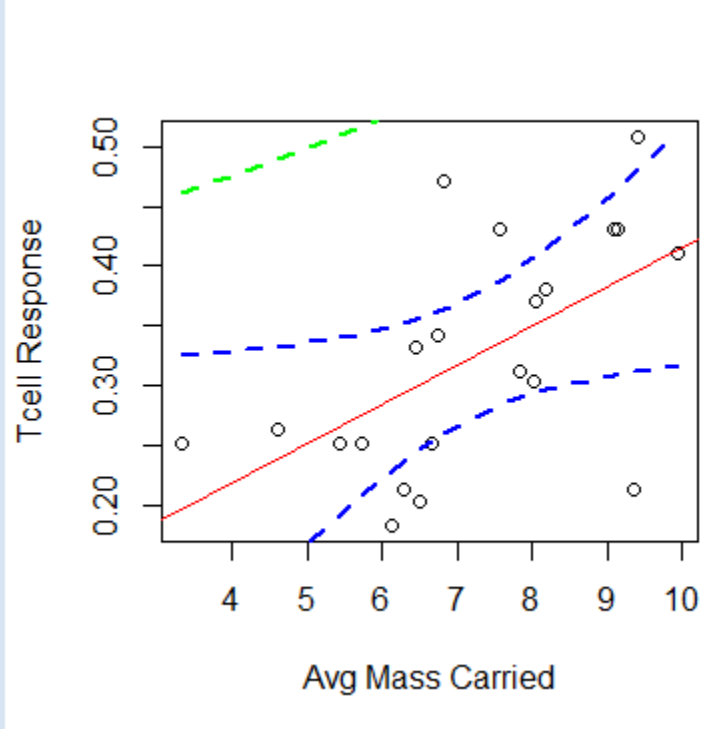
UNIT 10 HW

1. Black-eared wheatears are small birds of Spain and Morocco. Males of the species demonstrate an exaggerated sexual display by carrying many heavy stones to nesting cavities. This 35-gram bird transports, on average, 3.1 kg of stones per nesting season! Different males carry somewhat different sized stones, prompting a study of whether larger stones may be a signal of higher health status. M. Soler et al. calculated the average stone mass (g) carried by each of 21 male black-eared wheatears, along with T-cell response measurements (in mm) reflecting their immune system strengths. Analyze the data and write a statistical report (by answering the questions below); treat the T-cell as the response and the stone mass as the explanatory variable. You may assume all criteria for regression and related t-tests are met. You can find the data for this problem on 2DS. (Male Display Data Set)

   Analyze the data, providing the following:

   a. Provide a scatterplot with 99% confidence intervals of the regression line and 99% prediction intervals of the regression line. Please do this in R.



   b. Provide a table showing the t-statistics and p-values for the significance of the regression parameters $\beta_0 \; and \; \beta_1$ (as different from 0). Please do this in R.

```
> summary(Male.lm)

Call:
lm(formula = Tcell ~ Mass, data = Male)

Residuals:
     Min       1Q   Median       3Q      Max
-0.18138 -0.04673  0.01796  0.04219  0.15999

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.08750    0.07868   1.112  0.27996
Mass         0.03282    0.01064   3.084  0.00611 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08102 on 19 degrees of freedom
Multiple R-squared:  0.3336,    Adjusted R-squared:  0.2986
F-statistic: 9.513 on 1 and 19 DF,  p-value: 0.006105
```

c. Using the output in (b), show all 6 steps of ***each*** hypothesis test. (That's one test for $\beta_0 = 0$ and one test for $\beta_1 = 0$.) Find critical values in R. Your conclusion should include a confidence interval. Use alpha = 0.01

Slope

Step 1 -Hypotheses: $H_o: \beta_1 = 0$ $H_a: \beta_1 \neq 0$

Step 2 - Identification of Critical Value: $t_{0.995,\ 21-2} = \pm 2.861$

Step 3- Value of Test Statistic: First, calculate the standard error of $\hat{\beta}_1$**: 3.084**

Step 4- Give p-value: $p = 0.006105$

Step 5- Decision: Reject H_o

Step 6- Conclusion: There is sufficient evidence at the α=0.01 level of significance (p=0.006105) that there is a linear relationship between the stone mass of the bird and the t-cell. In other words, there is evidence that the slope of the regression equation is not equal to zero. Because this is an observational study, the results are limited to revealing the association between stone mass and t-cell, rather than causality.

Intercept

Step 1 -Hypotheses: $H_o: \beta_0 = 0$ $H_a: \beta_0 \neq 0$

Step 2 - Identification of Critical Value: $t_{0.995,\ 21-2} = \pm 2.861$

Step 3- Value of Test Statistic: First, calculate the standard error of $\hat{\beta}_0$**: 1.112**

Step 4- Give p-value: $p = 0.27996$

Step 5- Decision: Fail to Reject H_o

Step 6- Conclusion: There is not sufficient evidence at the α=0.01 level of significance (p=0.27996) to suggest that stone mass and t-cell are linearly related. Because this is an observational study, the results are limited to revealing the association between stone mass and t-cell, rather than causality.

d. State the regression equation. Be careful to use the mean Tcell or predicted Tcell, rather than just Tcell.

$$\mu_{\text{Tcell|Mass}} = 0.0875 + 0.0328 \cdot \text{Mass}$$

e. Interpret the slope in the model (regression equation).
Slope: for every increase of 1(g) in mass, the estimated mean T-cell response count increases by 0.0328mm.

f. Interpret the y-intercept in the model (regression equation).
Intercept: for a bird with no stone mass, the expected mean of the T-cell response is about 0.0875. Of course, this is extrapolation, as no stone mass scores of 0 are in the data set.

g. Find and interpret the 99% confidence interval for the mean t-cell response conditional on a stone mass of 4.5 grams. Please do this directly in R.

```
        fit       lwr       upr
1 0.2351937 0.1385665 0.3318209
```

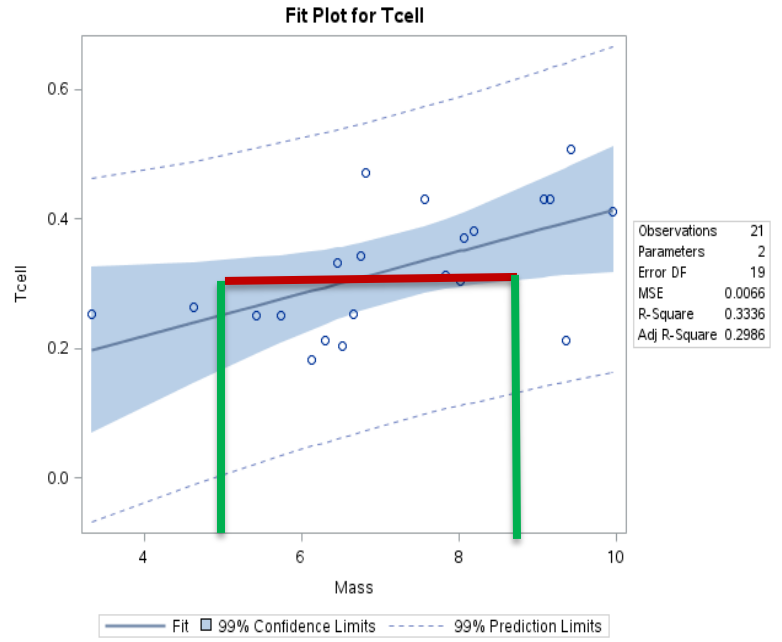
A 99% confidence interval for the t-cell response when observed at 4.5 grams of stone mass is between 0.0139 and 0.3318 mm.

h. Find and interpret the 99% prediction interval for the predicted t-cell response given a stone mass of 4.5 grams. Please do this directly in R.

```
        fit          lwr       upr
1 0.2351937 -0.01593192 0.4863193
```

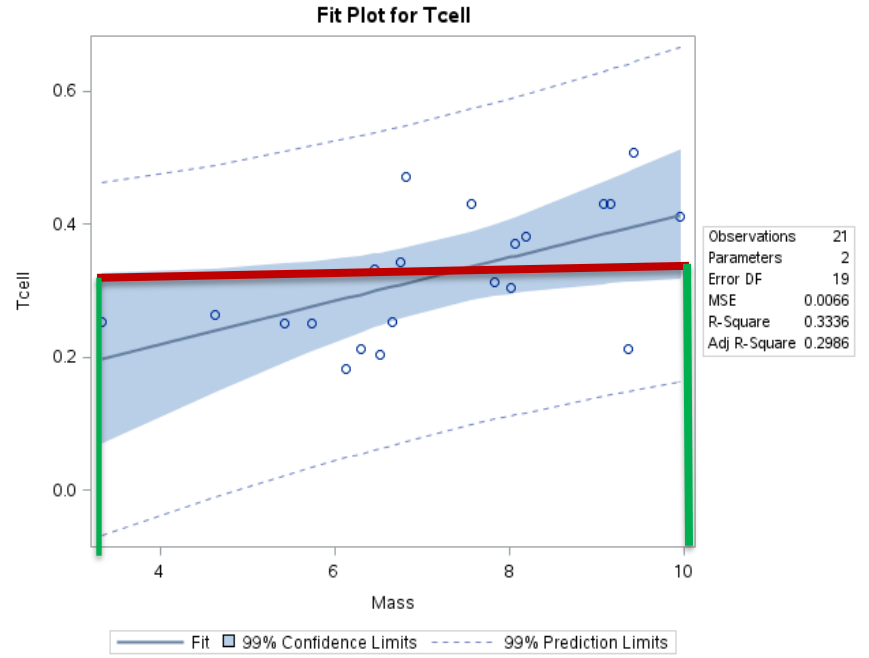
A 99% prediction interval for the t-cell response when observed at 4.5 grams of stone mass is between -0.0159 and 0.4683 mm.

i. Calibration intervals:
    1. Using the **graphical method** (using your best judgment using the graphs from part (a)), find the following using R, as part (a) was done in R.
        a. 99% calibration interval for the **mean** t-cell response of 0.3.



We are 99% confident that the estimated Stone Mass that would be needed to gross a mean of 0.3 T-cell is between 5 and 9 grams.

b. 99% calibration interval for a **single** t-cell response of 0.3.



We are 99% confident that the estimated Stone Mass that would be needed to gross 0.3 T-cell for an individual bird is between 3 and 10 grams.

2. Using **software directly**, find the following using R, as SAS does not provide calibration intervals directly. (R: package investr)

a. 99% calibration interval for the **mean** t-cell response of 0.3.
calibrate(Male.lm, y0 = 0.3, interval = "inversion", mean.response = TRUE, level = .99)

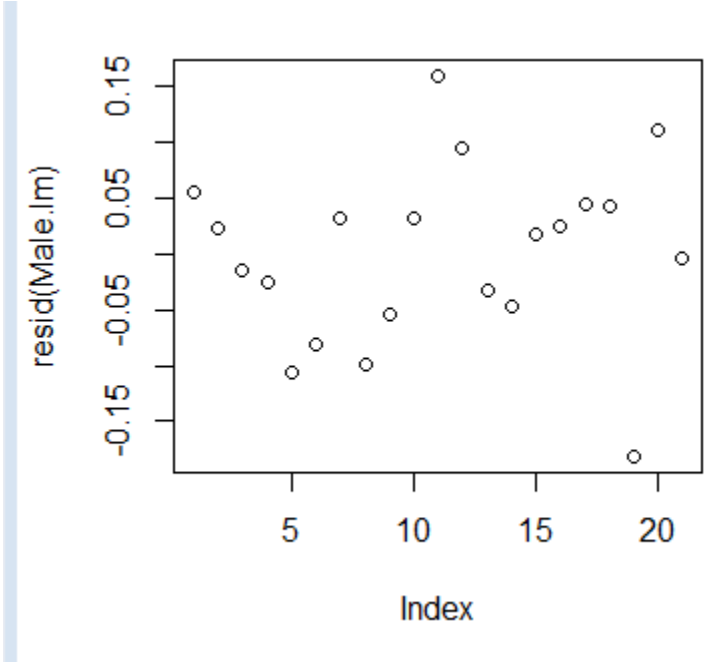
```
  estimate      lower      upper
6.474508 -4.389857  8.342649
```

b. 99% calibration interval for a **single** t-cell response of 0.3.
calibrate(Male.lm, y0 = 0.3, interval = "inversion", level = .99)

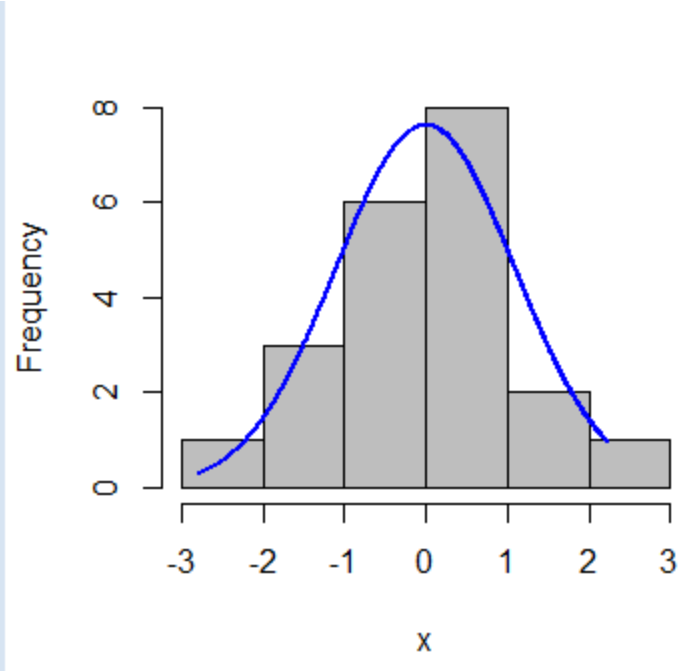
```
  estimate      lower      upper
6.474508 -17.968869  21.921661
```

3. **Interpret** the following using the results from (1) and (2) above.
   a. 99% calibration interval for the **mean** t-cell response of 0.3.
   We are 99% confident that the required AVG Stone Mass to indicate a mean t-cell of 0.3mm is between -4.389857 and 8.342649 grams.
   b. 99% calibration interval for a **single** t-cell response of 0.3.
   A 99% prediction interval for the required AVG Stone Mass to predict a t-cell of 0.3mm is between -17.968869 and 21.921661 grams.

j. Provide a scatterplot of residuals. Please do this in R.



k. Provide a histogram of residuals with a normal distribution superimposed. It might be helpful to use studentized residuals, rather than regular residuals, with a normal curve overlay. Use R. (You may need to research this, such as googling “histogram with normal curve in R.”)



l. Provide a measure of the **proportion** of variation in the response that is accounted for by the explanatory variable. **Interpret** this measure. Use R.
r2 = 0.3336.
It is estimated that mass explains about 33.36% of the variation in t-cell response for these birds.

2. Using the data for Black-eared Wheatears, calculate by "hand" (using Excel) the following elements. (An example of much of this was in the PowerPoints and in the videos below.)
   a. $\hat{\beta}_0 \; and \; \hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{1.902534286}{57.966114} = 0.032821$$

$$\hat{\beta}_0 = 0.323952 - 0.032821 \cdot 7.204286 = 0.087497$$

   b. The t-statistics and p-values for the hypothesis tests ($\beta_0 = 0$ and for $\beta_1 = 0$).

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{0.032821 - 0}{0.010641} = 3.084$$

$$t = \frac{\hat{\beta}_0 - 0}{SE(\hat{\beta}_0)} = \frac{0.0874977 - 0}{0.078677} = 1.112$$

P-value β_0 = .2800
P-value β_1 = .0061

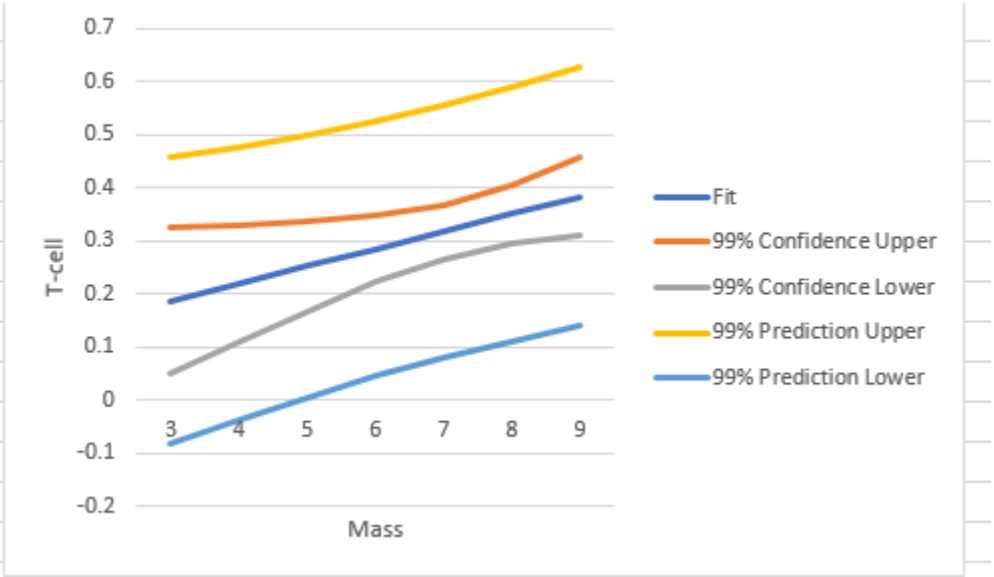
c. 99% confidence intervals for the mean of Y when X = {3,4,5,6,7,8,9} grams. You do NOT need to make a Bonferroni (or any other type of) multiple interval correction, as the primary purpose of these intervals is to be able to plot confidence interval bands.

Grams	Est Value	SE Est Val	Lower Lim	Upper Lim
3	0.18596	0.048106	0.048331	0.323589
4	0.218781	0.038409	0.108894	0.328668
5	0.251602	0.029374	0.167566	0.335638
6	0.284423	0.021836	0.221951	0.346895
7	0.317244	0.017813	0.266282	0.368206
8	0.350065	0.019603	0.293982	0.406148
9	0.382886	0.026033	0.308406	0.457366

d. 99% prediction intervals for the predicted Y when X = {3,4,5,6,7,8,9} grams. You do NOT need to make a Bonferroni (or any other type of) multiple interval correction, as the primary purpose of these intervals is to be able to plot prediction interval bands.

Grams	Est Value	SE Est Val	Lower Lim	Upper Lim
3	0.18596	0.094225	-0.08361	0.455532
4	0.218781	0.089663	-0.03774	0.475301
5	0.251602	0.08618	0.005047	0.498157
6	0.284423	0.08391	0.044361	0.524485
7	0.317244	0.082955	0.079916	0.554572
8	0.350065	0.083357	0.111585	0.588545
9	0.382886	0.085099	0.139422	0.62635

e. Provide a plot for the confidence intervals and prediction intervals using Excel. Fully label your graph. (Use the regression equation and parts (c) and (d) above to create the plot.)



f. Calibration intervals:
   i. Using the SE equations given in class and in the book (Version 3 page 194), find the following **analytically**. (Use Excel for calculations.)
      a. 99% calibration interval for the mean t-cell response of 0.3.

Mass	EST T-cell	SE EST Val	Lower Lim	Upper
6.474605	0.3	0.58834	4.980525	7.968684

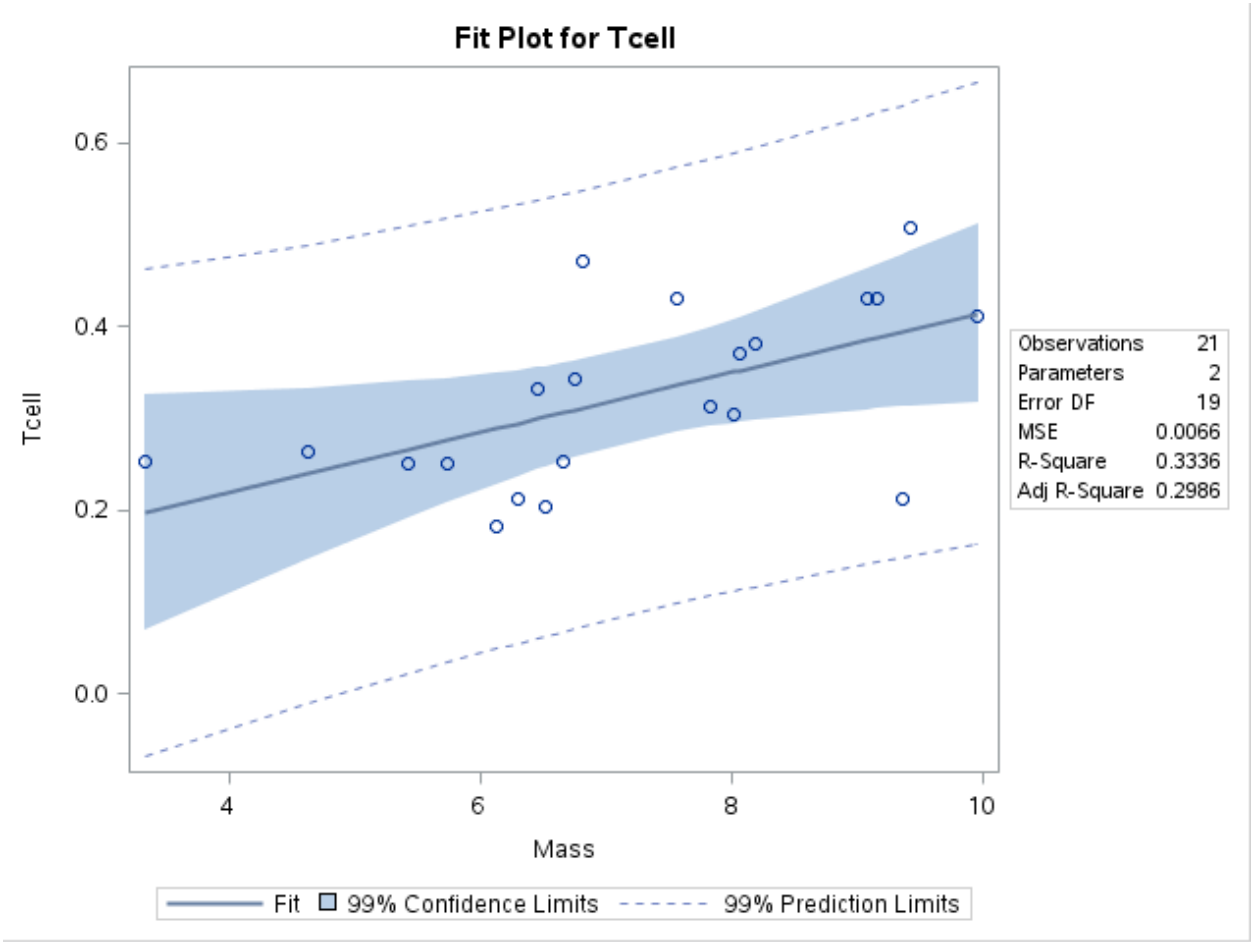
      b. 99% calibration interval for a single t-cell response of 0.3.

Mass	EST T-cell	SE EST Val	Lower Lim	Upper
6.474605	0.3	2.537668	0.030238	12.91897

   j. Using the **Excel graphs**, find the following. You may want to add data points to parts (c) and (d) so that the confidence and prediction limits extend well beyond the data range (although their interpretation is questionable outside the range).
      a. 99% calibration interval for the mean t-cell response of 0.3.

      b. 99% calibration interval for a single t-cell response of 0.3.

3. Bonus!
   a. Repeat 1 (a) using SAS.



b. Repeat 1 (b) using SAS.

The REG Procedure
Model: MODEL1
Dependent Variable: Tcell

Number of Observations Read	21
Number of Observations Used	21

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.06244	0.06244	9.51	0.0061
Error	19	0.12472	0.00656		
Corrected Total	20	0.18716			

Root MSE	0.08102	R-Square	0.3336
Dependent Mean	0.32395	Adj R-Sq	0.2986
Coeff Var	25.00969		

Parameter Estimates

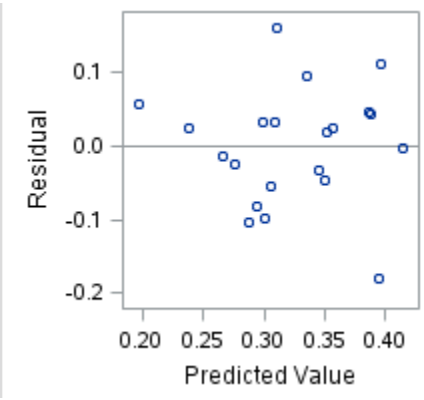
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.08750	0.07868	1.11	0.2800
Mass	1	0.03282	0.01064	3.08	0.0061

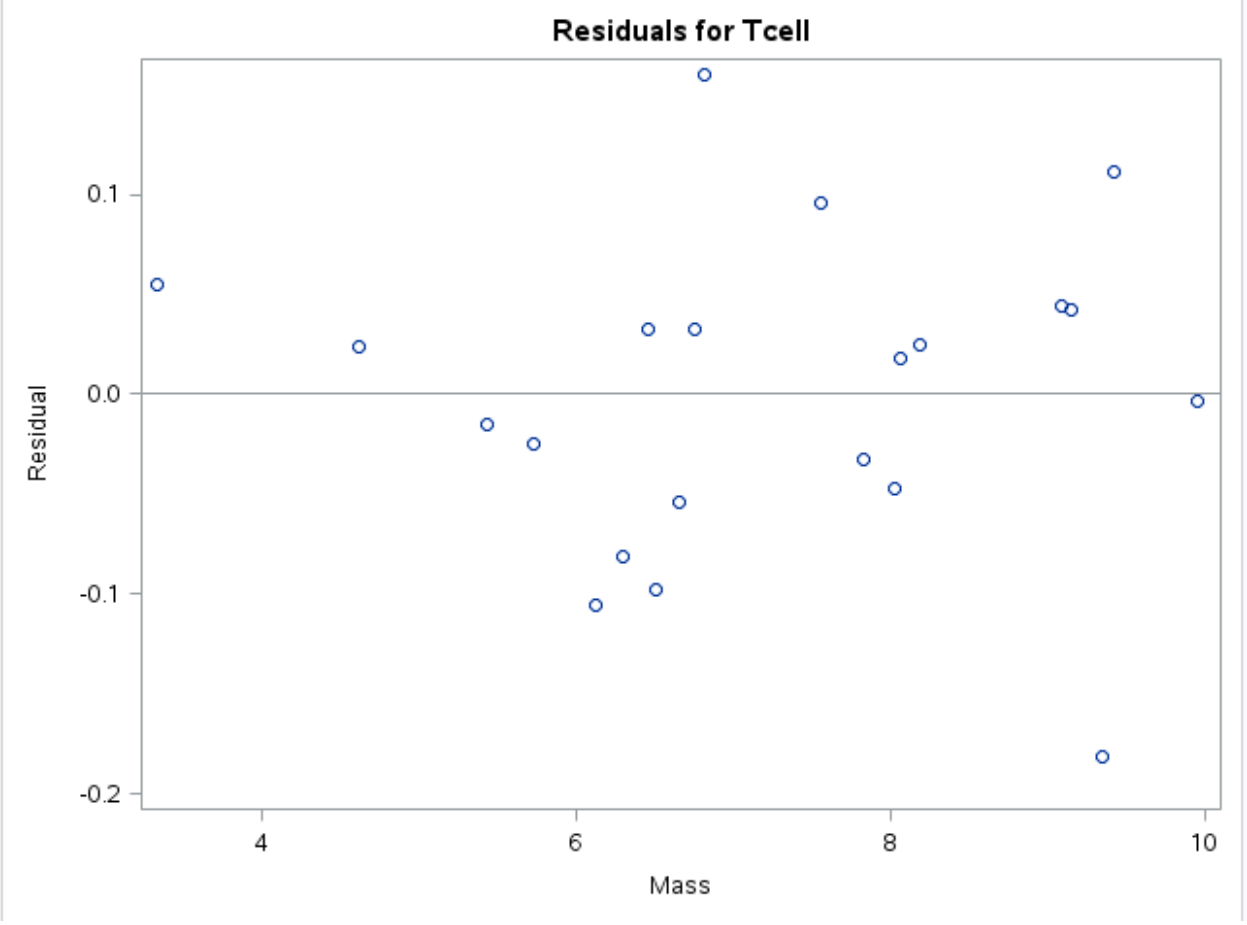
c. Repeat 1 (c) using SAS.

d. Repeat 1 (g) using SAS. & Repeat 1 (h) using SAS.

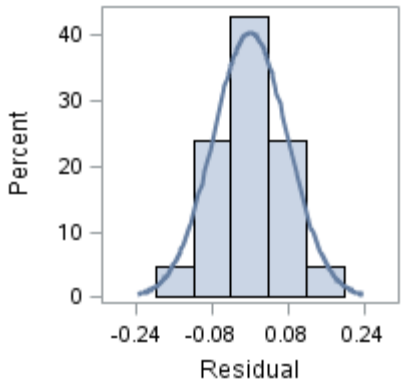
99% CL Mean		99% CL Predict	
0.0685	0.3251	-0.0682	0.4617
0.1456	0.3327	-0.0108	0.4891
0.1917	0.3397	0.0224	0.5090
0.2079	0.3432	0.0341	0.5170
0.2280	0.3488	0.0488	0.5279
0.2362	0.3517	0.0551	0.5328
0.2436	0.3547	0.0608	0.5376
0.2463	0.3560	0.0630	0.5394
0.2524	0.3591	0.0679	0.5436
0.2566	0.3615	0.0714	0.5467
0.2590	0.3630	0.0735	0.5486
0.2839	0.3874	0.0981	0.5731
0.2904	0.3985	0.1065	0.5825
0.2944	0.4071	0.1122	0.5893
0.2951	0.4089	0.1134	0.5907
0.2973	0.4146	0.1169	0.5951
0.3092	0.4618	0.1415	0.6295
0.3099	0.4657	0.1433	0.6323
0.3118	0.4770	0.1483	0.6405
0.3124	0.4810	0.1500	0.6433
0.3164	0.5118	0.1625	0.6656
0.1386	0.3318	-0.0159	0.4863

e. Repeat 1 (j) using SAS.





f. Repeat 1 (k) using SAS.



g. Repeat 1 (l) using SAS.

Root MSE	0.08102	R-Square	0.3336
Dependent Mean	0.32395	Adj R-Sq	0.2986
Coeff Var	25.00989		

Videos for using Excel:

$\hat{\beta}_0 \; and \; \hat{\beta}_1$ http://screencast.com/t/ztSxTImiOk6s

SE of $\hat{\beta}_0 \; and \; \hat{\beta}_1$ and RMSE: http://screencast.com/t/V9gnhSwb

Confidence Intervals: https://www.screencast.com/t/ELiUGTe7Kc

Prediction Intervals: https://www.screencast.com/t/ap8WETxsGUqN

CI and PI Plotting: https://www.screencast.com/t/efrpHrqgYZnG

Calibration Mean Gross: https://www.screencast.com/t/Yu7eqiiH0X

Calibration Single Movie: https://www.screencast.com/t/2vS1lGqtJ