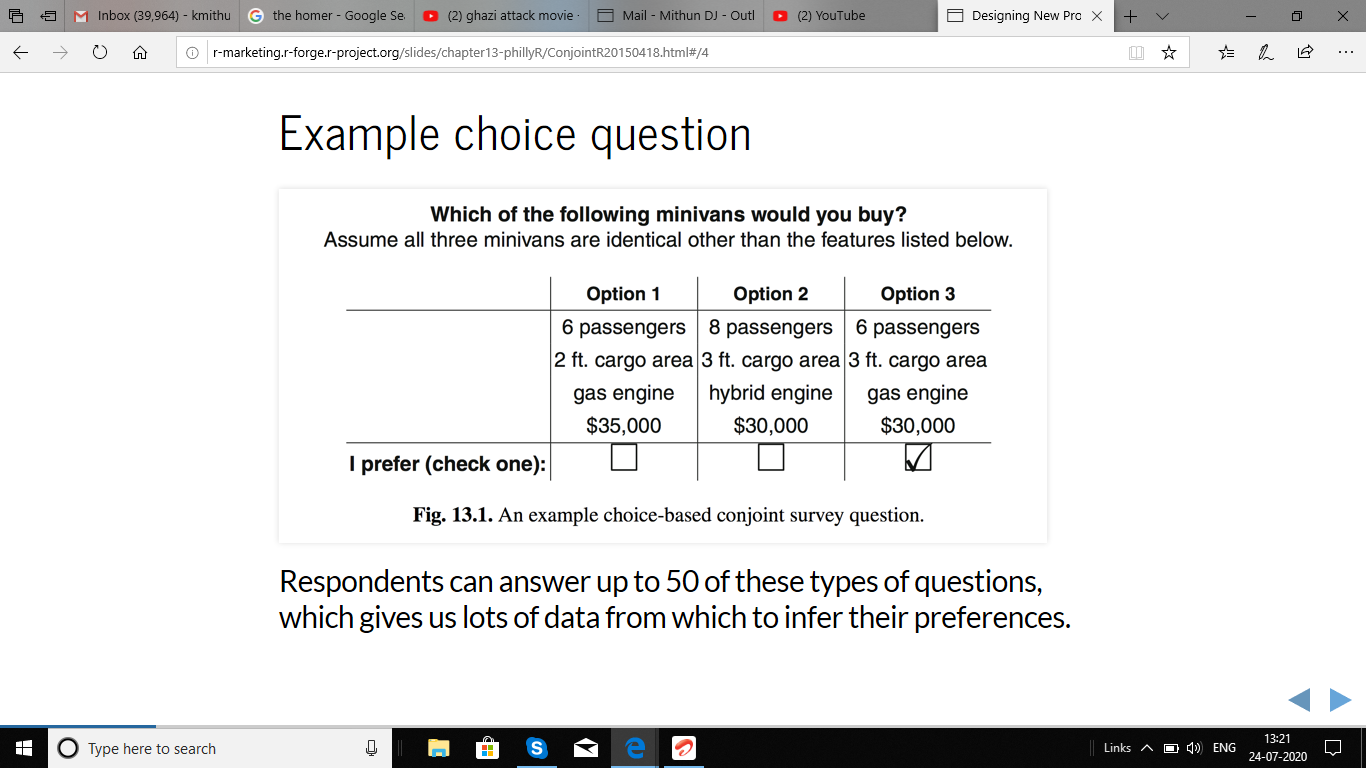
## Designing new products

When creating a new product like this Chevrolet Silverado, designers often face tough decisions.

* Should the truck have a smaller bed so that we can give more leg room to the passengers?
* Should we make the truck larger, even though the fuel economy will go down?
* Better designers spend time talking to potential customers about what they want and that is *sort-of* helpful.
* But customers typically want “everything” and if you listen to them you end up with **nothing**.

**Key idea**

1. Ask customers to choose from among alternative designs (something consumers do every day).
2. Use a choice model to *infer* their preferences from the choices.
3. *Predict* whether they will buy alternative designs using the model.



Respondents can answer up to 50 of these types of questions, which gives us lots of data from which to infer their preferences.

### Let's get started...

**Please follow along!**

### Load some choice data

cbc.df <-

read.csv("http://goo.gl/5xQObB",

colClasses = c(seat = "factor",

price = "factor"))

head(cbc.df[,c(-4, -5)])

resp.id ques alt cargo eng price choice

1 1 1 1 2ft gas 35 0

2 1 1 2 3ft hyb 30 0

3 1 1 3 3ft gas 30 1

4 1 2 1 2ft gas 30 0

5 1 2 2 3ft gas 35 1

6 1 2 3 2ft elec 35 0

Note that the dependant variable in the last column is *multinomial*…a choice from among multiple options.

**Always inspect the data before modeling!**

**Summarize the choices**

xtabs(choice ~ price, data=cbc.df)

price

30 35 40

1486 956 558

xtabs(choice ~ cargo, data=cbc.df)

cargo

2ft 3ft

1312 1688

**Summarize the choices**

xtabs(choice ~ seat, data=cbc.df)

seat

6 7 8

1164 854 982

xtabs(choice ~ eng, data=cbc.df)

eng

elec gas hyb

608 1444 948

**Why model?**

* As with any other multivariate data set, looking at univariate marginal summaries only tells part of the story.
* While we can see that customers tend to choose 6-seat minivans and tend to choose gas engines, it is hard to say whether seats or engines has a stronger influence on choice.
  + What if in the survey the 6-seat minivan options tended to have gas engines?
* As with many multivariate problems, the solution is a regression model (of a special type.)

**First load a package**

library(mlogit)

And do a bit of data formatting to tell mlogit which column is which in our choice data:

cbc.mlogit <-

mlogit.data(data=cbc.df, choice="choice",

shape="long", varying=3:6,

alt.levels=paste("pos",1:3),

id.var="resp.id")

It would be nice if mlogit used formula notation.

**Estimate a choice model**

m1 <- mlogit(choice ~ 0 + seat + cargo + eng + price, data = cbc.mlogit)

summary(m1)

Call:

mlogit(formula = choice ~ 0 + seat + cargo + eng + price, data = cbc.mlogit,

method = "nr", print.level = 0)

Frequencies of alternatives:

pos 1 pos 2 pos 3

0.32700 0.33467 0.33833

nr method

5 iterations, 0h:0m:0s

g'(-H)^-1g = 7.84E-05

successive function values within tolerance limits

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

seat7 -0.535280 0.062360 -8.5837 < 2.2e-16 \*\*\*

seat8 -0.305840 0.061129 -5.0032 5.638e-07 \*\*\*

cargo3ft 0.477449 0.050888 9.3824 < 2.2e-16 \*\*\*

enggas 1.530762 0.067456 22.6926 < 2.2e-16 \*\*\*

enghyb 0.719479 0.065529 10.9796 < 2.2e-16 \*\*\*

price35 -0.913656 0.060601 -15.0765 < 2.2e-16 \*\*\*

price40 -1.725851 0.069631 -24.7856 < 2.2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -2581.6

**Estimate a choice model (continued)**

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

seat7 -0.535280 0.062360 -8.5837 < 2.2e-16 \*\*\*

seat8 -0.305840 0.061129 -5.0032 5.638e-07 \*\*\*

cargo3ft 0.477449 0.050888 9.3824 < 2.2e-16 \*\*\*

enghyb -0.811282 0.060130 -13.4921 < 2.2e-16 \*\*\*

engelec -1.530762 0.067456 -22.6926 < 2.2e-16 \*\*\*

price35 -0.913656 0.060601 -15.0765 < 2.2e-16 \*\*\*

price40 -1.725851 0.069631 -24.7856 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Simulating choice shares: a little function**

predict.mnl <- function(model, data) {

data.model <-

model.matrix(

update(model$formula, 0 ~ .),

data = data)[,-1]

utility <- data.model%\*%model$coef

share <- exp(utility)/sum(exp(utility))

cbind(share, data)

}

**Simulating choice shares: output**

predict.mnl(m1, new.data)

share seat cargo eng price

8 0.44643895 7 2ft hyb 30

1 0.16497955 6 2ft gas 30

3 0.12150814 8 2ft gas 30

41 0.02771959 7 3ft gas 40

49 0.06030713 6 2ft elec 40

26 0.17904663 7 2ft hyb 35

Suppose you were designing the first minivan to compete against the other five. Looks like you have a pretty good design.

**Estimate a different choice model**

m3 <-

mlogit(choice ~ 0 + seat + cargo + eng

+ as.numeric(as.character(price)),

data = cbc.mlogit)

**Estimates for a different choice model**

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

seat7 -0.5345392 0.0623518 -8.5730 < 2.2e-16 \*\*\*

seat8 -0.3061074 0.0611184 -5.0084 5.488e-07 \*\*\*

cargo3ft 0.4766936 0.0508632 9.3721 < 2.2e-16 \*\*\*

enghyb -0.8107339 0.0601149 -13.4864 < 2.2e-16 \*\*\*

engelec -1.5291247 0.0673982 -22.6879 < 2.2e-16 \*\*\*

price -0.1733053 0.0069398 -24.9726 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Willingness-to-pay**

In a model like m3 where we estimate a single parameter for price, we can compute the average willingness-to-pay for a particular level of an attribute by dividing the coefficient for that level by the price coefficient.

coef(m3)["cargo3ft"]/

(-coef(m3)["as.numeric(as.character(price))"]/1000)

cargo3ft

2750.601

So, on average, customers are willing to pay $2750 more for 3ft of cargo space versus 2ft.

**Compare models**

lrtest(m1, m3)

Likelihood ratio test

Model 1: choice ~ 0 + seat + cargo + eng + price

Model 2: choice ~ 0 + seat + cargo + eng + as.numeric(as.character(price))

#Df LogLik Df Chisq Pr(>Chisq)

1 7 -2581.6

2 6 -2582.1 -1 0.9054 0.3413