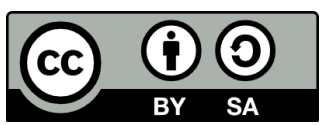
Applied Statistics

Correlation and Linear Regression

Cesar O. Aguilar
SUNY Geneseo

Portions of these notes were created from *Learning statistics with R* by Danielle Navarro, *Learning statistics with jamovi* by David Foxcroft, and *Introduction to Statistical Thinking* by Benjamin Yakir.



Chapter Goal

- The goal in this chapter is to introduce correlation and linear regression
- These are the standard tools that statisticians rely on when analysing the **relationship** between continuous predictors and continuous outcomes

A Motivating Example: Parenthood

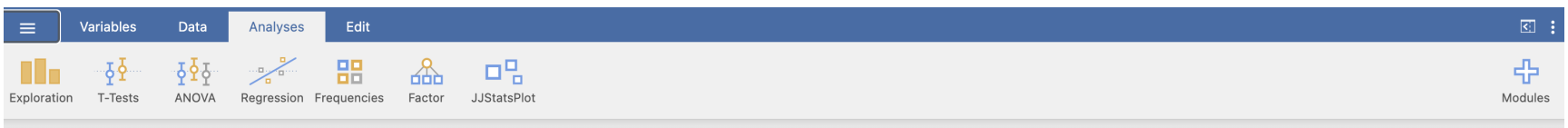
- Suppose that my name is Daniel and I'm a sleep-deprived parent
- Suppose that I'm curious to find out how much my infant son's sleeping habits affect my mood (I don't have an infant son, but let's pretend that I do)
- Let's say that I can rate my grumpiness very precisely, on a scale from 0 (not at all grumpy) to 100 (very very grumpy)
- And lets also assume that I've been measuring my grumpiness, my sleeping patterns and my son's sleeping patterns for the last 100 days
- I've recorded the data and saved it in the file `parenthood.csv`
- The variables in the data set are:
  - **dani.sleep** - the number of hours of sleep that I got
  - **baby.sleep** - the number of hours of sleep that my son got
  - **dani.grump** - my grumpiness rating (0-100)
  - **day** (the day of the observation)
  - **ID** (the ID of the observation)

A Motivating Example: Parenthood

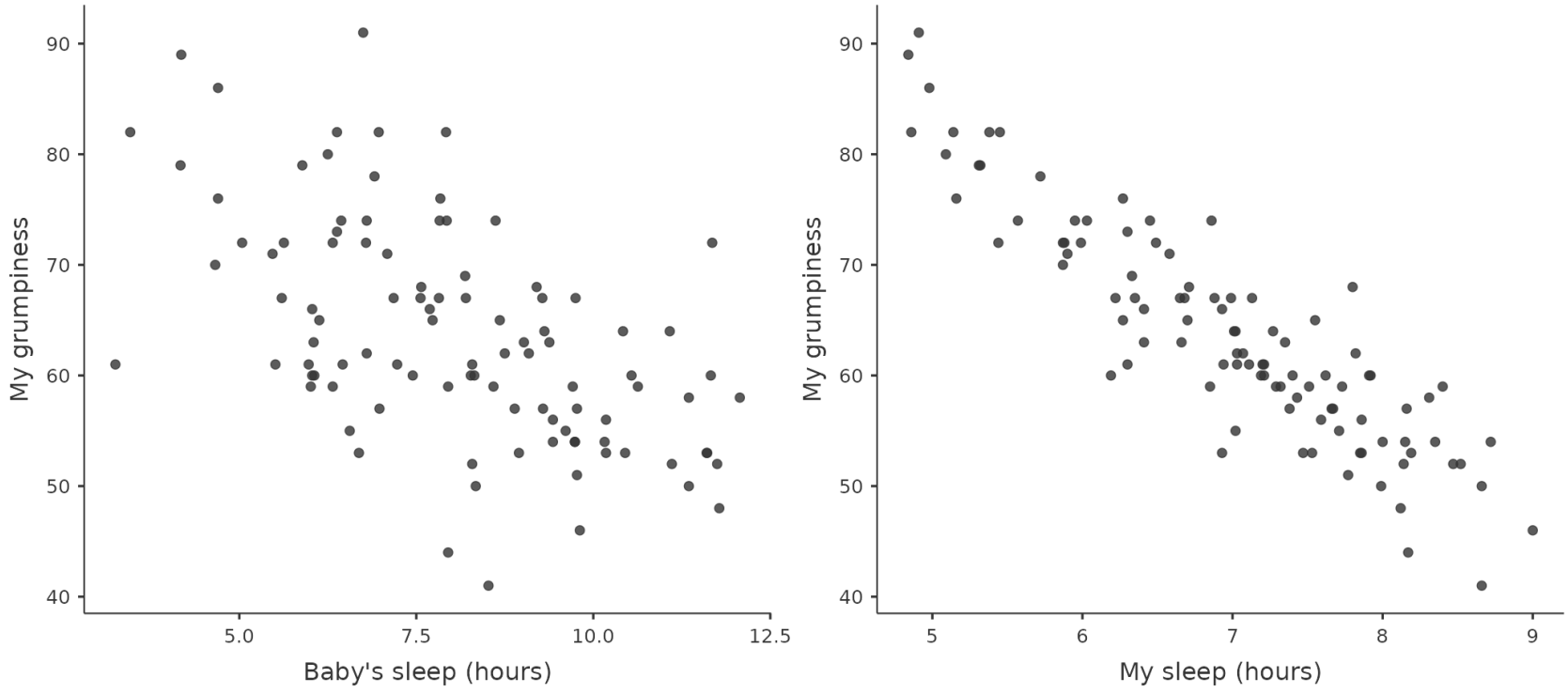
- Although I'm interested in finding out how my son's sleeping habits affect my mood, I'm also interested in finding out how my own sleeping habits affect my mood
- However, my mood may also affect my sleeping habits
- Thus, it is not always clear which variable is the true cause and which is the effect
- But I'm interested in how my son's sleep (or mine) affects my mood, not the other way around
- I would then call my son's sleep (or mine) the **predictor** variable and my mood the **outcome** variable
- A predictor variable is also called an **independent** variable and an outcome variable is also called a **dependent** variable
- Predictor variables are usually denoted by X and outcome variables by Y

A Motivating Example: Parenthood

- A quick way to learn about the relationship between two variables is to plot them
- A **scatterplot** is a plot of the outcome variable on the y-axis (vertical axis) and the predictor variable on the x-axis (horizontal axis)
- In `jamovi`, there is a module called `scatr` that can be used to create scatterplots
- In the **Analyses** menu, you can install a module by clicking on the large + sign

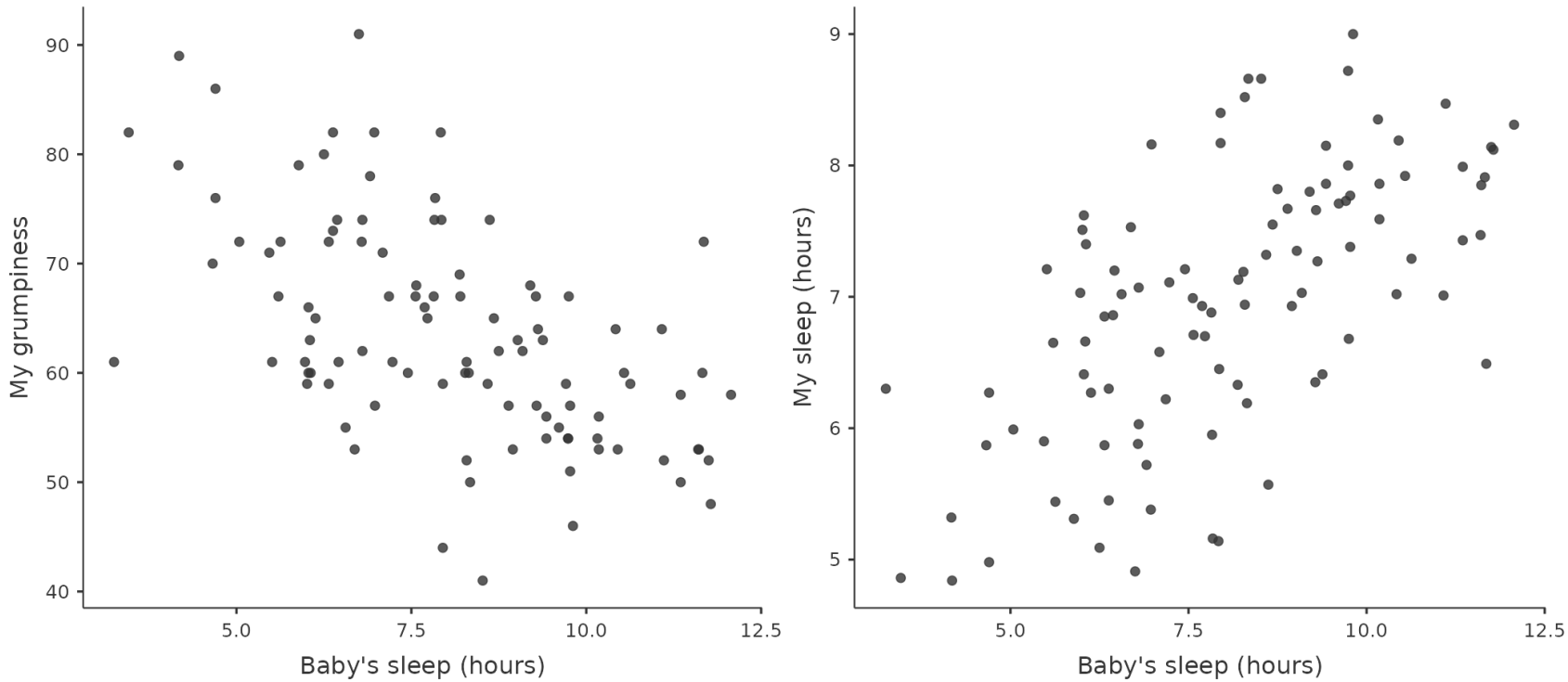


A Motivating Example: Parenthood



- In both plots, we observe that more sleep equals less grumpiness
- The relationship between **dani.sleep** and **dani.grump** is stronger than the relationship between **baby.sleep** and **dani.grump**

A Motivating Example: Parenthood

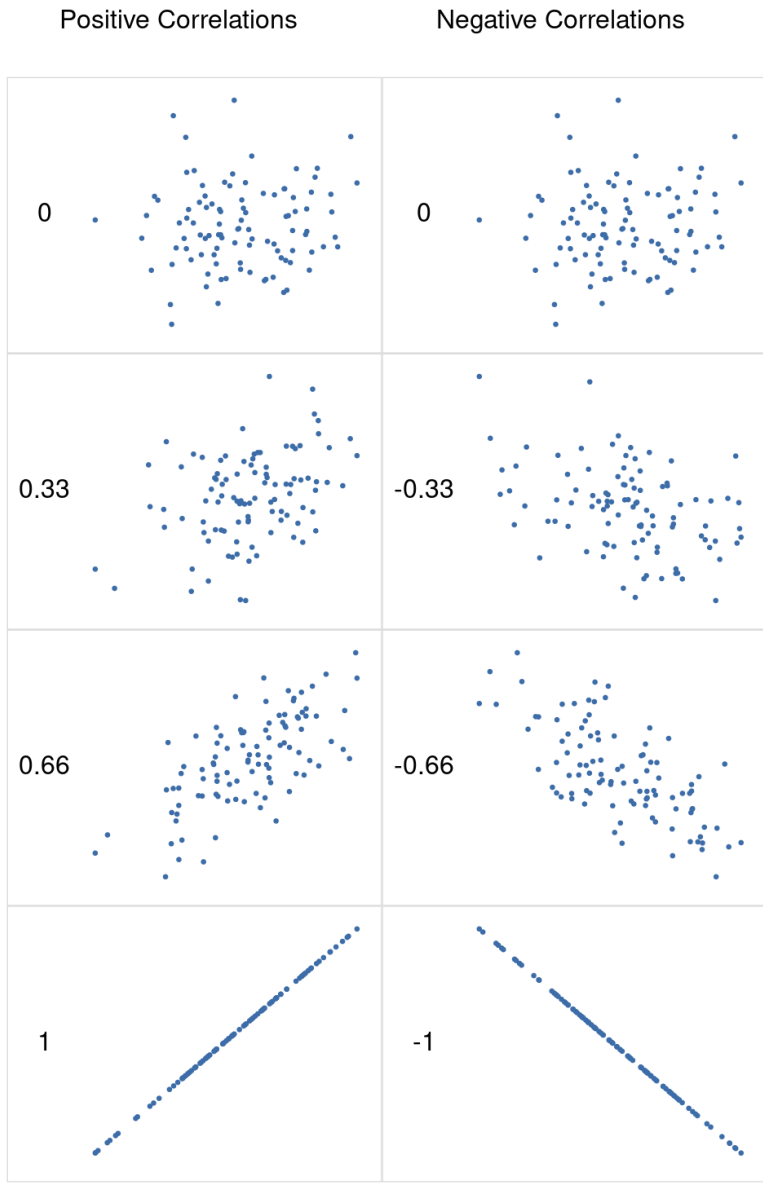


- The overall strength of the relationships "**baby.sleep** vs **dani.grump**" and "**baby.sleep** vs **dani.sleep**" are the same but the **directions** are different
- The relationship on the left is said to be a **negative** relationship and the relationship on the right is said to be a **positive** relationship

Correlation Coefficient

- A quantitative measure of the strength of the relationship between two variables is **Pearson's correlation coefficient** or just the **correlation coefficient**, denoted by r
- r varies between -1 and 1
- When $r = -1$ the relationship is said to be a **perfect negative** relationship
- When $r = 1$ the relationship is said to be a **perfect positive** relationship
- When $r = 0$ we say that the variables are **uncorrelated** or that there is no relationship between the variables

Correlation Coefficient



Correlation Coefficient

- To compute the correlation coefficient we first introduce the **covariance** between two variables X and Y
- The covariance between X and Y is denoted by $cov(X, Y)$ and is defined as

$$cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

- The covariance between X and Y is a measure of the joint variability of the two variables
- If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values (that is, the variables tend to show similar behavior), the covariance is positive
- In the opposite case, when the greater values of one variable mainly correspond to the lesser values of the other, (that is, the variables tend to show opposite behavior), the covariance is negative

Correlation Coefficient

- The covariance between two variables X and Y is a generalisation of the notion of the variance of one random variable
- Notice that

$$var(X) = cov(X, X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

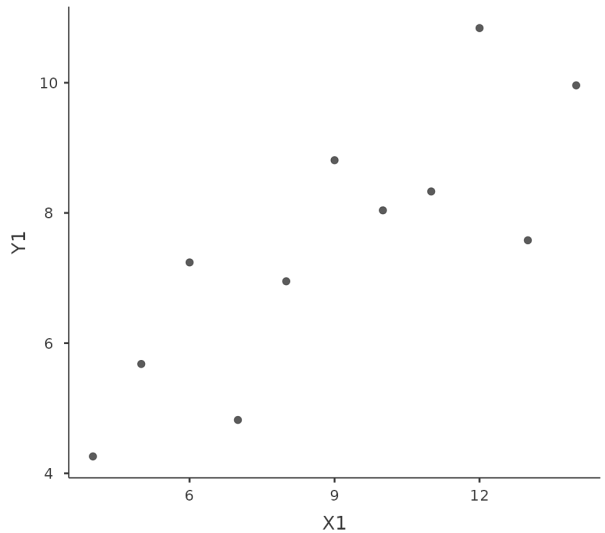
- By itself, the covariance is difficult to interpret since the units of $cov(X, Y)$ are the product of the units of X and Y
- To get a more useful standardized measure, we divide the covariance by the **standard deviations** of X and Y to get the correlation coefficient

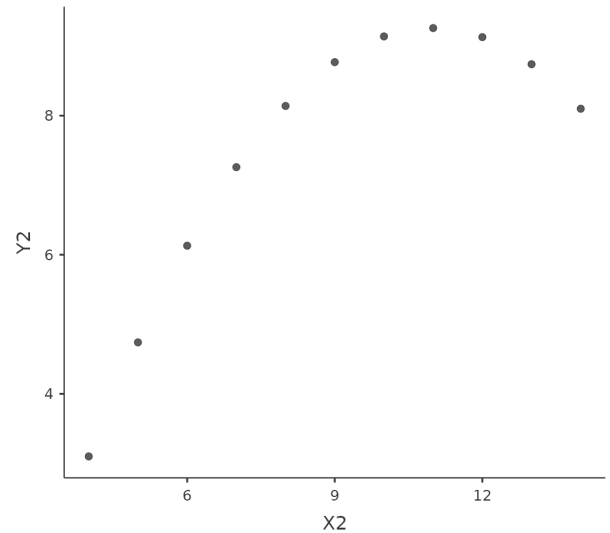
$$r = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

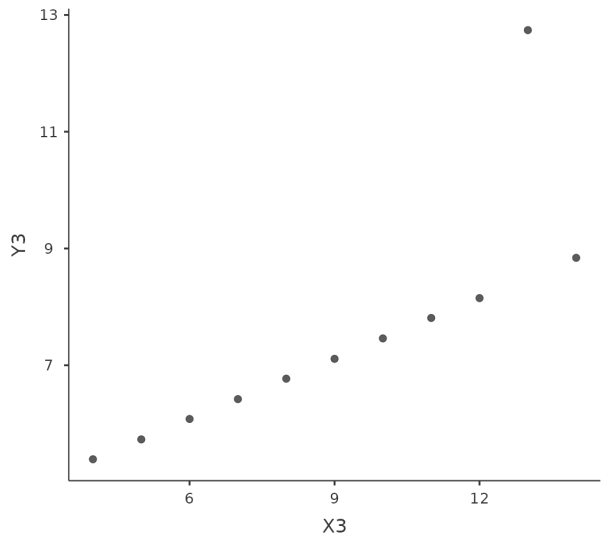
- In `jamovi`, use the **Correlation Matrix** in the **Regression** submenu under the **Analyses** menu to compute the correlation coefficient between any number of variables

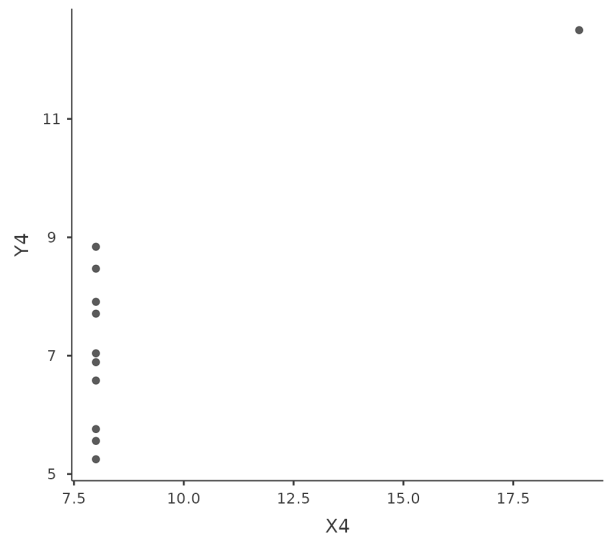
Anscombe's Quartet Scatterplots: $r = 0.816$

Francis Anscombe (1973)









Linear Regression Models

- Aside from the correlation coefficient, another way to quantify the relationship between variables is to fit a **linear regression model**
- A **linear regression model** between variables is a model that assumes that the relationship between the variables is linear
- The basic idea on how to build a linear regression model between X and Y is to find the line that best fits the data
- Recall that the equation of a line is given by

$$Y = \beta_0 + \beta_1 X$$

  where β_1 is the slope of the line and β_0 is the y-intercept of the line
- When $\beta_1 < 0$ the line has a negative slope and for every unit increase in X the line decreases by β_1 units
- When $\beta_1 > 0$ the line has a positive slope and for every unit increase in X the line increases by β_1 units

Linear Regression Models

- To find the line $Y = \beta_0 + \beta_1 X$ of best fit, we proceed as follows
- We are given the observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$
- Whatever value of β_0 and β_1 we choose, we can compute the predicted value $\hat{Y}_i$ for each observation X_i as

$$\hat{Y}_i = \beta_0 + \beta_1 X_i$$

- The **error** or **residual** between the observed value Y_i and the predicted value $\hat{Y}_i$ is given by

$$e_i = Y_i - \hat{Y}_i$$

- We want to find β_0 and β_1 such that the sum of the squared errors is minimized

$$SS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Linear Regression Models

- Using some calculus, it can be shown that the values of β_0 and β_1 that minimize SS are given by

$$\hat{\beta}_1 = \frac{cov(x, y)}{var(x)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Later we'll talk about why we are using hat notation for $\hat{\beta}_0$ and $\hat{\beta}_1$
- Recall that $r = \frac{cov(x,y)}{s_x s_y}$ and thus if r is known then

$$cov(x, y) = r \cdot \sigma_x \cdot \sigma_y$$

- In practice, we will use the sample standard deviations s_x and s_y to estimate the population standard deviations σ_x and σ_y

Linear Regression Models: Parenthood Example

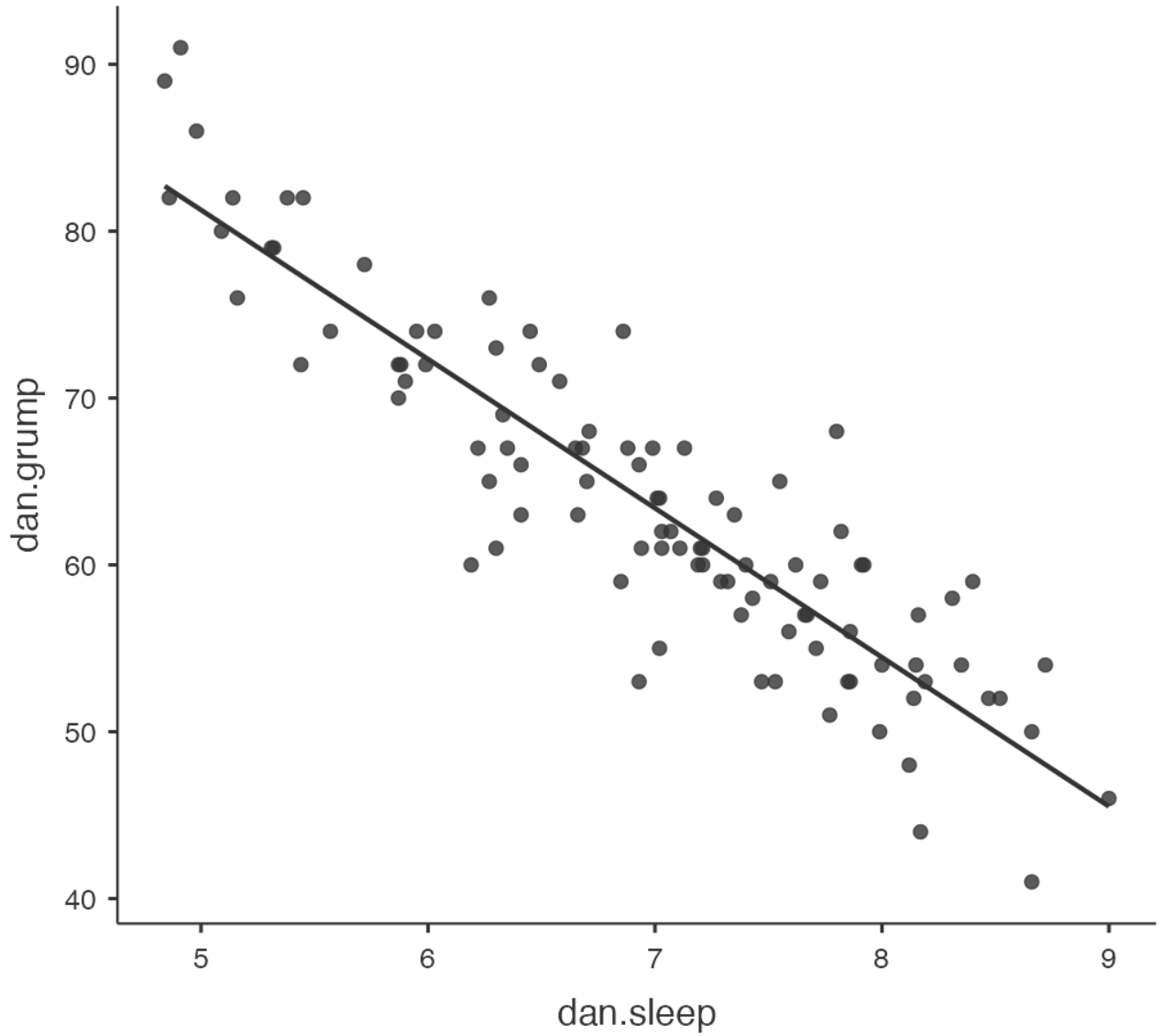
- To compute and graph a regression line for a data set, use the **Regression** submenu in the **Analyses** menu in `jamovi`
- Using **dan.grump** as the outcome variable and **dan.sleep** as the predictor variable (covariate) we find that

$$\hat{Y} = \beta_0 + \beta_1 X$$

$$= 125.96 - 8.94x$$

Linear Regression Models: Parenthood Example

$\hat{Y} = 125.96 - 8.94X$



Multiple Linear Regression Models

- Let's go back to the example of the **parenthood** data set
- My mood may be affected both by the amount of sleep that I get and the amount of my son's sleep
- More generally, the outcome variable Y may be affected by several predictor variables $X_1, X_2, \ldots, X_k$
- In this case we can use a **multiple linear regression model** to model the relationship between the outcome variable Y and the predictor variables $X_1, X_2, \ldots, X_k$
- For example, if X_1 denotes the amount of hours I sleep, X_2 the amount of hours my son sleeps, and Y denotes my grumpiness rating, the multiple linear regression model is given by

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- On the ith day, my grumpiness is predicted to be

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

Multiple Linear Regression Models

- As before, the residual associated with the actual ith observation Y_i is

$$e_i = Y_i - \hat{Y}_i$$

- And the sum of the squared errors is

$$SS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

- In the parenthood model, we need to find the values of $\beta_0, \beta_1, \beta_2$ that minimize the sum of the squared errors
- Multiple linear regression is conceptually the same as simple linear regression but we we just have more variables
- Multiple regression in `jamovi` is no different to simple regression; all we have to do is add additional variables to the **Covariates** box

Quantifying the Fit of the Regression Model: R^2

- We now need to quantify how good the regression model is
- This is done using the **coefficient of determination** R^2 given by

$$R^2 = \frac{SS_{tot} - SS}{SS_{tot}} = 1 - \frac{SS}{SS_{tot}}$$

- Where SS is the sum of the squared residuals and SS_{tot} is the total variation in the outcome variable Y:

$$SS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$SS_{tot} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

- R^2 measures the proportion of the variance in Y that is predicted by the regression model

Quantifying the Fit of the Regression Model: R^2

- From

$$R^2 = 1 - \frac{SS}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

  we can make the following observations

- A baseline model that always predicts the mean of Y, that is $\hat{Y}_i = \bar{Y}$, has $SS = SS_{tot}$ and then $R^2 = 0$
- If the regression model is a perfect fit, meaning that $\hat{Y}_i = Y_i$, then $SS = 0$ and $R^2 = 1$
- The closer R^2 is to 1, the better the regression model fits the data
- When there is only one predictor variable, $R^2 = r^2$ where r is the correlation coefficient between the predictor variable and the outcome variable

Significance of the Correlation Coefficient

- Recall that the correlation coefficient r is a measure of the strength and direction of the linear relationship between two variables
- We can perform a hypothesis test of the "significance of the correlation coefficient" to decide whether the linear relationship in the sample data is strong enough to use to model the relationship in the population
- Denote by ρ the true correlation coefficient in the population, which is unknown
- The null hypothesis is that the correlation coefficient in the population is zero, that is, there is no **significant** linear relationship between the two variables in the population:

$$H_0 : \rho = 0$$

- The alternative hypothesis is that the correlation coefficient in the population is not zero:

$$H_1 : \rho \neq 0$$

Significance of the Correlation Coefficient

- We perform a two-sided test using the test statistics is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

- This test statistic follows a t-distribution with $n-2$ degrees of freedom for simple linear regression
- If we reject the null hypothesis that $\rho = 0$ we conclude that there is a significant linear relationship between the two variables in the population
- If we fail to reject the null hypothesis that $\rho = 0$ we conclude that there is a significant linear relationship between the two variables in the population

Regression Analysis

- We now want to answer the following type of questions:
  - How good are the estimated values of β_0 and β_1 in a linear regression model $y = \beta_0 + \beta_1 x$?
  - How good is the estimate $\hat{y}$ of y using a linear regression model?
- The values of β_0 and β_1 are obtained using sample data and as such they are random variables
- In other words, if we were to take another sample, we would get different values of β_0 and β_1
- Similarly when estimating y using $\hat{y} = \beta_0 + \beta_1 x$, we get a different value of $\hat{y}$ for each sample
- To distinguish between the true values of β_0 and β_1 and the estimated values, we denote the true values by β_0 and β_1 and the estimated values by $\hat{\beta}_0$ and $\hat{\beta}_1$
- Similarly, we denote the estimated value of y by $\hat{y}$:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

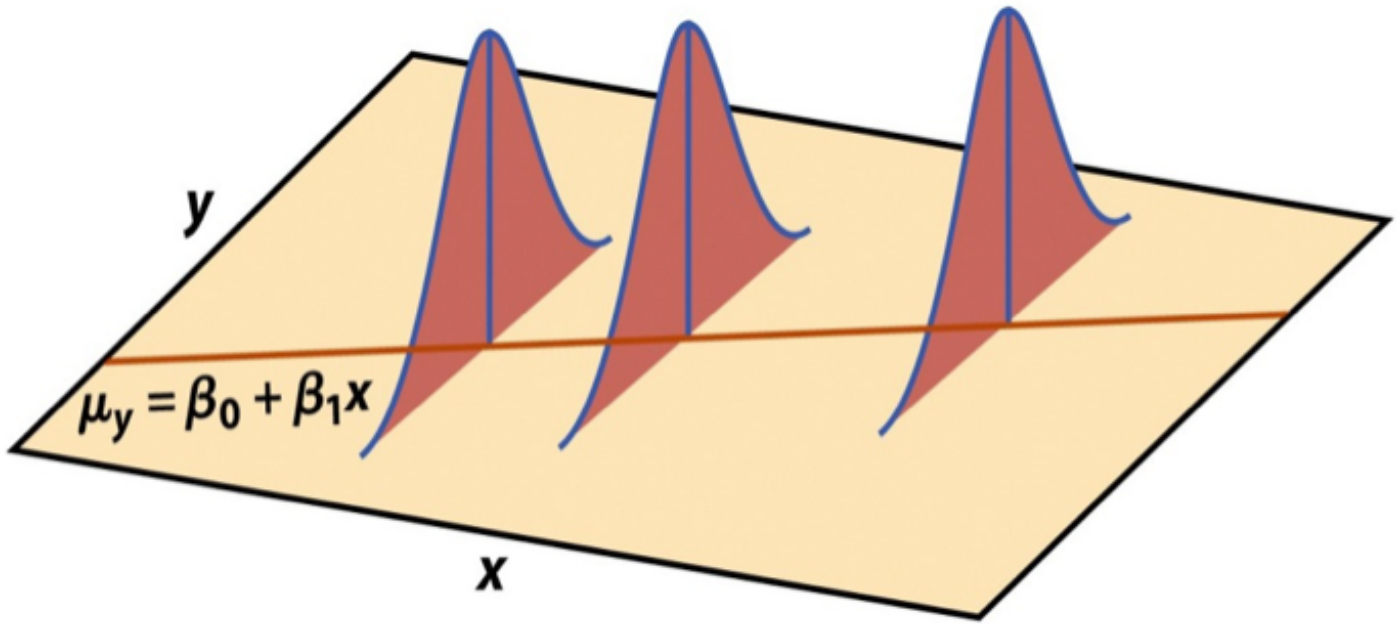
Regression Analysis: The True Regression Line

- When we first introduced the linear regression problem, we casted it as a problem of finding the line that best fits the data
- However, the idea of linear regression is that for each fixed value of x we have a random variable Y that follows some probability distribution and that the **mean** of Y is given by

$$\mu(Y) = \beta_0 + \beta_1 x$$

- For example, if x denotes the square footage of a house and Y denotes the sell price of the house, then $\mu(Y)$ is the average sell price of a house with square footage x
- We do not expect that **every** house with square footage x will sell for the same price, that is, there will be some variability in the sell price of houses with square footage exactly x

Regression Analysis: The True Regression Line



Source: https://www2.stat.duke.edu/courses/Fall19/sta210.001/

- The mean of Y for each value of x is given by the true regression line
- We assume that each distribution of Y has the same standard deviation σ
- When finding the line that best fits the data, we find estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 and $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is the **estimated** regression line

Regression Analysis: Coefficients

- One can prove that the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ obtained from minimizing SS are unbiased estimators of β_0 and β_1, respectively
- Assuming that each Y has normal distribution with standard deviation σ, the standard deviations of $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$\sigma_{\hat{\beta}_0} = \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \quad \text{and} \quad \sigma_{\hat{\beta}_1} = \frac{\sigma}{\sqrt{S_{xx}}}$$

- Where

$$S_{xx} = (n-1)s_x^2 = (n-1)var(x)$$

Regression Analysis: Coefficients

- In practice, we will not know the true value of σ and we will have to estimate it using an ubiased estimatore
- In this case, this estimator is given by

$$s_e = \sqrt{\frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}}$$

  and called the **standard error of the estimate**

- Where

$$S_{yy} = (n-1)var(y)$$

$$S_{xy} = (n-1)cov(x,y)$$

- As before if r is known then

$$cov(x,y) = r \cdot \sigma_x \cdot \sigma_y$$

Regression Analysis: Coefficients

- Now that we have the standard deviations of $\hat{\beta}_0$ and $\hat{\beta}_1$, we now need a statistic to make inferences on β_0 and β_1
- To make statistical inferences on β_0 we use the following statistics:

$$t = \frac{\hat{\beta}_0 - \beta_0}{\sigma_{\hat{\beta}_0}} = \frac{\hat{\beta}_0 - \beta_0}{s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}$$

- And for β_1 we use

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = \frac{\hat{\beta}_1 - \beta_1}{(s_e/\sqrt{S_{xx}})}$$

- Both of these statistics follow a t-distribution with $n - 2$ degrees of freedom
- In practice, the intercept β_0 is not of much interest, especially if $x = 0$ is not a meaningful value
- We are more interested in making inferences on β_1

Confidence Intervals for the Coefficients

- Recall that the general form of a confidence interval using the t-distribution is

$$\hat{\theta} \pm t_{\alpha/2} \cdot SE(\hat{\theta})$$

- Therefore, the confidence interval for β_1 is given by

$$\hat{\beta}_1 \pm t_{\alpha/2} \cdot \sigma_{\hat{\beta}_1}$$

or equivalently

$$\hat{\beta}_1 \pm t_{\alpha/2} \cdot \frac{s_e}{\sqrt{S_{xx}}}$$

- And the confidence interval for β_0 is given by

$$\hat{\beta}_0 \pm t_{\alpha/2} \cdot \sigma_{\hat{\beta}_0}$$

or equivalently

$$\hat{\beta}_0 \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

Example: Inferences on the Regression Coefficients

Example. For the data given below, find the regression line $y = \beta_0 + \beta_x$ and find a 95% confidence interval for the slope of the line. Then test the hypothesis that $\beta_1 = 0$ versus $\beta_1 \neq 0$.

$x = (47, 56, 116, 178, 19, 75, 160, 31, 12, 164, 43, 74)$

$y = (15.1, 14.1, 13.2, 12.7, 14.6, 13.8, 11.9, 14.8, 15.3, 12.6, 14.7, 14.0)$

- Using `jamovi` we obtain the following:

  $\bar{x} = 81.3, \quad \bar{y} = 13.9, \quad var(x) = 3471, \quad var(y) = 1.18, \quad r = -0.948$

- From these we can compute

  $$cov(x, y) = -60.67 \quad S_{XX} = 38181 \quad S_{YY} = 13 \quad S_{XY} = -667$$

- The coefficients are

  $$\hat{\beta}_1 = -0.0175 \quad \text{and} \quad \hat{\beta}_0 = 15.3$$

Example: Inferences on the Regression Coefficients

- To compute the standard error of $\hat{\beta}_1$ we need to compute s_e:

$$s_e = \sqrt{\frac{S_{YY} - \hat{\beta}_1 S_{XY}}{n-2}} = 0.36$$

- Then

$$\sigma_{\hat{\beta}_1} = \frac{s_e}{\sqrt{S_{XX}}} = 0.00186$$

- The critical value for a 95% confidence interval is $t_{\alpha/2} = 2.228$
- The confidence interval is:

$$\hat{\beta}_1 \pm t_{\alpha/2} \cdot \sigma_{\hat{\beta}_1} = -0.0175 \pm 2.228 \cdot 0.00186 = (-0.022, -0.013)$$

- Compare this with the answer obtained by `jamovi`

Example: Inferences on the Regression Coefficients

- We now test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

- The test statistic is

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = \frac{-0.0175 - 0}{0.00186} = -9.4$$

- Using the t-distribution with $n - 2 = 10$ degrees of freedom, the p-value is too small to report

- However, $t = -9.4$ lies outside the rejection region $\{t < 2.228 \quad \text{or} \quad t > 2.228\}$ and thus we reject the null hypothesis that $\beta_1 = 0$

- In other words, there is statistical evidence supporting that x and y are linearly related

Predictions for The Mean Response

- We now consider the problem of **estimating** the mean value of y for a given value of x_0
- The **estimated mean outcome** is given by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

- The $(1-\alpha)100\%$ confidence interval for the mean outcome is

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

- Where as before, $s_e = \sqrt{\frac{S_{YY} - \hat{\beta}_1 S_{XY}}{n-2}}$
- And $t_{\alpha/2}$ is the critical value for the t-distribution with $n-2$ degrees of freedom

Predictions for New Observations

- We now consider the problem of **predicting** the response y for a new observation with $x = x_0$
- This is different than estimating the mean response for a given value of x_0
- In this case, the $(1 - \alpha)100\%$ prediction interval is

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

- Notice that the only difference between the prediction interval and the confidence interval is the extra term 1 under the square root
- This term accounts for the uncertainty in the estimate of the mean response and the uncertainty in the prediction of a new observation
- Not surprisingly, the prediction interval is wider than the confidence interval

Example: Predictions for The Mean Response

Example. For the data given below, find the regression line $y = \beta_0 + \beta_1 x$.

- Find a 95% confidence interval for the mean response of y when $x = 1$.
- Find a 95% prediction interval for a new observation with $x = 2$.

$$x = (-2, -1, 0, 1, 2)$$
$$y = (0, 0, 1, 1, 3)$$

- We compute that $\bar{x} = 0$ and $\bar{y} = 1$
- Then $S_{xx} = 10$, $S_{yy} = 6$, and $S_{xy} = 7$
- Then

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{7}{10} = 0.7 \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1 - 0 = 1$$

- The regression line is $y = 1 + 0.7x$

Example: Predictions for The Mean Response

- Recall that the $(1-\alpha)100\%$ confidence interval for the mean response is

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

- We compute that

$$s_e = \sqrt{\frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}} = \sqrt{\frac{6 - 0.7 \cdot 7}{5-2}} = 0.6055$$

- And

$$\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = \sqrt{\frac{1}{5} + \frac{(1-0)^2}{10}} = 0.5477$$

- The critical value for the t-distribution with $n-2=3$ degrees of freedom and $\alpha/2 = 0.025$ is $t_{\alpha/2} = 3.182$

Example: Predictions for The Mean Response

- The 95% confidence interval for the mean response is then

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = 1.7 \pm 3.182 \cdot 0.6055 \cdot 0.5477$$

$$= (0.645, 2.755)$$

- For the prediction interval, we need to compute

$$\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = \sqrt{1 + \frac{1}{5} + \frac{(2-0)^2}{10}} = 1.2649$$

- In this case, $\hat{y} = \beta_0 + \beta_1 \times 2 = 2.4$

- The CI for the predicted response is therefore

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = 2.4 \pm 3.182 \cdot 0.6055 \cdot 1.2649$$

$$= (-0.7372, 4.1372)$$

- The intervals are pretty wide but only five data points were used

Summary of Formulas

$$\hat{\beta}_1 = \frac{cov(x,y)}{var(x)} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$S_{xx} = (n-1)var(x)$$

$$S_{yy} = (n-1)var(y)$$

$$S_{xy} = (n-1)cov(x,y)$$

$$cov(x,y) = r \cdot s_x \cdot s_y$$

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}}$$

$$t = \frac{\hat{\beta}_0 - \beta_0}{\sigma_{\hat{\beta}_0}}$$

$$\sigma_{\hat{\beta}_1} = \frac{s_e}{\sqrt{S_{xx}}}$$

$$\sigma_{\hat{\beta}_0} = s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

$$s_e = \sqrt{\frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}}$$