# Chapter 8 - Introduction to Linear Regression

Gabriel Campos

Nutrition at Starbucks, Part I. (8.22, p. 326)

The scatterplot below shows the relationship between the number of calories and amount of carbohydrates (in grams) Starbucks food menu items contain. Since Starbucks only lists the number of calories on the display items, we are interested in predicting the amount of carbs a menu item has based on its calorie content.

(a)
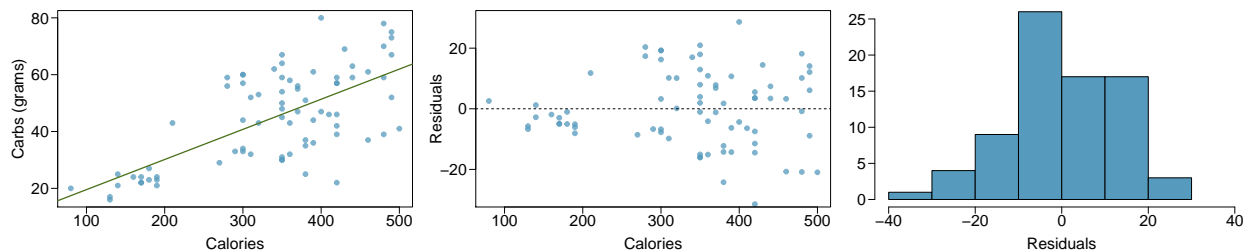
Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.

ANSWER:
The relationship between the number of calories and the amount of carbs in Starbucks food menu items shos a high variability with a linear positive relationship.

(b)

In this scenario, what are the explanatory and response variables?

ANSWER:
$Variable_{explanatory}$:Calories $[X - axis]$

$Variable_{explanatory}$ Carbohydrates (in grams) $[Y - axis]$

(c)

Why might we want to fit a regression line to these data?

ANSWER: To identify linear relationships, which can also assist in identify positive and moderate relationships as well.

(d)

Do these data meet the conditions required for fitting a least squares line? **ANSWER** (i) Residuals: normal & scattered (ii) No constant (iii) Data appears linear

$\therefore$ **I'd say yes**

Body measurements, Part I. (8.13, p. 316)

Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender for 507 physically active individuals. The scatterplot below shows the relationship between height and shoulder girth (over deltoid muscles), both measured in centimeters.

(a)
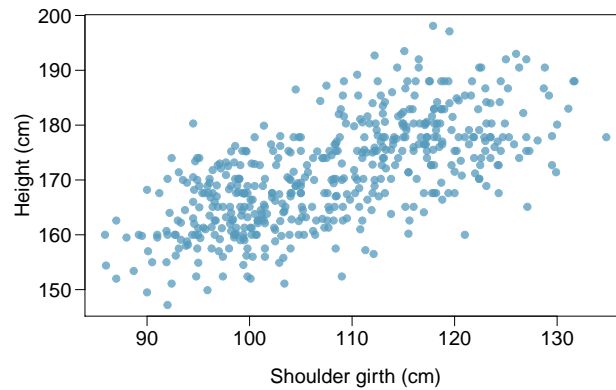
Describe the relationship between shoulder girth and height.

ANSWER:
As height increases so does girth $\therefore$ the relationship is positive.

(b)

How would the relationship change if shoulder girth was measured in inches while the units of height remained in centimeters?

ANSWER:
We would still see a positive relationship, but and larger increase for height, with and larger incremental value over time, making it steeper.

Body measurements, Part III.

(8.24, p. 326) Exercise above introduces data on shoulder girth and height of a group of individuals. The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

(a)

Write the equation of the regression line for predicting height.
ANSWER:

$\hat{hgt} = \beta_0 + \beta_1 \times x_{girth} = 105.9651 + 0.6079749 \times x_{girth}$ **where**

$\beta_0 = \hat{y}_{hgt} - \beta_1 \times \hat{x}_{girth}$ **AND**

$\beta_1 = \frac{S_{hgt}}{S_{girth}} \times R$

```
hgt_mean = 171.14
hgt_sd = 9.41
girth_mean = 107.2
girth_sd = 10.37
R = .67
slope_1 = (hgt_sd/girth_sd)*R
interc = hgt_mean - slope_1 * girth_mean
hgt_est = interc + slope_1 * girth_mean
slope_1
interc
hgt_est
```

(b)

Interpret the slope and the intercept in this context.
ANSWER:
$slope = 0.6079749$ $intercept = 105.9651$
indicating the slope increases a rate that of $\approx .6$ for each event where shoulder girth increases. The intercept is just telling us at what girth we can expect the height to be $0cm$.

(c)

Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.
ANSWER:
$R^2 = .67^2 = 0.4489 = 44.89\%$
So with respect to the relationship between height and shoulder girth, 44.89% of the variance we observe for height can be explained by the linear model.

```
R^2
```

(d)

A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.

$\hat{hgt} = \beta_0 + \beta_1 \times x_{girth} =$ **166.7626**

```
hgt_est_d = interc + slope_1 * 100
hgt_est_d
```

(e)

The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.
ANSWER:
$residual = actual\text{ - }estimate =$ 160 - 166.7626 = -6.762581 $\therefore$ We overstimated by $\approx$ 6.8 cm.

```
160-hgt_est_d
```

(f)

A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?
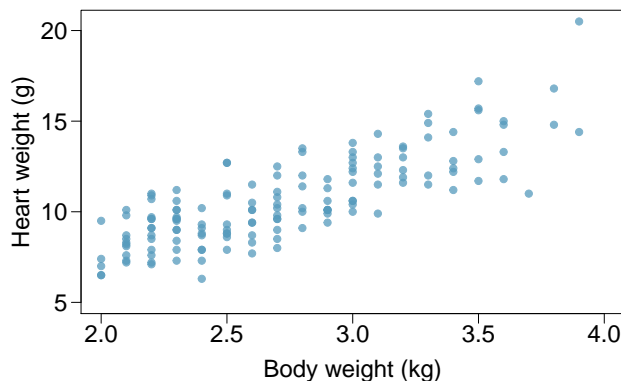**ANSWER:**
Original data set have to include this girth and although not provided in this HW exercise upon searching, the data that has been made available, had a range from 80cm and up, $\therefore$ I would most likely say no.

Cats, Part I.(8.26, p. 327)

The following regression output is for predicting the heart weight (in g) of cats from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cats.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.357	0.692	-0.515	0.607
body wt	4.034	0.250	16.119	0.000

$$s = 1.452 \qquad R^2 = 64.66\% \qquad R^2_{adj} = 64.41\%$$

(a)

Write out the linear model.
ANSWER:
β_0= -0.357 & β_1= 4.034
$\therefore \hat{y}=-0.357 + 4.034 \times x$

(b)

Interpret the intercept.
ANSWER:
Basically, as the cats heart weight will be in the negatives when its body weight is at 0kg. Although it is literally impossible for this to be true, its is a conclusion made if the trend hypothetically followed the linear model precisely.

(c)

Interpret the slope.
ANSWER:
The heart rate grows by 4.034 kg for every 1kg it grows in body weight is what the slope represents.

(d)

Interpret R^2.
ANSWER:
64.66% of the heart rate variation is captured or described in this linear model.

(e)

Calculate the correlation coefficient.
ANSWER:
Correlation Coefficient $= \sqrt{(R^2)}$ or $\sqrt{(.6466^2)}$ - 0.8041144

```
sqrt(.6466)
```

Rate my professor. (8.44, p. 340)

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. Researchers at University of Texas, Austin collected data on teaching evaluation score (higher score means better) and standardized beauty score (a score of 0 means average, negative score means below average, and a positive score means above average) for a sample of 463 professors. The scatterplot below shows the relationship between these variables, and also provided is a regression output for predicting teaching evaluation score from beauty score.

	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	4.010	0.0255	157.21	0.0000
beauty		0.0322	4.13	0.0000
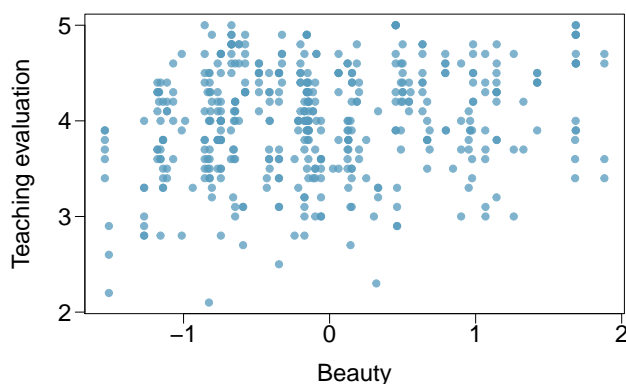
(a)

Given that the average standardized beauty score is -0.0883 and average teaching evaluation score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the information provided in the model summary table.

ANSWER:

$\hat{y} = \beta_0 + \beta_1 \times x$ or $\beta_1 = \frac{y - \beta_0}{x}$

$\therefore \frac{3.9983 - 4.010}{-0.0883} = 0.1325028$

```
(3.9983-4.010)/(-0.0883)
```

(b)

Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.
ANSWER:
Positive slope with a p-value close to 0, $\therefore$ I would say yes.

(c)

List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots.
ANSWER:
i. Linearity **ii.** Nearly Normal residual **iii.** Constant Variability **iv.** Independent Observations **i.** Scatterplot shows a linear relationship, so despite the lack of R & R^2 id say yes its satisfied. **ii.** 'Residuals' appears to be nearly normal. **iii.** 'Residuals' and scatterplot show variability. **iv.*** I'll assume independence unless evidence provided that Professors were sway results, one student filled out several ratings under false name, or some other factor that proves otherwise.