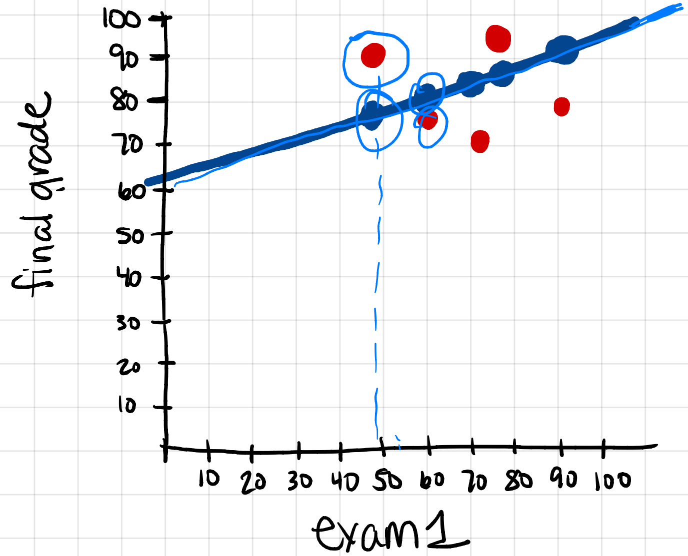
evaluating regression Models

What does it mean to "evaluate a model"?

$\hat{y} = b_0 + b_1 x$

Student ID	Exam1 x	Final grade y	Predicted final grade $\hat{y}$
1	48	90	77
2	60	76	80
3	71	70	84
4	76	94	87
5	90	80	92



Evaluating ⇒

How well did my model predict the students' final grades?

Why evaluate models?

Q1 Does this model add <u>ANY</u> value?

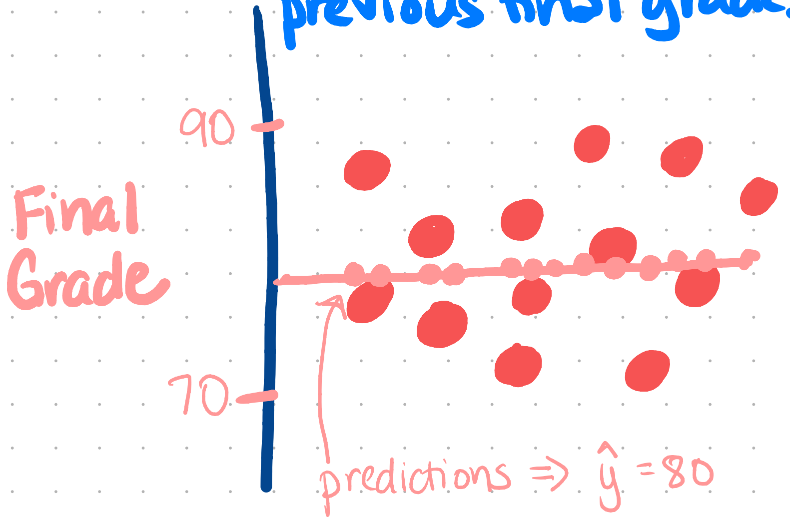
Q2 Which of these 2 models is better?

Q3 How much confidence should I have in this model?

Q1 Does my model add ANY value? Is it "Good Enough"?

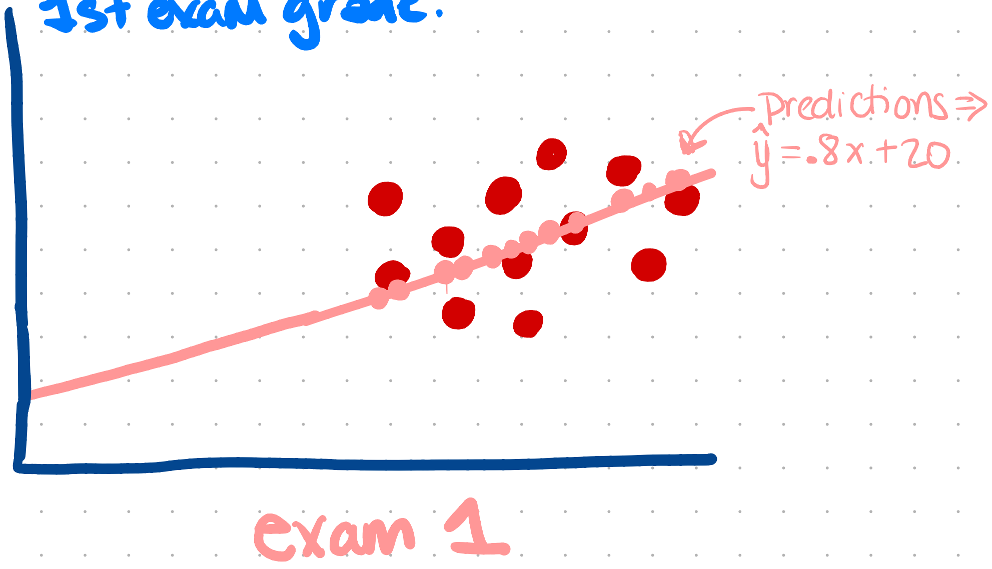
No Model ⇒

Predict final grade using only previous final grades.



My Model ⇒

predict final grade using the student's 1st exam grade.



Which is better?

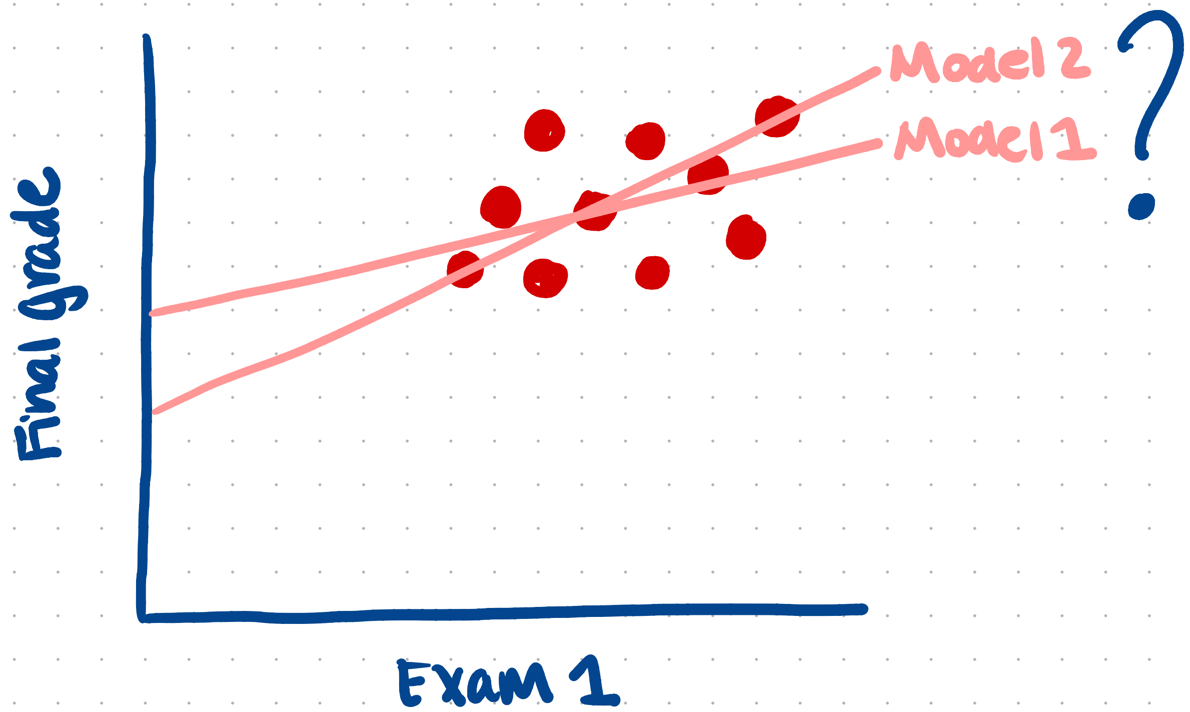
Predict final grade for each student to be 80!

??

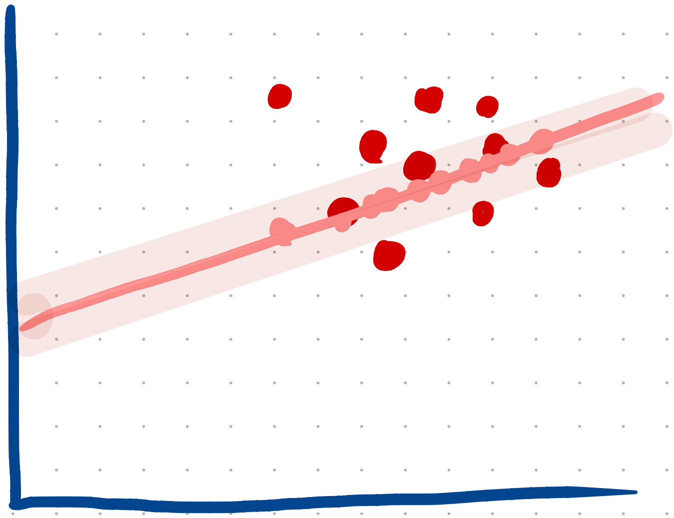
Predict the final grade for each student to be 80% of their 1st exam score + 20 points.

Q2: Which Model is better?

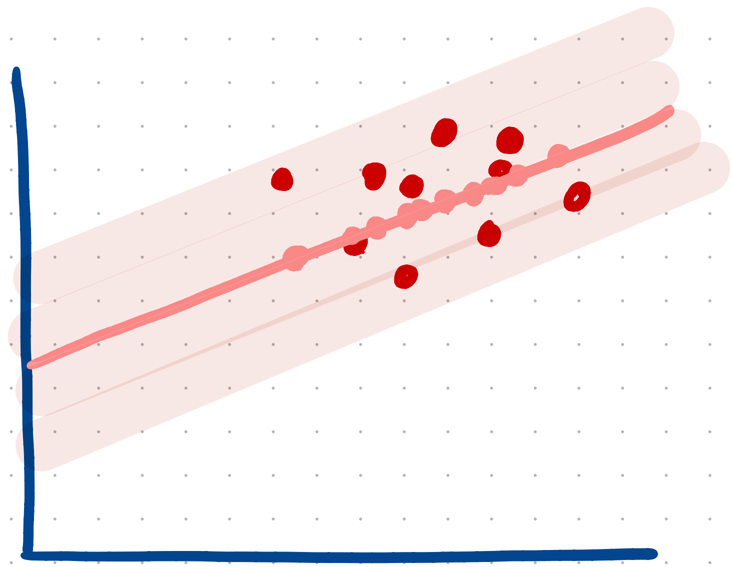
How can I determine which predicts better?



Q3: How much Confidence should I have in this model?



It's gonna be close!



Yeah, we really have no idea...

[Does My Model add value?] [Which Model is better?]

To Answer, we need a $\boxed{\text{single metric}}$ that measures how well a model predicts.

Common Metrics: [Root Mean Squared Error]
Mean Absolute Error
Median Absolute Error
[Mean Squared Error]

RMSE (via MSE)

Manual calculation

RESIDUALS Error of each prediction ⇒ **SSE** square each error then sum ⇒ **MSE** Avg of the squared errors ⇒ **RMSE** sqrt of the avg.

$\hat{y}_1 - y_1 = r_1$
$\hat{y}_2 - y_2 = r_2$
$\vdots$
$\hat{y}_n - y_n = r_n$

residuals

$$r_1^2 + r_2^2 + \dots + r_n^2 = \textbf{SSE} \rightarrow \frac{SSE}{n} = \textbf{MSE} \rightarrow \sqrt{MSE} = \textbf{RMSE}$$

n = # of observations/rows
r_1 = residual, error for the 1st row/observation
$\hat{y}_1$ = **predicted** value for the 1st row/observation
y_1 = **actual** value for the 1st row/observation

RMSE in plain english:
"our predictions are off by an average of 5 pts on the final grades."

RMSE: sklearn.metrics.mean_squared_error
math.sqrt

```
RMSE =
  sqrt(mean_squared_error(df.y, df.yhat))
```

Does my model work better than having no model?

[guessing? pred. mean? pred mode?]

* Applies to ordinary least squares algorithms such as statsmodels.api.ols & sklearn.linear_model.LinearRegression

Answer: run an f-regression test on R^2 [R^2 = the explained variance score = the coefficient of determination = (pearson's R)^2 (when simple univariate regression)]

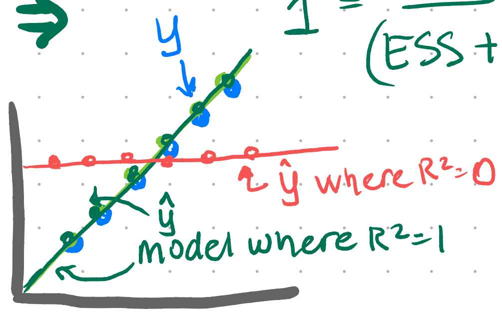
if the p-value of the f-statistic is $< \alpha$ ($\sim .05$), then our model is better than having no model.

$\boxed{R^2}$

$$R^2 = \frac{ESS}{TSS} = \frac{\text{Explained Sum of Squares}}{\text{Total Sum of Squares}} = \frac{ESS}{(ESS + SSE)} = \frac{\text{Explained SS}}{(\text{Explained SS} + \text{SS Error})}$$

Manual | python

$R^2 = 1 \Rightarrow$



$$1 = \frac{ESS}{(ESS + SSE)} \Rightarrow (\cancel{ESS} + SSE) = \cancel{ESS} \Rightarrow SSE = 0$$

<ALGEBRA>

NO ERROR!
Predictions = Actual
$\hat{y} = y$

$$R^2 = 0 \Rightarrow 0 = \frac{ESS}{(ESS + SSE)} \Rightarrow (ESS + SSE) \cdot 0 = ESS \Rightarrow 0 = ESS$$

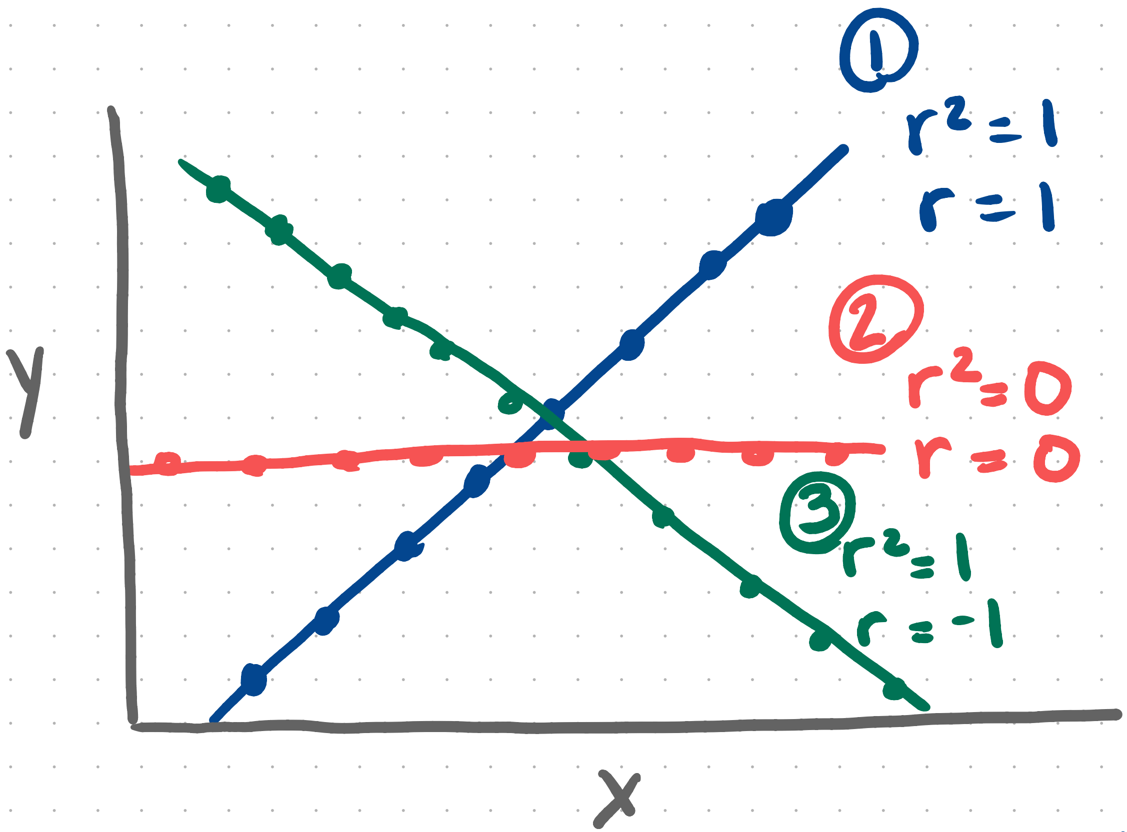
Nothing is explained! All error!

```
sklearn.metrics.explained_variance_score(df.y, df.yhat)
statsmodels.api.ols => my_ols_model.rsquared
```

$\boxed{\text{p-value}}$

```
my_ols_model.f_pvalue / sklearn.feature_selection.f_regression
```

r & r^2



1. As x↑, y↓ by a constant value
2. As x↑, y↓ by a constant value
3. As x↑, y doesn't change ⇒
   x does not affect y in ANY way

r → ∘ Perfect positive correlation
r^2 → ∘ The variance in x explains ALL of the variance in y

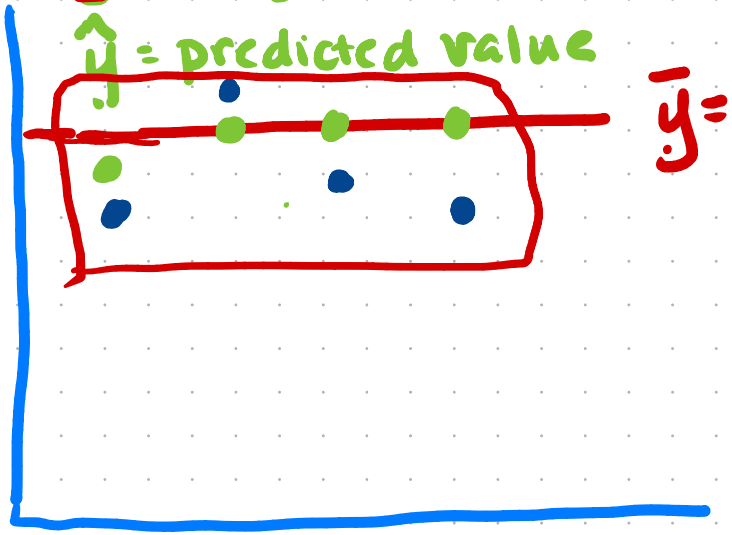
r → ∘ Perfect negative correlation
r^2 → ∘ The variance in x explains ALL of the variance in y

r → No correlation
r^2 → The variance in x explains NONE of the variance in y.

y = Actual value

$\bar{y}$ = Avg. value

$\hat{y}$ = predicted value



$\bar{y} = 80$

final grade

exam 1

Explained SS:

$$\hat{y}_1 - \bar{y} = \cancel{80}\ 75 - 80 = 0^2 \quad (-5)^2$$
$$\hat{y}_2 - \bar{y} = 80 - 80 = 0^2$$
$$\hat{y}_3 - \bar{y} = 80 - 80 = 0^2$$
$$\hat{y}_4 - \bar{y} = 80 - 80 = 0^2$$

$ESS = \cancel{0}\ 25$

$$R^2 = \frac{ESS}{TSS} = \frac{0}{TSS} = 0$$
$$= \frac{25}{270} = \boxed{.09}$$

$\hat{y}$ - pred.

$\bar{y}$ - mean of all y's

y - actual

Total SS:

$$\boxed{y_1 - \bar{y}} = 70 - 80 = (-10)^2$$
$$y_2 - \bar{y} = 85 - 80 = (5)^2$$
$$y_3 - \bar{y} = 72 - 80 = (-8)^2$$
$$y_4 - \bar{y} = 71 - 80 = (-9)^2$$

$TSS = 270$

Vocabulary

Model

baseline

residual

r - pearson's r

r^2 - coeff. of det / explained variance score

RMSE

TSS

ESS

SSE

$$\text{SSE} = (\underset{\text{Actual}}{y_1} - \underset{\text{Predicted}}{\hat{y}})^2 + (y_2 - \hat{y})^2 + \dots + (y_4 - \hat{y})^2$$

baseline predictions = predict every final grade to be the average of All = $\boxed{80}$

$$\text{SSE} = (70 - \boxed{\hat{80}})^2 + (88 - \hat{80})^2 + (72 - \hat{80})^2 + (71 - \hat{80})^2$$

$$= 270$$

$$\text{TSS} = \underbrace{(y_1 - \boxed{\bar{y}})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + (y_4 - \bar{y})^2}$$

$$(70 - \bar{80})^2 + \dots + (71 - \bar{80})^2$$

$$= 270$$

$\therefore$ SSE = TSS when predictions are the average of y's.

The ↑er ESS - the better the model!

If SSE is "close" to TSS, then model isn't great!

$$\boxed{\text{TSS} = \text{SSE} + \text{ESS}}$$

$$\cancel{270} = \cancel{270} + \text{ESS}$$

$$0 = \text{ESS}$$