Interpretation, Prediction and Confidence Intervals

Merlise Clyde

September 15, 2017

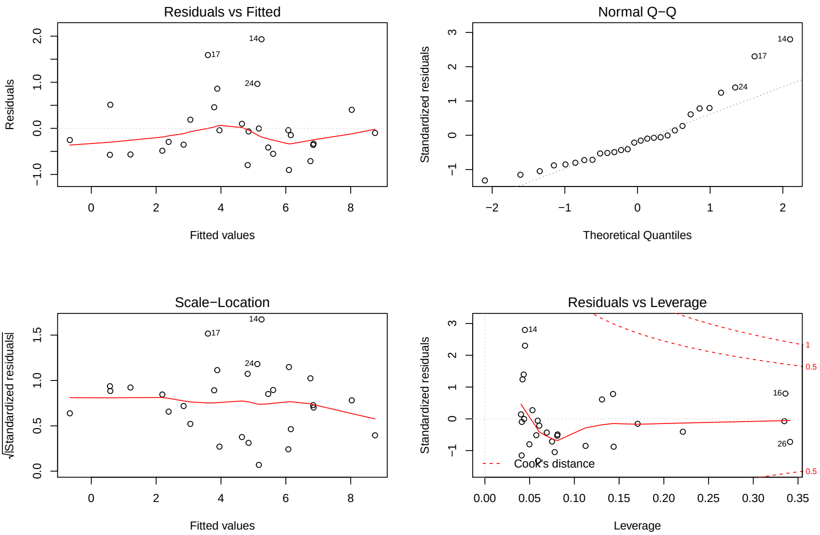
Last Class

Model for log brain weight as a function of log body weight

- Nested Model Comparison using ANOVA led to model with parallel lines
- Why does model with the 3 indicator variable contain the other models?

$$\log(\mathit{brain}) = \beta_0 + \log(\mathit{body})\beta_1 + \text{Dino.T}\beta_2 + \text{Dino.B}\beta_3 + \text{Dino.D}\beta_4 + \epsilon$$

Check residuals



Coefficient Summaries

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.16	0.19	11.09	0.00
log(body)	0.75	0.04	16.90	0.00
DinoTRUE	-5.22	0.53	-9.84	0.00

Distribution of Coefficients

- Joint Distribution under normality

$$\hat{\boldsymbol{\beta}} \mid \sigma^2 \sim \mathsf{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$$

- Distribution of SSE

$$\mathsf{SSE}/\sigma^2 \sim \chi^2(n-p)$$

- Marginal distribution

$$\frac{\hat{\beta}_j - \beta_j}{\mathsf{SE}(\hat{\beta}_j)} \sim \mathsf{St}(n-p)$$

$$\mathsf{SE}(\hat{\beta}_j) = \hat{\sigma}\sqrt{[\mathbf{X}^T\mathbf{X}]^{-1}]_{jj}]}$$

$$\hat{\sigma}^2 = \frac{\mathsf{SSE}}{n-p}$$

Confidence Intervals

$(1-\alpha/2)100\%$ Confidence interval for β_j

$$\hat{\beta}_j \pm t_{n-p,\alpha/2}\mathrm{SE}(\hat{\beta}_j)$$

```r
kable(confint(brain2.lm))
```

	2.5 %	97.5 %
(Intercept)	1.760198	2.5630848
log(body)	0.657346	0.8397591
DinoTRUE	-6.311226	-4.1274760

Converting to Original Units

- Model after exponentiating

$$
\begin{aligned}
\widehat{brain} &= e^{\hat{\beta}_0 + \log(body)\hat{\beta}_1 + \text{Dino}\hat{\beta}_2} \\
&= e^{\hat{\beta}_0} e^{\log(body)\hat{\beta}_1} e^{\text{Dino}\hat{\beta}_2} \\
&= e^{\hat{\beta}_0} body^{\hat{\beta}_1} e^{\text{Dino}\hat{\beta}_2}
\end{aligned}
$$

- 10% increase in body weight implies a

$$
\begin{aligned}
\widehat{brain}_{1.10} &= e^{\hat{\beta}_0} (1.10 * body)^{\hat{\beta}_1} e^{\text{Dino}\hat{\beta}_2} \\
&= 1.10^{\hat{\beta}_1} e^{\hat{\beta}_0} body^{\hat{\beta}_1} e^{\text{Dino}\hat{\beta}_2}
\end{aligned}
$$

- $1.1^{\hat{\beta}_1} = 1.074$ or a 7.4% increase in brain weight

95% Confidence interval

To obtain a 95% confidence interval, $(1.10^{CI} - 1) * 100$

	2.5 %	97.5 %
body	6.465603	8.332779

Interpretation of Intercept

- Evalutate model with predictors = 0

$$\widehat{\log(brain)} = \hat{\beta}_0 + \log(body)\hat{\beta}_1 + \text{Dino}\hat{\beta}_2$$

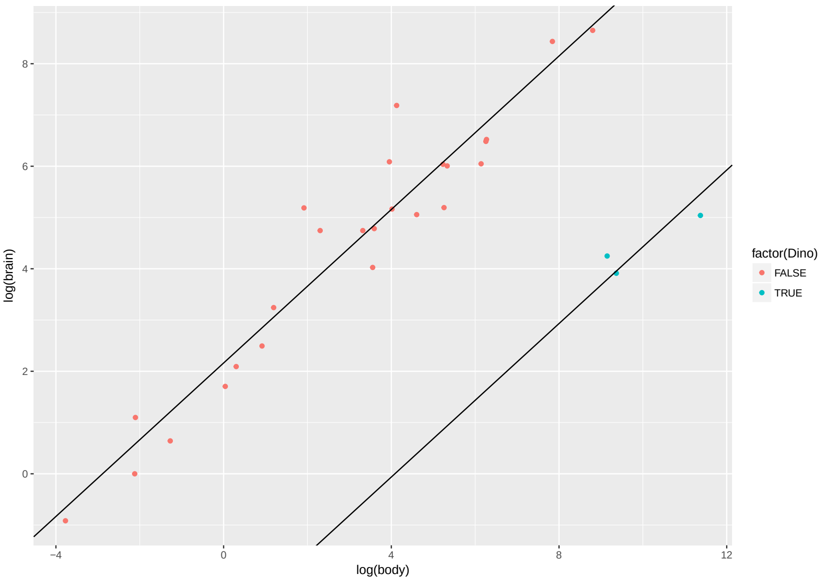
- For a non-dinosaur, if `log(body) = 0` (body weight = 1 kilogram), we expect that brain weight will be 2.16 log(grams) ???
- Exponentiate: predicted brain weight for non-dinosaur with a 1 kg body weight is

$$e^{\hat{\beta}_0} = 8.69 \text{ grams}$$

Plot of Fitted Values

```r
library(ggplot2)
beta= coef(brain2.lm)
gp = ggplot(Animals, aes(y=log(brain), x=log(body))) +
     geom_point(aes(colour=factor(Dino))) +
     geom_abline(aes(intercept=beta[1], slope=beta[2])) +
     geom_abline(aes(intercept=(beta[1]+beta[3]),
                     slope=beta[2]))
```

Plot of Fitted Values



Confidence Intervals for the $f(x)$

- Point Estimate

$$\widehat{f(x)} = \mathbf{x}^T \hat{\beta}$$

- Distribution of MLE given σ

$$\widehat{f(x)} \sim \mathrm{N}(f(x), \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x})$$

- Distribution of pivotal quantity

$$\frac{\widehat{f(x)} - f(x)}{\sqrt{\hat{\sigma}^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}}} \sim t(n-p)$$

- Confidence interval

$$\widehat{f(x)} \pm t_{\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}}$$

Prediction Intervals for Y_* at $\mathbf{x}_*$

- Model

$$Y_* = \mathbf{x}_*^T \beta + \epsilon_*$$

- Y_* independent of other Y's
- Prediction error

$$Y_* - \widehat{f(x)} = \mathbf{x}_*^T \beta - \widehat{f(x_*)} + \epsilon_*$$

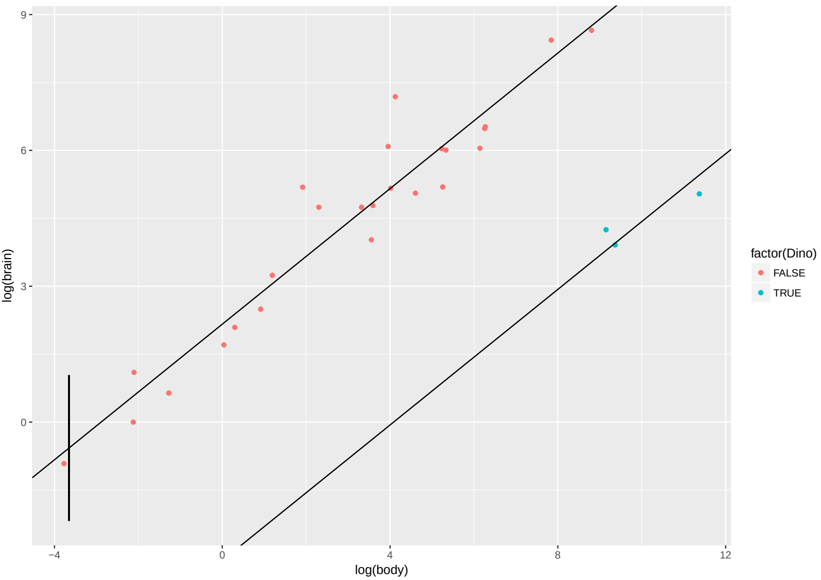
- Variance

$$\begin{aligned}
\mathrm{Var}(Y_* - \widehat{f(x)}) &= \mathrm{Var}(\mathbf{x}_*^T \beta - \widehat{f(x_*)}) + \mathrm{Var}(\epsilon_*) \\
&= \sigma^2 \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_* + \sigma^2 \\
&= \sigma^2 (1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*)
\end{aligned}$$

- Prediction Intervals

$$\widehat{f(x)} \pm t_{\alpha/2} \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*)}$$

Predictions for 259 gram cockatoo



Predictions in original units

- 95% Confidence Interval for $f(x)$

```
newdata = data.frame(body=.0259, Dino=FALSE)
fit = predict(brain2.lm, newdata=newdata,
              interval="confidence", se=T)
```

- 95% Prediction Interval for Brain Weight

```
pred = predict(brain2.lm, newdata=newdata,
               interval="predict", se=T)
```

CI/Predictions in original units for body=259 g

- 95% Confidence Interval for $f(x)$

```
exp(fit$fit)
```

```
##         fit       lwr      upr
## 1 0.5637161 0.2868832 1.107684
```

- 95% Prediction Interval for Brain Weight

```
exp(pred$fit)
```

```
##         fit       lwr     upr
## 1 0.5637161 0.1131737 2.80786
```

- 95% confident that the brain weight will be between 0.11 and 2.81 grams

Summary

- Linear predictors may be based on functions of other predictors (dummy variables, interactions, non-linear terms)
- need to change back to original units
- log transform useful for non-negative responses (ensures predictions are non-negative)
- Be careful of units of data
  - plots should show units
  - summary statements should include units
- Goodness of fit measure: R^2 and Adjusted R^2 depend on scale R^2 is percent variation in "Y" that is explained by the model

$$R^2 = 1 - SSE/SST$$

where $\text{SST} = \sum_i (Y_i - \bar{Y})^2$