Linear Regression Practice From Slides

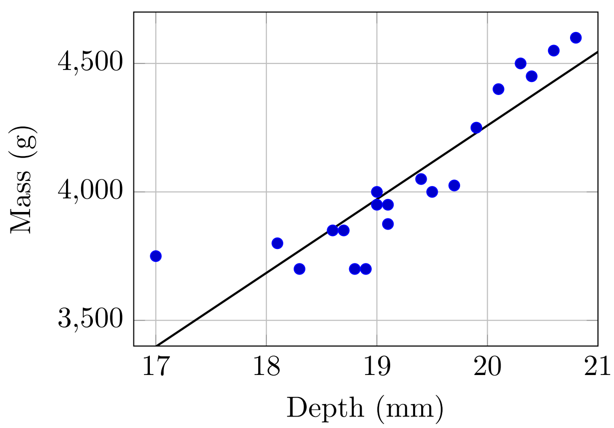
Mr. Merrick · AP statistics · October 21, 2025

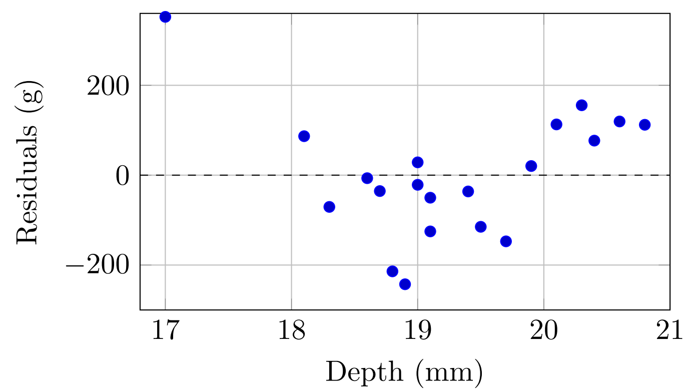
Part A — Standard Regression Examples

Example 1: Penguin Body Mass

Experimental data for Adelie penguins:

Culmen Depth (mm)	Mass (g)	Culmen Depth (mm)	Mass (g)
17.0	3750	19.1	3875
18.1	3800	19.4	4050
18.3	3700	19.5	4000
18.6	3850	19.7	4025
18.7	3850	19.9	4250
18.8	3700	20.1	4400
18.9	3700	20.3	4500
19.0	3950	20.4	4450
19.0	4000	20.6	4550
19.1	3950	20.8	4600





1. **Verify assumptions for linear regression (LINER):**
   (a) **Linearity:** Does the scatterplot suggest a roughly linear trend?
   Yes — it shows an approximately linear upward trend, though the residuals suggest slight curvature. Proceed with caution.
   (b) **Independence:** Are the observations reasonably independent?
   Treat as independent for this exercise — each penguin is measured once. The sample of penguins is small with respect to the population.
   (c) **Normality:** Do the residuals appear roughly normal?
   Residuals are fairly symmetric, with no extreme outliers. A histogram or normal probability plot would support that they are roughly symmetric with no strong skew.
   (d) **Equal variance:** Do the residuals show consistent spread across all fitted values?
   Not perfectly — residuals spread slightly wider at larger depths (mild heteroscedasticity). Proceed with caution.
   (e) **Random sampling:** Were the data collected randomly and representatively?
   Assume the penguins are sampled randomly and representative for this example.

2. Regression equation: $\widehat{\text{Mass}} = a + b \cdot (\text{Depth})$
$\widehat{\text{Mass}} = -1479.48 + 286.892\,(\text{Depth})$ g

3. Interpret a and b:

   $b \approx 286.9$ g/mm: each extra mm of depth predicts ~ 287 g higher mass. a centers the line (Depth $= 0$ not in range).

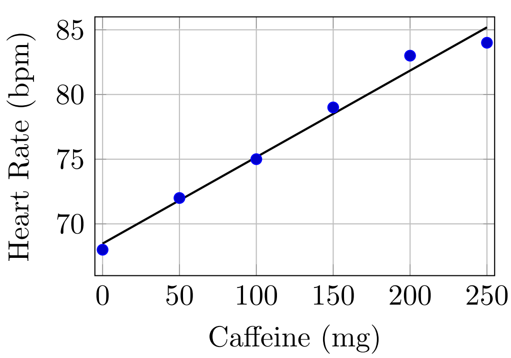
4. Compute and interpret R^2 and s:

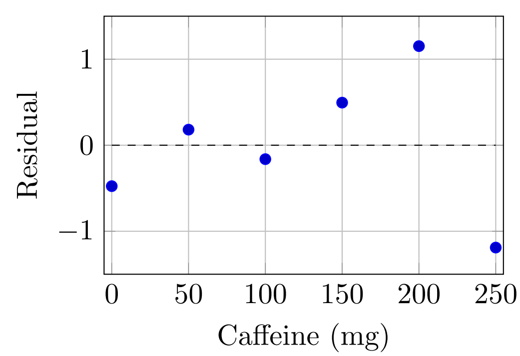
   $R^2 = 0.7874$ (78.7% explained); $s = 143.09$ g.

Example 2: Caffeine and Heart Rate

Experimental data:

Caffeine (mg)	**Heart Rate (bpm)**
0	68
50	72
100	75
150	79
200	83
250	84





1. Assume all conditions for linear regression are met. Determine the regression equation:
The least-squares regression line is $\widehat{\text{HR}} = 68.476 + 0.066857 \cdot \text{Caffeine}$.

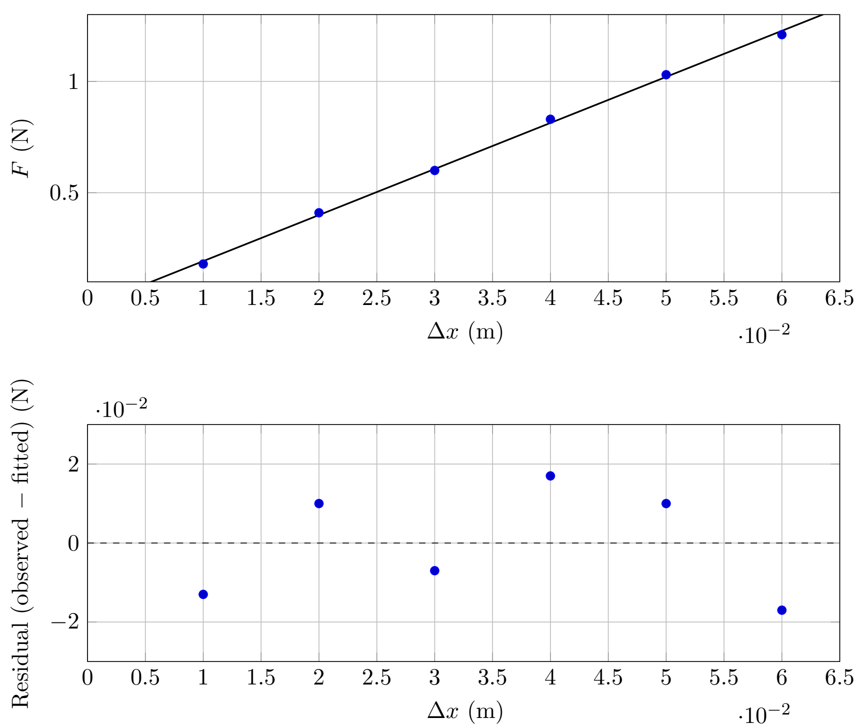
2. Interpret a and b:
The slope is about 0.0669 bpm per mg of caffeine. This means that, on average, the model predicts an increase of 0.0669 beats per minute in heart rate for each additional milligram of caffeine consumed. The intercept is about 68.48 bpm. This represents the predicted average resting heart rate when no caffeine is consumed (0 mg). While this is within the range of the data, interpretation should still be made with caution.

3. Compute and interpret R^2 and s:
The coefficient of determination is $R^2 = 0.9835$, which means that about 98.35% of the variation in heart rate is explained by the linear relationship with caffeine intake. The standard deviation of the residuals is $s = 0.905$ bpm, meaning that typical predictions of heart rate from the regression line are off by about 0.9 bpm.

Example 3: Hooke's Law

Hooke's law states $F = k\Delta x$. Data:

Displacement (m)	**Force (N)**
0.01	0.18
0.02	0.41
0.03	0.60
0.04	0.83
0.05	1.03
0.06	1.21



1. Regression equation: $\widehat{F} = a + b\Delta x$
   The least-squares regression line is $\widehat{F} = -0.0140 + 20.6857\,\Delta x$.

2. Interpret a and b:
   The slope is about 20.686 N/m. This means that, on average, the model predicts an increase of about 20.686 newtons of force for every additional meter of displacement of the spring. This value represents the spring constant k. The intercept is about -0.014 N. This represents the predicted force when displacement is 0 m. Because an ideal spring would give 0 N at 0 m, this small negative intercept likely reflects rounding or measurement error.

3. Compute and interpret R^2 and s:
   The coefficient of determination is $R^2 = 0.9987$, which means that about 99.87% of the variation in force is explained by the linear relationship with displacement. The standard deviation of the residuals is $s = 0.01563$ N, which means that typical predictions of force from the regression line are off by about 0.016 newtons.

Transformations and Linearization

Cars: Horsepower vs. Fuel Efficiency

Car	hp	mpg	Car	hp	mpg
Mazda RX4	110	21.0	Dodge Challenger	150	15.5
Mazda RX4 Wag	110	21.0	AMC Javelin	150	15.2
Datsun 710	93	22.8	Camaro Z28	245	13.3
Hornet 4 Drive	110	21.4	Pontiac Firebird	175	19.2
Hornet Sportabout	175	18.7	Fiat X1-9	66	27.3
Valiant	105	18.1	Porsche 914-2	91	26.0
Duster 360	245	14.3	Lotus Europa	113	30.4
Merc 240D	62	24.4	Ford Pantera L	264	15.8
Merc 230	95	22.8	Ferrari Dino	175	19.7
Merc 280	123	19.2	Maserati Bora	335	15.0
Merc 280C	123	17.8	Volvo 142E	109	21.4
Merc 450SE	180	16.4	Chrysler Imperial	230	14.7
Merc 450SL	180	17.3	Lincoln Continental	215	10.4
Merc 450SLC	180	15.2	Cadillac Fleetwood	205	10.4
Fiat 128	66	32.4	Toyota Corolla	65	33.9
Honda Civic	52	30.4	Toyota Corona	97	21.5

Dataset: `mtcars` (32 cars). Relationship between mpg and hp is nonlinear and decreasing.

We will test two models: **Model (i)**: mpg vs. ln(hp)   **Model (ii)**: ln(mpg) vs. ln(hp).

1. What are the regression equations for both models?

   (i) $\widehat{\text{mpg}} = 72.6405 - 10.7642 \ln(\text{hp})$

   (ii) $\widehat{\ln(\text{mpg})} = 5.545381 - 0.530092 \ln(\text{hp}) \quad (\Rightarrow \ \text{mpg} \approx 256.052\, \text{hp}^{-0.530092})$

2. Which model has the larger R^2 value? Which appears to fit better?

   $R^2_{(i)} = 0.7204, \quad R^2_{(ii)} = 0.7157$. Model (i) is slightly better.

3. After fitting each model, make a residual plot (residuals vs. fitted values). Which model shows residuals that look most like random noise (centered at 0, with constant spread and no clear pattern)?

   Model (i) looks slightly more random than Model (ii), consistent with its slightly larger R^2.

Integrated Rate Laws — Choose Order by R^2

For each reaction:

1) Make three transformed plots vs. time: $[A]$ (zeroth), $\ln[A]$ (first), and $1/[A]$ (second).

2) Fit a straight line to each and record its R^2.

3) Residual-plot diagnostic: For each transform, also draw a residual plot (residual vs. time). Favor the transform whose residuals look most like random noise centered at 0 with roughly constant spread and no structure.

4) Choose the order with the largest R^2 (and the best residual plot). Then read k from the slope (remember signs).

Reaction 1. $2\,N_2O(g) \rightarrow 2\,N_2(g) + O_2(g)$ (Pt).

Time (s)	0	10	20	30	40	50	60	70	80	90
$[A]$ **(mol/L)**	0.800	0.715	0.636	0.550	0.491	0.405	0.333	0.235	0.150	0.086

- Your work: compute R^2_{zero}, R^2_{first}, R^2_{second}; inspect residuals; choose the order; find k from slope.
  $R^2_{\text{zero}} = 0.99885$, $R^2_{\text{first}} = 0.91043$, $R^2_{\text{second}} = 0.68832 \Rightarrow$ **zeroth order**.
  Line: $[A] = 0.7980 - 0.0079533\,t$; $k = 7.9533 \times 10^{-3}$ mol L^{-1} s^{-1}. Residuals are pattern-free and centered near 0 with equal spread.

Reaction 2. $2\,H_2O_2(aq) \rightarrow 2\,H_2O(l) + O_2(g)$ (KI).

Time (s)	0	20	40	60	80	100	120	140	160	180
$[A]$ **(mol/L)**	1.000	0.815	0.665	0.531	0.441	0.352	0.285	0.230	0.185	0.145

- Your work: compute R^2_{zero}, R^2_{first}, R^2_{second}; inspect residuals; choose the order; find k from slope.
  $R^2_{\text{zero}} = 0.94112$, $R^2_{\text{first}} = 0.99956$, $R^2_{\text{second}} = 0.92276 \Rightarrow$ **first order**.
  Line: $\ln[A] = 0.0137308 - 0.0106545\,t$; $k = 0.0106545$ s^{-1}. Residuals are pattern-free and centered near 0 with equal spread.

Reaction 3. $2\,I^-(aq) + S_2O_8^{2-}(aq) \rightarrow I_2(aq) + 2\,SO_4^{2-}(aq)$.

Time (s)	0	10	20	30	40	50	60	70	80	90
$[A]$ **(mol/L)**	0.500	0.397	0.319	0.263	0.205	0.179	0.151	0.130	0.112	0.093

- Your work: compute R^2_{zero}, R^2_{first}, R^2_{second}; inspect residuals; choose the order; find k from slope.
  $R^2_{\text{zero}} = 0.90901$, $R^2_{\text{first}} = 0.99308$, $R^2_{\text{second}} = 0.97386 \Rightarrow$ **first order**.
  Line: $\ln[A] = -0.763590 - 0.0183543\,t$; $k = 0.0183543$ s^{-1}. Residuals are pattern-free and centered near 0 with equal spread.

TI-84 Quick Card: Regression & Diagnostic Plots

TI-84 Quick Commands

Before you start (do this once): Turn on r and R^2.
Keys: `2nd` `0 (CATALOG)` → type `D` to jump → select `DiagnosticOn` → `ENTER` `ENTER`.
You should see "Done." After this, regressions will display r and R^2.

a) Enter data ($L_1 = x$, $L_2 = y$).
Keys: `STAT` → `1:Edit...` → type x in L_1 and y in L_2.
If old data are in the way: `STAT` → `4:ClrList` `L1, L2` `ENTER` (then re-enter data).
Tip: If your x or y are in different lists, remember which lists they are in (e.g., L_3, L_4).

b) Run linear regression and store the equation in Y_1.
Keys: `STAT` → `CALC` → choose `LinReg(a+bx)`.
On the *command line*, type the lists and where to store the model:

`LinReg(ax+b) L1, L2,` Y_1

To paste Y_1: `VARS` → `Y-VARS` → `1:Function` → `1:`Y_1 → `ENTER`.
Then press: `ENTER` (*twice* if needed) to run.
What you should see: a (intercept), b (slope), r, R^2. The graphing screen now has $Y_1 = a + bX$ loaded.

c) Scatterplot with the fitted line on top.
Keys: `2nd` `Y=` (`STAT PLOT`) → `1:Plot1` → `On`.
Set **Type:** Scatter **Xlist:** L_1 **Ylist:** L_2 **Mark:** any.
Graph it: `ZOOM` → `9:ZoomStat`.
You should now see your points with the regression line Y_1 overlaid.

d) Residual plot (checks curvature and equal spread).
Residuals are automatically saved in list `RESID`.
Keys: `2nd` `Y=` (`STAT PLOT`) → `2:Plot2` → `On`.
Set **Type:** Scatter **Xlist:** L_1 **Ylist:** `RESID`.
(`RESID` lives in `2nd` `STAT (LIST)` → **NAMES** menu → scroll to `RESID`.)
Graph it: `ZOOM` → `9:ZoomStat`.
Interpretation: good residual plot ⇒ random cloud around 0, no curve/pattern, roughly constant vertical spread.

e) Normal probability plot of residuals (checks normality).
Keys: `2nd` `Y=` (`STAT PLOT`) → `3:Plot3` → `On`.
Set **Type:** *Normal Prob Plot* (last icon) **Data List:** `RESID`.
Graph it: `ZOOM` → `9:ZoomStat`.
Interpretation: points close to a straight line ⇒ residuals are roughly normal.

Tips & Troubleshooting

- If nothing shows: ensure `Plot1/Plot2/Plot3` are `On` and the correct lists (e.g., L_1, L_2, `RESID`) are selected.
- If r or R^2 are missing: redo `2nd` `0` → `DiagnosticOn` → `ENTER` `ENTER`.
- To model transformed variables (e.g., $\ln x$, $\ln y$): in `STAT` `1:Edit`, move to L_3, type `ln(` `2nd` `1` `)` then `ENTER`. Do the same for L_4 if needed, then run `LinReg(ax+b)` on L_3, L_4, storing to Y_1 again.
- Weird graphs? Clear extra equations in `Y=` and turn off unused Stat Plots.