small,scriptsize,tiny,p0.5linewidth

Question 1 Linear Regression

Import Data

```
table_2015f1 <- read_xlsx("qe_lab/Profits_2015f.xlsx")

table_2016s1 <- read_xlsx("qe_lab/RegressionSpr16.xlsx")[-1,]
table_2016s1$weight <- round(as.numeric(table_2016s1$weight), 2)
table_2016s1$age <- as.factor(table_2016s1$age)
table_2016s1$height <- round(as.numeric(table_2016s1$height), 2)
table_2016s1$male <- factor(table_2016s1$male, labels=c("female","male"))

table_2016f1 <- read_xlsx("qe_lab/cigcons.xlsx")
table_2016f1$State <- as.factor(table_2016f1$State)

# select a part
table_2018s1 <- read_xlsx("qe_lab/Problem1_ChildSmoking.xlsx")
table_2018s1_u6 <- table_2018s1[which(table_2018s1$age>5),]
table_2018s1_u6$age <- factor(table_2018s1_u6$age)
table_2018s1_u6$male <- factor(table_2018s1_u6$male, labels = c("female","male"))
table_2018s1_u6$smoker <- factor(table_2018s1_u6$smoker, labels = c("non-smoker","smoker"))
str(table_2018s1_u6)
summary(table_2018s1$height)

table_2018f1 <- read_xlsx("qe_lab/profits_2018f.xlsx")

# split to 2 parts
table_2019s1 <- read_xlsx("qe_lab/ModelBuildingData.xlsx")
dplyr::glimpse(table_2019s1)
table_2019s1_250 <- table_2019s1[1:250,]
table_2019s1_500 <- table_2019s1[251:500,]
str(table_2019s1_250)
str(table_2019s1_500)
```

Plot Data

```
ggplot(table_2015f1,aes(X2,Y, color=X1))+
  labs(x="adv",y="prof",color="month")+
  geom_point()+theme_light()

ggplot(table_2018s1_u6, aes(smoker,fill=male))+
  geom_bar()+facet_wrap(.~age,ncol = 7)+theme_light()

ggline(table_2016s1,"height","weight",add=c("mean","jitter"),color="age
ggline(table_2016s1,"height","weight",add=c("mean","jitter"),color="male")

ggpairs(table_2016f1[,-1]) # [c(1:10)]

ggpairs(table_2018s1_u6,mapping=aes(color =male,shape=male,alpha=0.3),
        columns=c("volume","height","smoker"))+theme_light()
```

Based on scatterplots and correlation, $Cor(y, X_1) = 0.866$. X1 have medium to strong positive linear relationship to the response variable (Correlation coefficient is more than 0.6). X2 has medium negative

Build Model

```
model_2015f1_1 <- lm(Y~X1*X2,table_2015f1)
```

$y = \beta_0 + \beta_1 X_1 + +\beta_2 X_2$

Regression Analysis

```
summary(model_2015f1_1)
ols_regress(model_2015f1_1)
```

$\hat{y} = 292.561 - 203.144 X_1 + 1055.782 X_2$

The fitted overall model is statistically significant at 5% significance level (p-value=$9.744 \times^{-06}$).
But most of the coefficients are not significent.
Coefficient of 511.713 in the full model suggests the average peak rate of flow increases by **511.713** cubic feet per second when the rainfall increases by 1 inch and other variables are constants.

ANOVA

```
anova(model_2015f1_1)
Anova(model_2015f1_1)
```

Accroding to the F test, the partial sum of squares explained by rainfall is **2209**, given that all the other regression coefficients are in the model.

elimination regression

```
vif(model_2016f1_1)
ols_vif_tol(model_2015f1_1)
```

According to the result of VIF test (variance inflation factor), the model does have **problems of multicollinearity**. The VIF of variables **X4 (105)** are larger than 10.

```
ols_step_both_aic(model_2015f1_1)
ols_step_both_p(model_2015f1_1)
ols_step_best_subset(model_2015f1_1)
```

Comparison

```
model_2015f1_2 <- lm(Y^2~X1*X2,table_2015f1)
table_2015f1_logy <- table_wf %>% mutate(logy=log(Y))
table_2015f1_logy$Y <- NULL
huxreg(model_2015f1_1, model_2015f1_2)
# Redo analysis
```

comparing to the old model, the new model has a higher (about by 6%) adjusted R square and higher (about by 5%) prediction R-square, which means it shows stronger predictive capability. All the coeficients in new model are statistically significant higher than 98% significance level (the maximum p-values are 0.019, respectively). PRESS statistic of the best model.

The value of PRESS is **6.53**. This model explains **90.8%** of variation in predicting the .

Check Adequacy

```
plot(model_2015f1_1)
```

Residual Diagnostics show some violations. The model didn't satisfied the OLS assumptions of random errors.
On the residual plot, there is a **funnel pattern**.
On the outlier and leverage plot, there are **two outliers**.
On the qq plot, most of points follow approximately straight line but have some **positive skew**.
Suggestion: using **natural log of response** to make a variance-stabilizing transformations.

Estmation

```
sqrt(predict(model_2015f1_2, newdata=data.frame(X1=20,X2=1500),
             interval="prediction", level=0.95 )) # "confidence"
confint(model_2015f1_2, level=0.05/2) # 1-(0.05/2)
```

partial sum of squares, estimated coefficients, standard errors, p-values, 95% Bonferroni joint confidence intervals for the coefficients of the best model.

```
model_2019s1_2 <- lm(table_2019s1_250,formula=log(y)~ x2+A+B)
model_2019s1_3 <- lm(table_2019s1_500,formula=log(y)~ x2+A+B)
rmse(table_2019s1_500$y,exp(predict(model_2019s1_2,table_2019s1_500)))
ols_press(model_2019s1_3)
# MPV::PRESS(model_2019s1_3)
sum((residuals(model_2019s1_3)/(1 - lm.influence(model_2019s1_3)$hat))^
ols_pred_rsq(model_2019s1_3)
# str(model_2019s1_3)
# From 564-lab caculate prediction power
deviation <- table_2019s1_500$y-mean(table_2019s1_500$y)
SST <- deviation%*%deviation
1-(MPV::PRESS(model_2019s1_3)/SST)
# by definition PRESS
sum((table_2019s1_500$y-exp(model_2019s1_2$fit))^2)
sum((table_2019s1_500$y-exp(predict(model_2019s1_2,table_2019s1_500)))^
# one method of RMSE
sqrt(mean(model_2019s1_3$residuals^2))

# remove outlier
table_2019s1_250[c(189,219,249),]
table_2019s1_250_noout <- table_2019s1_250[-c(189,219,249), ]
model_2019s1_noout <- lm(y ~ sqrt(!is.na(x1))+x2+x3+A+B, table_2019s1_2
,
```

Conclusion

Question 2 Factorial Design

Import Data

```
# gather colums
table_2015f2 <- read_xlsx("qe_lab/Springs_2015f.xlsx")
table_2015f2 <- gather(table_2015f2,'Height','...7','...8',
                       key = "1",value = "height" )[,-6]
str(table_2015f2)

# gather
DesignSpr16 <- read_excel("qe_lab/DesignSpr16.xlsx")
table_2016s2 <- gather(DesignSpr16[c(2:4,6:8),c(2:4,6:8,10:12)])
names(table_2016s2) <- c("machine","y")
table_2016s2 <- table_2016s2[c("y","machine")]
table_2016s2$machine <- as.factor(c(rep("machine1",18),
                                    rep("machine2",18),rep("machine3",18)))
table_2016s2$station <- as.factor(rep(c(rep("station1",6),
                                    rep("station2",6),rep("station3",6)),3))
table_2016s2$power<-as.factor(rep(c(rep("power1",3),rep("power2",3)),9))
str(table_2016s2)

# One-stage nested design
creek1 <- c(5.2, 5.4, 5.6, 5.7, 5.4, 5.4, 5.6, 5.5, 5.8, 5.5)
creek2 <- c(5.1, 5.3, 5.1, 5.0, 5.3, 5.2, 5.0, 5.0, 4.9, 5.1)
creek3 <- c(5.9, 5.8, 5.8, 5.8, 5.7, 5.8, 5.8, 5.9, 5.9, 5.9)
table_2016f2 <- gather(data.frame(creek1,creek2,creek3),creek,oxygen)
table_2016f2$creek <- as.factor(table_2016f2$creek)
table_2016f2$sample <- as.factor(c(rep("sample1",2),rep("sample2",2)
                ,rep("sample3",2),rep("sample4",2),rep("sample5",2)))
table_2016f2$rep <- as.factor(rep(c("rep1","rep2"),15))
str(table_2016f2)
## 'data.frame':    30 obs. of  4 variables:
##  $ creek : Factor w/ 3 levels "creek1","creek2",..: 1 1 1 1 1 1 1 1
##  $ oxygen: num  5.2 5.4 5.6 5.7 5.4 5.4 5.6 5.5 5.8 5.5 ...
##  $ sample: Factor w/ 5 levels "sample1","sample2",..: 1 1 2 2 3 3 4
##  $ rep   : Factor w/ 2 levels "rep1","rep2": 1 2 1 2 1 2 1 2 1 2 ...
table_2017sr1 <- read_xlsx("qe_lab/Profits_2017s.xlsx")

# BIBD
table_2017sd1 <- read_xlsx("qe_lab/NBalance.xlsx")
table_2017sd1$Block <- factor(table_2017sd1$Block,
labels=c("Blk1","Blk2","Blk3","Blk4","Blk5","Blk6","Blk7","Blk8","Blk9"
table_2017sd1$Animal <- factor(table_2017sd1$Animal,
                               labels = c("Ani1","Ani2","Ani3"))
table_2017sd1$Ration <- factor(table_2017sd1$Ration,
labels=c("Rat1","Rat2","Rat3","Rat4","Rat5","Rat6","Rat7","Rat8","Rat9"
str(table_2017sd1)

table_2018s2 <- read_xlsx("qe_lab/Problem2_Avocado.xlsx")
table_2018s2$Blk <- factor(table_2018s2$Block,
                           labels=c("Blk1","Blk2","Blk3"))
table_2018s2$Ship <- factor(table_2018s2$Shipping,
                            labels=c("Ship1","Ship2","Ship3"))
table_2018s2$Stor <- factor(table_2018s2$Storage,
                            labels=c("Stor1","Stor2"))
str(table_2018s2)

# Order
table_2018f2 <- read_xlsx("qe_lab/Springs_2018f.xlsx")
```

```
table_2018f2 <- table_2018f2[order(table_2018f2$D,
                  table_2018f2$C ,table_2018f2$B,table_2018f2$A),]
str(table_2018f2)
```

```
# [split-plot] [2019S2]
table_2019s2 <- read_xlsx("qe_lab/WoolShrink.xlsx")
str(table_2019s2)
glimpse(table_2019s2)
table_2019s2$Run <- factor(table_2019s2$Run,
                  labels=c("Day1","Day2","Day3","Day4"))
table_2019s2$Trt <- factor(table_2019s2$Trt,
                  labels=c("Untrt","15Sec","4Min","15Min"))
table_2019s2$Rev <- as.factor(table_2019s2$Rev)
str(table_2019s2)
glimpse(table_2019s2)
```

Plot Data

```
ggline(table_2018s2,"Stor","Y",add=c("mean","jitter"),color="Ship",
  shape="Ship",linetype="Ship",ylab="acceptability",facet.by="Blk")
```

The above plots show that:
Not all the lines are parallel in the interaction plot. Therefore, in the model, there is the interaction effect of source level and technicians nested in the lab.
There is not much difference in the average shrink from different days. The average shrink are lower when the treatment is longer. The average shrink are higher when the revolutions are faster.

```
kable(favstats(Y ~ Ship, data=table_2018s2),format="latex",booktabs=T)
      %>%kable_styling("striped", full_width = F,font_size = 8)
      %>%column_spec(7,background ="#EAFAF1")
kable(favstats(Y ~ Stor, data=table_2018s2),format="latex",booktabs=T)
      %>%kable_styling("striped", full_width = F,font_size = 8)
      %>%column_spec(7,background ="#EAFAF1")
kable(favstats(Y ~ Blk, data=table_2018s2),format="latex",booktabs=T)
      %>%kable_styling("striped", full_width = F,font_size = 8)
      %>%column_spec(7,background ="#EAFAF1")
```

The Tables show the same thing with the numerical summaries for each factor level and their combinations.

Build Model

```
# 5k1p Fractional Factorial Design
model_2015f2 <- aov(height~A*B*C*D, table_2015f2)
model_2015f2$coefficients
```

This is a neseted and crossed design. Three fixed sources apply on all random selected technicians nested in fixed labs.
$y = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{k(ij)}, i = 1,2,3; j = 1,2,3,4,5; k = 1,2$

```
# Br in Af * Cf [2016S2] [2017F2]
model_2016s2_1 <- lm(y~machine*station*power, table_2016s2)
```

$y = \mu + \beta_1 \ln(H) * Age * Male + \beta_2 Smoker + \varepsilon$

```
model_2018s1_2<-lm(log(volume)~log(height):age:male+smoker,table_2018s1_2)
```

$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{j(i)k} + \varepsilon_{(ijk)l}$
for $i = 1,2$; $j = 1,2,3$; $k = 1,2,3$; $l = 1,2,3$;
μ is the overall true mean response;
τ_i is the fixed main effect of i^{th} level of labs;
$\beta_{j(i)}$ is the random effect of j^{th} level of technicians nested in i^{th} level of labs;
γ_k is the main fixed effect of k^{th} level of sources;
$(\tau\gamma)_{ik}$ is the interaction effect of i^{th} level of labs and k^{th} level of sources;
$(\beta\gamma)_{j(i)k}$ is the interaction random effect of k^{th} level of sources and j^{th} level of technicians nested in i^{th} level of labs.
y_{ijkl} is response value for the l^{th} replication for j^{th} level of technicians nested in i^{th} level of labs when k^{th} level of sources is applied;
$\varepsilon_{(ijk)l}$ is random error for the l^{th} replication for j^{th} level of technicians nested in i^{th} level of labs when k^{th} level of sources is applied.
Assumptions: Usually, the technicians in a lab are skillful. From the above plots, the technicians' performances are stable. The covariance between two observations from the same level of the random factor can be either positive or negative. Thus we assume this is an **restricted model**. $\varepsilon_{(ijk)l}$, $\beta_{j(i)}$, and $(\beta\gamma)_{j(i)k}$ are independent.

$$\varepsilon_{(ijk)l} \sim iidN(0,\sigma^2) \quad \left|\sum_{i=1}^{2}\tau_i = 0 \quad \right| \sum_{k=1}^{3}\gamma_k = 0 \quad \left|\beta_{j(i)} \sim iidN(0,\sigma^2_\beta)\right.$$
$$\sum_{i=1}^{2}(\tau\gamma)_{ik} = 0 \quad \left|\sum_{k=1}^{3}(\tau\gamma)_{ik} = 0\right| \sum_{i=1}^{2}(\beta\gamma)_{j(i)k} = 0 \left|(\beta\gamma)_{j(i)k} \sim iidN(0,\tfrac{2-1}{2}\sigma^2_{\beta\gamma})\right.$$

This is a simple Split-Plot design model (fat is whole-plot factor and temperature is split-plot factor)

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijk}$$

for $i = 1,2,3,4$; $j = 1,2,3$; $k = 1,2,3,4$
μ is the overall true mean response;
τ_i is the effect of i^{th} replication of days;
β_j is the main effect of j^{th} level of temperature (effect of split-plot factor);
$(\tau\beta)_{ij}$ is the interaction effect of i^{th} replication and j^{th} level of temperature;
γ_k is the main effect of k^{th} level of fat (effect of whole-plot factor);
$(\tau\gamma)_{ik}$ is the interaction effect of i^{th} replicatin and k^{th} level of fat(whole-plot error);
$(\beta\gamma)_{jk}$ is the interaction effect of j^{th} level of temperature and k^{th} level of fat;
$(\tau\beta\gamma)_{ijk}$ is the interaction effect of i^{th} replicatin, j^{th} level of temperature and k^{th} level of fat (sub-plot error);
y_{ijk} is response value for the i^{th} replication when j^{th} level of temperature and k^{th} level of fat are applied;
ε_{ijk} is random error for the i^{th} replication when j^{th} level of temperature and k^{th} level of fat are applied.
Assumptions: For an experienced baker, he/she will try to let the recipe and temperature are accurate in each day. the covariance between two observations from the same level of the random factor can be either positive or negative. Thus, we assume this is a **restricted model**.

$$\varepsilon_{ijk} \sim iidN(0,\sigma^2) \quad \left|\tau_i \sim iidN(0,\sigma^2_\tau)\right.$$
$$\sum_{j=1}^{3}\beta_j = 0 \quad \left|\sum_{j=1}^{3}(\tau\beta)_{ij} = 0\right. \quad (\tau\beta)_{ij} \sim iidN(0,\tfrac{3-1}{3}\sigma^2_{\tau\beta})$$
$$\sum_{k=1}^{4}\gamma_k = 0 \quad \left|\sum_{k=1}^{4}(\tau\gamma)_{ik} = 0\right. \quad (\tau\gamma)_{ik} \sim iidN(0,\tfrac{4-1}{4}\sigma^2_{\tau\gamma})$$
$$\sum_{j=1}^{3}(\beta\gamma)_{jk} = 0 \quad \left|\sum_{k=1}^{4}(\beta\gamma)_{jk} = 0\right.$$
$$\sum_{j=1}^{3}(\tau\beta\gamma)_{ijk} = 0 \left|\sum_{k=1}^{4}(\tau\beta\gamma)_{ijk} = 0\right| (\tau\beta\gamma)_{ijk} \sim iidN(0,\tfrac{(3-1)(4-1)}{3\times4}\sigma^2_{\tau\beta\gamma})$$

ε_{ijk}, τ_i, $(\tau\beta)_{ij}$, $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$, and $(\tau\beta\gamma)_{ijk}$ are independent.
Since this is a simple replicated factorial design, I use $(\tau\beta\gamma)_{ijk}$ to compute SSE and df.

Regression Analysis

```
summary(model_2016s2_1)
```

```
# When some factors are random
table_2019s2$Run_r <- as.random(table_2019s2$Run)
table_2019s2$Trt_f <- as.fixed(table_2019s2$Trt)
table_2019s2$Rev_f <- as.fixed(table_2019s2$Rev)
model_2019s2_1<-aov(Shrink ~ Run_r+Trt_f +
  Trt_f%in%Run_r+ Rev_f%in%Run_r + Rev_f + Trt_f:Rev_f,table_2019s2)
pander(gad(model_2019s2_1))
```

The results show all the main effects and the interaction effect of Runs and Recolutions are significant at 0.05 significance level (P-value=0.5082).

ANOVA

```
pander(anova(model_2016s2_1))
```

The ANOVA table shows that only sources have significant effects on the average purity of a chemical compound synthesized at 0.05 significance level (p-value=).

```
# When some factors are random
model_2017f2_3<-lmer(y~(1|machine)+station+power+
  (1|machine:station)+ (1|machine:station:power),table_2017f2,REML=TRUE
summary(model_2017f2_3)$varcor
pander(confint(model_2019s2_2)[1:4,1:2])
```

$\hat{\sigma}^2_{\tau\beta} =;$
$\hat{\sigma}^2_{\tau} =;$
$\hat{\sigma}^2 =;$
$CI_{\sigma^2_{\tau\beta}} : (,);$
$CI_{\sigma^2_{\tau}} : (,);$
$CI_{\sigma^2} : (,);$
The results of variance components show the variance of interaction term of Runs and revolutions is negligible and hence dropping interaction term of them.
Similarly, there is a significant interaction effect from the fat and temperature, on average amount of force (g) (p-value=).
This means that the effects of day v.s.temperature and fat v.s.temperature on the force are not independent. Hence, the simple effects must be tested.

elimination regression

```
# 5k1p Fractional Factorial Design [2015F2] [2018F2]
library(daewr)
halfnorm(model_2015f2$coefficients[2:8],alpha=1)
summary(model_2015f2)
library(gghalfnorm)
gghalfnorm(x =model_2015f2$coefficients[2:8],
  labs = names(model_2015f2$coef) , nlab = 8)+theme_light()
```

I=ABCD, AB=CD, AC=BD, AD=BC; A=BCD, B=ACD, C=ABD, D=ABC; III [2015f2]
I=ABCD, AB=CD, AC=BD, BC=AD; A=BCD, B=ACD, C=ABD, D=ABC; III [2018F2]

```
model_2016s2_2 <- lm(y~power:machine:station, table_2016s2)
model_2016f1_2<-lm(log(Scig)~perfem:Income:Age+log(price),table_2016f1)
# Redo analysis
```

The ANOVA table of new model shows that the interaction effects are significant. This means that the effects of day v.s.revolutions and treatment v.s.revolutions on the shrink are not independent. Hence, the simple effects must be tested.
The results of variance components and condidence intervals show that none of the effects related with technician has significant variance on average value of purity at 0.05 significance level. The variance of interaction effect between sources and technicians nested in labs is zero with confidence intervals $(0, 1.539^2)$ at 0.05 significance level. The variance of technicians nested in labs is zero with confidence intervals $(0, 1.603^2)$ at 0.05 significance level.

Comparison

The Tables below show the summary of all those simple effect comparison tests.

```
kable(pairs(lsmeans(model_2016f2_1,~creek,adjust=c("tukey"))))
kable(test(lsmeans(model_2016f2_1,~creek,adjust=c("tukey"))))
kable(TukeyHSD(model_2016f2_3,conf.level=0.95)$creek_f)

model_2017sd1 <- aov(Nitrogen~Animal+Ration, table_2017sd1)
TukeyHSD(model_2017sd1,conf.level = 0.95)
```

$H_0 : \beta_2 = 0, H_1 : \beta_2 \neq 0$

```
Rev_Trt <- pairs(lsmeans(model_2019s2_1,~ Rev|Trt))
Trt_Rev <- pairs(lsmeans(model_2019s2_1,~ Trt|Rev))
Trt_Run <- pairs(lsmeans(model_2019s2_3,~ Trt_f|Run_r))
Run_Trt <- pairs(lsmeans(model_2019s2_3,~ Run_r|Trt_f))
kable(test(rbind(Trt_Run,Run_Trt),adjust="tukey"),format="latex")%>%
  kable_styling("condensed",full_width=F,font_size = 8)%>%
  row_spec(c(10,26:27),bold =T)%>%
  row_spec(c(10,26:27),background ="#EAFAF1")
kable(test(rbind(Rev_Trt,Trt_Rev),adjust="tukey"),format="latex")%>%
  kable_styling("condensed",full_width=F,font_size = 8)%>%
  row_spec(c(21,85),bold =T)%>%
  row_spec(c(21,85),background ="#EAFAF1")
```

When the day2, the mean shrinks between the 15-Sec and 4-Min treatment don't have significant difference. For all the rest of days, the mean shrinks are significantly different between any different treatment.

The changes of days for a given treatment don't give consistent results.
For untreated cases, the mean shrinks are not significantly different between 1200 and 1400 revolutions. For all the rest of treatments, the mean shrinks are significantly different between any different revolutions.
For a given revolution, 15-Sec and 4-Min treatment don't have significant differece on the mean shrinks.
Conclusion: As the main effects of sources shown in the above tables, the average purity is different with sources. The average purity from source 1 is lowest (12.72222). The the average purity from source 2 and 3 are 21.38889 and 20.27778 respectively. The selections of labs and technicians don't change this result.

Check Adequacy

```
plot(model_2016s2_2)

model_2017sr1 <- lm(Y^2~X1+X2, table_2017sr1)
plot(model_2017sr1,c(1,3,5))
residual_2017sr1 <- rstudent(model_2017sr1)
qqnorm(residual_2017sr1)
qqline(residual_2017sr1)
olsrr::ols_plot_resid_hist(model_2017sr1)
hist(residual_2017sr1)
```

In the plots of residuals versus predicted value of shrink, there is no significant pattern on this plot. Therefore, the fitted model is good enough to describe the relationship between the mean value of shrink and the days, revolutions, and treatment. The residuals in this plot are almost symmetrically distributed about zero and hence zero mean assumption is not violated. Further, the vertical deviation of the residuals from zero is about same for each predicted value and hence the constant variance assumption is not violated.
The points are along the straight line in the normal qq plot shown at bottom left and the histogram of residuals shown at the top right is about normal. These plots show no violation of normal distribution assumption of residuals.

Conclusion

Choosing a higher revolution for a given treatment can get a larger shrink.
In most of the cases, longter alcoholic potash have less shrink. This effect will be more significant when higher revolution.
For a given temperature, almost all of the amount of fat cannot change the texture but different days don't give consistent results. If the baker wants to examine the texture for a specific temperature, he/she should check what other factors in different days may affect the results and redo the experiment.