Confidence Sets for Regression Coefficients

Econ 440 - Introduction to Econometrics

Patrick Toche, ptoche@fullerton.edu

19 April 2022

Dataset:

```
library(readxl)
df <- read_xlsx("caschool.xlsx", trim_ws=TRUE)
head(df)
```

```
## # A tibble: 6 x 18
##   `Observation Num~` dist_cod county district gr_span enrl_tot teachers calw_pct
##                <dbl>    <dbl> <chr>  <chr>    <chr>      <dbl>    <dbl>    <dbl>
## 1                  1    75119 Alame~ Sunol G~ KK-08        195     10.9    0.510
## 2                  2    61499 Butte  Manzani~ KK-08        240     11.1    15.4
## 3                  3    61549 Butte  Thermal~ KK-08       1550     82.9    55.0
## 4                  4    61457 Butte  Golden ~ KK-08        243     14      36.5
## 5                  5    61523 Butte  Palermo~ KK-08       1335     71.5    33.1
## 6                  6    62042 Fresno Burrel ~ KK-08        137      6.40    12.3
## # ... with 10 more variables: meal_pct <dbl>, computer <dbl>, testscr <dbl>,
## #   comp_stu <dbl>, expn_stu <dbl>, str <dbl>, avginc <dbl>, el_pct <dbl>,
## #   read_scr <dbl>, math_scr <dbl>
```

Regression

Rename variables to match textbook notation

```
names(df)[names(df) == "testscr"] <- "TestScore"
names(df)[names(df) == "expn_stu"] <- "Expn"
names(df)[names(df) == "str"] <- "STR"
names(df)[names(df) == "el_pct"] <- "PctEL"
df$Expn <- df$Expn/1000

m1 <- lm(TestScore ~ STR + Expn + PctEL, data=df)
summary(m1)
```

```
##
## Call:
## lm(formula = TestScore ~ STR + Expn + PctEL, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.34  -10.11    0.29   10.32   43.18
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 649.5779    15.2057   42.72   <2e-16 ***
## STR           -0.2864     0.4805   -0.60   0.5515
## Expn           3.8679     1.4121    2.74   0.0064 **
## PctEL         -0.6560     0.0391  -16.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.4 on 416 degrees of freedom
## Multiple R-squared:  0.437,  Adjusted R-squared:  0.433
## F-statistic:  107 on 3 and 416 DF,  p-value: <2e-16
```

Confidence Ellipses

Draw confidence ellipses for regression coefficients on *STR* and *Expn*.

```
library(ellipse)
```

```
##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##     pairs
```

```
library(latex2exp)
conf.ellipse <- data.frame(ellipse(m1, which=c(2,3)))
ggplot(conf.ellipse, aes(x=STR, y=Expn)) +
  geom_path() +
  geom_polygon(fill="cornflowerblue", alpha=0.5, color="darkblue", size=1) +
  scale_x_continuous(breaks=seq(-2,2,0.5), limits=c(-1.5,1.5)) +
  scale_y_continuous(breaks=seq(-1,9,1)) +
  geom_point(x=m1$coefficient[2], y=m1$coefficient[3], shape=3, size=2) +
  geom_point(x=0, y=0, shape=3, size=2) +
  labs(x="Coefficient on STR", y="Coefficient on Expn",
       title="95% Confidence Set for Coefficients on STR and Expn") +
  annotate("segment", x=0.75, xend=0.05, y=0, yend=0) +
  annotate("segment",
           x=m1$coefficient[2]+0.05, xend=0.75,
           y=m1$coefficient[3]-0.1, yend=1) +
  annotate("text", x=0.8, y=1,
           label=TeX("($\\hat{\\beta}_{1}, \\hat{\\beta}_{2})=(-0.29,3.87)$"),
           parse=TRUE, hjust=0) +
  annotate("text", x=0.8, y=0,
           label=TeX("($\\beta_{1}, \\beta_{2})=(0,0)$"),
           parse=TRUE, hjust=0) +
  theme_classic()
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'

## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```

95% Confidence Set for Coefficients on STR and Expn

$$(\hat{\beta}_1, \hat{\beta}_2) = (-0.29, 3.87)$$

$$(\beta_1, \beta_2) = (0, 0)$$

Coefficient on STR
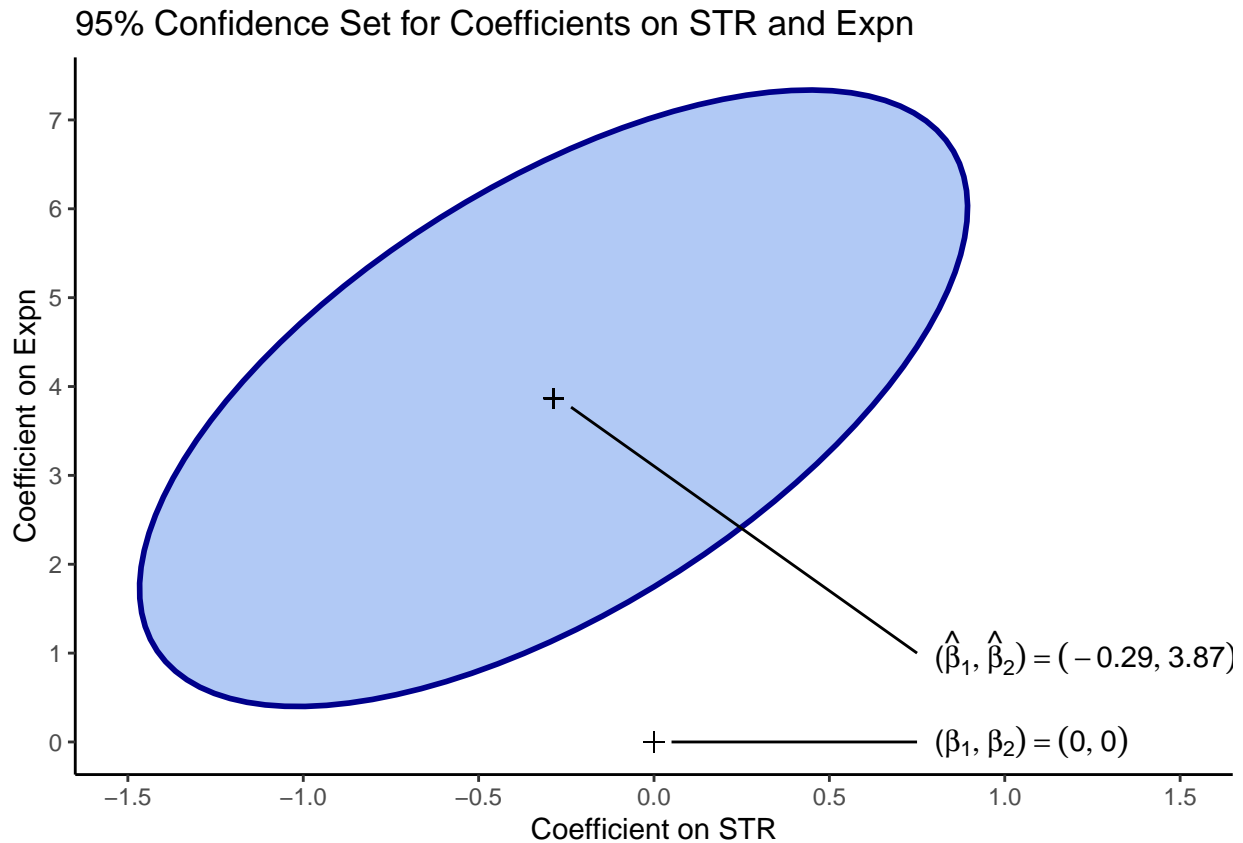
```
ggsave(last_plot(), file="plot-regression-ellipses.pdf", width=8, height=5)
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```

The 95% confidence set Coefficient on STR (β_1) and $Expn$ (β_2) is an ellipse. The ellipse contains the pairs of values of β_1 and β_2 that cannot be rejected using the F-statistic at the 5% significance level. The point (β_1, β_2=(0,0)) is not contained in the confidence set, so the null hypothesis $H_0 : \beta_1 = 0 \text{and} \beta_2 = 0$ is rejected at the 5% significance level.

Joint Hypothesis test

Joint test of hypothesis $H_0 : STR = 0$ and $Expn = 0$.

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:ellipse':
##
##      ellipse
```

```
## The following object is masked from 'package:dplyr':
##
```

```
##       recode
```

```
H0 <- c("STR=0", "Expn=0")
tidy(linearHypothesis(m1, H0))
```

```
## # A tibble: 2 x 6
##   res.df    rss    df sumsq statistic   p.value
##    <dbl>  <dbl> <dbl> <dbl>     <dbl>     <dbl>
## 1    418 89000.    NA    NA        NA  NA
## 2    416 85700.     2 3300.      8.01  0.000386
```