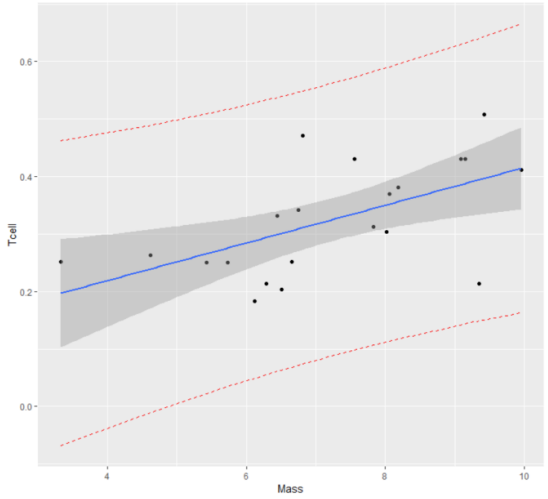
Javier Saldana

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.




```
#create model
model <- lm(Tcell ~ Mass, data = Male.Display.Data.Set)
#create predictions
pred.int <- predict(model, interval = "prediction", level = 0.99)
#combine predictions to data
mydata <- cbind(Male.Display.Data.Set, pred.int)
#plot data
ggplot(mydata, aes(Mass, Tcell)) +
  geom_point() +
  stat_smooth(method = lm) +
  geom_line(aes(y = lwr), color = "red", linetype = "dashed") +
  geom_line(aes(y = upr), color = "red", linetype = "dashed")
```

   b. Provide a table showing the t-statistics and p-values for the significance of the regression parameters β_0 *and* β_1 (as different from 0). Please do this in R.

```
> summary(model)

Call:
lm(formula = Tcell ~ Mass, data = Male.Display.Data.Set)

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.

	$\beta_0 = 0$	$\beta_1 = 0$
Hypotheses	$H_o: \beta_0 = 0$ $H_a: \beta_0 \neq 0$	$H_o: \beta_1 = 0$ $H_a: \beta_1 \neq 0$

Critical Value	±2.860935
	`> qt(0.995, 19)`
	`[1] 2.860935`
		±2.860935
		`> qt(0.995, 19)`
		`[1] 2.860935`

Test Statistic	1.112	3.084

Javier Saldana

p-value	0.2800111
	`> 2*pt(1.112, 19, lower.tail = F)`
	`[1] 0.2800111`
		0.006109041
		`> 2*pt(3.084, 19, lower.tail = F)`
		`[1] 0.006109041`

Decision	Fail to reject the null hypothesis	Reject the null hypothesis
Interpretation	There is insufficient evidence at the alpha = 0.01 level of significance (p value = 0.280) that the predicted tcell count is not zero when the average stone mass is 0. In other words, there is insufficient evidence that the regression equation does not pass through the origin. We do not have sufficient evidence on data collection to generalize to a broader population. Considering this was an observational study, the results may not establish causality.	There is evidence at the alpha = 0.01 level of significance (p value = 0.006) that there is a linear relationship between the mean mass of the stones carried and tcells in the birds. In other words, there is evidence that the slope of the regression equation is not equal to zero. We do not have sufficient evidence on data collection to generalize to a broader population. Considering this was an observational study, the results may not establish causality.

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

The mean of the predicted Tcell = mean mass of stones × 0.3282 + 0.0875

e. Interpret the slope in the model (regression equation).

For every 1 kg increase in the mean mass of the stones, the predicted Tcell mean increases by 0.3282.

f. Interpret the y-intercept in the model (regression equation).

When the birds have 0 mean mass of stones, the mean of the predicted Tcell count is 0.0875.

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.

```
> predict(model, newdata = data.point, interval = "confidence", level = 0.99)
        fit       lwr       upr
1 0.2351937 0.1385665 0.3318209
```

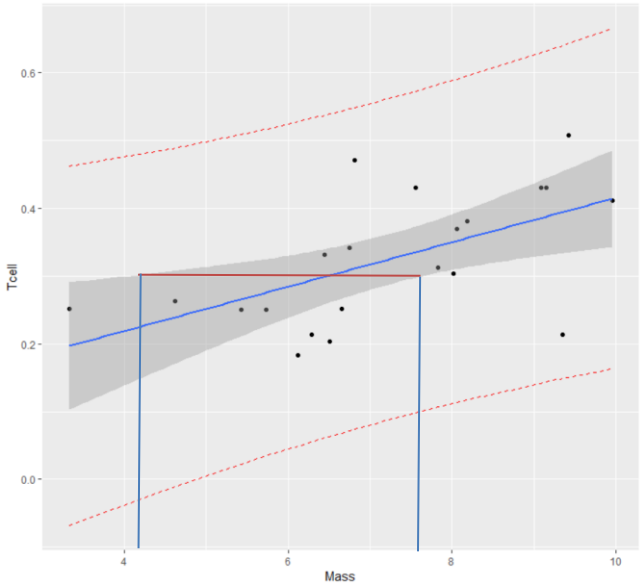
h. Find and interpret the 99% prediction interval for the predicted t-cell response given a stone mass of 4.5 grams. Please do this directly in R.

```
> predict(model, newdata = data.point, interval = "prediction", level = 0.99)
        fit         lwr       upr
1 0.2351937 -0.01593192 0.4863193
```

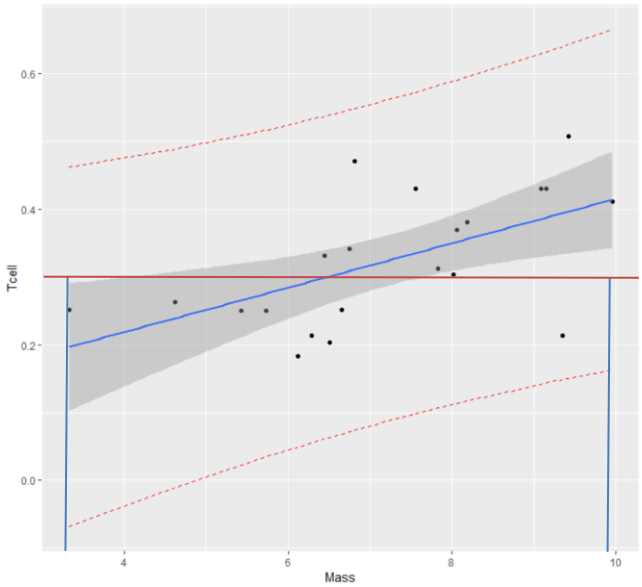
i. Calibration intervals:

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.

Javier Saldana



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



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(model, y0 = 0.3, interval = "inversion", mean.response = TRUE, level = 0.99)
 estimate     lower     upper
 6.474508 -4.389857  8.342649
```

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

```
> calibrate(model, y0 = 0.3, interval = "inversion", level = 0.99)
  estimate      lower      upper
  6.474508 -17.968869  21.921661
```

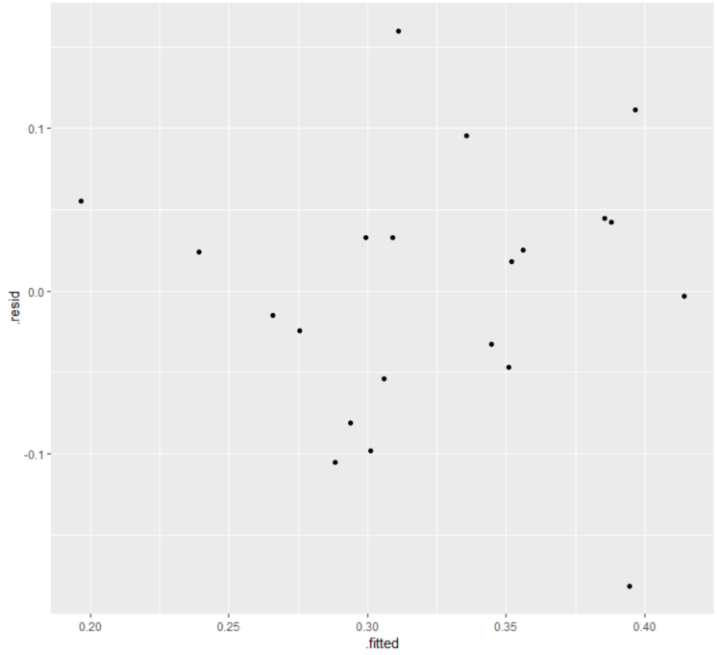
3. **Interpret** the following using the results from (1) and (2) above.
   a. 99% calibration interval for the **mean** t-cell response of 0.3.

We are 99% confident that the estimated mean mass of stones carried by birds with a 0.3 Tcell response is between 0 kg and 8.34 kg.

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

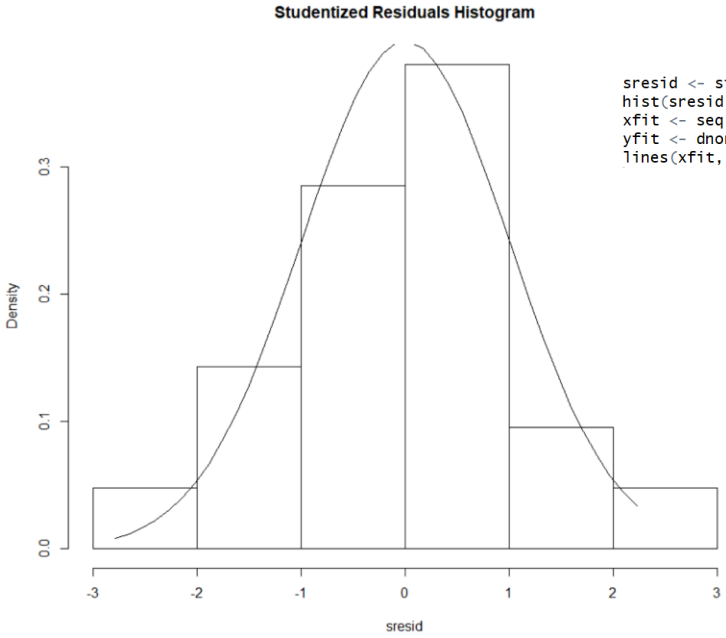
We are 99% confident that the mean mass of stones carried by a single bird with a Tcell response of 0.3 is between 0 kg and 21.9 kg.

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




```
ggplot(model) +
  geom_point(aes(x=.fitted, y=.resid))
```

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.")




```
sresid <- studres(model)
hist(sresid, freq = FALSE, main = "Studentized Residuals Histogram")
xfit <- seq(min(sresid),max(sresid), length = 40)
yfit <- dnorm(xfit)
lines(xfit, yfit)
```

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.

Javier Saldana

```
> summary(model)

Call:
lm(formula = Tcell ~ Mass, data = Male.Display.Data.Set)

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
```

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. $\hat{\beta}_0 \; and \; \hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0.03821 \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 0.087497$$

sx	xbar	ybar	B1	Bo	MSE	sigma hat	SE(B1 hat)	SE(Bo hat)
1.702441	7.204286	0.323952	0.032821	0.087497	0.006564	0.081019	0.010641	0.078677

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

$$t_{statistic} = t_{0.975,21-2} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = 3.084296 \qquad t_{statistic} = t_{0.975,21-2} = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} = 1.11211$$

   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.

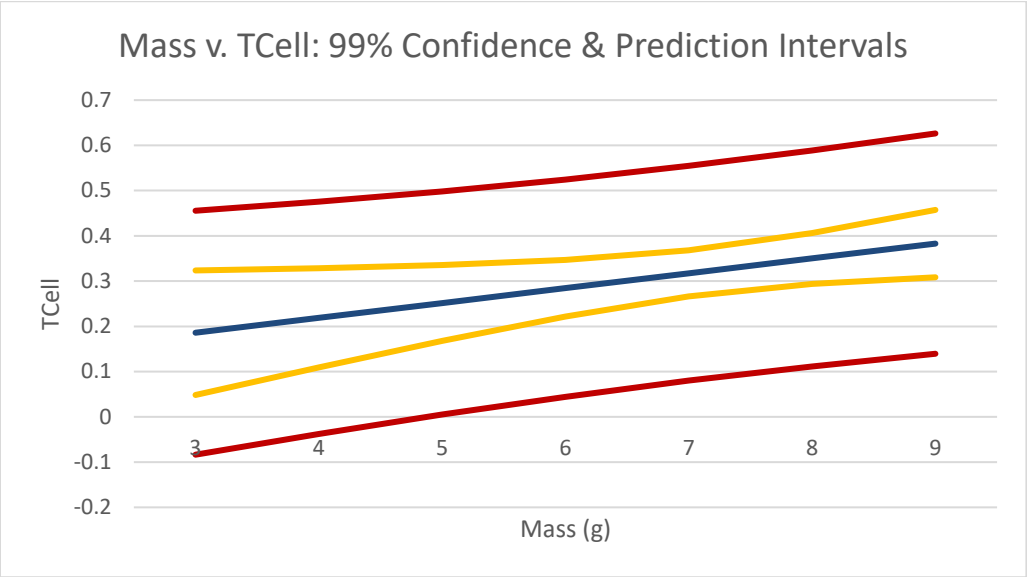
Mass (x)	Estimate	SE Mean	Lower	Upper
3	0.185961	0.048106	0.048332	0.323591
4	0.218783	0.038409	0.108896	0.32867
5	0.251604	0.029374	0.167569	0.33564
6	0.284426	0.021836	0.221954	0.346897
7	0.317247	0.017813	0.266285	0.368209
8	0.350069	0.019603	0.293986	0.406152
9	0.38289	0.026033	0.308411	0.45737

   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.

Mass (x)	Estimate	SE Predict	Lower	Upper
3	0.185961	0.094225	-0.08361	0.455534
4	0.218783	0.089663	-0.03774	0.475303
5	0.251604	0.08618	0.00505	0.498159
6	0.284426	0.08391	0.044364	0.524488
7	0.317247	0.082955	0.07992	0.554575
8	0.350069	0.083357	0.111589	0.588549
9	0.38289	0.085099	0.139427	0.626354

Javier Saldana

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.

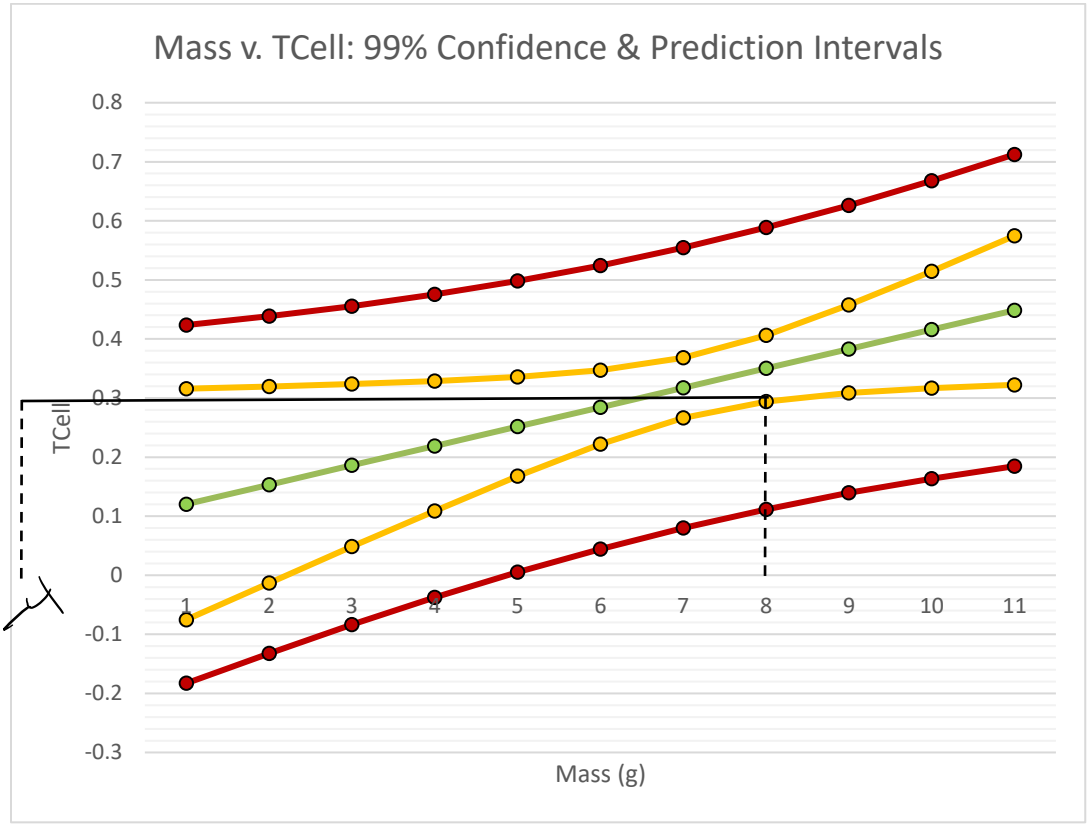
b. 99% calibration interval for a single t-cell response of 0.3.

		Calibration Interval for Mean Tcell				Calibration Interval for Single Tcell			
Tcell (y)	Est Mass	SE Mean	SE Est. Mas	Lower	Upper	SE Predict	SE Est. Mas	Lower	Upper
0.3	6.474508	0.01931	0.588344	4.791296	8.157721	0.083289	2.537633	-0.7855	13.73451

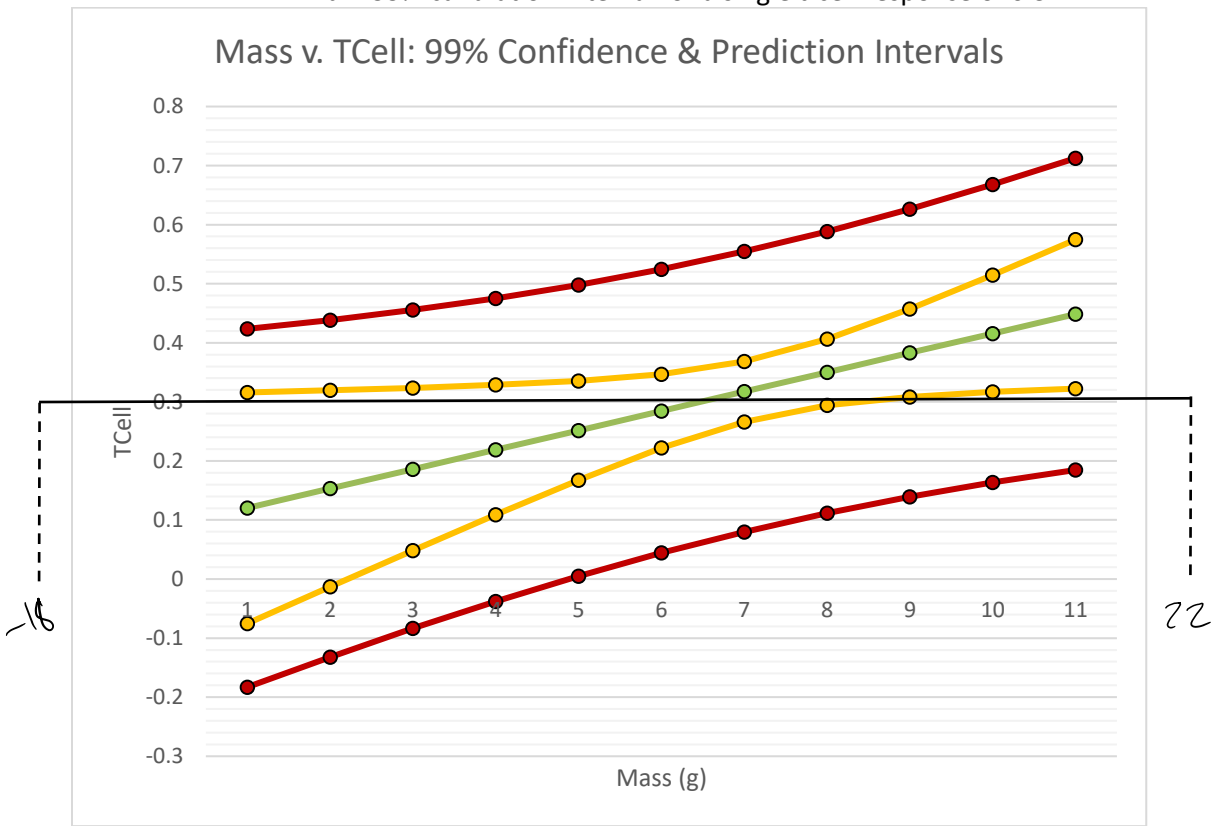
j. Using the **Excel graphs**, find the following. You may want to add data points to parts (c) and (d) so that the confidence and prediction limits extend well beyond the data range (although their interpretation is questionable outside the range).

a. 99% calibration interval for the mean t-cell response of 0.3.

Javier Saldana



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.

Javier Saldana

c. Repeat 1 (c) using SAS.
d. Repeat 1 (g) using SAS.
e. Repeat 1 (h) using SAS.
f. Repeat 1 (j) using SAS.
g. Repeat 1 (k) using SAS.
h. Repeat 1 (l) using SAS.

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