MATH 4740/MSSC 5740
Chapter 9 Problem Solving # *, Example 9.7, Example 9.9

9.6 Summary

Correlation Coefficient: r	$$r = \frac{\text{cov}(x,y)}{\sqrt{s_x^2 s_y^2}}$$ $$\text{cov}(x,y) = \frac{1}{n-1}\left[\sum XY - \frac{1}{n}\left(\sum Y\right)\left(\sum X\right)\right]$$ $$s_x^2 = \frac{1}{n-1}\left[\sum X^2 - \frac{1}{n}\left(\sum X\right)^2\right], \quad s_y^2 = \frac{1}{n-1}\left[\sum Y^2 - \frac{1}{n}\left(\sum Y\right)^2\right]$$
Linear Regression: $\hat{y} = b_0 + b_1 x$	$$b_1 = r\frac{s_y}{s_x}, \qquad b_0 = \bar{Y} - b_1\bar{X}$$
Logistic Regression: $\hat{p} = \frac{1}{1+e^{-b_0 - b_1 x_1 - \ldots - b_p x_p}}$	$$ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = b_0 + b_1 x_1 + \ldots + b_p x_p$$ $$\hat{OR} = e^{\hat{\beta}_1\Delta_1 + \ldots + \hat{\beta}_p\Delta_p}$$

MATH 4740/MSSC 5740
Chapter 9 Problem Solving # *, Example 9.7, Example 9.9

9.6 Practice Problems

* Given (x,y) points $(1,2),(2,1),(3,4),(4,3)$,

a) Plot the points.

b) Find r, b_0 and b_1 by hand with sums.

c) Draw the fitted regression line on the same graph as points.

d) What do b_0 and b_1 mean?

MATH 4740/MSSC 5740
Chapter 9 Problem Solving # *, Example 9.7, Example 9.9

Example 9.7 (page 216)

An observational study is conducted to investigate risk factors associated with infant weight. The study involves $n=832$ pregnant women. Investigators wish to determine whether there are any differences in birth weight by infant sex, gestational age, mothers age, and mother's race/ethnicity. A multiple regression analysis is performed.

The model is:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6$$

$H_0: \beta_j=0$ vs. $H_0: \beta_j\neq 0$

y =birth weight (grams)
x_1=Infant sex (1=male, 0=female),
x_2=Gestational age (in weeks),
x_3=Mothers age (in years), and
x_4=Black race/ethnicity (1=yes, 0=no)
x_5=Hispanic race/ethnicity (1=yes, 0=no)
x_6=Other race/ethnicity (1=yes, 0=no)

$$t_j = \frac{b_j - 0}{\sqrt{\text{var}(b_j)}}$$

$df=n-p-1$
$t_{\alpha/2}=t_{0.025,825}=1.96$

	Independent Variable	Regression Coefficient	t	p-value
b_0	Intercept	−3850.92	−11.56	0.0001
b_1	Male infant	174.79	6.06	0.0001
b_2	Gestational age (weeks)	179.89	22.35	0.0001
b_3	Mother's age (years)	1.38	0.47	0.6361
b_4	Black race/ ethnicity	−138.46	−1.93	0.0535
b_5	Hispanic race/ ethnicity	−13.07	−0.37	0.7103
b_6	Other race/ ethnicity	−68.67	−1.05	0.2918

With b_0, b_1, b_2, b_3, b_4, b_5, b_6 calculated using a software program such as ***R***.

$$\hat{y} = -3850.92 + 174.79x_1 + 179.89x_2 + 1.38x_3 - 138.46x_4 - 13.07x_6 - 68.67x_6$$

a) What does b_0=-3850.92 mean?

b) What does b_2=179.89 mean?

c) Keeping x_1, x_2, x_3, fixed, what does a change from x_4=1, x_5=0, x_6=0 to x_4=0, x_5=1, x_6=0 mean?

d) Which variables are important (β_j coefficient statistically significant from 0)?

MATH 4740/MSSC 5740
Chapter 9 Problem Solving # *, Example 9.7, Example 9.9

Example 9.8

Assume that a logistic regression model were fit to relate obesity to probability of CVD

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -2.367 + 0.658(Obesity)$$

where 1=obese and 0=not obsess.

a) What does b_1=0.685 mean?

b) To look at statistical significance,

Independent Variable	Regression Coefficient	χ^2	*p*-value
Intercept	–2.367	307.38	0.0001
Obesity	0.658	9.87	0.0017

H_0: β_j=0 vs. H_0: $\beta_j \neq 0$

df=1

Assume that age group is added to the model.

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -2.367 + 0.415(Obesity) + 0.655(Age\ Group)$$

c) What does b_1=0.415 mean?