Simple Linear Regression
Part 5: Propellant Example

STAT 705: Regression and Analysis of Variance

Example

NASA is interested in the shear strength of the bond between propellants in a rocket motor and whether it may be related to age of the propellant batch. A sample of 20 propellant batches is collected and their shear strengths (in pounds per square inch) and ages (in weeks) are recorded.

Source: Montgomery, et.al, 2006

Obs (i)	Shear Strength (Y)	Age (X)
1	2158.7	15.5
2	1678.15	23.75
3	2316	8
4	2061.3	17
5	2207.5	5.5
6	1708.3	19
7	1784.7	24
8	2575	2.5
9	2357.9	7.5
10	2256.7	11
11	2165.2	13
12	2399.55	3.75
13	1779.8	25
14	2336.75	9.75
15	1765.3	22
16	2053.5	18
17	2414.4	6
18	2200.5	12.5
19	2654.2	2
20	1753.7	21.5

Analyze the Propellant Data

- Questions of interest
  1. Is the shear strength linearly related to the age of the propellant? If so, quantify the relationship.
  2. Estimate the mean shear strength for propellants that are 10 weeks old.
  3. Predict the shear strength for a propellant that is 10 weeks old.
- Before we can answer the questions, we need to
  - Define and fit a linear model
  - Assess the validity of the model

SAS Code, Part 1

- Create a temporary SAS data set called ‘nasa’
- The last observation has a missing value (a period) for Shear Strength and value 10 for Age. This is for estimation and prediction at 10 weeks.

```
data nasa;
   input Obs ShearStrength Age;
   cards;
1    2158.7  15.5
2    1678.15 23.75
3    2316     8
4    2061.3  17
5    2207.5  5.5
6    1708.3  19
7    1784.7  24
8    2575.   2.5
9    2357.9  7.5
10  2256.7  11
11  2165.2  13
12  2399.55 3.75
13  1779.8  25
14  2336.75 9.75
15  1765.3  22
16  2053.5  18
17  2414.4  6
18  2200.5  12.5
19  2654.2  2
20 1753.7  21.5
21      .   10
;
run;
```

SAS Code, Part 2

Always print the data before ANY analysis	`proc print data=nasa;`
	`     var Obs ShearStrength Age;`
	`run;`

Summary statistics ('sanity' check)	`proc means data=nasa n sum mean stdev min max;`
	`     var ShearStrength Age;`
	`run;`

Scatterplot	`symbol1 v=dot i=none c=blue;`
	`proc gplot data=nasa;`
	`     plot ShearStrength*Age;`
	`run;`

Regression	`proc reg Data=nasa;`
	`     model ShearStrength = Age/clb p clm cli alpha=0.05;`
	`     output out=Predict p=Predicted r=Residual`
	`           uclm=uclm  lclm=lclm  ucl=ucl lcl=lcl;`
	`run;`

SAS Code, Part 3

Diagnostic information	`* Construct a residual plot;`
	`symbol1 v=dot     i = none    c = blue;`
	`proc gplot data=Predict;`
	`     plot Residual*Predicted/vref=0;`
	`run;`
	`* Obtain normal probability plot and tests for normality;`
	`proc univariate data=Predict normal plots;`
	`     var Residual;`
	`run;`

Generate a scatterplot with the confidence intervals and prediction intervals	`* Plot the data, fitted line, CI and PI... all in one;`
	`proc sort data = Predict; by Predicted; run;`
	`symbol1 v=dot  i=none  c=black;`
	`symbol2 v=none i=join  c=red;`
	`symbol3 v=none i=join  c=blue;`
	`symbol4 v=none i=join  c=blue;`
	`symbol5 v=none i=join  c=green;`
	`symbol6 v=none i=join  c=green;`
	`proc gplot data=Predict;`
	`     plot ShearStrength*Age Predicted*Age uclm*Age`
	`          lclm*Age ucl*Age lcl*Age / overlay;`
	`run;`

Output: PROC PRINT

Obs	Obs	ShearStrength	Age
1	1	2158.70	15.50
2	2	1678.15	23.75
3	3	2316.00	8.00
4	4	2061.30	17.00
5	5	2207.50	5.50
6	6	1708.30	19.00
7	7	1784.70	24.00
8	8	2575.00	2.50
9	9	2357.90	7.50
10	10	2256.70	11.00
11	11	2165.20	13.00
12	12	2399.55	3.75
13	13	1779.80	25.00
14	14	2336.75	9.75
15	15	1765.30	22.00
16	16	2053.50	18.00
17	17	2414.40	6.00
18	18	2200.50	12.50
19	19	2654.20	2.00
20	20	1753.70	21.50
21	21	.	10.00

Always print the data the first time you read it into SAS.

If SAS had difficulty reading the data, you would most likely see a lot of cells with a single period. (In SAS, a single period indicates missing data.)

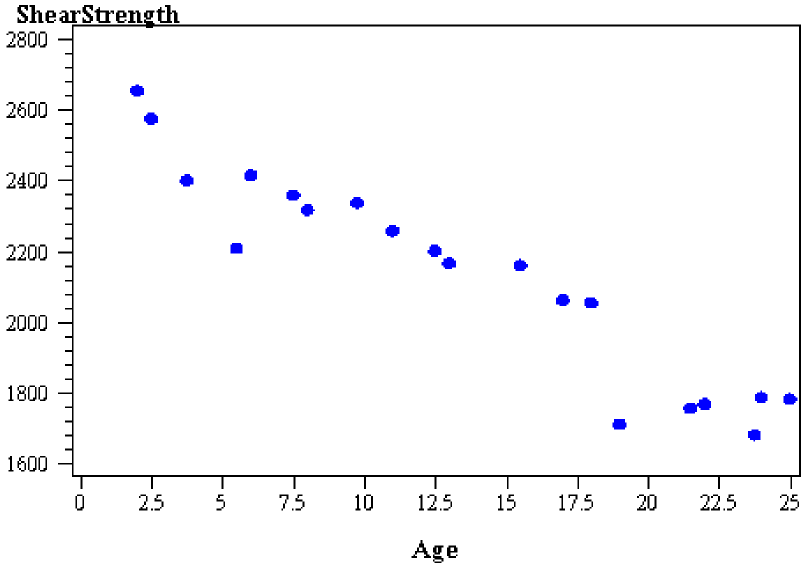
We see only one cell with missing data. This is for ShearStrength in observation number 21 . This is the observation we added for the purpose of generating a confidence interval and a prediction interval for Age = 20.

SAS has read this data correctly. ✓

Output: Means and Scatterplot

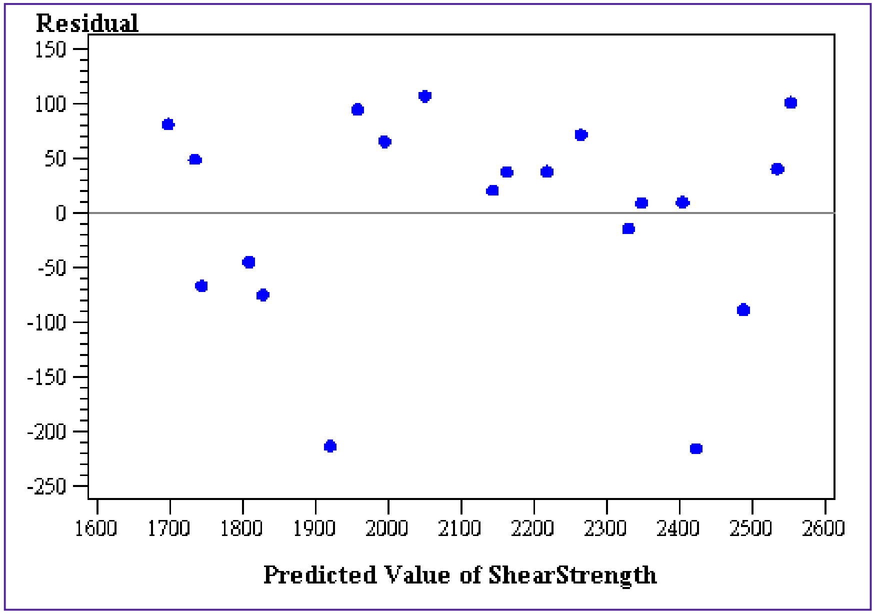
The MEANS Procedure

Variable	N	Sum	Mean	Std Dev	Minimum	Maximum
ShearStrength	20	42627.15	2131.36	298.5700660	1678.15	2654.20
Age	21	277.2500000	13.2023810	7.4743808	2.0000000	25.0000000

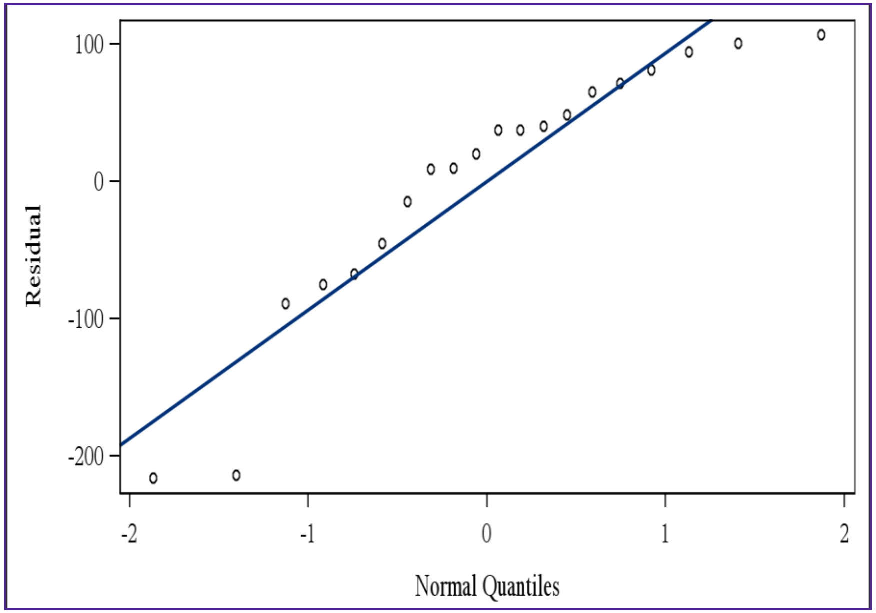


- The summary statistics appear reasonable.
- The scatterplot shows a linear trend.
- A simple linear regression model seems appropriate.

Output: Model Diagnostics



Fitted vs. Residual



Normal Probability Plot

- On left: no discernible pattern
- On right: points follow the line
- Overall: Scant evidence that model assumptions are violated.

Partial Output: PROC REG

Number of Observations Read	21
Number of Observations Used	20
Number of Observations with Missing Values	1

one missing?

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	2627.82236	44.18391	59.47	<.0001	2534.99540	2720.64931
Age	1	-37.15359	2.88911	-12.86	<.0001	-43.22338	-31.08380

$$\text{ShearStrength} = 2627.8 - 37.15 \times \text{Age}$$

Answer Question 1

- Are Shear Strength and Age linearly related?
  - Yes. We reject the hypothesis that the slope is 0
    - Test statistic is t = -12.86
    - p-value < .0001
- Quantify the relationship.
  - For each additional week that the propellant ages, the shear strength of the bond between propellants is reduced, on average, by 37.15 psi.

Confidence and Prediction Intervals

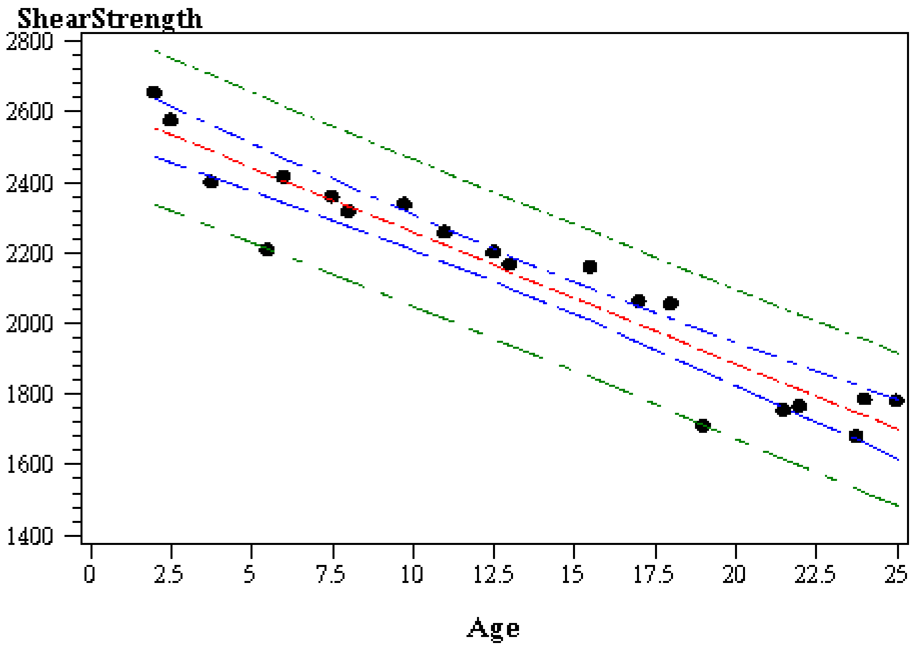
Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	2159	2052	22.3597	2005	2099	1845	2259	106.7583
2	1678	1745	36.9114	1668	1823	1529	1962	-67.2746
3	2316	2331	26.4924	2275	2386	2121	2540	-14.5936
4	2061	1996	23.9220	1946	2046	1788	2204	65.0887
5	2208	2423	31.2701	2358	2489	2211	2636	-215.9776
6	1708	1922	26.9647	1865	1979	1712	2132	-213.6041
7	1785	1736	37.5010	1657	1815	1519	1953	48.5638
8	2575	2535	38.0356	2455	2615	2318	2752	40.0616
9	2358	2349	27.3623	2292	2407	2139	2559	8.7296
10	2257	2219	22.5479	2172	2267	2012	2427	37.5671
11	2165	2145	21.5155	2100	2190	1938	2352	20.3743
12	2400	2488	35.1152	2415	2562	2274	2703	-88.9464
13	1780	1699	39.9031	1615	1783	1480	1918	80.8174
14	2337	2266	23.8903	2215	2316	2058	2474	71.1752
15	1765	1810	32.9326	1741	1880	1597	2024	-45.1434
16	2054	1959	25.3245	1906	2012	1750	2168	94.4423
17	2414	2405	30.2370	2341	2468	2193	2617	9.4992
18	2201	2163	21.6340	2118	2209	1956	2370	37.0975
19	2654	2554	39.2360	2471	2636	2335	2772	100.6848
20	1754	1829	31.8519	1762	1896	1616	2042	-75.3202
21	.	2256	23.5837	2207	2306	2048	2464	.

Answer Questions 2 and 3

For propellants that are 10 weeks old, we are 95% confident that the mean shear strength will be between 2207 and 2306 psi.

For a single propellant that is 10 weeks old, we predict that its shear strength will be between 2048 and 2464 psi.

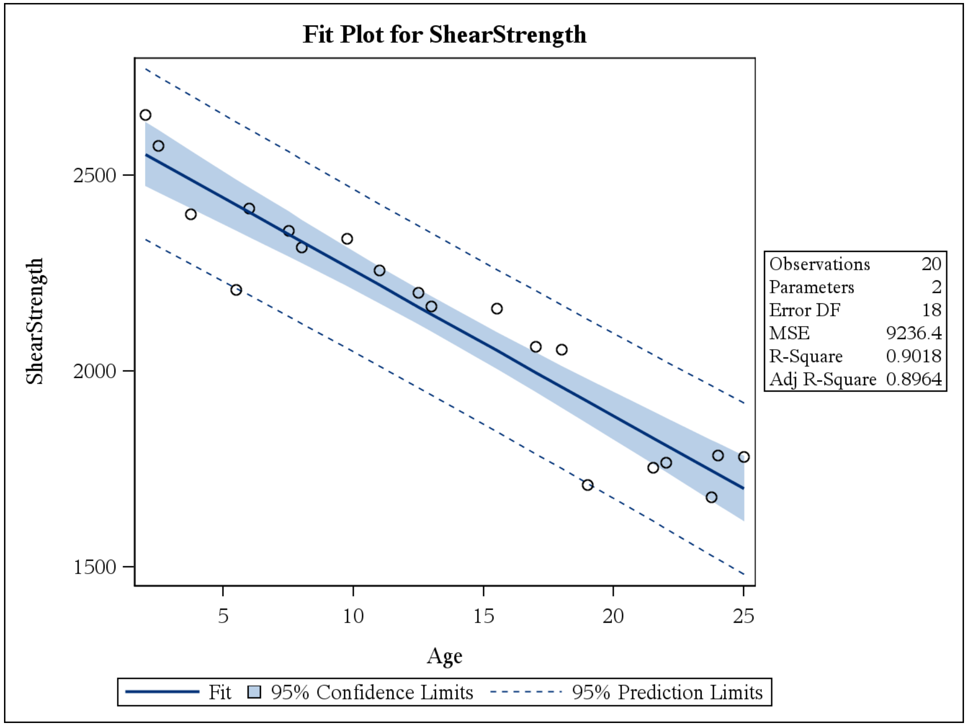
Graph of CI and PI



This is the graph we generated with PROC GPLOT.

A better graph is on the next slide.

Graph of CI and PI



- Graph is generated by PROC REG
- Solid blue band is the confidence interval (for the mean)
- Dashed lines show the prediction interval (for an individual Shear Strength)

Confidence and Prediction Intervals

- Both types of intervals change when the value for Age (X) changes
- For any value of X, confidence intervals are more narrow than prediction intervals. (We can estimate a mean more precisely than we can predict an individual value.)
- If you just look at confidence intervals
  - The interval gets narrower when X is close to the mean, and wider when X is far away from the mean
- If you just look at prediction intervals
  - The interval gets narrower when X is close to the mean, and wider when X is far away from the mean

Other Questions

- There are many other questions that can be answered by either reading the SAS output or using values from the output to perform hand calculations. For example,
  - Find a 95% confidence interval for the population slope.
  - Find a 95% interval estimate for the shear strength of a new propellant (age = 0 weeks).
  - What proportion of the variability in shear strength is explained by this regression model?

Things You Should Know

- The SAS code and complete output for this model are provided on the course website.
  - file name: NASAexample.sas
- Run the code
- Use the output to answer the previous questions