

R instructions for the 9th seminar

R Instructions for the problem 1:

- Get familiar with data (means and variances in groups,useful graphs). What the graphs suggest about interaction? Which factor looks to be significant? Do variances in particular groups look similar? Do the means in particular groups look similar? Why is it important?

```
tapply(Crop$Yield,list(Crop$Irrigation,Crop$Fertilizer),mean)
tapply(Crop$Yield,list(Crop$Irrigation,Crop$Fertilizer),var)

!library(lattice)
xyplot(Yield~Irrigation:Fertilizer,data=Crop)
interaction.plot(Crop$Irrigation,Crop$Fertilizer,Crop$Yield)
interaction.plot(Crop$Fertilizer,Crop$Irrigation,Crop$Yield)

!library(sciplot) whiskers represent standard errors for means
lineplot.CI(Fertilizer,Yield,group=Irrigation,data=Crop)
lineplot.CI(Irrigation,Yield,group=Fertilizer,data=Crop)
lineplot.CI(Irrigation,Yield,data=Crop)
lineplot.CI(Fertilizer,Yield,data=Crop)
```

R Instructions for the problem 2:

- Perform a Shapiro-Wilks normality test for each group.

```
tapply(Crop$Yield,list(Crop$Irrigation,Crop$Fertilizer),shapiro.test)[1:6]
```

R Instructions for the problem 3:

- Perform the Levene's Test for Homogeneity of Variances.

```
!library(DescTools)
LeveneTest(Yield~Irrigation*Fertilizer, data=Crop,center=mean)
```

R Instructions for the problem 4:

- Create an object in R bearing all essential results of two-way ANOVA.

```
AResults1<-lm(Yield~Irrigation+Fertilizer+Irrigation:Fertilizer,Crop)
```

(It is vital to remember which level of Irrigation and Fertilizer is set to be reference level. In R it is the first one. Check it by

```
levels(Crop$Irrigation), the first in order is the reference one.
levels(Crop$Fertilizer)
```

R Instructions for the problem 5:

- Create an ANOVA table. Which factors/interactions are significant?

```
anova(AResults1)
```

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Irrigation	1	4.3	4.27	0.3582	0.552
Fertilizer	2	4994.1	2497.07	209.6418	< 2.2e-16 ***
Irrigation:Fertilizer	2	810.1	405.07	34.0075	2.764e-10 ***
Residuals	54	643.2	11.91		

(Irrigation is insignificant, however interaction of Irrigation with Fertilizer as well as Fertilizer itself are significant. See the p-values in a last column.)

R Instructions for the problem 6:

- Predict values in particular groups.

```
tapply(predict(AResults1),list(Crop$Irrigator,Crop$Fertilizer),mean)
```

```
      littleF mediumF plentyF
littleI    17.8    30.2    46.4
plentyI    21.4    37.8    36.8
```

(Remark: results are the same like descriptive means. It is because we had balanced data.)

R Instructions for the problem 7:

- Create a linear model with *treatment* parametrization and interpret coefficients. (Tricky - for volunteers!) Symbols of α and β have different meaning then in lecture!!!!

```
summary(AResults1)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.800	1.091	16.310	< 2e-16 ***
IrrigationplentyI	3.600	1.543	2.332	0.0234 *
FertilizermediumF	12.400	1.543	8.034	8.70e-11 ***
FertilizerplentyF	28.600	1.543	18.530	< 2e-16 ***
IrrigationplentyI:FertilizermediumF	4.000	2.183	1.833	0.0724 .
IrrigationplentyI:FertilizerplentyF	-13.200	2.183	-6.047	1.43e-07 ***

	α	$+$	$\beta_1 \cdot I_{pl}$	$+$	$\beta_2 \cdot F_{med}$	$+$	$\beta_3 \cdot F_{pl}$	$+$	$\beta_4 \cdot I_{pl} \cdot F_{med}$	$+$	$\beta_5 \cdot I_{pl} \cdot F_{pl}$
I:little F:little	α	$+$	$\beta_1 \cdot 0$	$+$	$\beta_2 \cdot 0$	$+$	$\beta_3 \cdot 0$	$+$	$\beta_4 \cdot 0$	$+$	$\beta_5 \cdot 0$
I:little F:medium	α	$+$	$\beta_1 \cdot 0$	$+$	$\beta_2 \cdot 1$	$+$	$\beta_3 \cdot 0$	$+$	$\beta_4 \cdot 0$	$+$	$\beta_5 \cdot 0$
I:little F:plenty	α	$+$	$\beta_1 \cdot 0$	$+$	$\beta_2 \cdot 0$	$+$	$\beta_3 \cdot 1$	$+$	$\beta_4 \cdot 0$	$+$	$\beta_5 \cdot 0$
I:plenty F:little	α	$+$	$\beta_1 \cdot 1$	$+$	$\beta_2 \cdot 0$	$+$	$\beta_3 \cdot 0$	$+$	$\beta_4 \cdot 0$	$+$	$\beta_5 \cdot 0$
I:plenty F:medium	α	$+$	$\beta_1 \cdot 1$	$+$	$\beta_2 \cdot 1$	$+$	$\beta_3 \cdot 0$	$+$	$\beta_4 \cdot 1$	$+$	$\beta_5 \cdot 0$
I:plenty F:plenty	α	$+$	$\beta_1 \cdot 1$	$+$	$\beta_2 \cdot 0$	$+$	$\beta_3 \cdot 1$	$+$	$\beta_4 \cdot 0$	$+$	$\beta_5 \cdot 1$
I:little F:little	α										
I:little F:medium	$\alpha + \beta_2$										
I:little F:plenty	$\alpha + \beta_3$										
I:plenty F:little	$\alpha + \beta_1$										
I:plenty F:medium	$\alpha + \beta_1 + \beta_2 + \beta_4$										
I:plenty F:plenty	$\alpha + \beta_1 + \beta_3 + \beta_5$										

R Instructions for the problem 8:

- Asses whether assumptions of homogeneity of variances and normality of residuals are met.

```
par(mfrow=(c(1:2)))
plot(AResults1,which=1:2)
```

The first graph depicts dependence of residuals on predicted values (notice that residuals look to have homogeneous variance); the second suggests small problems with normality of residuals.

```
hist(residuals(AResults1))
```

R Instructions for the problem 9:

- Find out the confidence intervals for population means in particular groups. To process

that, first we have to perform two-way ANOVA in *textbook parametrization*. The easiest way to get this parametrization is to create new factor with 6 levels corresponding to 6 combinations of previous two factors. This will lead to "one way" ANOVA form without intercept. The new factor denote **both**

```
both<-paste(Crop$Irrigation,Crop$Fertilizer)
both<-factor(both)
Crop$both<-both
AResults2<-lm(Yield~both-1,Crop) "-1" is a symbol that the model is without intercept
anova(AResults2)
summary(AResults2)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
bothlittleI littleF	17.800	1.091	16.31	<2e-16 ***
bothlittleI mediumF	30.200	1.091	27.67	<2e-16 ***
bothlittleI plentyF	46.400	1.091	42.52	<2e-16 ***
bothplentyI littleF	21.400	1.091	19.61	<2e-16 ***
bothplentyI mediumF	37.800	1.091	34.63	<2e-16 ***
bothplentyI plentyF	36.800	1.091	33.72	<2e-16 ***

In the column "Estimate" there are estimated population means in particular group and the p-value in a last column has clear interpretation: Is the population mean significantly different from zero? The 95% confidence interval for any population mean can be obtained by

```
confint(AResults2)
```

	2.5 %	97.5 %
bothlittleI littleF	15.61191	19.98809
bothlittleI mediumF	28.01191	32.38809
bothlittleI plentyF	44.21191	48.58809
bothplentyI littleF	19.21191	23.58809
bothplentyI mediumF	35.61191	39.98809
bothplentyI plentyF	34.61191	38.98809

Homework:

File *Fuel_Consumption.RData* contains data about the effects of two types of gas (factor A) and three different additives (factor B) on overall fuel consumption. The experiment is carried in a balanced design (equal number of observations for all factor combinations) and is repeated twice for each combination of factors. Consider a model with interaction and repeat the same exercises as before.