

Övning 3 Statistisk översikskurs

Ulf Högnäs

April 2025

För problem med udda nummer, se lösnignarna i boken. Kontakta mig om ni har frågor!

Kapitel 8

12. (a) Det är lite oklart vad som avses med “the equation”, så jag ger två svar här. Här den modell som antas stämma med verkligheten. Denna ekvation skapas innan vi skattar koefficienterna med hjälp av vår insamlade data:

$$\begin{aligned} Y = & \beta_0 + \beta_1 \cdot \text{bill_length_mm} + \beta_2 \cdot \text{bill_depth_mm} \\ & + \beta_3 \cdot \text{flipper_length_mm} + \beta_4 \cdot \text{sexmale} \\ & + \beta_5 \cdot \text{speciesChinstrap} + \beta_6 \cdot \text{speciesGentoo} + \varepsilon \end{aligned}$$

Sedan ekvationen för att skatta pingvinvikter, med skattade koefficienter:

$$\begin{aligned} \hat{y} = & -1461.0 + 18.2 \cdot \text{bill_length_mm} + 67.2 \cdot \text{bill_depth_mm} \\ & + 16.0 \cdot \text{flipper_length_mm} + 389.9 \cdot \text{sexmale} \\ & + -251.5 \cdot \text{speciesChinstrap} + 1014.6 \cdot \text{speciesGentoo} \end{aligned}$$

Den första ekvationen är en teoretisk modell av sambandet mellan förklaringsvariablerna och responsvariabeln. Den säger att det finns ett (genomsnittligt) linjärt samband mellan Y , pingvinernas vikt, och varje förklaringsvariabel. Kring dessa genomsnitt finns det variation, som representeras av symbolen ε .

| Variable | Interpretation |
|--------------------------------|---|
| Intercept | Baseline body weight for a female Adelie penguin with all numeric predictors equal to 0. |
| <code>bill_length_mm</code> | Each additional mm in bill length is associated with an increase of 18.2 g in body weight, holding other variables constant. |
| <code>bill_depth_mm</code> | Each additional mm in bill depth is associated with an increase of 67.2 g in body weight, holding other variables constant. |
| <code>flipper_length_mm</code> | Each additional mm in flipper length is associated with an increase of 16.0 g in body weight, holding other variables constant. |
| <code>sexmale</code> | Male penguins weigh on average 389.9 g more than female penguins, all else equal. |
| <code>speciesChinstrap</code> | Chinstrap penguins weigh on average 251.5 g less than Adelie penguins, all else equal. |
| <code>speciesGentoo</code> | Gentoo penguins weigh on average 1014.6 g more than Adelie penguins, all else equal. |

Table 1: Interpretation of regression coefficients predicting penguin body weight

- (b) Jag har gjort den fina tabellen ovan med hjälp av ChatGPT. Först, lägg märke till tolkningen av interceptet. Eftersom både `speciesChinstrap` och `speciesGentoo` är noll får vi en Adelie-pingvin, eftersom detta är vår base category. Det samma gäller `sexmale`. Men eftersom alla andra variabler är noll får vi något absurt: en pingvin utan näbb och vingar. Interceptet har ingen rimlig tolkning här.

För övriga variabler ser vi fraserna “holding other variables konstant” och “all else equal”. Detta är en viktig poäng. En extra mm näbb längd är förknippad med 18.2 g extra vikt, **holding other variables konstant**. Vi jämför alltså två pingviner där den enda skillnaden är 1 mm näbb längd. De ska vara samma art, kön, ha samma vinglängd och samma näbbdjup. I så fall så ska pingvinen med längre näbb väga 18.2 g mer i snitt. Kommer man inte ihåg detta så kommer man inte att förstå vad koefficienterna betyder.

- (c)

$$\begin{aligned}
 \hat{y} &= -1461.0 + 18.2 \cdot 39.1 + 67.2 \cdot 18.7 \\
 &\quad + 16.0 \cdot 181 + 389.9 \cdot 1 \\
 &\quad + -251.5 \cdot 0 + 1014.6 \cdot 0 \\
 &= 3793.16
 \end{aligned}$$

Vår residual beräknas som den faktiska vikten minus den skattade vikten

$$3750 - 3793.16 = -43.16$$

Denna pingvins vikt överskattades av modellen.

- (d) Vår modell förklarar 87.2% av variationen i vikt hos dessa tre pingvinarter.



Figure 1: En Adeliepingvin. Sue Flood / Getty Images.

14. Här gäller det bara att hitta den variabel för vilken R^2_{adj} förbättras mest när vi tar bort just den variabeln. Men det står ju “The adjusted R^2 of the full model is 0.9.” och i samtliga fall sjunker R^2_{adj} när vi tar bort en förklaringsvariabel. Vi kan inte förbättra modellen genom att ta bort någon variabel från modellen i detta fall.