REGRESSION WITH MULTIPLE PREDICTORS

1. Reading

- [1, Section 10.1]
  - Example 10.1.2. This is the linear regression model with multiple predictors.
  - This entire section is useful and all the exercises are good practice.
- [2, Section 10.2]
- [2, Section 10.5]
- [1, Section 3.2]

2. Learning objectives

- Understand how to simulate and fit linear regression models with multiple predictors.
- Interpret regression coefficients in linear regression with multiple predictors.

3. Multiple predictor linear regression

- The real power of regression comes when we work with models of the form

$$Y = \beta_0 + \sum_{i=1}^{K} \beta_i X_i + \epsilon \tag{1}$$

$$\epsilon \sim \text{Normal}(0, \sigma^2) \tag{2}$$

where X_i is a set of K predictor variables. Alternatively, we can write

$$Y|(X_1 = x_1, \ldots, X_K = x_K) \sim \text{Normal}\left(\beta_0 + \sum_{i=1}^{K} \beta_i X_i, \sigma^2\right) \tag{3}$$

I will also use the shorthand,

$$Y \sim \text{LR}(X, \beta, \sigma^2). \tag{4}$$

I will also use the notation

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{K} \hat{\beta}_i X_i$$

to refer to the predicted value of $E[Y|X]$ after fitting a regression model.

In these notes, our goal is to answer the following questions

(1) What are estimators of the parameters in this model?
(2) How do we interpret the **regression coefficients** β_i?
(3) Precisely what are the assumptions we are making when we use a linear regression model?
(4) How do we access the model assumptions?

Example 1 (Simulating and fitting a regression with two predictors)**.** *See colab notebook.*

- The output from the regression with multiple predictors is basically the same as for single-predictor, except now we have multiple rows for the difference regression coefficients. In each case, the interpretation of the p-value and confidence intervals are nearly the same as they were for the single predictor case. However, for the p-value, we need to remember that this is the p-value testing the hypothesis that a particular predictor is zero. The F-statistic is used to test the hypothesis that all predictors are zero, although I won't go into much more detail because I don't place a big emphasis on hypothesis testing in this course.
- The interpretation of R^2 is the same as before, except that now we are considering the ratio of the variance conditioned on ALL predictors to the overall variation in Y; that is,

$$R^2 = 1 - \frac{\sum_i r_i^2}{\sum_i (y_i - \bar{y})^2} \approx 1 - \frac{\text{var}(Y|X_1, X_2)}{\text{var}(Y)}$$

where in the multi-predictor case

$$r_i = Y_i - \left(\hat{\beta}_0 + \sum_{k}^{m} \hat{\beta}_k X_{i,k}\right). \tag{5}$$

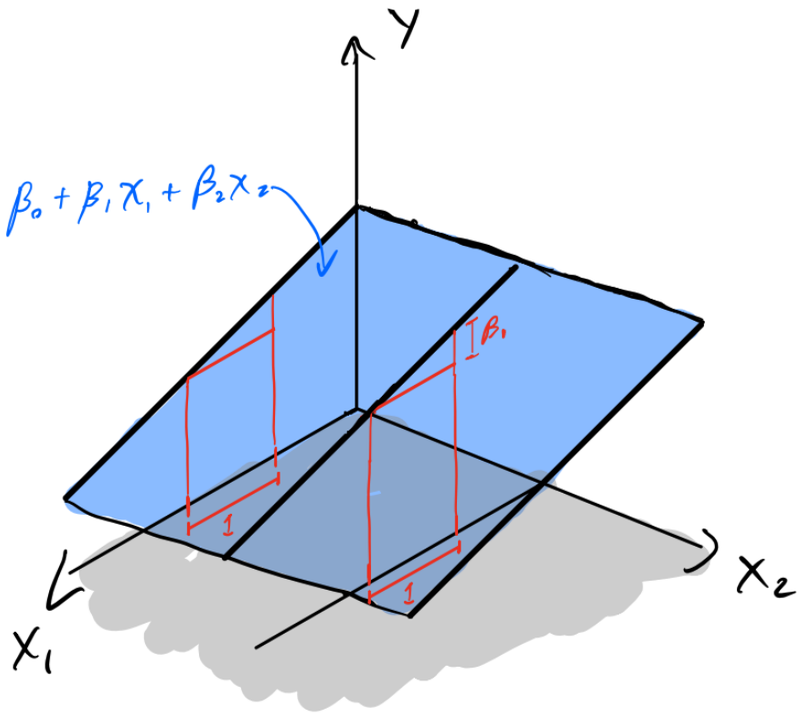

Figure 1. The function $y(x_1, x_2)$

3.1. **Basic interpretation and estimation of the parameters.**

- In order to interpret the parameters, it's easiest to work with just two predictors like we have in the example above. The formula for the conditional expectation of Y is

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{6}$$

where I'm using the shorthand

$$E[Y|X] = E[Y|(X_1, X_2)]$$

to mean the expected value of Y conditioned on both predictors.

Equation 6 is the equation for a flat surface in two dimensions:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{7}$$

A drawing of y is shown in 2.

- If we make a slice through the surface in the x_1 direction and look it at from the side, we see a line with slope β_1 (and similarly for x_2). This leads to the following interpretation of β_i:

$$\beta_1 \text{ is the slope of } E[Y|X] \text{ vs. } X_1 \text{ for fixed } X_2.$$

Notice that in the statement above, even though we are conditioning on both variables, the slope β_1 is independent of which value of X_2 we condition on. We can obtain the interpretation of β_2 by flipping the role of X_1 and X_2. The fact that is doesn't matter which value of X_2 (respectively X_1) we have conditioned on is one of the core model assumption of linear regression with multiple predictors, which we do not encounter in the single predictor case. Another way of articulating it is to say: the "effect" of X_1 and X_2 are not dependent on the other predictors value.

- In the case of children's test scores, we are saying that the association between the mother's high school education and test scores is not influenced by the mother's IQ. that is, If we compare two random children whose mothers have the same IQ, differ in whether they attended high school, then the average *difference* between their test scores will not depend on the IQ of their mothers, although the average magnitude of their test scores will depend on the mother's IQ.

Example 2 (Test scores)**.** *Consider the example of children's test scores, but now we will have two predictors: X_{iq} representing the mother iq and X_{hs} representing whether the mother went to high school.*

Question: Fit the data to a linear regression model with two predictors and answer the questions

(a) What are the regression coefficients and the interpretations?

(b) Based on this regression analysis, which factor, IQ or high school education do we believe is more predictive of test scores?

(c) Overall, how well do high school education and IQ os methods do at predicting the test scores of children?

(d) What is the chance a student whose mother has an IQ of 90 and did not go to high school does better than a student whose mother has an IQ of 110 and did go to high school?

Solutions: We get the following output from `statsmodels` *in the colab notebook:*

```
>                            OLS Regression Results
> ==============================================================================
> Dep. Variable:                      y   R-squared:                       0.214
> ==============================================================================
>                  coef    std err          t      P>|t|      [0.025      0.975]
> ------------------------------------------------------------------------------
> const         25.7315      5.875      4.380      0.000      14.184      37.279
> mom_hs         5.9501      2.212      2.690      0.007       1.603      10.297
> mom_iq         0.5639      0.061      9.309      0.000       0.445       0.683
>
>
```

(a) For the regression coefficients we find the follow:
- $\beta_{\text{hs}} \approx 5.95$. *This means that among students whose mothers* **have the same IQ**, *a student whose mother attended high school will, on average, have a score that is* 5.95 *points higher than a student whose mother did not.*
- $\beta_{\text{iq}} \approx 0.56$. *This means that among students whose mother's* **have the same high school education** *(either they all attended or did not attend high school), the difference between scores of students whose mothers IQ differs by one point is, on average,* 0.56 *points.*

- $\hat{\beta}_0 \approx 26$. *Mathematically, this tells us the average score of students whose mother did not attend high school and have zero IQ, but this is not a meaningful quantity since noone has zero iq. We can therefore ignore it when it comes to interpreting the output.*

(b) Clearly β_{hs} is smaller, but we need to remember that are comparing quantities that have different units. X_{iq} takes values from around 70 to 130, while X_{hq} is either zero or 1. What is actually more useful is to compare how much a difference in one standard deviation of the predictor makes. For example, $\beta_{\text{iq}}\sigma_{\text{iq}}$ is the average difference in test scores between students whose mothers have the same high school education, but whose mother's IQ differ by one standard deviation. To this end, we can compute the following measures of effects

$$\hat{\beta}_{\text{hs}}\hat{\sigma}_{\text{hs}} \approx 2.44$$
$$\hat{\beta}_{\text{iq}}\hat{\sigma}_{\text{iq}} \approx 8.44.$$

The association between IQ and scores is actually larger. Note that the comparison is not perfect, since X_{hs} is a binary variable, but it still gives us a generally idea of the effects.

(c) The R^2 value is 0.214, *so about* 20% *of the variation in test scores is explained by the variation in high school education and IQ of mothers.*

(d) In the colab notebook we calculate this to be about 25%.

3.2. Interpretation of regression coefficient: a deeper look.

- We can express the regression coefficients explicitly in terms of conditional averages as

$$\beta_1 = E[Y|X_1 = (x+1), X_2] - E[Y|X_1 = x, X_2]. \tag{8}$$

Now let's think about how the regression coefficients are related to covariance. One guess would be that, just as in the single-predictor case, β_1 is given by $\text{cov}(Y, X_1)/\sigma^2_{x_1}$. After all, if we look a slice of the 2D planer function $y(x_1, x_2)$ along the x_1 direction, we get the same slope for all x_2. It stands to reason that if we look at only the points in the x_1-y plane our regression slope would be β_1. However, **this argument assumes that when we change x_1, x_2 does not also change**. This is best understood with an example.

Example 3 (Test scores with multiple vs. single predictors)**.** *Here we will consider once again the example of children's test scores and compare using both predictors in the sample above to the results we obtain we using only one predictor (high school education).*

Question: What is the difference between the coefficient of X_{hs} when this is the only predictor and the coefficient when X_{iq} is also used? How is the coefficient in the multiple predictor case related to coefficient in the single predictor case?

Solution: When we performed the regression using only the mother's high school education as a predictor, we obtained a coefficients of about $\hat{\beta}'_{\text{hs}} \approx 12$ and $\hat{\beta}'_0 \approx 78$ (i'll use β' indicate coefficients in the single predictor model, as opposed to the multiple predictor model). The fitted model is

$$\hat{y} = 12X_{\text{hs}} + 78$$

while when also using X_{iq} as a predictor, the coefficient is about half that.

In the model with one predictor, the regression coefficient of 12 *means that on average a student whose mother went to high school will do* 12 *points better than one whose mother did not. That is, we are predicting*

$$E[Y|X_{\text{hs}} = 1] - E[Y|X_{\text{hs}} = 0] = \beta'_{\text{hs}} \approx 12$$

Let's compare this to what we would predict in the model with two predictors. In that case, the average test score of student whose mother went to high school is

$$\begin{aligned}\hat{y}_{\text{hs}} &\approx E[Y|X_{\text{hs}}=1]\\ &= E[\beta_0+\beta_{\text{hs}}+\beta_{\text{iq}}X_{\text{iq}}|X_{\text{hs}}=1]\\ &= \beta_0+\beta_{\text{hs}}+\beta_{\text{iq}}E[X_{\text{iq}}|X_{\text{hs}}=1]\\ &\approx 6\times 1+26+0.6\overline{X}_{\text{iq|hs}}\end{aligned}$$

where

$$\overline{X}_{\text{iq|hs}} = \textit{sample average IQ of mother who attended high school} \approx E[X_{\text{iq}}|X_{\text{hs}}=1]$$

On the other hand

$$\hat{y}_{\text{no-hs}} = 6\times 0+26+0.6\overline{X}_{\text{iq,no-hs}}$$

where

$$\overline{X}_{\text{iq|no-hs}} = \textit{sample average IQ of mother who DID NOT attend high school} \approx E[X_{\text{iq}}|X_{\text{hs}}=0]$$

Thus, according to the model with two predictors, the average difference in test scores between the `hs` *and* `no-hs` *groups is*

$$\Delta\hat{y}_{\text{hs}} = 6+0.6(\overline{X}_{\text{iq|hs}}-\overline{X}_{\text{iq|no-hs}})$$

or written in terms of more probabilistic notation

$$E[Y|X_{\text{hs}}=1]-E[Y|X_{\text{hs}}=0] = \beta_{\text{hs}}+\beta_{\text{iq}}(E[X_{\text{iq}}|X_{\text{hs}}=1]-E[X_{\text{iq}}|X_{\text{hs}}=0])$$

We can compute $\overline{X}_{\text{iq|hs}}-\overline{X}_{\text{iq|no-hs}}\approx 10.3$*, which gives* $\Delta\hat{y}_{\text{hs}}\approx 12$*. Thus, we have calculated the single-predictor regression coefficient from the multiple predictor case.*

- The important thing is that the two predictors are in independent. If they were, then $\overline{X}_{\text{iq|hs}}-\overline{X}_{\text{iq|no-hs}}$ would be zero, and it would have to be that the coefficient of X_{hs} is the same in both cases. We can generalize this to any model where X_1 is a binary predictor to obtain a relationship between the regression coefficient for β_1 with and without the second predictor; that is,

$$\beta_1' = \beta_1+\beta_2(E[X_2|X_1=1]-E[X_2|X_2=0])$$

  where β_1' is the regression coefficient without using X_2 as a predictor in our model.
- Now we will dig deeper into the underlying, math, with the goal of better understanding how the relationship between predictors shapes the regression coefficients. A byproduct of this exploration will be formulas for the regression coefficients in terms of covariances between the predictors, and covariance between the predictors and the response variable. These formulas generalize the relationship $\text{cov}(X,Y)=\beta_1\sigma_x^2$, which we discovered to hold in the single predictors case.

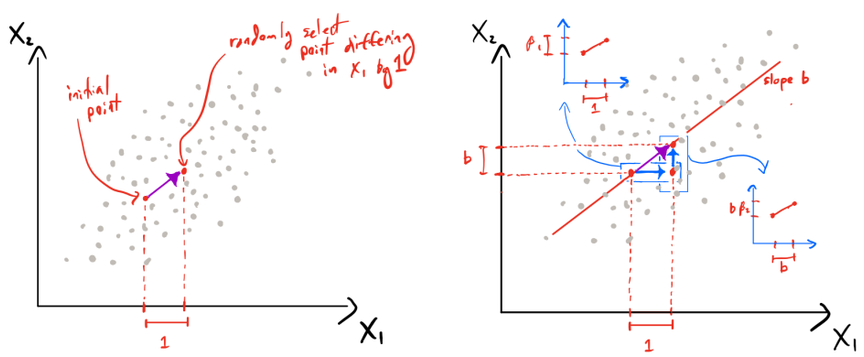

Figure 2. When we increase x_1 by 1, x_2 changes by b (which is the slope between x_1 and x_2 here, not the intercept.)

- Consider a generic linear regression model with two predictors. We will set $\beta_0 = E[X_1] = E[X_2] = 0$ for simplicity, since these cancels out in the end. We start by computing $\text{cov}(X_1,Y)$, which is simply $E[X_1Y]$ since $E[X_1]=E[Y]=0$. Just as we did for the single-predictor case (week 3), we write

$$\begin{aligned}\text{cov}(X_1,Y) = E[X_1Y] &= E[X_1E[Y|X_1]]\\ &= E[X_1(\beta_1X_1+\beta_2X_2)] = \beta_1E[X_1^2]+\beta_2E[X_1X_2]\\ &= \beta_1\sigma_{x_1}^2+\beta_2\text{cov}(X_1,X_2)\end{aligned}$$

  where we have used that, since $E[X_1]=E[X_2]=0$, $\text{var}(X_1)=E[X_1^2]$ and $\text{cov}(X_1,X_2)=E[X_1X_2]$. If we do the same for X_2, we get two equations

$$\begin{aligned}\text{cov}(X_1,Y) &= \beta_1\sigma_{x_1}^2+\beta_2\text{cov}(X_1,X_2)\\ \text{cov}(X_2,Y) &= \beta_2\sigma_{x_2}^2+\beta_1\text{cov}(X_1,X_2)\end{aligned}$$

  As with the single-predictor case, it is very useful to represent β_1 and β_2 as expectations which can be computed as averages over our data points. In addition to providing some insight into the meaning of the regression coefficients, this will yield candidates for our estimators of these quantities. Since everything in

the equation can be represented as a some sort of expectation except the coefficients β_1 and β_2, we just need to solve for these coefficients. Solving the linear system

$$\begin{bmatrix} \text{cov}(X_1, Y) \\ \text{cov}(X_2, Y) \end{bmatrix} = \begin{bmatrix} \sigma^2_{x_1} & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \sigma^2_{x_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

yields

$$\begin{aligned} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} &= \begin{bmatrix} \sigma^2_{x_1} & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \sigma^2_{x_2} \end{bmatrix}^{-1} \begin{bmatrix} \text{cov}(X_1, Y) \\ \text{cov}(X_2, Y) \end{bmatrix} \\ &= \frac{1}{\sigma^2_{x_2}\sigma^2_{x_1} - \text{cov}(X_1, X_2)^2} \begin{bmatrix} \sigma^2_{x_2} & -\text{cov}(X_1, X_2) \\ -\text{cov}(X_1, X_2) & \sigma^2_{x_1} \end{bmatrix} \begin{bmatrix} \text{cov}(X_1, Y) \\ \text{cov}(X_2, Y) \end{bmatrix} \end{aligned}$$

After using the formula for the inverse of 2×2 matrix, we obtain

$$\beta_1 = \frac{\text{cov}(X_1, Y)\sigma^2_{x_2} - \text{cov}(X_2, Y)\text{cov}(X_1, X_2)}{\sigma^2_{x_2}\sigma^2_{x_1} - \text{cov}(X_1, X_2)^2} \tag{9}$$

The formula is particularly revealing if all the variances are set to one

$$\beta_1 = \frac{1}{1 - \rho^2_{1,2}}(\rho_1 - \rho_{1,2}\rho_2)$$

where $\rho_{1,2}$ is the correlation coefficient between X_1 and X_2. Notice that if X_1 and X_2 are uncorrelated ($\rho_{1,2} = 0$), we obtain the usual connection between the regression coefficient and the correlation coefficient between X_1 and X_2.

Example 4 (Correlated predictors)**.** *Consider the model*

$$\begin{aligned} X_1 &\sim \text{Normal}(0, 1) \\ X_2|X_1 &\sim \text{Normal}(bX_1, 1 - b^2) \\ Y|(X_1, X_2) &\sim \text{Normal}(\beta_1 X_1 + \beta_2 X_2, \sigma^2). \end{aligned}$$

Question:
(a) Show that $\text{var}(X_1) = \text{var}(X_2) = 1$ *and* $\text{cov}(X_1, X_2) = b$
(b) Expression β_1 *as a function of* b.
(c) Test Equation 9 with simulations.

Solution:
(a) By definition of the model $\text{var}(X_1) = 1$ *and*

$$\begin{aligned} \text{var}(X_2) &= b^2\text{var}(X_1) + 1 - b^2 = b^2 + 1 - b^2 = 1 \\ \text{cov}(X_1, X_2) &= b\text{var}(X_1) = b \end{aligned}$$

(b) We can write Equation 9 as

$$\beta_1 = \frac{\text{cov}(X_1, Y) - \text{cov}(X_2, Y)b}{1 - b^2}$$

(c) See colab notebook.

- This can all be generalized to the situation where we have many predictors. The general formula for the regression coefficient in terms of expectation is

$$\begin{aligned} \beta_i = E[Y|X_1, \ldots, X_{i-1}, X_i = x_i + 1, X_{i+1}, \ldots, X_K] \\ - E[Y|X_1, \ldots, X_{i-1}, X_i = x_i, X_{i+1}, \ldots, X_K] \end{aligned}$$

Note how this is a very natural extension of Equation 8. We get a more complex expression for the coefficients but the idea is the same.

4. Collinearity and sloppiness

- **The sample distribution of coefficients** Just as before, we want to understand what the sample distribution of the coefficients looks like. In the multiple predictor case, we need to think about the joint distribution of $(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_M)$. We will start by focusing on the two predictor case.

Example 5 (Predictor sample distribution)**.** *Consider the model in 4. Let's look at the sample distribution by fitting many simulated replicates.*

Question: Write a function to generate a dataframe containing samples from the sample distribution of $(\hat{\beta}_1, \hat{\beta}_2)$*. Make a scatter plot and explore the structure of the sample distribution, in particular it dependence on b, which controls the correlations between* X_1 *and* X_2.

Solution: See colab notebook

To better understand what is going on, imagine X_1 and X_2 are very highly correlated (if they are perfectly correlated we say they are **colinear**). We can then write

$$\begin{aligned} Y &= \beta_1 X_1 + \beta_2 X_2 + \epsilon \approx \beta_1 X_1 + \beta_2 X_1 + \epsilon \\ &\approx (\beta_1 + \beta_2) X_1 + \epsilon \end{aligned}$$

There are many ways to select β_1 and β_2 so that the surface $\beta_1 x_1 + a_2\beta_2$ is close to the lines, since a change in β_1 can be compensated by a change in β_2. This means that **if we estimate** β_1 **and** β_2 **and then generated**

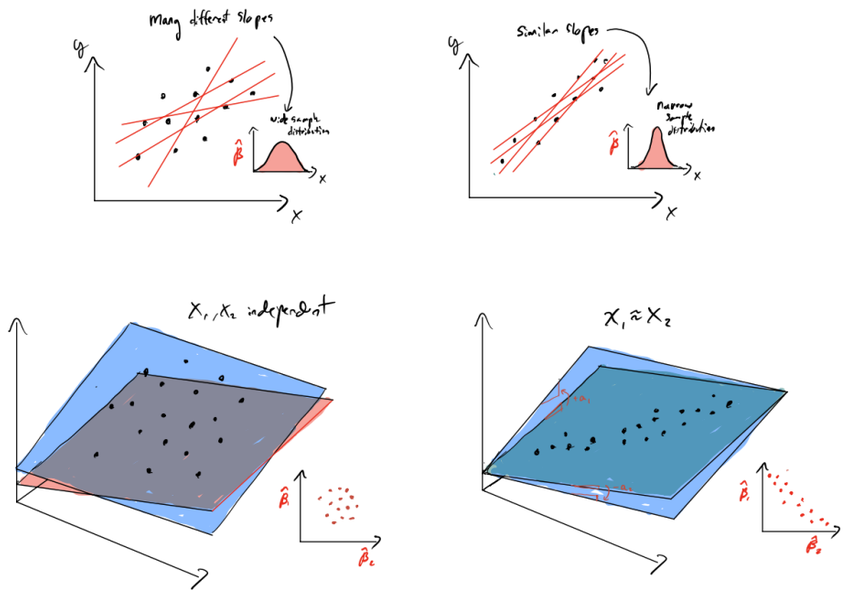

Figure 3. (top) In the single-predictor case, the width of the sample distribution measures how confident we are of a particular slope. It will be narrow if a replicate of our data is likely to produce a very similar slope. These means we get a rough idea of the width of sample distribution by seeing much we can change our regression line and still obtain something that appears to pass through our data. (bottom) In the two predictor case, we have a regression plane and changing a_1 and a_2 will "wiggle" the plane by tilting it in the x_1 and x_2 directions (there is also the intercept which can shift the plane up and down, but I'm not illustrating that). If X_1 and X_2 are uncorrelated, it doesn't matter which way we wiggle it, the fit will be similar, but if X_1 and X_2 are strongly correlated, wiggling the plane in the direction perpendicular to the points has a much smaller effect that parallel to them.

new data, it would be possible to get a VERY different value of $\hat{\beta}_1$ and $\hat{\beta}_2$, so long as $\hat{\beta}_1 + \hat{\beta}_2$ is close to what we got before. This is illustrated in Figure 3 and Figure 4. The following exercises explored in more depth what this means for the sample distribution.

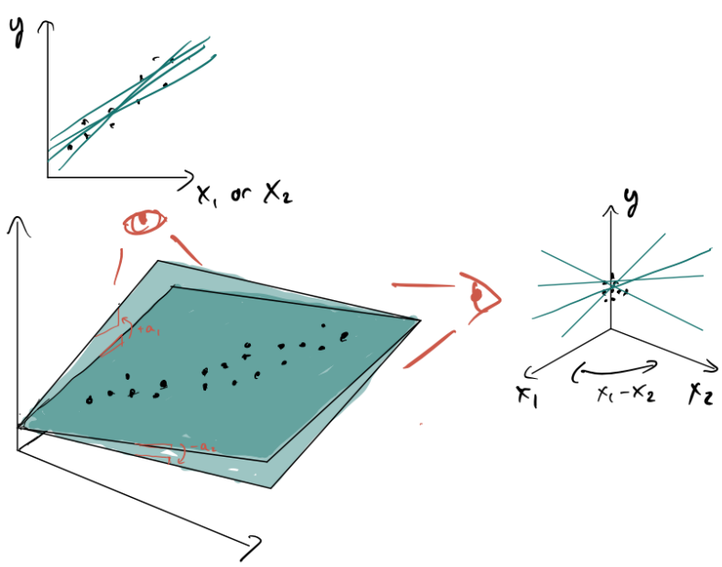

Figure 4. Different views of the data in the case when X_1 and X_2 are correlated. If we look at the data from the side, or along the $X_1 = X_2$ direction, then all our regression planes appear similar; however, when looked at from the "front" as shown in the right panel, we see that the places actually have very different slopes in the other direction.

- **Changing variables.** At this point, you should understand that the sample distribution is related to correlations between x_1 and x_2. Indeed, for a large enough sample, one can show that

$$\hat{\beta}_1 \sim \text{Normal}\left(\beta_1, \sqrt{\frac{\sigma_\epsilon^2 \sigma_{x_1}^2}{\text{cov}(X_1, X_2)^2 - \sigma_{x_1}^2 \sigma_{x_2}^2}}\right) \tag{10}$$

Here, we can see explicitly what happens when X_1 and X_2 become highly correlated – the standard deviation of the sample distribution blows up. When this happens, we will say the model is **sloppy**. How do we deal with this situation? One approach is to use different predictor variables, for example, if $X_1 \approx X_2$, we might simply work with $X_1 + X_2$ as our predictor.

5. Dealing with categorical data

- One situation in which models with multiple predictors frequently arrises is when trying to predict a Y variable based on categorical predictors, such as race. In this case, we need to transform the categories into numerical values. For example, if there are two categories (e.g. YES and NO) we map our variable to 0 or 1. If we have 3 categories (e.g. White, Black, Other), we might first think to map them to 0, 1 and 2. This has a problem though: A chance from 1 to 2 should not necessarily correspond to a change from 0 to 1. In other words, **there is no clear ordering of the x values**. Sometimes we refer to such predictors and **qualitative** rather than **quantitive**, since they express a quality of our data points instead of a numerical quantity.
- To address this issue, we create **dummy variables**. In particular, In order to take a categorical variable and transform it into a set of indicator variables in python, we use the python function `get_dummies`. The usage of this is illustrated in the following example.

Example 6 (Racial disparities in earnings)**.** *Here we will fit the earnings data to a model with race as a predictor. In particular, we want to know: What is the association between race and earnings among adults in the US? We will start with a model using only race as a predictor. One way to approach this would be to simply use a binary predictor and consider only 2 race categories (e.g. White and non-White). This is limiting though. Instead, we can create a variable for each rate category we are interested in. In the dataset there are 4 race categories*

$$\{\text{Black}, \text{White}, \text{Hispanic}, \text{Other}, \text{White}\}$$

In principle, we could create a binary variable for each one (these are what we call dummy variable), to obtain a model like

$$Y = \beta_0 + \beta_{\text{black}} X_{\text{black}} + \beta_{\text{hispanic}} X_{\text{hispanic}} + \beta_{\text{other}} X_{\text{other}} + \beta_{\text{white}} X_{\text{white}} + \epsilon$$

This is problematic though, since at least one of the predictors above MUST be 1. This means that the first 3 of the predictors are perfectly correlated with the other one. By the default, python will drop the first predictor (in alphabetical order), leaving us with the model

$$Y = \beta_0 + \beta_{\text{hispanic}} X_{\text{hispanic}} + \beta_{\text{other}} X_{\text{other}} + \beta_{\text{white}} X_{\text{white}} + \epsilon.$$

Question: Fit the data to the model above. What is the expected disparity in earnings between someone who is white and someone who is hispanic.

Solution: See colab notebook. To answer the question posed above, we begin with the interpretations of the regression coefficients. In terms of conditional expectation, these are

$$\begin{aligned}
\beta_{\text{white}} &= E[Y|X_{\text{white}} = 1, X_{\text{hispanic}} = X_{\text{other}} = 0] - E[Y|X_{\text{white}} = 0, X_{\text{hispanic}} = X_{\text{other}} = 0] \\
&= E[Y|\textit{someone is white}] - E[Y|\textit{someone is black}] \approx 4.9 \\
\beta_{\text{hispanic}} &= E[Y|X_{\text{hispanic}} = 1, X_{\text{white}} = X_{\text{other}} = 0] - E[Y|X_{\text{hispanic}} = 0, X_{\text{white}} = X_{\text{other}} = 0] \\
&= E[Y|\textit{someone is hispanic}] - E[Y|\textit{someone is black}] \approx -0.7
\end{aligned}$$

Our goal however is to compute

$$\begin{aligned}
&E[Y|\textit{someone is white}] - E[Y|\textit{someone is hispanic}] \\
&= E[Y|X_{\text{white}} = 1, X_{\text{hispanic}} = 0, X_{\text{other}} = 0] - E[Y|X_{\text{white}} = 0, X_{\text{hispanic}} = 1, X_{\text{other}} = 0] \\
&= \beta_0 + \beta_{\text{white}} - \beta_0 - \beta_{\text{hispanic}} \\
&= \beta_{\text{white}} - \beta_{\text{hispanic}}
\end{aligned}$$

References

[1] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning (python version)*, volume 112. Springer, 2013.

[2] John Tabak. *Probability and statistics: The science of uncertainty*. Infobase Publishing, 2014.