

Section 14

Simple linear regression.

Let us look at the 'cigarette' dataset from [1] (available to download from journal's website) and [2]. The cigarette dataset contains measurements of tar, nicotine, weight and carbon monoxide (CO) content for 25 brands of domestic cigarettes. We are going to try to predict CO as a function of tar and nicotine content. To visualize the data let us plot each of these variable against others, see figure 14.1. Since the variables seem to have a linear relationship we fit a least-squares line, which we will explain below, to fit the data using Matlab tool 'polytool'. For example, if our vectors are 'nic' for nicotine, 'tar' for tar and 'carb' for CO then, for example, using

```
polytool(nic,carb,1)
```

will produce figure 14.1 (a), etc. We can also perform statistical analysis of these fits, in a sense that will gradually be explained below, using Matlab 'regress' function. For carbon monoxide vs. tar:

```
[b,bint,r,rint,stats]=regress(carb,[ones(25,1),tar]);
```

```
b = 2.7433      bint = 1.3465  4.1400
     0.8010           0.6969  0.9051
```

```
stats = 0.9168  253.3697  0.000  1.9508,
```

for carbon monoxide vs. nicotine

```
[b,bint,r,rint,stats]=regress(carb,[ones(25,1),nic]);
```

```
b = 1.6647      bint = -0.3908  3.7201
     12.3954           10.2147  14.5761
```

```
stats = 0.8574  138.2659  0.000  3.3432
```

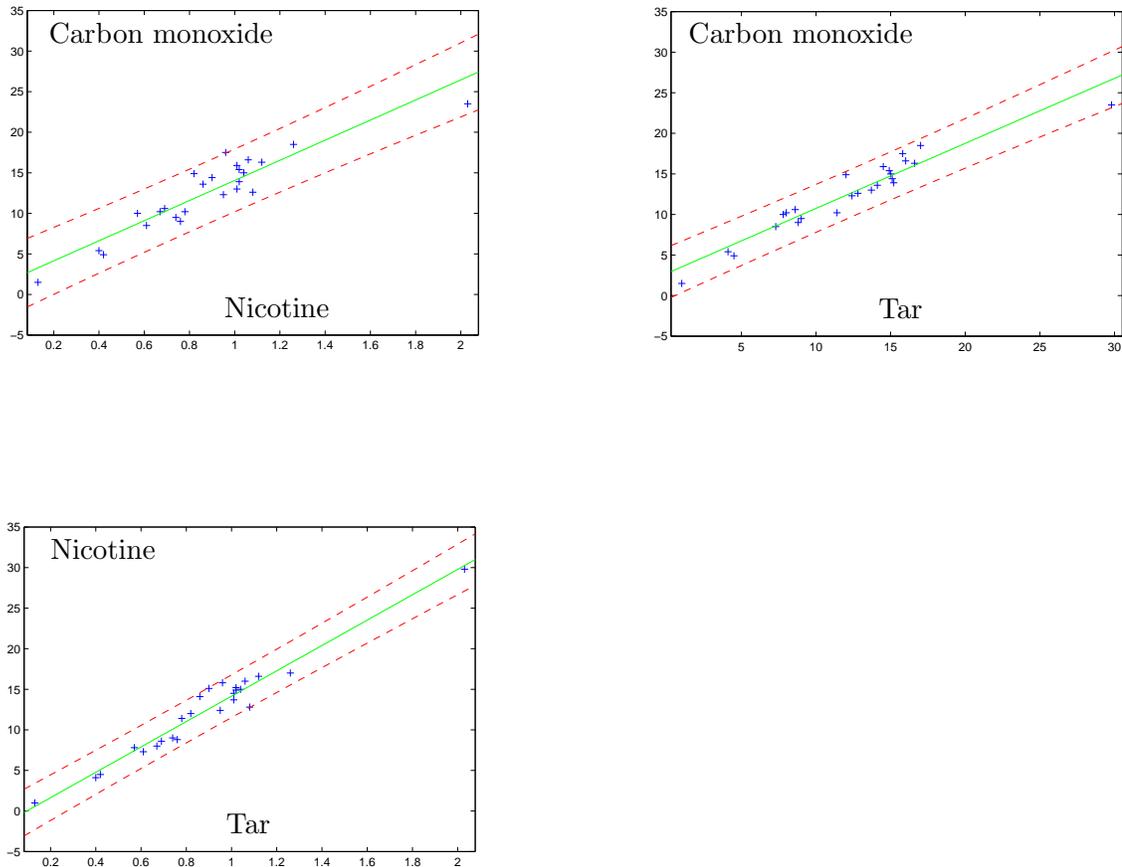


Figure 14.1: Least-squares line (solid line). (a) Carbon monoxide content (mg.) vs. nicotine content (mg.). (b) Carbon monoxide vs. tar content. (c) Tar content vs. nicotine content.

and for nicotine vs. tar

```
[b,bint,r,rint,stats]=regress(tar,[ones(25,1),nic]);
```

```
b = -1.4805    bint = -2.8795    -0.0815
     15.6281           14.1439    17.1124
```

```
stats = 0.9538  474.4314  0.000  1.5488
```

The output of 'regress' gives a vector 'b' of parameters of a fitted least-squares line, 95% confidence intervals 'bint' for these parameters, and 'stats' contains in order:

*R*² statistic, *F* statistic, *p*-value of *F* statistic, MLE $\hat{\sigma}^2$ of the error variance.

All of these will be explained below.

□

Simple linear regression model.

Suppose that we have a pair of variables (X, Y) and a variable Y is a linear function of X plus random noise:

$$Y = f(X) + \varepsilon = \beta_0 + \beta_1 X + \varepsilon,$$

where a random noise ε is assumed to have normal distribution $N(0, \sigma^2)$. A variable X is called a predictor variable, Y - a response variable and a function $f(x) = \beta_0 + \beta_1 x$ - a linear regression function.

Suppose that we are given a sequence of pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ that are described by the above model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. $N(0, \sigma^2)$. We have three unknown parameters - β_0, β_1 and σ^2 - and we want to estimate them using a given sample. The points X_1, \dots, X_n can be either random or non random, but from the point of view of estimating linear regression function the nature of X s is in some sense irrelevant so we will think of them as fixed and non random and assume that the randomness comes from the noise variables ε_i . For a fixed X_i , the distribution of Y_i is equal to $N(f(X_i), \sigma^2)$ with p.d.f.

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-f(X_i))^2}{2\sigma^2}}$$

and the likelihood function of the sequence Y_1, \dots, Y_n is:

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - f(X_i))^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2}.$$

Let us find the maximum likelihood estimates of β_0, β_1 and σ^2 that maximize this likelihood function. First of all, it is obvious that for any σ^2 we need to minimize

$$L := \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$$

over β_0, β_1 . The line that minimizes the sum of squares L is called the *least-squares line*. To find the critical points we write:

$$\begin{aligned} \frac{\partial L}{\partial \beta_0} &= - \sum_{i=1}^n 2(Y_i - (\beta_0 + \beta_1 X_i)) = 0 \\ \frac{\partial L}{\partial \beta_1} &= - \sum_{i=1}^n 2(Y_i - (\beta_0 + \beta_1 X_i))X_i = 0 \end{aligned}$$

If we introduce the notations

$$\bar{X} = \frac{1}{n} \sum X_i, \quad \bar{Y} = \frac{1}{n} \sum Y_i, \quad \bar{X}^2 = \frac{1}{n} \sum X_i^2, \quad \bar{X}\bar{Y} = \frac{1}{n} \sum X_i Y_i$$

then the critical point conditions can be rewritten as

$$\beta_0 + \beta_1 \bar{X} = \bar{Y} \quad \text{and} \quad \beta_0 \bar{X} + \beta_1 \bar{X}^2 = \bar{X}\bar{Y}.$$

Solving for β_0 and β_1 we get the MLE

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \text{ and } \hat{\beta}_1 = \frac{\bar{X}\bar{Y} - \bar{X}\bar{Y}}{\bar{X}^2 - \bar{X}^2}.$$

These estimates are used to plot least-squares regression lines in figure 14.1. Finally, to find the MLE of σ^2 we maximize the likelihood over σ^2 and get:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.$$

The differences $r_i = Y_i - \hat{Y}_i$ between observed response variables Y_i and the values predicted by the estimated regression line

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

are called the *residuals*. The R^2 statistic in the examples above is defined as

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

The numerator in the last sum is the sum of squares of the residuals and the denominator is the variance of Y and R^2 is usually interpreted as the proportion of variability in the data explained by the linear model. The higher R^2 the better our model explains the data. Next, we would like to do statistical inference about the linear model.

1. *Construct confidence intervals for parameters of the model β_0, β_1 and σ^2 .*
2. *Construct prediction intervals for Y given any point X (dotted lines in figure 14.1).*
3. *Test hypotheses about parameters of the model. For example, F -statistic in the output of Matlab function 'regress' comes from a test of the hypothesis $H_0 : \beta_0 = 0, \beta_1 = 0$ that the response Y is not 'correlated' with a predictor variable X .*

In spirit all these problems are similar to statistical inference about parameters of normal distribution such as t -tests, F -tests, etc. so as a starting point we need to find a joint distribution of the estimates $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}^2$.

To compute the joint distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ is very easy because they are linear combinations of Y_i s which have normal distributions and, as a result, $\hat{\beta}_0$ and $\hat{\beta}_1$ will have normal distributions. All we need to do is find their means, variances and covariance, which is a straightforward computation. However, we will obtain this as a part of a more general computation that will also give us joint distribution of all three estimates $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}^2$. Let us denote the sample variance of X s by

$$\sigma_x^2 = \bar{X}^2 - \bar{X}^2.$$

Then we will prove the following:

1. $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{n\sigma_x^2}\right)$, $\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{X}^2}{n\sigma_x^2}\right)\sigma^2\right) = N\left(\beta_0, \frac{\sigma^2}{n\sigma_x^2}\bar{X}^2\right)$,
 $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{X}\sigma^2}{n\sigma_x^2}$.
2. $\hat{\sigma}^2$ is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.
3. $\frac{n\hat{\sigma}^2}{\sigma^2}$ has χ_{n-2}^2 distribution with $n - 2$ degrees of freedom.

Remark. Line 1 means that $(\hat{\beta}_0, \hat{\beta}_1)$ have jointly normal distribution with mean (β_0, β_1) and covariance matrix

$$\Sigma = \frac{\sigma^2}{n\sigma_x^2} \begin{pmatrix} \bar{X}^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}.$$

Proof. Let us consider two vectors

$$a_1 = (a_{11}, \dots, a_{1n}) = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$$

and

$$a_2 = (a_{21}, \dots, a_{2n}) \text{ where } a_{2i} = \frac{X_i - \bar{X}}{\sqrt{n\sigma_x^2}}.$$

It is easy to check that both vectors have length 1 and they are orthogonal to each other since their scalar product is

$$a_1 \cdot a_2 = \sum_{i=1}^n a_{1i}a_{2i} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \bar{X}}{\sqrt{n\sigma_x^2}} = 0.$$

Let us choose vectors a_3, \dots, a_n so that a_1, \dots, a_n is orthonormal basis and, as a result, the matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ a_{12} & \cdots & a_{n2} \\ \vdots & \vdots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix}$$

is orthogonal. Let us consider vectors

$$Y = (Y_1, \dots, Y_n), \mu = \mathbb{E}Y = (\mathbb{E}Y_1, \dots, \mathbb{E}Y_n)$$

and

$$Y' = (Y'_1, \dots, Y'_n) = \frac{Y - \mu}{\sigma} = \left(\frac{Y_1 - \mathbb{E}Y_1}{\sigma}, \dots, \frac{Y_n - \mathbb{E}Y_n}{\sigma}\right)$$

so that the random variables Y'_1, \dots, Y'_n are i.i.d. standard normal. We proved before that if we consider an orthogonal transformation of i.i.d. standard normal sequence:

$$Z' = (Z'_1, \dots, Z'_n) = Y'A$$

then Z'_1, \dots, Z'_n will also be i.i.d. standard normal. Since

$$Z' = Y'A = \left(\frac{Y - \mu}{\sigma}\right)A = \frac{YA - \mu A}{\sigma}$$

this implies that

$$YA = \sigma Z' + \mu A.$$

Let us define a vector

$$Z = (Z_1, \dots, Z_n) = YA = \sigma Z' + \mu A.$$

Each Z_i is a linear combination of Y_i s and, therefore, it has a normal distribution. Since we made a specific choice of the first two columns of the matrix A we can write down explicitly the first two coordinates Z_1 and Z_2 of vector Z . We have,

$$Z_1 = \sum_{i=1}^n a_{i1}Y_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sqrt{n}\bar{Y} = \sqrt{n}(\hat{\beta}_0 + \hat{\beta}_1\bar{X})$$

and the second coordinate

$$\begin{aligned} Z_2 &= \sum_{i=1}^n a_{i2}Y_i = \sum_{i=1}^n \frac{(X_i - \bar{X})Y_i}{\sqrt{n\sigma_x^2}} \\ &= \sqrt{n\sigma_x^2} \sum_{i=1}^n \frac{(X_i - \bar{X})Y_i}{n\sigma_x^2} = \sqrt{n\sigma_x^2}\hat{\beta}_1. \end{aligned}$$

Solving these two equations for $\hat{\beta}_0$ and $\hat{\beta}_1$ we can express them in terms of Z_1 and Z_2 as

$$\hat{\beta}_1 = \frac{1}{\sqrt{n\sigma_x^2}}Z_2 \text{ and } \hat{\beta}_0 = \frac{1}{\sqrt{n}}Z_1 - \frac{\bar{X}}{\sqrt{n\sigma_x^2}}Z_2.$$

This easily implies claim 1. Next we will show how $\hat{\sigma}^2$ can also be expressed in terms of Z_i s.

$$\begin{aligned} n\hat{\sigma}^2 &= \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \sum_{i=1}^n \left((Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X}) \right)^2 \quad \{\text{since } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}\} \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2 - 2\hat{\beta}_1 n\sigma_x^2 \underbrace{\frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{n\sigma_x^2}}_{\hat{\beta}_1} + \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2 - \hat{\beta}_1^2 n\sigma_x^2 = \sum_{i=1}^n Y_i^2 - \underbrace{n(\bar{Y})^2}_{Z_1^2} - \underbrace{\hat{\beta}_1^2 n\sigma_x^2}_{Z_2^2} \\ &= \sum_{i=1}^n Y_i^2 - Z_1^2 - Z_2^2 = \sum_{i=1}^n Z_i^2 - Z_1^2 - Z_2^2 = Z_3^2 + \dots + Z_n^2. \end{aligned}$$

In the last line we used the fact that $Z = YA$ is an orthogonal transformation of Y and since orthogonal transformation preserves the length of a vector we have,

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n Y_i^2.$$

If we can show that Z_3, \dots, Z_n are i.i.d. with distribution $N(0, \sigma^2)$ then

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \left(\frac{Z_3}{\sigma}\right)^2 + \dots + \left(\frac{Z_n}{\sigma}\right)^2 \sim \chi_{n-2}^2$$

has χ^2 -distribution with $n - 2$ degrees of freedom, because $Z_i/\sigma \sim N(0, 1)$. Since we showed above that

$$Z = \mu A + \sigma Z' \Rightarrow Z_i = (\mu A)_i + \sigma Z'_i,$$

the fact that Z'_1, \dots, Z'_n are i.i.d. standard normal implies that Z_i s are independent of each other and $Z_i \sim N((\mu A)_i, \sigma^2)$. Let us compute the mean $\mathbb{E}Z_i = (\mu A)_i$:

$$\begin{aligned} (\mu A)_i &= \mathbb{E}Z_i = \mathbb{E} \sum_{j=1}^n a_{ji} Y_j = \sum_{j=1}^n a_{ji} \mathbb{E}Y_j = \sum_{j=1}^n a_{ji} (\beta_0 + \beta_1 X_j) \\ &= \sum_{j=1}^n a_{ji} (\beta_0 + \beta_1 \bar{X} + \beta_1 (X_j - \bar{X})) \\ &= (\beta_0 + \beta_1 \bar{X}) \sum_{j=1}^n a_{ji} + \beta_1 \sum_{j=1}^n a_{ji} (X_j - \bar{X}). \end{aligned}$$

Since the matrix A is orthogonal its columns are orthogonal to each other. Let $a_i = (a_{1i}, \dots, a_{ni})$ be the vector in the i th column and let us consider $i \geq 3$. Then the fact that a_i is orthogonal to the first column gives

$$a_i \cdot a_1 = \sum_{j=1}^n a_{j1} a_{ji} = \sum_{j=1}^n \frac{1}{\sqrt{n}} a_{ji} = 0$$

and the fact that a_i is orthogonal to the second column gives

$$a_i \cdot a_2 = \frac{1}{\sqrt{n\sigma_x^2}} \sum_{j=1}^n (X_j - \bar{X}) a_{ji} = 0.$$

This show that for $i \geq 3$

$$\sum_{j=1}^n a_{ji} = 0 \text{ and } \sum_{j=1}^n a_{ji} (X_j - \bar{X}) = 0$$

and this proves that $\mathbb{E}Z_i = 0$ for $i \geq 3$ and $Z_i \sim N(0, \sigma^2)$ for $i \geq 3$. As we mentioned above this also proves that $n\hat{\sigma}^2/\sigma^2 \sim \chi_{n-2}^2$.

Finally, $\hat{\sigma}^2$ is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$ because $\hat{\sigma}^2$ can be written as a function of Z_3, \dots, Z_n and $\hat{\beta}_0$ and $\hat{\beta}_1$ can be written as functions of Z_1 and Z_2 . □

Statistical inference in simple linear regression. Suppose now that we want to find the confidence intervals for unknown parameters of the model β_0, β_1 and σ^2 . This is

straightforward and very similar to the confidence intervals for parameters of normal distribution. For example, using that $n\hat{\sigma}^2/\sigma^2 \sim \chi_{n-2}^2$, if we find the constants c_1 and c_2 such that

$$\chi_{n-2}^2(0, c_1) = \frac{1-\alpha}{2} \text{ and } \chi_{n-2}^2(c_2, +\infty) = \frac{1-\alpha}{2}$$

then with probability α we have $c_1 \leq n\hat{\sigma}^2/\sigma^2 \leq c_2$. Solving this for σ^2 we find the α confidence interval:

$$\frac{n\hat{\sigma}^2}{c_2} \leq \sigma^2 \leq \frac{n\hat{\sigma}^2}{c_1}.$$

Similarly, we find the α confidence interval for β_1 . Since

$$(\hat{\beta}_1 - \beta_1) / \sqrt{\frac{\sigma^2}{n\sigma_x^2}} \sim N(0, 1) \text{ and } \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2$$

then

$$(\hat{\beta}_1 - \beta_1) \sqrt{\frac{n\sigma_x^2}{\sigma^2}} / \sqrt{\frac{1}{n-2} \frac{n\hat{\sigma}^2}{\sigma^2}} \sim t_{n-2}$$

has Student t_{n-2} -distribution with $n-2$ degrees of freedom. Simplifying, we get

$$(\hat{\beta}_1 - \beta_1) \sqrt{\frac{(n-2)\sigma_x^2}{\hat{\sigma}^2}} \sim t_{n-2}. \quad (14.0.1)$$

Therefore, if we find c such that $t_{n-2}(-c, c) = \alpha$ then with probability α :

$$-c \leq (\hat{\beta}_1 - \beta_1) \sqrt{\frac{(n-2)\sigma_x^2}{\hat{\sigma}^2}} \leq c$$

and solving for β_1 we obtain the α confidence interval:

$$\hat{\beta}_1 - c \sqrt{\frac{\hat{\sigma}^2}{(n-2)\sigma_x^2}} \leq \beta_1 \leq \hat{\beta}_1 + c \sqrt{\frac{\hat{\sigma}^2}{(n-2)\sigma_x^2}}.$$

Similarly, to find the confidence interval for β_0 we use that

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\left(\frac{1}{n} + \frac{\bar{X}^2}{n\sigma_x^2}\right)\sigma^2}} / \sqrt{\frac{1}{n-2} \frac{n\hat{\sigma}^2}{\sigma^2}} = (\hat{\beta}_0 - \beta_0) / \sqrt{\frac{\hat{\sigma}^2}{n-2} \left(1 + \frac{\bar{X}^2}{\sigma_x^2}\right)} \sim t_{n-2} \quad (14.0.2)$$

and α confidence interval for β_0 is:

$$\hat{\beta}_0 - c \sqrt{\frac{\hat{\sigma}^2}{n-2} \left(1 + \frac{\bar{X}^2}{\sigma_x^2}\right)} \leq \beta_0 \leq \hat{\beta}_0 + c \sqrt{\frac{\hat{\sigma}^2}{n-2} \left(1 + \frac{\bar{X}^2}{\sigma_x^2}\right)}.$$

□

We can now construct various t -tests based on t -statistics (14.0.1) and (14.0.2).

Linear combinations of parameters. More generally, let us compute the distribution of a linear combination

$$c_0\hat{\beta}_0 + c_1\hat{\beta}_1$$

of the estimates. This will allow us to construct confidence intervals and t -tests for linear combinations of parameters $c_0\beta_0 + c_1\beta_1$. Clear, the distribution of this linear combination will be normal with mean

$$\mathbb{E}(c_0\hat{\beta}_0 + c_1\hat{\beta}_1) = c_0\beta_0 + c_1\beta_1.$$

We compute its variance:

$$\begin{aligned} \text{Var}(c_0\hat{\beta}_0 + c_1\hat{\beta}_1) &= \mathbb{E}(c_0\hat{\beta}_0 + c_1\hat{\beta}_1 - c_0\beta_0 - c_1\beta_1)^2 = \mathbb{E}(c_0(\hat{\beta}_0 - \beta_0) + c_1(\hat{\beta}_1 - \beta_1))^2 \\ &= c_0^2 \underbrace{\mathbb{E}(\hat{\beta}_0 - \beta_0)^2}_{\text{variance of } \hat{\beta}_0} + c_1^2 \underbrace{\mathbb{E}(\hat{\beta}_1 - \beta_1)^2}_{\text{variance of } \hat{\beta}_1} + 2c_0c_1 \underbrace{\mathbb{E}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)}_{\text{covariance}} \\ &= c_0^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{n\sigma_x^2} \right) \sigma^2 + c_1^2 \frac{\sigma^2}{n\sigma_x^2} - 2c_0c_1 \frac{\bar{X}\sigma^2}{n\sigma_x^2} \\ &= \sigma^2 \left(\frac{c_0^2}{n} + \frac{(c_0\bar{X} - c_1)^2}{n\sigma_x^2} \right). \end{aligned}$$

This proves that

$$c_0\hat{\beta}_0 + c_1\hat{\beta}_1 \sim N\left(c_0\beta_0 + c_1\beta_1, \sigma^2 \left(\frac{c_0^2}{n} + \frac{(c_0\bar{X} - c_1)^2}{n\sigma_x^2} \right)\right). \quad (14.0.3)$$

Using $(c_0, c_1) = (1, 0)$ or $(0, 1)$, will give the distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$. □

Prediction Intervals. Suppose now that we have a new observation X for which Y is unknown and we want to predict Y or find the confidence interval for Y . According to simple regression model,

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

and it is natural to take $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ as the prediction of Y . Let us find the distribution of their difference $\hat{Y} - Y$. Clearly, the difference will have normal distribution so we only need to compute the mean and the variance. The mean is

$$\mathbb{E}(\hat{Y} - Y) = \mathbb{E}\hat{\beta}_0 + \mathbb{E}\hat{\beta}_1 X - \beta_0 - \beta_1 X - \mathbb{E}\varepsilon = \beta_0 + \beta_1 X - \beta_0 - \beta_1 X - 0 = 0.$$

Since a new pair (X, Y) is independent of the prior data we have that Y is independent of \hat{Y} . Therefore, since the variance of the sum or difference of independent random variables is equal to the sum of their variances, we get

$$\text{Var}(\hat{Y} - Y) = \text{Var}(\hat{Y}) + \text{Var}(Y) = \sigma^2 + \text{Var}(\hat{Y}),$$

where we also used that $\text{Var}(Y) = \text{Var}(\varepsilon) = \sigma^2$. To compute the variance of \hat{Y} we can use the formula above with $(c_0, c_1) = (1, X)$

$$\text{Var}(\hat{Y}) = \text{Var}(\hat{\beta}_0 + X\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{X} - X)^2}{n\sigma_x^2} \right).$$

Therefore, we showed that

$$\hat{Y} - Y \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(\bar{X} - X)^2}{n\sigma_x^2}\right)\right).$$

As a result, we have:

$$\frac{\hat{Y} - Y}{\sqrt{\sigma^2\left(1 + \frac{1}{n} + \frac{(\bar{X} - X)^2}{n\sigma_x^2}\right)}} \bigg/ \sqrt{\frac{1}{n-2} \frac{n\hat{\sigma}^2}{\sigma^2}} \sim t_{n-2}$$

and the $1 - \alpha$ prediction interval for Y is

$$\hat{Y} - c\sqrt{\frac{\sigma^2}{n-2}\left(n+1 + \frac{(\bar{X} - X)^2}{\sigma_x^2}\right)} \leq Y \leq \hat{Y} + c\sqrt{\frac{\sigma^2}{n-2}\left(n+1 + \frac{(\bar{X} - X)^2}{\sigma_x^2}\right)}.$$

These are the dashed curves created by Matlab 'polytool' function.

□

Simultaneous confidence set for (β_0, β_1) and F -test. We will now construct a statistic that will allow us to give a confidence set for both parameters β_0, β_1 at the same time and test the hypothesis of the type

$$H_0 : \beta_0 = 0 \text{ and } \beta_1 = 0. \quad (14.0.4)$$

The values $(0, 0)$ could be replaced by any other predetermined values. Looking at the proof of the joint distribution of the estimates, as an intermediate step we showed that estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ can be related to

$$Z_1 = \sqrt{n}(\hat{\beta}_0 + \hat{\beta}_1 X) \quad \text{and} \quad Z_2 = \sqrt{n\sigma_x^2}\hat{\beta}_1$$

where normal random variables Z_1, Z_2 are independent of each other and independent of

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2.$$

Also, Z_1 and Z_2 have variance σ^2 . Standardizing these random variables we get

$$A = \frac{\sqrt{n}}{\sigma}((\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)\bar{X}) \sim N(0, 1) \quad \text{and} \quad B = \frac{\sqrt{n\sigma_x^2}}{\sigma}(\hat{\beta}_1 - \beta_1) \sim N(0, 1)$$

which implies that $A^2 + B^2 \sim \chi_2^2$ -distribution. By definition of F -distribution,

$$\frac{n-2}{2}(A^2 + B^2) \bigg/ \frac{n\hat{\sigma}^2}{\sigma^2} \sim F_{2, n-2}.$$

Simplifying the left-hand side we get

$$F := \frac{n-2}{2\hat{\sigma}^2} \left((\hat{\beta}_0 - \beta_0)^2 + \bar{X}^2(\hat{\beta}_1 - \beta_1)^2 + 2\bar{X}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) \right) \sim F_{2, n-2}.$$

This allows us to obtain a joint confidence set (ellipse) for parameters β_0, β_1 . Given a confidence level $\alpha \in [0, 1]$ we define a threshold c by $F_{2,n-2}(0, c) = \alpha$ then with probability α we have

$$F := \frac{n-2}{2\hat{\sigma}^2} \left((\hat{\beta}_0 - \beta_0)^2 + \bar{X}^2(\hat{\beta}_1 - \beta_1)^2 + 2\bar{X}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) \right) \leq c.$$

This inequality defines an ellipse for (β_0, β_1) . To test the hypothesis (14.0.4), we use the fact that under H_0 the statistic

$$F := \frac{n-2}{2\hat{\sigma}^2} (\hat{\beta}_0^2 + \bar{X}^2\hat{\beta}_1^2 + 2\bar{X}\hat{\beta}_0\hat{\beta}_1) \sim F_{2,n-2}$$

and define a decision rule by

$$\delta = \begin{cases} H_0 : & F \leq c \\ H_1 : & F > c, \end{cases}$$

where $F_{2,n-2}(c, \infty) = \alpha$ - a level of significance.

F -statistic output by Matlab 'regress' function will be explained in the next section. □

References.

- [1] "Using Cigarette Data for An Introduction to Multiple Regression." by Lauren McIntyre, *Journal of Statistics Education* v.2, n.1 (1994).
- [2] Mendenhall, W., and Sincich, T. (1992), *Statistics for Engineering and the Sciences* (3rd ed.), New York: Dellen Publishing Co.