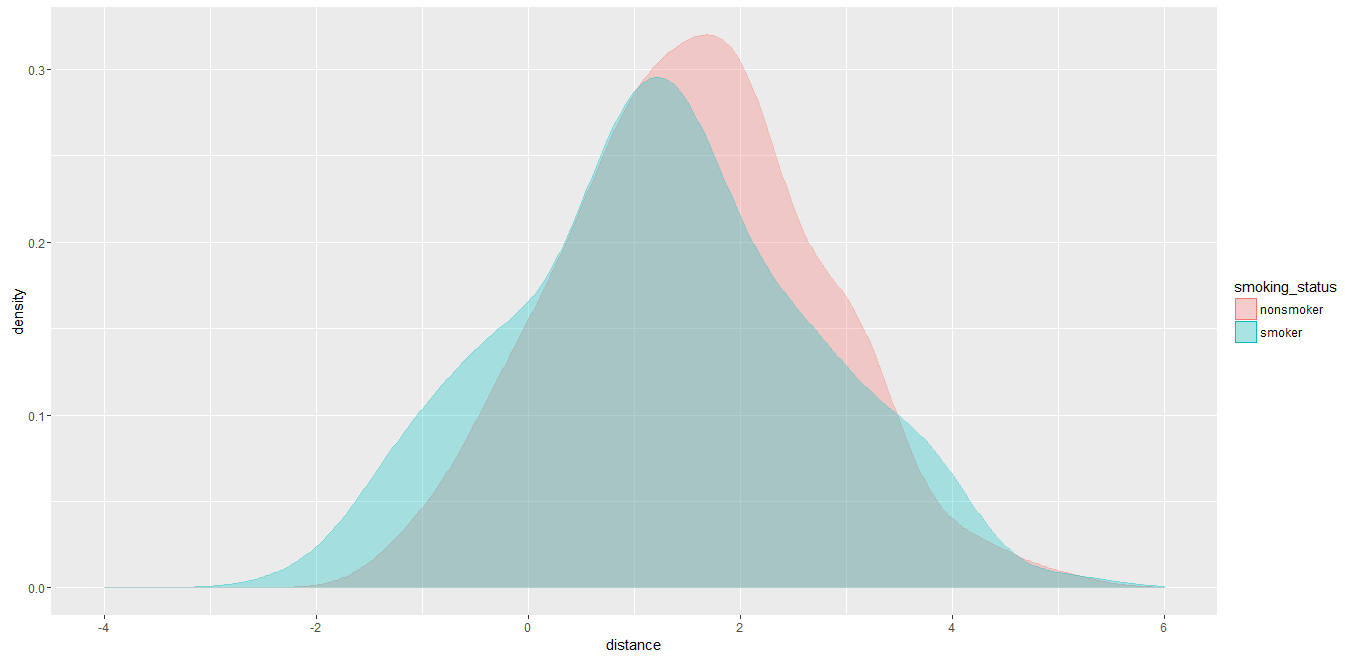
**(a) Write down the lme4 model syntax you plan to use for the model. Do not run it yet.**

**(b) Understand your data by generating the following plots:**

* A densityplot showing the overall distribution of your dependent variable

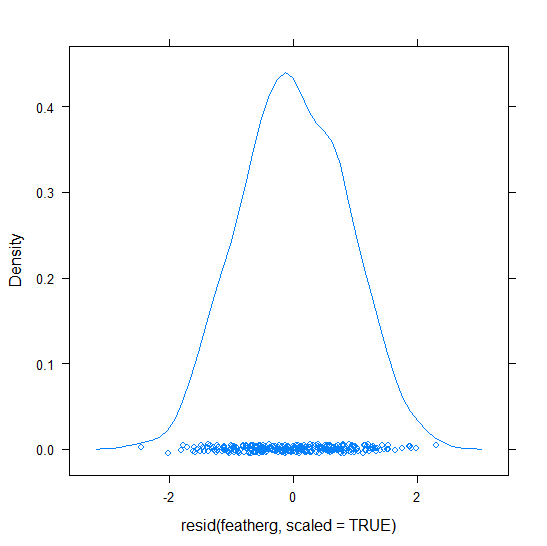
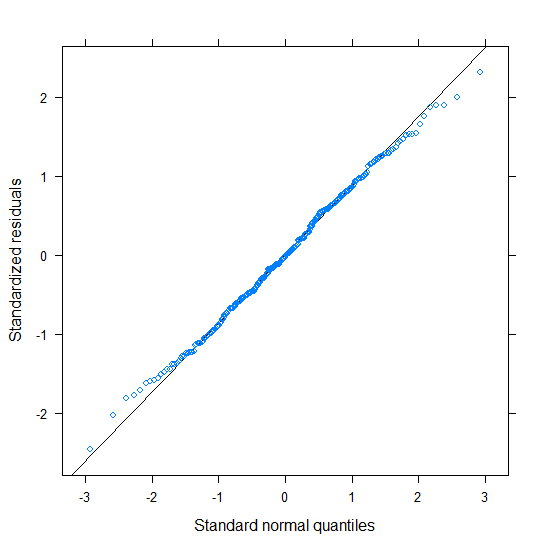
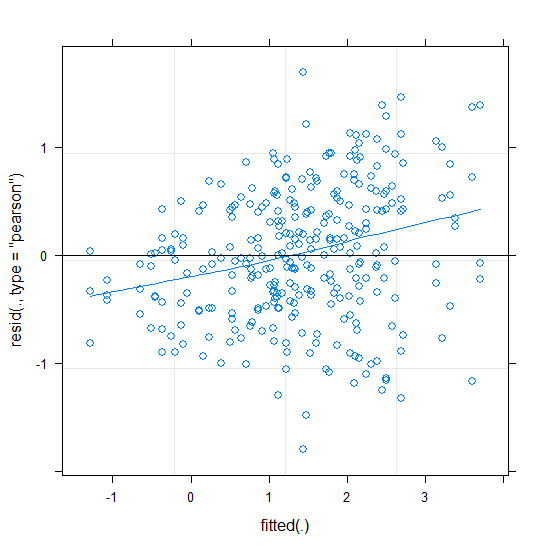
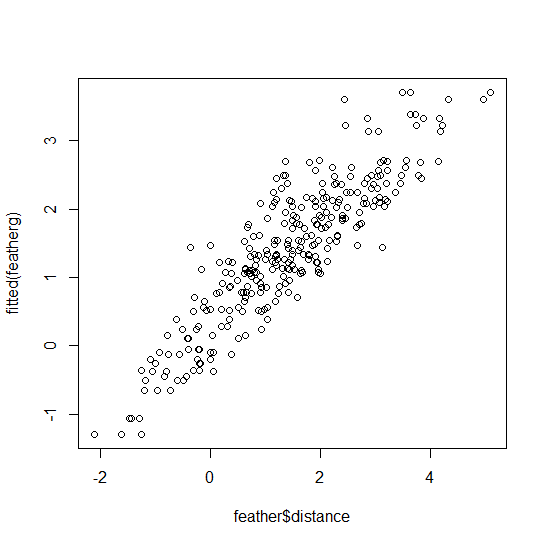
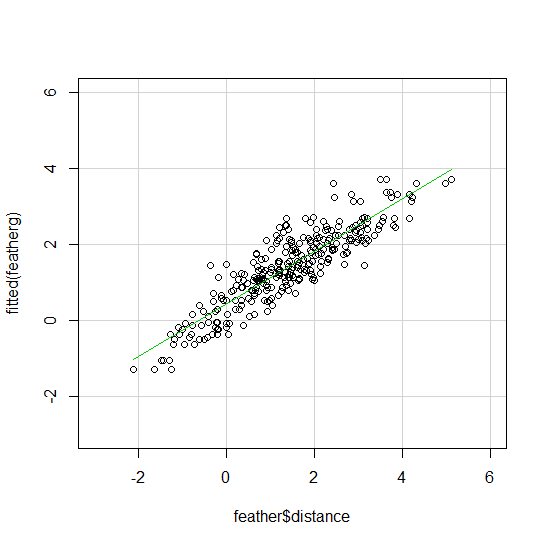


* A multi-panel densityplot showing the smokers and the nonsmokers in different panels (or as two different curves in the same panel; either is fine)



**(c) Run your model**

**(d) Do model diagnostics and check the summary output**

* Create the following diagnostic plots and comment whether things look ok or problematic:
  + A densityplot of the scaled model residuals
  + 
  + A q-q plot of the residuals; you can simply use qqmath(MyModel)
  + 
  + A scatterplot showing the fitted vs. the residuals to check for homo/heteroskedasticity
  + 
  + A scatter plot showing the observed vs. the fitted values
  + 
  + Add a regression line to the previous scatter plot (there are different ways how to do this; I like to use the function scatterplot() from the library car).
  + 
  + Compute the following numbers: How many standardized residuals (i) > +/- 3, (ii) > +/- 2.5, (iii) > +/- 2 are there? **In order: 0, 0 , 3**
  + What do you conclude in terms of the model based on these graphs and numbers? Is it ok to continue and make inferences (such as computing the CIs in the next step) or are there problems, and if so, what are they. **The diagnostics seem OK except from the heteroskedasticity, which appears to be funneled with this shape “<” indicating a possible violation of the assumption.**
* What is the variance associated with the random participant-intercept? **1.2119**
* What's the residual variance? **.5352**
* What is the estimate, standard error, and t value for the overall (i.e., fixed) intercept? What does that number mean? **the estimate is 1.3945, this means that the overall distance of blowing a feather is 1.39**
* What is the estimate, standard error, and t value for the smoking status effect? What does that number mean? **the estimate = .1433, SE = .1179, t = 1.215, non-smokers blow the feather .1433 distance more than the average of 1.39, and smokers blow the feather .1433 less than the average of 1.39**
* So, based on these estimates (intercept and smoking\_status), what do you think is the average distance for smokers and the average distance for nonsmokers? **1.5378 for non-smokers, and 1.2512 for smokers.**
* Compare this by using describeBy from the library psych to get the raw means for smokers and nonsmokers.
* **> by(feather$distance, feather$smoking\_status, mean)**
* **feather$smoking\_status: nonsmoker**
* **[1] 1.53777**
* **------------------------------------------------------------**
* **feather$smoking\_status: smoker**
* **[1] 1.251166**
* Was your guess based on the intercept and slope correct or not? If not, why not? **Yes they have the same values.**

**(e) Confidence intervals**

* Compute and report the 90%, 95%, and 99% confidence intervals for the intercept and the smoking\_status effect using the "profile" and "boot" methods (use the tricks I showed in class, so you don't have to compute the 90%, 95%, and 99% CIs separately; for the boot method, you can again use all 3 boot "types" (norm, basic, perc) or just one or two of them).

**Profile**

**> confint(featherprofci, level = 0.95)**

**2.5 % 97.5 %**

**sd\_(Intercept)|pid 0.92982791 1.2824279**

**sigma 0.66532302 0.8095500**

**(Intercept) 1.16347679 1.6254592**

**smoking\_status1 -0.08768961 0.3742928**

**> confint(featherprofci, level = 0.99)**

**0.5 % 99.5 %**

**sd\_(Intercept)|pid 0.8857330 1.3530605**

**sigma 0.6465709 0.8369299**

**(Intercept) 1.0887495 1.7001865**

**smoking\_status1 -0.1624169 0.4490201**

**> confint(featherprofci, level = 0.999)**

**0.05 % 99.95 %**

**sd\_(Intercept)|pid 0.8372813 1.4419402**

**sigma 0.6259014 0.8706508**

**(Intercept) 0.9997582 1.7891778**

**smoking\_status1 -0.2514082 0.5380114**

**bootstrap**

**> featherboot95**

**BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS**

**Based on 10000 bootstrap replicates**

**Intervals :**

**Level Normal Basic Percentile**

**95% (-0.0916, 0.3719 ) (-0.0921, 0.3697 ) (-0.0831, 0.3787 )**

**Calculations and Intervals on Original Scale**

**> featherboot99**

**BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS**

**Based on 10000 bootstrap replicates**

**Intervals :**

**Level Normal Basic Percentile**

**99% (-0.1645, 0.4448 ) (-0.1727, 0.4356 ) (-0.1490, 0.4593 )**

**Calculations and Intervals on Original Scale**

**> featherboot999**

**BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS**

**Based on 10000 bootstrap replicates**

**Intervals :**

**Level Normal Basic Percentile**

**99.9% (-0.2490, 0.5293 ) (-0.2648, 0.5092 ) (-0.2226, 0.5514 )**

**Calculations and Intervals on Original Scale**

**Some basic intervals may be unstable**

**Some percentile intervals may be unstable**

**(f) Smokers vs. non-smokers**

Based on the CIs you computed in (e), what do you conclude with respect to your research question?

**That the difference between smokers and non-smokers regarding blowing a feather should be true with a 99% chance, given the sample conditions, of being in between the ranges calculated by a bootstrap method with three types: normal, basic and percentile respectively (-0.1645, 0.4448 ), (-0.1727, 0.4356 ) and (-0.1490, 0.4593 ) and by a profile method (-0.1624169 , 0.4490201).**

**(g) Feathers vs. Cherry Pits**

Based on the CIs you computed in (e) and based on the CIs you computed for the Cherry Pit Spit data further up, do you think you can or cannot make any conclusions about whether the achieved average distances in the feather versus the cherry pit contests differ? If you think you can make conclusions, what are they?

Wald 1.135488 6.117477

profile 0.9716735 6.271297

bootstrap 1.028825 6.214145

**females-males 3.621**

1.13 < Pa-Pb <6.11

.9716 < Pa-Pb <6.27

1.02 < Pa-Pb <6.21

**(-0.1645, 0.4448 ),**

**(-0.1727, 0.4356 )**

**(-0.1490, 0.4593 )**

**(-0.1624169 , 0.4490201)**

**Non’smokers’smokers = .1433**

-.16 < Pa-Pb <.44

-.17 < Pa-Pb <.44

-.15 < Pa-Pb <.46

Here the null hypothesis is if pa-pb =0, because the ranges include that value, then we can asume that this difference is significant given that if it where we are implying that 0 is a reasonable possibility for the true value of the difference.