Chapter 2

Problem with significance testing:
Study one: $\widehat{\theta} = 25$ with SE $= 10$. Two-sided p-value $= .012$
Study two: $\widehat{\theta} = 10$ with SE $= 10$. p-value $= .30$
Compare study one to study two: $\widehat{\Delta} = 15$ with SE $= \sqrt{25^2 + 10^2} = 26.9$ for 2-sided p-value $= .35$

3.1 SLR

Regress kids test score on mom's high school (binary: 1 or 0). Interpret the intercept estimate (78) and slope estimate (12).

```
> summary(lm(kid.score ~ mom.hs, kids))$coef
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)       78        2.1    37.7 1.4e-138
mom.hs            12        2.3     5.1  6.0e-07
```

Regress kids test score on mom's IQ (continuous). Interpret the intercept estimate (26) and slope estimate (0.6).

```
> summary(lm(kid.score ~ mom.iq, kids))$coef
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    25.80      5.917     4.4  1.6e-05
mom.iq          0.61      0.059    10.4  7.7e-23
```


3.2 MLR

Using both predictors. Slopes are change when all other factors are held constant (and are conditional on the other terms being in the model). $\widehat{\beta} = (26 \;\; 6 \;\; 0.6)^T$
Holding other predictors constant does not make sense in many settings (e.g. polynomials, interactions).

```
> summary(lm(kid.score ~ mom.hs + mom.iq, kids))$coef
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    25.73      5.875     4.4  1.5e-05
mom.hs          5.95      2.212     2.7  7.4e-03
mom.iq          0.56      0.061     9.3  6.6e-19
```


Interpretations:

- Prediction: If we look at another person or group where IQ is one pt higher, but HS was the same, how does predicted score change? or Change in prediction for HS vs nonHS at the same IQ?
- Counterfactuals: Imagine rewinding the clock, and change moms HS from 0 to 1. What is the effect on the same kid's score? It's a thought experiment. I can imagine it better if we're assigning treatments. More in Chapters 9-10.

3.3 Interactions

Effect of IQ depends on the level of HS.
Plug in 0's for HS to get estimates for HS grads, 1's give adjustments to intercept and slope. What if HS were a factor with levels "nohs" and "hs"? How would R code them?

Note advice about finding interactions: tend to appear when main effects are large. Really??

Helps to center predictors about their mean.

3.4 Inference

Terminology

- Units – of analysis (not of measurement)
- Predictors
- Outcome (response) variable
- Matrix notation
- Distributions: $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{\Omega})$ with, for now, $\mathbf{\Omega} = \mathbf{I}$.
- G & H have an R package called `arm` containing the data and functions like `display`

OLS estimator

Variance of $\widehat{\beta} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$ is $\mathrm{Var}(\widehat{\beta}) = \sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1} = \sigma^2\mathbf{V}_\beta$
If two predictors are linearly independent, what will the covariance of their estimated coefficients be?

```
> summary(fit1)$coef
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    25.80      5.917     4.4  1.6e-05
mom.iq          0.61      0.059    10.4  7.7e-23
```

```
> summary(fit2)$coef
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    86.80      0.877      99 5.1e-299
centeredIQ      0.61      0.059      10  7.7e-23
```

```
> cbind(summary(fit1, cor=TRUE)$cor, summary(fit2, cor=TRUE)$
```

```
            (Intercept) mom.iq (Intercept) centeredIQ
(Intercept)        1.00  -0.99     1.0e+00    2.7e-16
mom.iq            -0.99   1.00     2.7e-16    1.0e+00
```


Centering

```
> anova(fit1, fit2)
```

```
Analysis of Variance Table

Model 1: kid.score ~ mom.iq
Model 2: kid.score ~ centeredIQ
  Res.Df    RSS Df Sum of Sq F Pr(>F)
1    432 144137
2    432 144137  0  5.82e-11
```

```
> summary(predict(fit1) - predict(fit2))
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-5.7e-14 -4.3e-14 -2.8e-14 -3.4e-14 -2.8e-14  0.0e+00
```

$$\mathbf{1}^\mathsf{T}(\mathbf{x} - \mathbf{1}\overline{x}) = \mathbf{1}^\mathsf{T}\mathbf{x} - \mathbf{1}^\mathsf{T}\mathbf{1}\overline{x} = \sum x_i - n\overline{x} = 0$$

Welcome to Bayesian Inference

Note in Figure 3.7, $\widehat{\beta}$ is known and labeled on the x axis. Frequentists would consider unknown β the center for the sampling distribution of $\widehat{\beta}$, i.e. $\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 \mathbf{V}_\beta)$. Our authors use a Bayesian interpretation, $\boldsymbol{\beta} \sim N(\widehat{\boldsymbol{\beta}}, \sigma^2 \mathbf{V}_\beta)$, our ignorance about the location of true $\boldsymbol{\beta}$ is expressed as a posterior distribution.

Variances

For any estimable linear combination of the coefficients, the variance is:

$$\mathrm{Var}(\boldsymbol{\lambda}^\mathsf{T}\widehat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{\lambda}^\mathsf{T}\mathrm{Var}(\widehat{\boldsymbol{\beta}})\boldsymbol{\lambda}$$

If $\boldsymbol{\lambda}$ is a single column what is the dimension?
If $\boldsymbol{\Lambda}$ has four columns (each making an estimable linear combination) what are the dimensions and what are the components of

$$\mathrm{Var}(\boldsymbol{\Lambda}^\mathsf{T}\widehat{\boldsymbol{\beta}}) =$$

Comparing Means

Compute pairwise comparisons of breakage by tension level.

```
> coef(warp.fit <- lm(breaks ~ tension , warpbreaks))
```

```
(Intercept)     tensionM     tensionH
        36          -10          -15
```

```
> Lambda <- matrix(c( 0,1,0, 0,0,1, 0,1,-1),byrow=TRUE,3,3)
> as.numeric(Lambda %*% coef(warp.fit))
```

```
[1] -10.0 -14.7   4.7
```

```
> sqrt(diag( Lambda %*% summary(warp.fit)$cov.unscaled %*%
+            t(Lambda))) * summary(warp.fit)$sigma
```

```
[1] 4 4 4
```


Another Generalized Inverse

```
> coef(warp.fit2 <- lm(breaks ~ tension -1, warpbreaks))
```

```
tensionL tensionM tensionH
      36       26       22
```

```
> Lambda <- matrix(c( 1,-1,0, 1,0,-1, 0,1,-1),byrow=TRUE,3,3)
> as.numeric(Lambda %*% coef(warp.fit2))
```

```
[1] 10.0 14.7  4.7
```

```
> sqrt(diag( Lambda %*% summary(warp.fit2)$cov.unscaled %*%
+            t(Lambda))) * summary(warp.fit2)$sigma
```

```
[1] 4 4 4
```

See how the two var-cov matrices compare:
`summary(warp.fit)$cov.unscaled*summary(warp.fit)$sigma^2`