Module 3 - SLR models

STAT 401

Section 1: SLR models

For n pairs of observations, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ a simple linear regression model is defined as

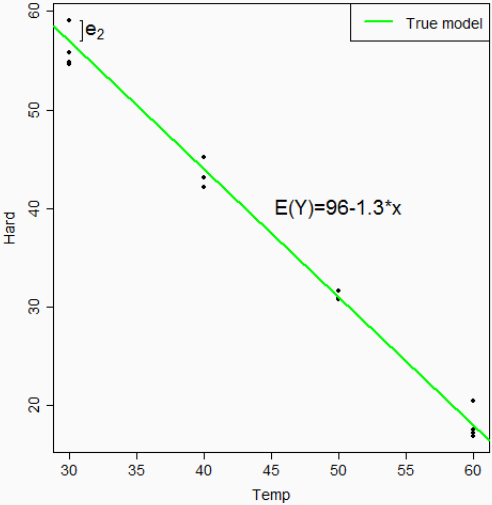
$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

where

- $E(e_i) = 0$ or $E(Y_i) = \beta_0 + \beta_1 x_i$ for all i
- $V(Y_i) = V(e_i) = \sigma^2$ for all i
- e_i and e_j are independent for all $i \neq j$

For example, consider the predictor "quench bath temperature" and the response "coil spring hardness" where 14 observations were made:

$$
\begin{aligned}
55.8 &= \beta_0 + \beta_1 * (30) + e_1 \\
59.1 &= \beta_0 + \beta_1 * (30) + e_2 \\
\vdots &= \vdots \\
16.9 &= \beta_0 + \beta_1 * (60) + e_{14}
\end{aligned}
$$

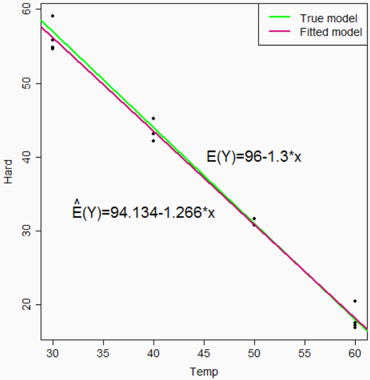


True model
e_2
E(Y)=96-1.3*x
Hard
Temp
60
50
40
30
20
30
35
40
45
50
55
60

We interpret $\beta_0 = 96$ as the mean (or expected) response when $x = 0$.

We interpret $\beta_1 = -1.3$ as the change in mean response when x increases by 1 unit.

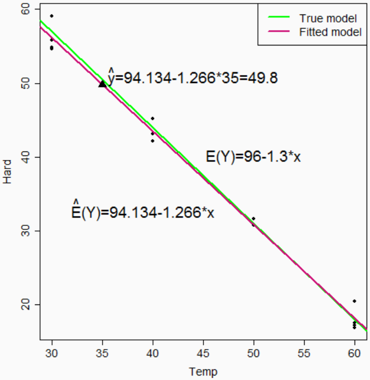
We must estimate β_0 and β_1 from the data. This estimation is known as *fitting the model*. We obtain the estimates: $\hat{\beta}_0 = 94.134$ and $\hat{\beta}_1 = -1.266$.



For a given x, the estimated expected value of Y is

$$\hat{E}(Y) = \hat{y} = 94.134 - 1.266 * x$$

If $x = 35$:

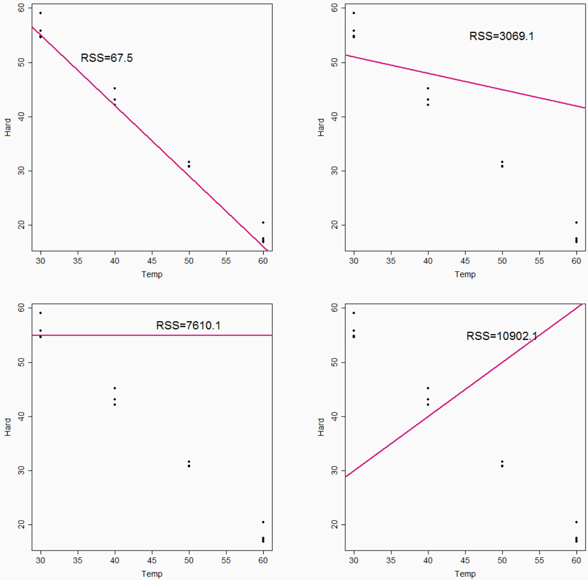


Section 2: Fitting SLR models

There are three parameters in the typical simple linear regression model: $\beta_0, \beta_1,$ and σ^2. Several reasonable estimation procedures have been proposed for estimating these. Most common is *Ordinary Least Squares* or OLS. We begin by considering the function

$$RSS(b_0, b_1) = \sum_{i=1}^{n}(y_i - b_0 - b_1x_i)^2$$

RSS (residual sum of squares) is a function of two variables, b_0 and b_1. For any choice of these two variables, RSS is a summary measure of how poorly the resulting line fits the sample data. Larger values of RSS correspond to a worse fit.


RSS=67.5
RSS=3069.1
RSS=7610.1
RSS=10902.1
Hard
Temp

If we find some values of b_0 and b_1 which make RSS small, we can expect those values to be good estimators of β_0 and β_1, respectively. This is the idea behind ordinary least squares. We can find the values of b_0 and b_1 which minimize RSS using calculus:

$$\frac{\partial RSS}{\partial b_0} = -\sum 2(y_i - b_0 - b_1 x_i) \overset{\text{set}}{=} 0$$

$$\frac{\partial RSS}{\partial b_1} = -\sum 2(y_i - b_0 - b_1 x_i) x_i \overset{\text{set}}{=} 0$$

Setting these expressions equal to 0, we have what are known as the *normal equations*. We then solve for b_0 and b_1:

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{\sum x_i y_i - \bar{y}\bar{x}n}{\sum x_i^2 - n\bar{x}^2}$$

It can be shown that $\sum x_i y_i - \bar{y}\bar{x}n = \sum(x_i - \bar{x})(y_i - \bar{y})$ and also that $\sum x_i^2 - n\bar{x}^2 = \sum(x_i - \bar{x})^2$ so that

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

For conciseness, we define the following quantities:

$$SXX = \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_x^2 = \frac{SXX}{n-1}$$

$$SXY = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad s_{xy} = \frac{SXY}{n-1}$$

$$SYY = \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad s_y^2 = \frac{SYY}{n-1}$$

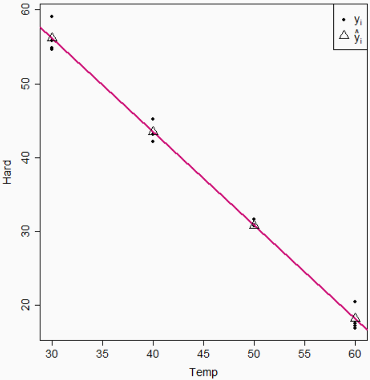
Since these values of b_0 and b_1 are those which "best" estimate β_0 and β_1, we designate them as $\hat{\beta}_0$ and $\hat{\beta}_1$. In other words, in OLS estimation of a SLR model (and plugging in the quantities defined on the previous slide):

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

$$\hat{\beta}_1 = \frac{SXY}{SXX}$$

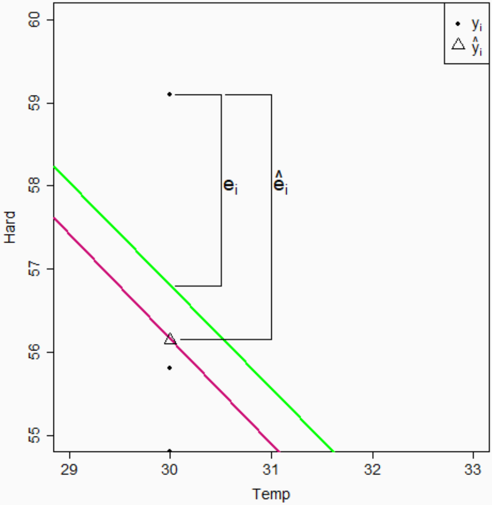
The estimated mean function is then $\hat{E}(Y) = \hat{\beta}_0 + \hat{\beta}_1 x = \hat{y}$

For the x_is in the data set, the corresponding y_is are known as *fitted values* because they are the values of y predicted by the fitted model at the x_is.



The vertical distances from the fitted model and the data points represent variation in Y that is not accounted for by the model. Since the model estimates the true mean function $E(Y) = \beta_0 + \beta_1 x$, these leftover distances should estimate the true errors e.

We call these estimates *residuals* and notate them $\hat{e}$.



The formula for a residual can be written several ways:

$$\begin{aligned}\hat{e}_i &= y_i - \hat{E}(Y) \\ &= y_i - \hat{y}_i \\ &= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\end{aligned}$$

For example, let x = time to recovery of a disease and let y = duration of remission. The data are:

x	y
4	6
2	6
3	7
4	5
2	6

First some preliminary calculations:

$$\bar{x} = 3$$

$$\bar{y} = 6$$

$$SXX = (4-3)^2 + \cdots + (2-3)^2 = 4$$

$$SYY = (6-6)^2 + \cdots + (6-6)^2 = 2$$

$$SXY = (4-3)(6-6) + \cdots + (2-3)(6-6) = -1$$

Now we can fit a simple linear regression model using the OLS estimators:

$$\hat{\beta}_1 = \frac{SXY}{SXX} = -0.25$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 6.75$$

For patients whose disease recovery time is $x = 3$, the expected remission duration is the fitted value

$$\hat{E}(Y) = \hat{y}_3 = 6.75 - 0.25 * (3) = 6$$

Since the patient in our sample whose recovery time was 3 had a remission duration of 7, the residual becomes

$$\hat{e}_3 = 7 - 6 = 1$$

$\hat{\beta}_0 = 6.75$ is interpreted as:

When time to recovery is 0, the mean remission duration is 6.75

$\hat{\beta}_1 = -0.25$ is interpreted as:

> *For every additional unit of time to recovery, the mean remission duration decreases 0.25 units*

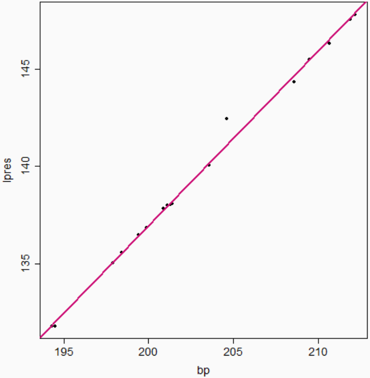
For the Forbes data in `alr4`, where $Y =$ `lpres` and $X =$ `bp`,

$$\hat{E}(Y) = \hat{y} = -42.1378 + 0.8955 * x.$$

```
summary(lm(lpres ~ bp, data = Forbes))

##
## Call:
## lm(formula = lpres ~ bp, data = Forbes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32220 -0.14473 -0.06664  0.02184  1.35978
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -42.13778    3.34020  -12.62 2.18e-09 ***
## bp            0.89549    0.01645   54.43  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.379 on 15 degrees of freedom
## Multiple R-squared:  0.995,^^IAdjusted R-squared:  0.9946
## F-statistic:  2963 on 1 and 15 DF,  p-value: < 2.2e-16
```

When `bp=0`, the mean `lpres` is -42.1378. For every additional unit degree of `bp`, the mean `lpres` increases 0.8955 units.



Section 3: Equivalent expressions

At this point we should review some mathematical equivalencies. We already claimed that

$$\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}n = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

and, as a special case

$$\sum_{i=1}^{n} x_i^2 - \bar{x}^2 n = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

We have also defined

$$SXX = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$SYY = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$SXY = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

so that

$$s_{xy} = \frac{SXY}{n-1}, \; s_x^2 = \frac{SXX}{n-1}, \text{ and } s_y^2 = \frac{SYY}{n-1}$$

Notice now that

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{SXY/(n-1)}{\sqrt{\frac{SXX}{n-1} \cdot \frac{SYY}{n-1}}} = \frac{SXY}{SXX \cdot SYY}$$

Finally, we learned that

$$\hat{\beta}_1 = \frac{SXY}{SXX}$$

This last expression can be manipulated using earlier ones:

$$\hat{\beta}_1 = \frac{SXY}{SXX} = \frac{s_{xy} * (n-1)}{s_x^2 * (n-1)} = \frac{r_{xy} * s_x * s_y * (n-1)}{s_x^2 * (n-1)} = r_{xy} * \frac{s_y}{s_x}$$

Going one step further:

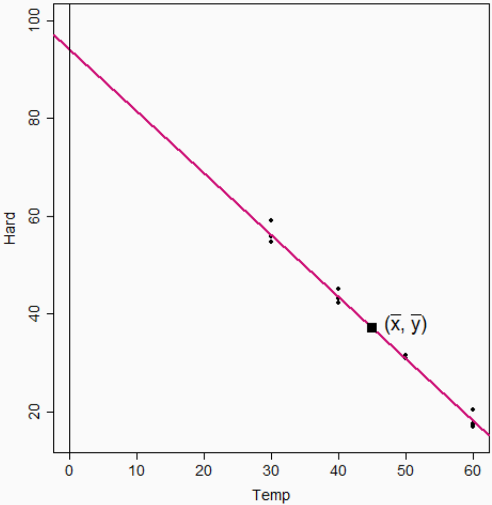
$$\hat{\beta}_1 = r_{xy} * \frac{s_y}{s_x} = r_{xy} * \frac{\sqrt{SYY/(n-1)}}{\sqrt{SXX/(n-1)}} = r_{xy} * \sqrt{\frac{SYY}{SXX}}$$

These two identities emphasize a connection between $\hat{\beta}_1$ and r_{xy}. Also, they provide formulas for easily computing $\hat{\beta}_1$ in cases where certain quantities are readily available but SXY is not.

One additional item of interest is that the OLS regression line will always cross the point $(\bar{x}, \bar{y})$. This is evident from the form of the estimator for β_0:

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

This form indicates that in order for the best-fit line to get to a height of $\bar{y}$, it must have traveled exactly $\bar{x}$ distance from 0 along the x-axis.


Hard
Temp
100
80
60
40
20
0
10
20
30
40
50
60
$(\bar{x}, \bar{y})$

Section 4: Estimating variances

We have O.L.S. estimators for β_0 and β_1 but still not $\sigma^2 = V(e_i) = E(e_i^2)$ for all i. In Stat 200, we repeatedly estimated parameters with their corresponding statistics. We do the same here, since $E(e_i^2)$ is a parameter. The sample statistic version of this is

$$\sum_{i=1}^{n} \frac{\hat{e}_i^2}{n} = \sum_{i=1}^{n} \frac{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n} = \frac{RSS(\hat{\beta}_0, \hat{\beta}_1)}{n}$$

This seems like a good estimator, but it needs two adjustments. Notationally, we will shorten $RSS(\hat{\beta}_0, \hat{\beta}_1)$ to simply RSS when it is evaluated at $\hat{\beta}_0$ and $\hat{\beta}_1$. This quantity is the minimum value possible from the $RSS(b_0, b_1)$ function for the given data set.

Second, it is not correct to assume that RSS contains n independent pieces of information; in reality, it contains $n-2$ independent pieces of information because one piece is "used up" to estimate each of β_0 and β_1. Hence, the denominator of our sample mean should be $n-2$. This results in the estimator:

$$\hat{\sigma}^2 = \frac{RSS}{n-2}$$

The $n-2$ is known as the *degrees of freedom* of RSS.

In the disease remission data,

x	y	$\hat{y}$
4	6	5.75
2	6	6.25
3	7	6
4	5	5.75
2	6	6.25

Hence,

$RSS = 0.25^2 + (-0.25)^2 + 1^2 + (-0.75)^2 + (-0.25)^2 = 1.75$

and $\hat{\sigma}^2 = \frac{1.75}{3} = 0.583$

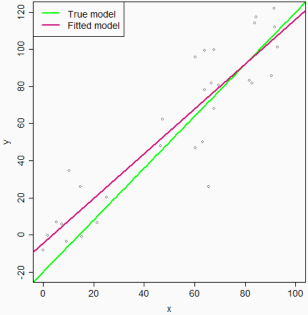
Some people like to report $\hat{\sigma}$ instead of $\hat{\sigma}^2$. Here, that would be $\sqrt{0.5833} = 0.764$. R calls this quantity *Residual standard error*. It is in the same units as the response variable.

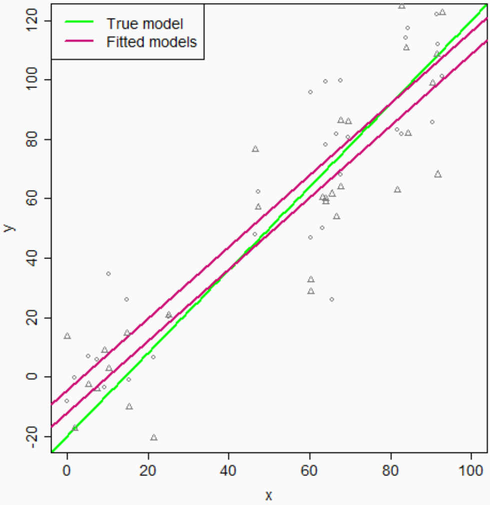
```r
x <- c(4, 2, 3, 4, 2)
y <- c(6, 6, 7, 5, 6)
summary(lm(y ~ x))
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     1     2     3     4     5
##  0.25 -0.25  1.00 -0.75 -0.25
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.7500     1.1955   5.646    0.011 *
## x            -0.2500     0.3819  -0.655    0.559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7638 on 3 degrees of freedom
## Multiple R-squared:  0.125,^^IAdjusted R-squared:  -0.1667
## F-statistic: 0.4286 on 1 and 3 DF,  p-value: 0.5594
```

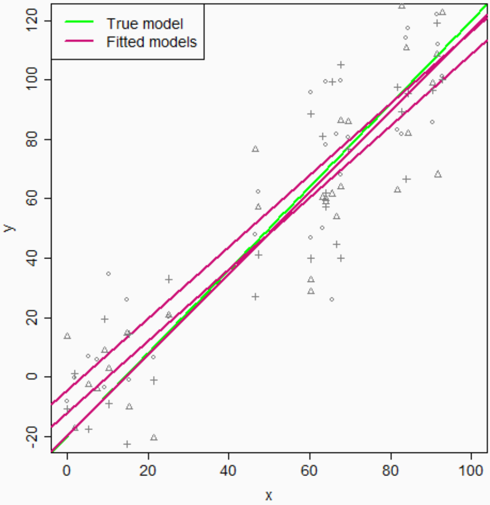
Section 5: Sampling properties of OLS estimators

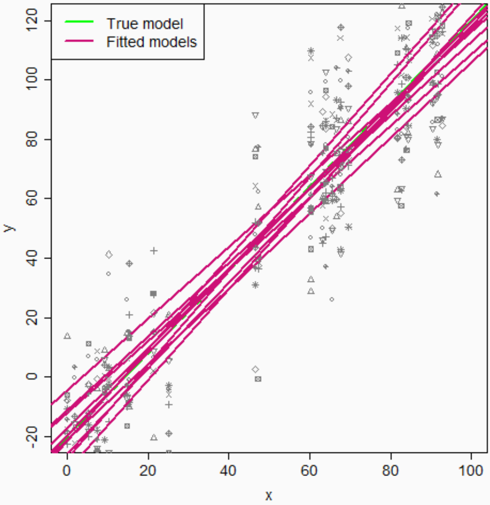
The estimators $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}^2$ are functions of the random variables $Y_1, Y_2, \ldots, Y_n$, which makes these estimators random variables, also. Every time a data set is drawn, a these estimators will turn out somewhat different, even though β_0 and β_1 are fixed constants.

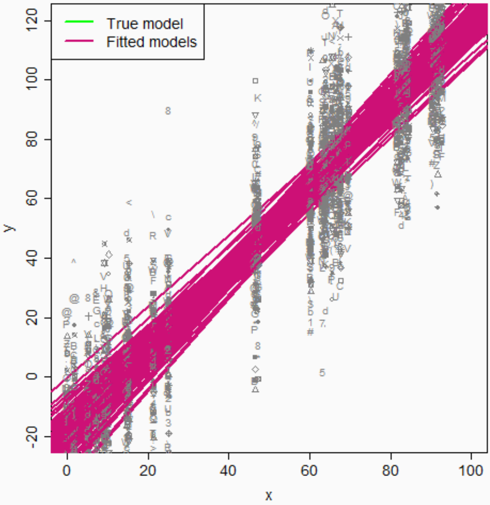




True model
Fitted models
y
x
120
100
80
60
40
20
0
-20
0
20
40
60
80
100


True model
Fitted models
y
x
-20
0
20
40
60
80
100
120
0
20
40
60
80
100


True model
Fitted models
y
x
-20
0
20
40
60
80
100
120
0
20
40
60
80
100



True model
Fitted models
y
x
120
100
80
60
40
20
0
-20
0
20
40
60
80
100

The estimators possess *sampling variability*. But their random nature is constrained by certain properties, which can be mathematically proven.

$$E(\hat{\beta}_0) = \beta_0$$

$$E(\hat{\beta}_1) = \beta_1$$

$$E(\hat{\sigma}^2) = E\left(\frac{RSS}{n-2}\right) = \sigma^2 \neq E\left(\frac{RSS}{n}\right)$$

This is the *unbiasedness* property of an estimator. In other words, $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}^2$ are unbiased estimators.

$$V(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)$$

$$V(\hat{\beta}_1) = \sigma^2 \frac{1}{SXX}$$

Among a certain class of estimators in SLR (linear, unbiased), these variances have the smallest variance possible. That is, the OLS estimators are the *best*.

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\bar{x}}{SXX}$$

$$\rho(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$$

If the model assumptions of the S.L.R. model are:

- $E(e_i) = 0$ or $E(Y_i) = \beta_0 + \beta_1 x_i$ for all i
- $V(Y_i) = V(e_i) = \sigma^2$ for all i
- e_i and e_j are independent for all $i \neq j$
- $e_i \sim N(0, \sigma^2)$ for all i

then $\hat{\beta}_0$ and $\hat{\beta}_1$ are themselves normally distributed and

$$(n-2) * \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

and $\hat{\sigma}^2$ is independent of both $\hat{\beta}_0$ and $\hat{\beta}_1$.

Section 6: Functions of OLS estimators

Since the fitted values $\hat{y}$ are functions of the Y_is, they, too are random. It can be shown that:

$$E(\hat{y}) = \beta_0 + \beta_1 x$$

We can regard the $\hat{y}$s as one of two things:

1. The estimated mean of Y at some value of X, $\hat{E}(Y)$
2. The predicted value $\tilde{y}$ of a new Y drawn from the population at some value of X, y^*

In other words,

$$\hat{y} = \hat{E}(Y) = \tilde{y}$$

However, the variances of $\hat{E}(Y)$ and $\tilde{y}$ are not calculated equally. In the first case,

$$\hat{E}(Y) = \hat{\beta}_0 + \hat{\beta}_1 x$$

and so

$$V(\hat{E}(Y)) = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right)$$

In the second case, we will actually care only about the variance of the prediction error $\tilde{y} - y^*$.

$$
\begin{aligned}
V(\tilde{y} - y^*) &= V(\tilde{y}) + V(y^*) - 2Cov(\tilde{y}, y^*) \\
&= \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right) \\
&= \sigma^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right)
\end{aligned}
$$

The residuals, too have expected values and variance. For now, we consider only the expected values:

$$E(\hat{e}) = E(y - \hat{\beta}_0 - \hat{\beta}_1 x) = 0$$

Section 7: Proofs of sampling properties

We will prove some of the results from the last two sections. To do this, we start with another identity:

$$SXY = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

Notice now what happens in the expression for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{SXY}{SXX} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{SXX} = \sum_{i=1}^{n} \frac{1}{SXX}(x_i - \bar{x})y_i = \sum_{i=1}^{n} c_i y_i$$

where $c_i = \frac{1}{SXX}(x_i - \bar{x})$ is a constant (non-random) quantity. Thus $\hat{\beta}_1$ is a *linear* estimator (in the y_is).

This makes expectation and variance calculations very simple:

$$E(\hat{\beta}_1) = E\left(\sum_{i=1}^{n} c_i Y_i\right) = \sum_{i=1}^{n} c_i E(Y_i)$$
$$= \sum_{i=1}^{n} c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i$$

For variances:

$$V(\hat{\beta}_1) = V\left(\sum_{i=1}^{n} c_i Y_i\right) = \sum_{i=1}^{n} c_i^2 V(Y_i) = \sigma^2 \sum_{i=1}^{n} c_i^2 = \frac{\sigma^2}{SXX}$$

The estimator $\hat{\beta}_0$ is also a linear combination of the y_is, and so is also a linear estimator:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \sum_{i=1}^{n} \left(\frac{y_i}{n} - \frac{c_i y_i \bar{x}}{n} \right)$$
$$= \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{c_i \bar{x}}{n} \right) y_i = \sum_{i=1}^{n} d_i y_i$$

where $d_i = (1/n - c_i x_i)$ is also a constant quantity. A similar approach to above will now prove that $\hat{\beta}_0$ is unbiased for β_0 and give the variance formula from the previous section.

Not only are $\hat{\beta}_0$ and $\hat{\beta}_1$ linear functions of the y_is, but so are the fitted values:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \sum_{i=1}^{n} (d_i + c_i x_i) y_i$$

Another consequence of the fact that $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{y}$ are linear combinations of y_is is that, if $e_i \sim N(0, \sigma^2)$ for all i, and therefore

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2),$$

then $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{y}$ are linear combinations of normally distributed random variables. As we discussed in our Stat 200 review, linear combinations of normal random variables are themselves normal random variables.

The facts that

$$(n-2) * \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

and $\hat{\sigma}^2$ is independent of both $\hat{\beta}_0$ and $\hat{\beta}_1$ follow using more theory.