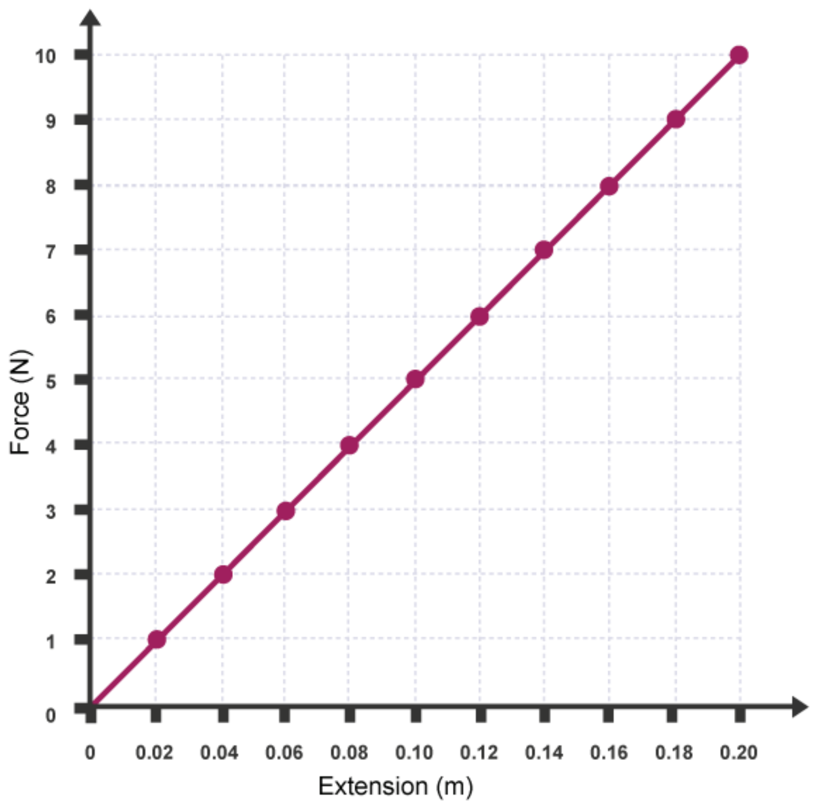
Fitting a Linear Regression with a Fixed Intercept

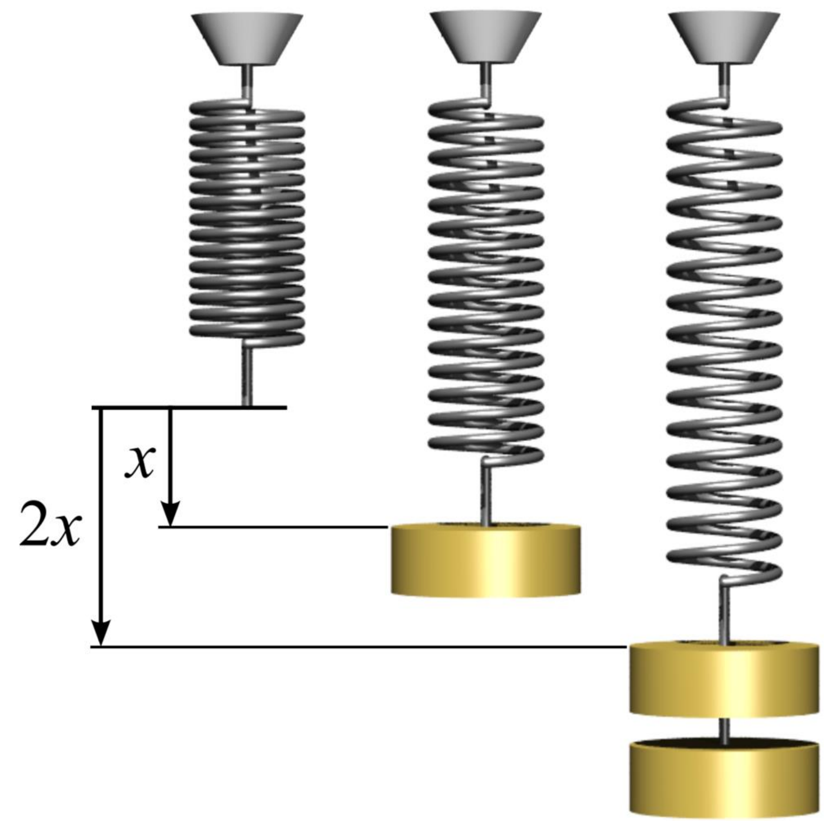


STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

When Does it Make Sense to Use Zero Intercept?

- When you are sure that the intercept is zero.
- Example: Hooke's Law
  - The spring extends more for heavier weights
  - Doesn't extend at all for weight zero
  - $\Delta X = \beta_1 w$
  - w=0 must mean ΔX=0



Qualitative Variables: A convenient way to estimate the mean

```r
summary(lm(Percentage~0+Pair, data=fish))
```

```
##
## Call:
## lm(formula = Percentage ~ 0 + Pair, data = fish)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.429  -8.414   0.247  10.859  28.871
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## PairPair1   56.406      3.864   14.60   <2e-16 ***
## PairPair2   60.886      4.131   14.74   <2e-16 ***
## PairPair3   62.429      3.749   16.65   <2e-16 ***
## PairPair4   67.000      4.131   16.22   <2e-16 ***
## PairPair5   64.211      5.152   12.46   <2e-16 ***
## PairPair6   63.336      4.131   15.33   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.46 on 78 degrees of freedom
## Multiple R-squared:  0.9458, Adjusted R-squared:  0.9416
## F-statistic: 226.8 on 6 and 78 DF,  p-value: < 2.2e-16
```

Would have to add the value of the intercept to the coefficients to get the sample means if the intercept weren't 0

ANOVA when the intercept is 0

```
anova(lm(Percentage~0+Pair, data=fish))
```

```
## Analysis of Variance Table
##
## Response: Percentage
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Pair       6 325175   54196  226.83 < 2.2e-16 ***
## Residuals 78  18637     239
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Percentage~Pair, data=fish))
```

```
## Analysis of Variance Table
##
## Response: Percentage
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Pair       5   938.7  187.75  0.7858  0.563
## Residuals 78 18636.7  238.93
```

$$SSE = \sum_{ij} \left(y_{ij} - \hat{y}_i\right)^2$$

$$SST = \sum_{ij} \left(y_{ij} - 0\right)^2$$

$$SSR = SST - SSE = \sum_{ij} (\hat{y}_i - 0)^2$$

$$SSE = \sum_{ij} \left(y_{ij} - \hat{y}_i\right)^2$$

$$SST = \sum_{ij} \left(y_{ij} - \bar{y}\right)^2$$

$$SSR = SST - SSE = \sum_{ij} (\hat{y}_i - \bar{y})^2$$

F-test

$$SSE = \sum_{ij} \left(y_{ij} - \hat{y}_i\right)^2$$

$$SST = \sum_{ij} \left(y_{ij} - 0\right)^2$$

$$SSR = SST - SSE = \sum_{ij} (\hat{y}_i - 0)^2$$

Null Hypothesis: $\mu_1 = \mu_2 = \cdots = 0$
F-statistic: $\frac{MSR}{MSE}$
$MSR = SSR/df_{SSR}$ *is potentially huge*

$$SSE = \sum_{ij} \left(y_{ij} - \hat{y}_i\right)^2$$

$$SST = \sum_{ij} \left(y_{ij} - \bar{y}\right)^2$$

$$SSR = SST - SSE = \sum_{ij} (\hat{y}_i - \bar{y})^2$$

Null Hypothesis: $\mu_1 = \mu_2 = \cdots = 0$
F-statistic: $\frac{MSR}{MSE}$

A Fix

- Suppose we ran ANOVA fixing the intercept at 0
- What we want is to test the hypothesis that *Null Hypothesis:* $\mu_1 = \mu_2 = \cdots = \bar{y}$
- Compute the right SSR by hand!
- (in R)