STAT 425

Multiple Linear Regression. Part 4

Confidence Intervals for the β_j's

- A $(1-\alpha)$ CI for β_j is given by

$$\left(\hat{\beta}_j \ \pm \ t_{n-p}^{(\alpha/2)} \mathsf{se}(\hat{\beta}_j)\right) \ = \ \left(\hat{\beta}_j \ \pm \ t_{n-p}^{(\alpha/2)} \hat{\sigma} \sqrt{\left[(\mathbf{X}^t\mathbf{X})^{-1}\right]_{jj}} \ \right)$$

where $t_{n-p}^{(\alpha/2)}$ is the $(1-\alpha/2)$ percentile of the student T-dist with $(n-p)$ degree-of-freedom.

In R we can use the function confint(.)

Use the command **confint** to obtain confidence intervals for regression coefficients.

```
confint(fullmodel)
```

```
##                    2.5 %       97.5 %
## (Intercept) 13.753330728 43.378842354
## pop15       -0.752517542 -0.169868752
## pop75       -3.873977955  0.490982602
## dpi         -0.002212248  0.001538444
## ddpi         0.014533628  0.804856227
```

```
confint(fullmodel, 'pop15', level=0.99)
```

```
##            0.5 %      99.5 %
## pop15 -0.8502207 -0.07216559
```

Confidence Region

Just as we can use estimated standard errors and t-stats to form confidence intervals for a single parameter, we can also obtain a $(1-\alpha) \times 100\%$ confidence region for the entire vector $\boldsymbol{\beta}$. In particular:

$$\boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2(\mathbf{X^t X})^{-1})$$

Thus, the quadratic form:

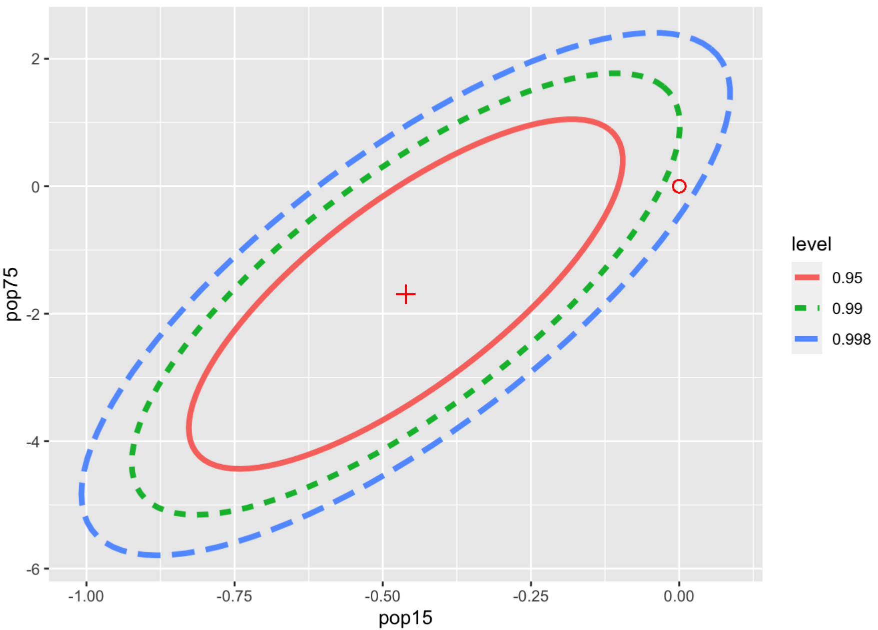
$$\frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t \mathbf{X^t X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{p\hat{\sigma}^2} \sim F_{p,n-p}$$

Then we can construct a $(1-\alpha)\times 100\%$ confidence region for $\boldsymbol{\beta}$ to be all the points in the following ellipsoid

$$\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^t\mathbf{X}^{\mathbf{t}}\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{p\hat{\sigma}^2} < F(\alpha;p,n-p)$$

where $F(\alpha;p,n-p)$ is defined to be the point such that:

$$Pr[F_{p,n-p} > F(\alpha;p,n-p)] = \alpha$$



pop75
2
0
-2
-4
-6
-1.00
-0.75
-0.50
-0.25
0.00
pop15
level
0.95
0.99
0.998

Confidence Interval and Prediction Interval at a future observation $\mathbf{x}^*$

- We are interested in obtaining an estimate $E[Y|\mathbf{x}^*] = \mu^* = (\mathbf{x}^*)^t\boldsymbol{\beta}$ and a prediction for a future observation Y^* at $\mathbf{x}^*$. We also want to have a CI for μ^* and a PI for y^*.
- The Gauss-Markov theorem tells us that the BLUE (Best Linear Unbiased Estimate) of μ^* is:

$$\hat{\mu}^* = (\mathbf{x}^*)^t\hat{\boldsymbol{\beta}} = (\mathbf{x}^*)^t(\mathbf{X^tX})^{\mathbf{-1}}\mathbf{X^ty}$$

This is just a linear transformation of $\mathbf{y}$, so we can easily derive its variance, and find its standard error.

It can be shown that:

$$se(\hat{\mu}^*) = \hat{\sigma}\sqrt{(\mathbf{x}^*)^{\mathbf{t}}(\mathbf{X}^{\mathbf{t}}\mathbf{X})^{-\mathbf{1}}\mathbf{x}^*}$$

- A Confidence Interval for μ^* is given by:

$$(\hat{\mu}^* - t_{n-p}^{(\alpha/2)} se(\hat{\mu}^*), \hat{\mu}^* + t_{n-p}^{(\alpha/2)} se(\hat{\mu}^*))$$

- The best estimate for y^* at a future observation $\mathbf{x}^*$ is also

$$\hat{y}^* = (\mathbf{x}^*)^t \hat{\boldsymbol{\beta}}$$

- In order to find a prediction interval (PI), we need to consider the variance due to $\hat{\boldsymbol{\beta}}$ in addition to the variance associated with a new observation, which is σ^2.
- The standard error of a prediction estimate $\hat{y}^*$ is:[1]

$$se(\hat{y}^*) = \hat{\sigma}\sqrt{1 + (\mathbf{x}^*)^{\mathbf{t}}(\mathbf{X}^{\mathbf{t}}\mathbf{X})^{-\mathbf{1}}\mathbf{x}^*}$$

- A $(1-\alpha)100\%$ PI for a new observation Y^* at $\mathbf{x}^*$ is given by:

$$(\hat{y}^* - t_{n-p}^{(\alpha/2)} se(\hat{y}^*), \hat{y}^* + t_{n-p}^{(\alpha/2)} se(\hat{y}^*))$$

[1]Note that no matter how large the sample size becomes, the width of a PI, unlike a CI, will never approach 0.

```r
# create a data frame on which you'd like to predict
meanvalue=apply(savings[,2:5],2,mean)
meanvalue
```

```
##     pop15     pop75       dpi      ddpi
##   35.0896    2.2930 1106.7584    3.7576
```

```r
x=data.frame(t(meanvalue))
predict.lm(fullmodel,x,interval="confidence")
```

```
##     fit      lwr      upr
## 1 9.671 8.587858 10.75414
```

```r
predict.lm(fullmodel,x,interval="prediction")
```

```
##     fit      lwr      upr
## 1 9.671 1.935822 17.40618
```

Simultaneous CIs and PIs

- Consider a simple linear regression $y_i = \beta_0 + \beta_1 x_i + e_i$
- Given the values of x^*, the $(1-\alpha)100\%$ Confidence Interval for $\mu^* = E[y|x^*] = \beta_0 + \beta_1 x_i$ is:

$$I(x^*) = (\hat{\mu}^* \pm t_{n-2}^{(\alpha/2)} se(\hat{\mu}^*) \qquad (1)$$

where

$$\hat{\mu}^* = \hat{\beta}_0 + \hat{\beta}_1 x^* \ \text{ and } \ se(\hat{\mu}^*) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

- If we want CIs at multiple points $(x_1^*, x_2^*, \ldots, x_m^*)$, we can use formula (1) to have CIs at the m points: $I(x_1^*), I(x_2^*), \ldots, I_m(x^*)$

Bonferroni Correction

- We know that:

$$Pr[\mu_i^* \in I(x_i^*)] = (1 - \alpha)$$

  This is the point-wise coverage probability for μ_i^* and formula (1) gives the point-wise CI.

- What about the simultaneous coverage probability? i.e.:

$$Pr[\mu_i^* \in I(x_i^*),\ \text{for}\ \ i = 1, \ldots, m] =?$$

- To make sure that (for example):

$$Pr[\mu_i^* \in I(x_i^*),\ \text{for}\ \ i = 1, \ldots, m] = .95$$

  we need to set $\alpha = 5\%/m$, which is known as the Bonferroni correction

Bonferroni Correction

Let A_k denotes the event that the *k*th confidence interval covers μ_k^* with:

$$Pr(\mathrm{A}_k) = (1 - \alpha)$$

Then we can show:

$$\begin{aligned}
&Pr(\text{All CIs cover the corresponding } \mu_k^* \text{ values}) \\
&= Pr(\mathrm{A}_1 \cap \mathrm{A}_2 \ldots \cap \mathrm{A}_m) \\
&= 1 - Pr(\mathrm{A}_1^c \cup \mathrm{A}_2^c \ldots \cup \mathrm{A}_m^c) \\
&\geq 1 - Pr(\mathrm{A}_1^c) - \ldots - Pr(\mathrm{A}_m^c) \\
&= 1 - m\alpha
\end{aligned}$$

If we choose α/m instead of α, the simultaneous coverage probability will be $(1 - \alpha)$.