Inferences concerning the regression line

- Estimating the mean of the Y’s for a particular value of X, say \( X_0 \)
  
  - Example: what is the average per capita health spending for a country with per capita gross domestic product 10 PPP

\[
E[Y | X_0] = \hat{Y}_{X_0} = a + bX_0
\]

- estimated standard error of \( E[Y | X_0] \)

\[
\hat{\sigma}_{\hat{Y}_{X_0}} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}
\]

- A \((1 - \alpha)\)% confidence interval for \( E[Y | X_0] \) is given by:

\[
\hat{Y}_{X_0} \pm \left( t_{1-\alpha/2, df=n-2} \right) \left( \hat{\sigma}_{\hat{Y}_{X_0}} \right)
\]
Prediction interval for a new individual’s Y given that we know their value of X, say $X_0$

- Point estimate is the group mean
  $$\hat{Y}_{new} = a + bX_0$$

- Would you expect an individual’s response to be more or less variable than the group’s mean response?

- two sources of variability in prediction of individual Y
  - uncertainty of the group mean $E[Y|X_0]$
  - variability of individual responses around the group mean

* Example: The Netherlands has per capita PCGDP of 13, but is unlikely to have PCH of exactly the mean of all possible countries with the same GDP.

- estimated standard error of $\hat{Y}_{new}$
  $$\hat{\sigma}_{\hat{Y}_{new}} = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

- A $(1 - \alpha)\%$ prediction interval for $Y_{new}$ is given by:
  $$\hat{Y}_{new} \pm (t_{1-\alpha/2, df=n-2})(\hat{\sigma}_{\hat{Y}_{new}})$$

```r
proc reg data = nations;
model killed = power / cl;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: killed

```
  Output Statistics
Dep Var Predicted  Std Error
 Gini killed Value Mean Predict 95% C. Predict Residual

  1 12.0000 14.3627 1.3629 4.1636 24.6050  1.3627
  2 21.0000 16.0059 1.7974 5.8985 26.1130  4.9941
  3 26.0000 18.4280 1.0976 8.6836 28.5745  5.3730
  4 16.0000 20.7697 1.4528 10.3142 30.3341  4.7807
  5 26.0000 22.6236 1.3420 12.5832 32.3991  1.3704
  6 26.0000 23.4987 1.3688 13.7230 32.2887  2.4987
  7 10.0000 24.2686 1.3622 14.5320 32.3916  2.4680
  8 36.0000 28.3072 1.1482 18.7198 38.0147  5.6326
  9 30.0000 31.6137 1.2650 21.2606 41.3707  1.3863
 10 30.0000 35.2346 1.2909 25.9792 44.2673  2.2346
 11 30.0000 39.1034 1.5187 29.2178 48.9929  6.1054
 12 43.0000 42.8012 1.7974 32.7462 52.9882  0.1688
 13 50.0000 47.3682 2.1763 36.6937 57.6607  2.6538
 14 47.0000 48.3651 2.3647 37.8038 58.8884  1.3451
```

Sum of Residuals  0
Sum of Squared Residuals  238.44091
Predicted Residual SS (PRESS)  281.76275