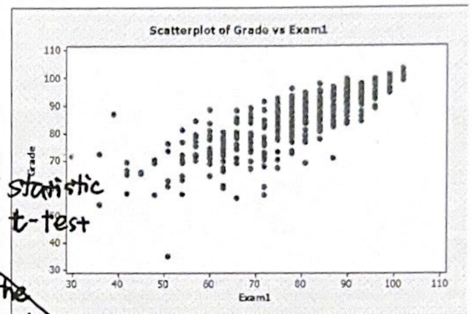
EXAMPLE: Can we predict your final grade in the class from your 1st exam score?

Regression Analysis: Grade versus Exam1




```
The regression equation is
Grade = 36.8 + 0.614 Exam1
```

standard error of b and a.

test statistic for t-test

	Predictor	Coef	SE Coef	T	P
(α, a)	Constant	36.832	1.655	22.26	0.000
(β, b)	Exam1	0.61352	0.02060	29.78	$\boxed{0.000}$

```
S = 5.76575   R-Sq=67.1%   R-sq(adj)=67.1%
```

$\sqrt{MSE}$

Take into account the number of predictors (later)

Should be exactly the same

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	29480	29480	886.77	$\boxed{0.000}$
Residual Error	434	14428	33		
Total	435	43908			

Unusual Observations

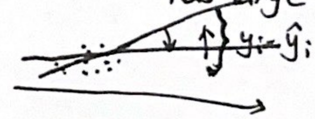
standardized residual

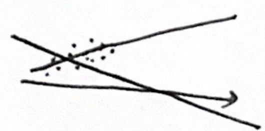
Obs	Exam1	Grade	Fit	SE Fit	Residual	St Resid
1	72	57.250	81.005	0.313	-23.755	-4.13R
2	42	58.000	62.600	0.814	-4.600	-0.81 X
4	72	60.313	81.005	0.313	-20.693	-3.59R
5	51	34.813	68.121	0.643	-33.309	-5.81R
6	36	53.720	58.919	0.932	-5.199	-0.91 X
7	63	60.000	75.484	0.433	-15.484	-2.69R
10	54	57.750	69.962	0.588	-12.212	-2.13R
13	48	57.500	66.281	0.699	-8.781	-1.53 X
15	75	67.500	82.846	0.289	-15.346	-2.66R
24	81	69.000	86.527	0.279	-17.527	-3.04R
33	45	65.250	64.440	0.756	0.810	0.14 X
39	78	72.750	84.686	0.277	-11.936	-2.07R

Etc….

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

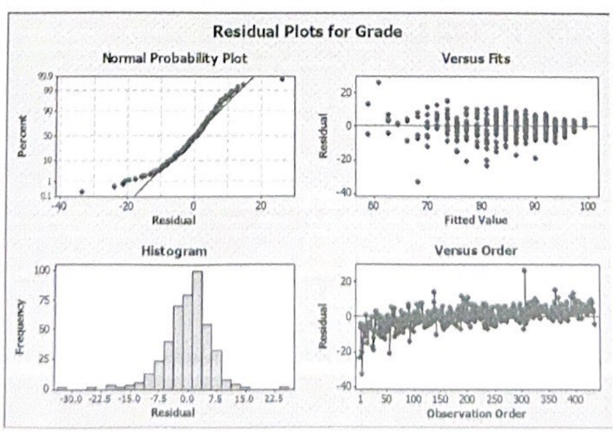
has large leverage





Predicted Values for New Observations

New Obs	Exam1	Fit	SE Fit	95% CI	95% PI
1	80.0	85.913	0.277	(85.370, 86.457)	(74.568, 97.259)



Using the Minitab output from the previous page, complete the following:

1. Write the full model, the assumptions, and identify the parameters and estimators.

Model: $y=\alpha+\beta x+\varepsilon$

Assumption: $\varepsilon \overset{iid}{\sim} N(0,\sigma^2)$

In words:

- SRS of students/final scores
  436 students need to be representative of students taking STA3024 — ok for the same instructor.
- Normal distribution of errors/final scores (Check outliers)
  There are a couple, but n=436 large so it is ok.
- constant variance of errors/final scores
  Ok.

Parameters: α β σ

Estimators: a b $s=\sqrt{MSE}$

$a = 36.8$ $b = 0.614$ $s = 5.76575$

2. Construct the 95% confidence interval for the slope and interpret.

$b \pm t_{n-2,\frac{5\%}{2}} \cdot s.e._b$

$\cancel{0.614}$ 0.61352 $\pm$ 1.96 . 0.02060 $= (0.57, 0.65)$

df=434, so we can use $z_{2.5\%} = 1.960$

- We are 95% confident that the true slope of the line that is used to predict Final score from Ex 1 score in population is somewhere between 0.57 and 0.65.
- Is β diff/sig from zero? Yes because zero is NOT included. ⇒ x is a good predictor of y.

3. Conduct the t-test for the slope and interpret.

$H_0: \beta=0$ $H_a: \beta\neq 0$

TS: $t=\frac{b}{s.e._b}=\frac{0.61352}{0.02060}=29.78$

pval ≈ 0 Very strong evidence to say:

- β is sig diff from zero
- Ex1 grade is a good predictor for Final score.
- Line is useful for prediction
- x and y are associated.

4. Conduct the ANOVA test for the slope and interpret.

$H_0: \beta=0$ $H_a: \beta\neq 0$

TS: $F=886.77$

pval ≈ 0 same conclusion as in 3.

Confidence and Prediction Intervals for Response