Data Analysis In-Class Worksheet #09: Regression 1 of 2

Regression: Definitions

Does temperature predict the ice cream sales?

Temperature is the ☑ independent variable ☐ dependent variable

Ice cream sales is the ☐ independent variable ☑ dependent variable

Business Question #1

Does the number of BEDs predict the number of BATHs?

BED is the ☑ independent ☐ dependent variable.

BATH is the ☐ independent ☑ dependent variable.

To answer this question, we can rephrase it as a pair of hypotheses:

H_0: No of beds don't predict number of baths.

H_1: No of beds predict number of baths.

Use the FROZEN class data set to answer the following questions.

The regression line that best fits the BED/BATH data from class is

BATH_hat = _0.584_____ + ____0.824____ BED

The first parameter is called ☑ intercept ☐ slope.

The first parameter is called ☑ β_0 ☐ β_1.

The first parameter is called ☐ weight ☑ bias ☐ hyperparameter in AI.

Is this parameter's value significantly different from zero? ☐ yes ☑ no

The second parameter is called ⬜ intercept ✅ slope.

The second parameter is called ⬜ β_0 ✅ β_1.

The second parameter is called ✅ weight ⬜ bias ⬜ hyperparameter in AI.

The t test score value for the second parameter = $\frac{0.824}{0.140}$ = ____5.887_____.

Is this parameter's value significantly different from zero? ✅ yes ⬜ no

Each extra BED is associated with an increase of _0.824__ BATH.

A house with 0 BED would have _0.584___ BATH, based on this prediction model.

How does the standardized coefficient (slope) differ from the unstandardized coefficient (slope)?

- **Unstandardized slope** → talks in the *real-world units*.
    - Example: "Each extra bedroom adds about **0.824 bathrooms**."
    - It uses bedrooms and bathrooms directly.
- **Standardized slope** → talks in *standard units* (standard deviations).
    - Example: "If bedrooms go up by 1 standard chunk, bathrooms go up by **0.524 standard chunks**."
    - This doesn't use bedrooms/bathrooms directly — it uses a common scale so we can compare different predictors (like bedrooms vs. square feet).

In short:

- **Unstandardized** = real-world effect in actual units (easy to explain).

- **Standardized** = effect in a common scale (easy to compare across variables).

How does BATH_hat differ from BATH?

BATH (actual) → the **real number of bathrooms** a house has (observed data).

BATH_hat (predicted) → the **number of bathrooms our regression model predicts** based on bedrooms (using the formula 0.584+0.824×BED).

Calculate your own BATH_hat

BATH_hat = _____ + _____ BED

Calculate your own residual:

Residual = BATH - BATH_hat = _____ - _____ = _____

The model ⬜over-estimated ⬜ under-estimated my number of bathrooms.

If **Residual > 0** → the model **under-estimated** your bathrooms.

If **Residual < 0** → the model **over-estimated** your bathrooms.

Write up this finding in APA style:

Regression analysis was used to test if the number of bedrooms (BED) significantly predicted the number of bathrooms (BATH).

Results indicated that the predictor explained __27.4_____% of the variance (Adjusted R^2 =__0.266_____), F(___1__ , __92___) = ___34.65_____, p __<0.001_____.

Business Question #2

Does the number of BEDs predict Price?

BED is the ☑ independent ☐ dependent variable.

Price is the ☐ independent ☑ dependent variable.

To answer this question, we can rephrase it as a pair of hypotheses:

H_0: The number of bedrooms (BED) does **not** significantly predict house price.

H_1: The number of bedrooms (BED) **does** significantly predict house price.

Use the FROZEN class dataset to answer the following questions.

The regression line that best fits the data from class is

Price_hat = __59,285_____ + _217,187_____ BED

Pearson's correlation r between Price and BED is ___0.315_____.

The model's R^2 is ____0.099_____, which is ___$(0.315)^2$.

Each extra BED is associated with an increase of _$217,187_____ in Price.

A house with 0 BED would be estimated at Price of _$59,285_____, based on this prediction model.

Calculate your own Price_hat:

Price_hat = _____ + _____ BED

Calculate your own residual:

Residual = _____ - _____ = _____

The model ☐over-estimated ☐ under-estimated my Price.

Write up this finding in APA style:

Regression analysis was used to test if the number of bedrooms (BED) significantly predicted house price. Results indicated that the predictor explained ___9.9____% of the variance (Adjusted R^2 =_0.089__), F(___1__ , _92__) = ___10.14__, p _=0.002_.