Matrix Notation

It is very important to be familiar with the matrix notation.

A matrix as a rectangular collection (array) of numbers (or symbols or expressions) arranged into a fixed number of rows and columns.

$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{Ki} + \varepsilon_i$ (LRM)

Here, $i = 1, \ldots, N$. Note that there are N equations where in each equation the dependent variable y is postulated to be linearly related to x_k where $k = 1, \ldots, K$.

Let $\boldsymbol{X}$ represent a N by K matrix of explanatory variables.

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & x_{22} & \cdots & x_{K2} \\ \vdots & \vdots & & \vdots \\ x_{1N} & x_{2N} & \cdots & x_{KN} \end{bmatrix}$$

Linear Regression Models

Let's revisit the ***College*** data from the previous lecture.

The data represent 116 colleges on annual post-college earnings (Earnings in $), the average annual cost (Cost in $), the graduation rate (Grad in %), the percentage of students paying down debt (Debt in %), and the location variable (Location = City or Non-city).

School	Earnings	Cost	Grad	Debt	Location
St. Ambrose Univ.	44800	22920	62	88	City
Albion College	45100	23429	73	92	Non-city
⋮	⋮	⋮	⋮	⋮	⋮
Wittenburg Univ.	42700	26616	64	90	City

For each variable (column), we have 116 observations (rows). The above data information is in a matrix format.

For the X matrix, the first row and the first column would be deleted. Also, the location variable will be converted into a 1/0 dummy variable.

Linear Regression Models

In most regression contexts, the first column is a vector of 1s, which allows for the intercept term β_1 in the model.

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{21} & \dots & x_{K1} \\ 1 & x_{22} & \dots & x_{K2} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{2N} & \dots & x_{KN} \end{bmatrix}$$

To avoid confusion, we will use boldfaced $\boldsymbol{x}$ to denote a column or row of $\boldsymbol{X}$. Subscript k will be used to denote a column (variable) of $\boldsymbol{X}$ and subscript i will be used to denote a row (observations) of $\boldsymbol{X}$.

$$\boldsymbol{x}_k = \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kN} \end{bmatrix} \text{ and } \boldsymbol{x}_i = \begin{bmatrix} 1 \\ x_{2i} \\ \vdots \\ x_{Ki} \end{bmatrix}$$

Here $\boldsymbol{x}_k$ is a $N \times 1$ column vector and $\boldsymbol{x}_i$ is a $K \times 1$ column vector that is the transpose of the i^{th} (1 by K) row of $\boldsymbol{X}$. Therefore, $\boldsymbol{x}_i' = [1 \quad x_{2i} \quad \dots \quad x_{Ki}]$ is simply the i^{th} (1 by K) row of $\boldsymbol{X}$.

Linear Regression Models

Similarly, let $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ be N by 1 column vectors and $\boldsymbol{\beta}$ be a K by 1 column vector of parameters.

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

We specify the linear regression model in a matrix notation as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & \dots & x_{K1} \\ 1 & x_{22} & \dots & x_{K2} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{2N} & \dots & x_{KN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Note that $y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \varepsilon_i$. Alternatively, $y_i = \boldsymbol{x_i'\beta} + \varepsilon_i$

where $\boldsymbol{x_i'\beta} = [x_{1i} \quad x_{2i} \quad \dots \quad x_{Ki}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \varepsilon_i$

The Principle of Least Squares

In order to estimate the parameters $\boldsymbol{\beta}$ under the classical assumptions of a linear regression model, we need a rule, method, or criterion.

For ease of exposition, we use the vector $\boldsymbol{b}$ to denote both the estimator and an estimate.

The least squares regression method, also referred to as ordinary least squares, estimates the regression parameters whereby the residual sum of squares $\sum e_i^2$ is minimized. The residual are defined as $e_i = y_i - \hat{y}_i$, where $\hat{y}_i = \boldsymbol{x_i'b}$ is the predicted value of the dependent variable using $\boldsymbol{b}$ as the estimated coefficient vector.

Linear Regression Models

Residuals: $e_i = y_i - \hat{y}_i = y_i - \boldsymbol{x_i'b}$

In the matrix notation, $\boldsymbol{e = y - \hat{y} = y - Xb}$

The residual sum of squares $\sum e_i^2$ can be written as $\boldsymbol{e'e}$ where

$$\boldsymbol{e'e} = [e_1 \;\; e_2 \;\ldots\; e_N] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = e_1^2 + e_2^2 + \cdots + e_N^2 = \sum e_i^2$$

Therefore, we need to solve a minimization problem where we minimize $S = \boldsymbol{e'e} = (\boldsymbol{y - Xb})'(\boldsymbol{y - Xb})$ with respect to $\boldsymbol{b}$.

Linear Regression Models

Recall single variable optimization with calculus.

Suppose we want to find the relative maximum or minimum of $y = f(x)$.

1st Order Condition: Set the 1st derivative $f'(x) = \frac{dy}{dx} = 0$ and solve for x^*.

2nd Order Condition: Evaluate the 2nd derivative $f''(x)$ at x^*.

- Relative minimum if $f''(x*) > 0$
- Relative maximum if $f''(x*) < 0$

The optimization rule can be extended to multiple variables. Recall that with multiple variables, the 1st order condition sets the partial derivatives w.r.t. the variables to zero.

Here, we use matrix calculus to find the OLS estimates.

Linear Regression Models

Let $\boldsymbol{b}$ be the OLS estimates of $\boldsymbol{\beta}$.

The objective is to minimize $S = \boldsymbol{e'e} = (\boldsymbol{y} - \boldsymbol{Xb})'(\boldsymbol{y} - \boldsymbol{Xb})$ with respect to $\boldsymbol{b}$.

$$S = \boldsymbol{y'y} - \boldsymbol{y'Xb} - \boldsymbol{b'X'y} + \boldsymbol{b'X'Xb}$$

1st Order Condition: Set $\frac{\partial S}{\partial \boldsymbol{b}} = \boldsymbol{0}$

$$\frac{\partial S}{\partial \boldsymbol{b}} = -\boldsymbol{X'y} - \boldsymbol{X'y} + 2\boldsymbol{X'Xb} = \boldsymbol{0}$$

$$-2\boldsymbol{X'y} + 2\boldsymbol{X'Xb} = \boldsymbol{0}; \boldsymbol{b} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}.$$

2nd Order Condition: Evaluate $\frac{\partial^2 S}{\partial \boldsymbol{b}\partial \boldsymbol{b'}}$ at $\boldsymbol{b}$.

$\frac{\partial^2 S}{\partial \boldsymbol{b}\partial \boldsymbol{b'}} = 2\boldsymbol{X'X}$, which is positive definite if $\boldsymbol{X}$ has full rank.

Therefore, minimum at $\boldsymbol{b} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$.

Monte Carlo Simulations

Monte Carlo simulation is a computational technique used to simulate random variables and their interactions. We will simulate data and then estimate the OLS estimates using the formulae. By using the for-loop function in R, we can later verify the sampling properties of the OLS estimators.

Consider X consisting of two column vectors x_1 and $x_2$where x_1 is a vector of 1's and x_2 is drawn randomly from (say) a normal distribution with mean of 80 and standard deviation of 6. Let the sample size be 50 observations. We can set the intercept of 20 and slope of 1.2 (population parameters). Let the error term be normally distributed with mean zero and standard deviation 6.

Estimate the model using the lm function and also using the following two formulae:

$b = (X'X)^{-1}X'y$ (Use the solve function for matrix inversion)

$\hat{\sigma} = \sqrt{\frac{e'e}{N-K}}$, where $e = y - Xb$

We should expect b and $\hat{\sigma}$ to be very close to the chosen population parameters; they should be identical to the ones estimated by the lm function. With the loop, say with 1000 replications, their mean should be almost identical to the chosen population parameters. This is a classic example of Monte Carlo simulation, which is a useful tool in data analytics.

Later, we can use the above Monte Carlo set up to verify the theoretical variance covariance matrix of the OLS estimator derived in class.

Monte Carlo Simulations

Monte Carlo simulation is a computational technique used to simulate random variables and their interactions.

```
n <- 50
x1 <- rep(1, n); x2 <- rnorm(n, 80, 6); x <- cbind(x1,x2)
b1 <- 20; b2 <- 1.2; b <- c(b1,b2)
u <- rnorm(n, 0, 6)
y <- x %*% b + u
Model <- lm(y ~ x2); summary(Model)
bols <- solve((t(x) %*% x)) %*% (t(x) %*% y)
eols <- y - x%*%bols
se <- sqrt(t(eols) %*% eols/(n-2))
par <- t(c(bols,se)); par

# Using for loop
n <- 50
x1 <- rep(1, n); x2 <- rnorm(n, 80, 6); x <- cbind(x1,x2)
b1 <- 20; b2 <- 1.2; b <- c(b1,b2);   nrep <- 1000
 TMat <- numeric()  # empty data object
 for(i in 1:nrep)
 {
  u <- rnorm(n,0,6)
  y <- x %*% b + u
  bols <- solve((t(x) %*% x)) %*% (t(x) %*% y)
  eols <- y - x%*%bols
  se <- sqrt(t(eols) %*% eols/(n-2))
  par <- (c(bols,se))
  TMat <- c(TMat,par)
 }
TMat <- matrix(TMat,nrow=nrep,ncol=3,byrow=TRUE)
colMeans(TMat)
```

Linear Regression Models

The population parameter vector $\boldsymbol{\beta}$ that quantifies the linear relationship is unknown. How do we estimate $\boldsymbol{\beta}$? In order to address this issue, further assumptions are necessary.

Assumptions of the Linear Regression Model

1. The regression model, $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{Ki} + \varepsilon_i$, is linear in the parameters and is correctly specified.

- Note that Assumption 1 requires linearity in the parameters $\beta_1, \beta_2, \ldots, \beta_K$, but not the variables, $y, x_1, x_2, \cdots, x_K$.

- The assumption is not overly restrictive because it allows us to accommodate common nonlinear specifications by making simple transformations of the variables. These transformations include squares and natural logarithms, which capture interesting nonlinear relationships while still allowing easy estimation within the framework of a linear regression model.

- Assumption 1 also requires that we correctly specify the model.

Linear Regression Models

2. The matrix $\boldsymbol{X}$ has full rank; there is no perfect multicollinearity.

- The above assumption implies that there is no exact linear relationship among any of the explanatory variables in the model.

- As we will see later, this assumption is necessary for estimation of the parameters of the model.

- In most applications, some degree of correlation exists between the explanatory variables. The problem with (non-perfect) multicollinearity is similar to that of small samples.

- Multicollinearity (non-perfect) does not violate any of the assumptions; however, its presence results in imprecise estimates of the slope coefficients. In other words, multicollinearity makes it difficult to disentangle the separate influences of the explanatory variables on the dependent variable.

Linear Regression Models

3. The explanatory variables are exogenous; that is, conditional on the explanatory variables, the error term has a zero expected value.

- Using the matrix notation, $\mathrm{E}(\boldsymbol{X}'\boldsymbol{\varepsilon}) = 0$.

Ideally, we would like $\boldsymbol{X}$ to be non-stochastic, which is rare in business and economics applications.

The above less restrictive assumption implies that the error term is not correlated with the explanatory variables.

A common cause for this assumption breaking down is when important explanatory variables are excluded. If one or more of the relevant explanatory variables are excluded, then the resulting OLS estimators are biased.

The extent of the bias depends on the degree of the correlation between the included and the excluded explanatory variables. More later.

Linear Regression Models

4. There is no heteroskedasticity and/or serial correlation.

- The above assumption implies that conditional on $\boldsymbol{X}$, the random error term ε_i is independent and identically distributed (*iid*). In other words, ε_i has the same finite variance σ^2 for all values of i and is uncorrelated with ε_j for all $i \neq j$.
- Using the matrix notation, $E(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}$, where **I** is a $N \times N$ diagonal matrix with 1 diagonal values and 0 non-diagonal values.
- If this assumption breaks down, it does not bias the estimators, but the confidence intervals and significance tests are no longer valid.

5. The error term ε_i is normally distributed.

- The above assumption allows us to construct confidence intervals and conduct tests of significance. If ε_i is not normally distributed, then the confidence intervals and significance tests are valid only in large samples (Central Limit Theorem).

We are now ready to discuss the estimation rule for model parameters.

LRM Properties

As mentioned earlier, for convenience, we will discontinue the boldface convention for matrices and vectors.

$y = X\beta + \varepsilon$ (LRM)

$b = (X'X)^{-1} X'y$ (OLS estimator)

$b = (X'X)^{-1} X'(X\beta + \varepsilon)$

$= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon$

$= \beta + (X'X)^{-1} X'\varepsilon$

- b is a linear transformation of the random variable ε.
- As mentioned earlier, the estimator b is a random variable.
- A particular value of b, using a random sample, is an estimate.

LRM Properties

Properties of b

Unbiasedness

An estimator is unbiased if its expected value equals the population parameter.

$$\mathrm{E}(b) = \mathrm{E}(\beta + (X'X)^{-1} X'\varepsilon)$$

$$= \beta + \mathrm{E}((X'X)^{-1} X'\varepsilon) = \beta + (X'X)^{-1}\mathrm{E}(X'\varepsilon)$$

$$= \beta \text{ (refer to Assumption 3)}$$

b is an unbiased estimator of β.

Unbiasedness is a very important property of an estimator that holds irrespective of whether the estimate is based on a small or a large sample. Intuitively, an unbiased estimator is 'right on target'.

Efficiency: An unbiased estimator is deemed efficient if its variability between samples is smaller than that of other unbiased estimators. In other words, if you take all the unbiased estimators of the unknown population parameter, the efficient estimator will have the least variance.

LRM Properties

$$\mathrm{Var}(b) = \mathrm{E}((b-\beta)(b-\beta)')$$
$$= E\left((X'X)^{-1}(X'\varepsilon)\,\varepsilon'X(X'X)^{-1}\right)$$
$$= (X'X)^{-1}X'X(X'X)^{-1}E(\varepsilon\,\varepsilon')$$
$$= \sigma^2(X'X)^{-1} \text{ (refer to Assumption 4)}$$

It can be shown mathematically that under the classical linear model assumptions (1-4), the OLS estimator is BLUE (Best Linear Unbiased Estimator). In a class of all linear unbiased estimators, OLS is best (efficient) since it has the smallest variance (or standard error).

Consistency. Another desirable property, which is often considered a minimum requirement for an estimator, is consistency. An estimator is consistent if it approaches the population parameter of interest as the sample size increases. Formally, an estimator is said to be consistent if it converges in probability to the unknown parameter; that is, $plim\ b_j = \beta_j$. Again, it can be shown that the OLS estimators are consistent.

Sampling Distribution

Statistical inference is based on the assumption that the random error term ε is normally distributed.

If the normality assumption cannot be justified, then inferences are valid when the sample size is sufficiently large (CLT).

Recall that $b = \beta + (X'\,X\,)^{-1}\,X'\varepsilon$ (a linear function of ε).

Assuming that $\varepsilon \sim N(0_{N\times 1}, \sigma^2 \mathrm{I}_{N\times N})$, then $b \sim N(\beta, \sigma^2 (X'\,X\,)^{-1})$

If σ^2 were known, we could make inferences based on the standard normal or the chi-square distribution.

It can be shown that $\frac{\sum e_i^2}{N-K}$ is an unbiased estimator of σ^2.

With $\frac{\sum e_i^2}{N-K}$, we generally use the t_{df} distribution or the $F_{(df_1, df_2)}$ distribution for making inferences. [We will not derive these distributions]

Monte Carlo Simulation Continued

```
n <- 50
x1 <- rep(1, n); x2 <- rnorm(n, 80, 6)
x <- cbind(x1,x2)
b1 <- 20; b2 <- 1.2
b <- c(b1,b2)
 nrep <- 1000
 TMat <- numeric()  # empty data object
 for(i in 1:nrep)
 {
  u <- rnorm(n,0,6)
  y <- x %*% b + u
  bols <- solve((t(x) %*% x)) %*% (t(x) %*% y)
  eols <- y - x%*%bols
  se <- sqrt(t(eols) %*% eols/(n-2))
  par <- (c(bols,se))
  TMat <- c(TMat,par)
 }
TMat <- matrix(TMat,nrow=nrep,ncol=3,byrow=TRUE)
colMeans(TMat)

b1Mat <- TMat[1:nrep,1]
b2Mat = TMat[1:nrep,2]
sMat = TMat[1:nrep,3]
sd(b1Mat); sd(b2Mat)

VarCov <- (mean(sMat))^2 * solve((t(x) %*% x))
sqrt(diag(VarCov))

Model <- lm(y~x2); summary(Model)
```

Sampling Distribution

Let U be a $m \times 1$ vector of random variables. If $U \sim N(\mu, \Sigma)$ where Σ is the $m \times m$ variance-covariance matrix of U, then

- $Q = AU + b \sim N(A\mu + b, A\Sigma A')$

Chi-square distribution:

If $U \sim N(\mu, \Sigma)$, then $(U - \mu)'\Sigma^{-1}(U - \mu)$ has a χ^2_{df} distribution with $df = m$.

Student t_{df} Distribution

Let Z have a standard normal distribution and Q have a χ^2_{df} distribution. If Z and Q are distributed independently, then $t = \frac{Z}{\sqrt{Q/m}}$ has a t_{df} distribution with $df = m$.

$F_{(df_1, df_2)}$ Distribution

Let Q_1 and Q_2 be independent χ^2_{df} variables with degrees of freedom, $df = m_1$ and m_2, respectively. Then, $F = \frac{Q_1/m_1}{Q_2/m_2}$ has a $F_{(df_1, df_2)}$ distribution with degrees of freedom, $df_1 = m_1$ and $df_2 = m_2$.

Confidence and Prediction Intervals

$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \varepsilon_i$

We can estimate the above model as $\hat{y}_i = b_1 x_{1i} + b_2 x_{2i} + \cdots + b_K x_{Ki}$

Let $x_1^0, x_2^0, \ldots, x_K^0$ denote specific values of the explanatory variables. At these specific values, $\hat{y}_i^0 = b_1 x_{1i}^0 + b_2 x_{2i}^0 + \cdots + b_K x_{Ki}^0$.

- $\hat{y}_i^0$ is the point estimate for both y_i and $E(y_i)$.
- Point estimators are subject to sampling variations. In other words, the estimates will change if we use a different sample to estimate the regression model.
- We use the point estimate along with the margin of error to construct two types of interval estimates.

Confidence Interval: Interval estimate for $E(y_i)$

Prediction Interval: Interval estimate for y_i

The prediction interval is always wider that the confidence interval because it also accounts for the variability due to the random error term ε_i.

Confidence and Prediction Intervals

For specific values $x_1^0, x_2^0, \ldots, x_K^0$, the $100(1-\alpha)\%$ **confidence interval** for the expected value of y_i is:

$$\hat{y}_i^0 \pm t_{\alpha/2,df} se(\hat{y}_i^0)$$

- $\hat{y}_i^0 = b_1 x_{1i}^0 + b_2 x_{2i}^0 + \cdots + b_K x_{Ki}^0$; $se(\hat{y}_i^0)$ is the standard error of $\hat{y}_i^0$
- $t_{\alpha/2,df}$ is the value associated with the probability $\alpha/2$ in the upper tail of the t distribution with $df = N - K$.

For specific values $x_1^0, x_2^0, \ldots, x_K^0$, the $100(1-\alpha)\%$ **prediction interval** for an individual value of y_i is:

$$\hat{y}_i^0 \pm t_{\alpha/2,df} \sqrt{(se(\hat{y}_i^0))^2 + s_e^2}$$

- $\hat{y}_i^0 = b_1 x_{1i}^0 + b_2 x_{2i}^0 + \cdots + b_K x_{Ki}^0$; $se(\hat{y}_i^0)$ is the standard error of $\hat{y}_i^0$
- $t_{\alpha/2,df}$ is defined as above with $df = N - K$
- s_e is the standard error of the estimate

Confidence and Prediction Intervals

Let us revisit the ***College*** data example.

Dependent Variable: Annual post-college earnings (Earnings in $)

Explanatory Variables:

The average annual cost (Cost in $; x_2)
The graduation rate (Grad in %; x_3)
The percentage of students paying down debt (Debt in %; x_4)
If a college is located in a city (City equals 1 if a city location, 0 otherwise; x_5)

LRM: $y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_5 x_{5i} + \varepsilon_i$, where $i = 1,\ldots,116$.

Question: Estimate the model to predict Earnings if Cost = $25,000, Grad = 60, Debt = 80, and City = 1.

$\widehat{\text{Earnings}} = 10{,}004.97 + 0.434 \times 25{,}000 + 178.10 \times 60 + 141.48 \times 80 + 2{,}526.79 \times 1 = 45{,}408.7991$ or about $45,409

$45,409 is the point estimate.

Confidence and Prediction Intervals

R offers an easy way to find the confidence and prediction intervals.

```
> # Import College data for confidence/prediction intervals
> Model <- lm(Earnings ~ Cost + Grad + Debt + City, data = myData)
> predict(Model, data.frame(Cost=25000,Grad=60,Debt=80,City=1), level=0.95, interval="confidence")
      fit      lwr      upr
1 45408.8 43392.99 47424.61
> predict(Model, data.frame(Cost=25000,Grad=60,Debt=80,City=1), level=0.95, interval="prediction")
      fit      lwr      upr
1 45408.8 34041.05 56776.55
```

For the specific values of the explanatory variables:

- The 95% confidence interval is [$43,393, $47,425]
- The 95% prediction interval is [$34,041, $56,777]
- As expected, the prediction interval is wider because it also accounts for the variability caused by the random error term.
- At the 90% level, the confidence and the prediction intervals are [$43,721, $47,096] and [$35,893, $54,924].
- Why are the intervals narrower at 90% confidence?
- $t_{0.025,111} = 2.200985$ and $t_{0.05,111} = 1.795885$ [qt(α/2, *df*, lower.tail=FALSE)

Confidence and Prediction Intervals

Practice Problem: Revisit the **Revenue** data to develop a regression model to predict the revenue for a subscription firm. Explanatory variables include the marketing expenditure (MktPct, in percentage) and the type of sales channel (Channel 1, Channel 2, and Channel 3) used to sell the firm's products.

Estimate an appropriate linear regression model to construct the 95% confidence and prediction intervals for the revenue for Chanel 3 with a marketing expenditure of 20%. Which interval is wider and why?