STATS 201 Experimental Class 3

Runze Liao, 222020321102007

Pizza size

In 2012, an Australian pizza company launched a marketing campaign claiming that their pizzas were larger than their rival's. We can test these claims using data collected by an independent company.

Each row in pizza.txt corresponds to a pizza that was purchased and measured.

The variables are

• size: the diameter of the pizza in cm.

• store: a factor indicating the store the pizza was purchased from, with levels A and B (the store names have been anonymised).

• crust: a factor describing the crust of the pizza, with levels thick, thin, and mid. The claims were made by Store A.

We are interested in a few questions. Does Store A actually make bigger pizzas? Is the type of crust related to the size of the pizza? Does the difference in pizza size between stores depend on the crust type? Conduct a full analysis, and include Methods and Assumption Checks along with an Executive Summary, answering the above questions.

Code and output

```
## Loading in the data.
library(s20x)

## Warning: package 's20x' was built under R version 4.0.5

library(emmeans)

pizza.df = read.table(file = "pizza.txt", header = TRUE)
pizza.df$store = as.factor(pizza.df$store)
pizza.df$crust = as.factor(pizza.df$crust)
## Plot the data.
## INSERT CODE HERE.
## Analyse the data.
## INSERT CODE HERE.

interactionPlots(size~store+crust, data = pizza.df)
```

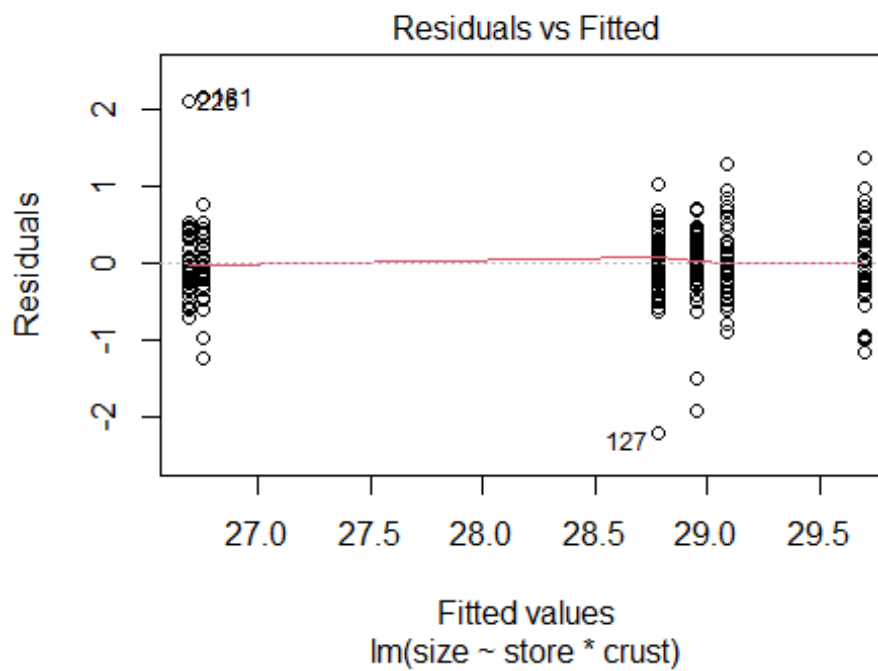## Plot of 'size'
by levels of 'store' and 'crust'

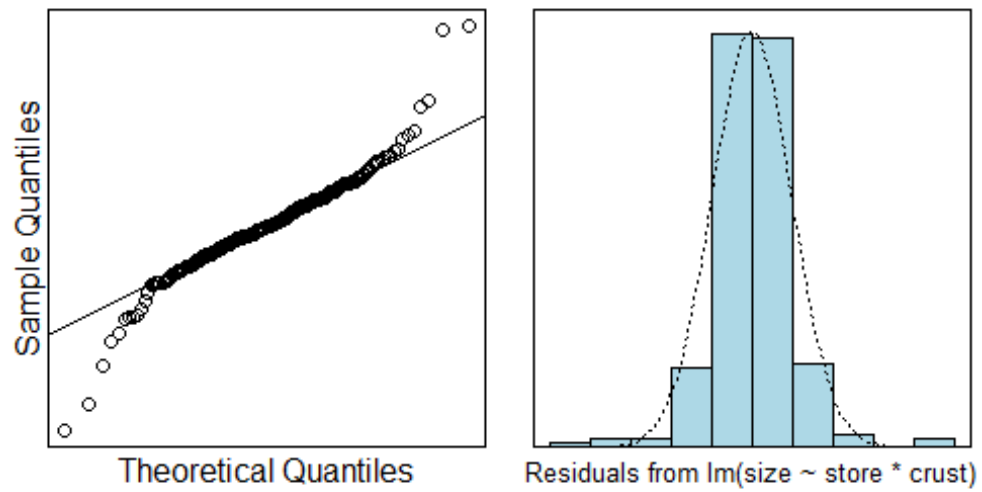Note we do not have parallel lines, so we indicate that there may have some interaction between the two factor.

```
pizza.fit = lm(size~store*crust, data = pizza.df)

plot(pizza.fit, which = 1)
```
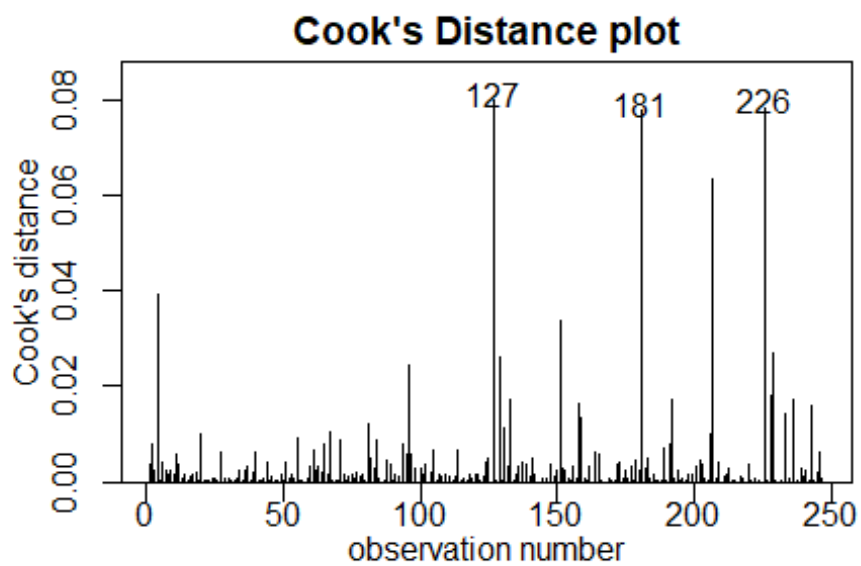
Residuals vs Fitted

Fitted values
lm(size ~ store * crust)

```
normcheck(pizza.fit)
```


Theoretical Quantiles — Sample Quantiles

Residuals from lm(size ~ store * crust)

```
cooks20x(pizza.fit)
```


Cook's Distance plot

The residual plot seems good, EOV assumption is satisfied. The normality assumption seems good too, other than a lack of negative residuals.

We see the cooks distance, all points < 0.4, no strong influential data points. However, points 117, 181,226 seems strange, but we will keep it.

All assumptions were satisfied, we conduct we can trust our fit model, Let us see the summary.

```
summary(pizza.fit)

##
## Call:
## lm(formula = size ~ store * crust, data = pizza.df)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.20209 -0.24224 -0.00099  0.27235  2.14667
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       28.78209    0.07574 380.035  < 2e-16 ***
## storeB            -2.02876    0.10774 -18.830  < 2e-16 ***
## crustthick         0.30721    0.10711   2.868  0.00449 **
## crustthin          0.91842    0.10982   8.363 4.86e-15 ***
## storeB:crustthick -0.37054    0.15333  -2.417  0.01641 *
## storeB:crustthin   1.28093    0.15475   8.277 8.54e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4966 on 242 degrees of freedom
## Multiple R-squared:  0.8487, Adjusted R-squared:  0.8455
## F-statistic: 271.4 on 5 and 242 DF,  p-value: < 2.2e-16
```

All p-vlaue of coefficients and interaction coefficients are less than 0.05, it seemd that we can trust our models.

```
#emmeans(pizza.fit, specs = store~crust)$emmeans
summary2way(pizza.fit, page = "interaction")

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = fit)
##
## $`Comparisons within c("store", "crust")`
##                     diff           lwr       upr      p adj
## A:thick-A:mid   0.30720930 -0.0004735879 0.6148922 0.0506094
## A:thin-A:mid    0.91841980  0.6029462339 1.2338934 0.0000000
## B:thick-B:mid  -0.06333333 -0.3785250980 0.2518584 0.9924384
## B:thin-B:mid    2.19934959  1.8861327930 2.5125664 0.0000000
```

```
## A:thin-A:thick  0.61121049  0.2957369316 0.9266841 0.0000010
## B:thin-B:thick  2.26268293  1.9456216810 2.5797442 0.0000000
##
## $`Comparisons between c("store", "crust")`
##                      diff       lwr        upr p adj
## B:mid-A:mid     -2.0287597 -2.338269 -1.7192508     0
## B:thick-A:thick -2.3993023 -2.712701 -2.0859035     0
## B:thin-A:thin   -0.7478299 -1.066942 -0.4287177     0
```

Methods and Assumption Checks

We have 2 explanatory factors variable, namely the store and crust, the store can be A or B, the crust could be thin, mid and thick.So we fitted a two-way ANOVA model with interaction between the crust and store.

Through the interaction plot we can see that the two variables have interactions, so we build a linear model with interaction by having tow factors explanatory variables, the EOV check is good, the Normal Check is good, and no other influence points, all the assumptions were satisfied by our final model.

Our final model is:

$$Size_i = \beta_0 + \beta_1 \times store_i + \beta_2 \times crustthick_i + \beta_3 \times crustthin_i \times store_i + \beta_4 \times storeB{:}crustthick_i + \beta_5 \times storeB{:}crustthin_i + \epsilon_i$$

where $\epsilon_i \sim iid. N(0, \sigma^2)$. And one of the indicators of store takes it value 1 when the store is B, the indicators of crustthin take value of 1 when the crust is thin, crustthick take value of 1 when the crust is thick, otherwise, it is mid.

Our model explains about 49.7% of the variability in pizza size.

Executive Summary

We wanted to have a model to explain the the pizza's size influenced by its crust and its store methods. so we build a linear model with interaction by having tow factors explanatory variables. There was evidence to suggest that the explanatory variables were related to the response variable, we We have evidence that the difference in pizza size between stores depend on the crust type.

We estimate that:

- With the Same store A, the different crust thin is larger 0.60 between 1.23 than mid.

- With the same Store B, the different crust thin is larger 1.89 between 2.51 than mid.

- With the Same store A, the different crust thin is larger 0.29 between 0.93 than thick.

- With the same Store B, the different crust thin is larger 1.95 between 2.58 than thick.

- With the same crust mid, the different store A is larger than 1.72 between 2.34 than B.

- With the same crust thick, the different store A is larger than 2.09 between 2.34 than B.

- With the same crust thin, the different store A is larger than 0.43 between 1.07 than B.

- with the same crust, the size of Store A make bigger than B for 2.29 on average.

So we can say that the store A actually make bigger pizzas than B.

Does Store A actually make bigger pizzas?

Yes, and we can see that the Store A make bigger than B 2.29.

Is the type of crust related to the size of the pizza?

Yes, the coefficient of crust is all larger than 0.05, it is in the 95% confindence interval, so we have the evidence that the type of crust is related to the size of the pizza.

Does the difference in pizza size between stores depend on the crust type?

Yes, firstly, we can see the interaction Plot, it is lack of parallel lines. Secondly, we have the evidence that the pizza size depends on curst type by having the p-value is smaller than 0.05. So they have interaction.