Week 4 progress report for upgrading thyrosim

benjamin chu

4/25/2017

Feldschuh's paper

Two weeks ago we realized that steady state concentrations of T3/T4 in skinny patients are way too high. This is rather strange because our model works for normal and obese patients. After some literature search, we realized that Nadler's paper[1] which first proposed the formula we used to predict blood volume did not include raw data. Therefore, we hypothesize that perhaps Nadler's data does not contain enough skinny patients, so the formula they proposed were poorly extrapolated for this kind of patients.

To test whether our model is predicting the correct blood volume for skinny patients, we found real data published by Feldschuh[2]. They have enough data for patients of all sizes, but was not available in a downloadable format. So we used an online tool called WebPlotDigitizer, which can extract data from figures. In addition, Feldschuh's model is based on deviation from ideal weight, which we obtained through the online tool too.

Extracting the ideal weight model.

```r
male <- read.csv("feldschuh-male-idealweight.csv")
female <- read.csv("feldschuh-female-idealweight.csv")

height1 <- male$h
height2 <- height1^2
quadratic1.model <- lm(male$w ~ height1 + height2)
summary(quadratic1.model)
```

```
##
## Call:
## lm(formula = male$w ~ height1 + height2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41248 -0.10495  0.00975  0.10451  0.46944
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.003e+02  3.487e+00   57.44   <2e-16 ***
## height1     -2.557e+00  4.213e-02  -60.69   <2e-16 ***
## height2      1.064e-02  1.267e-04   84.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1895 on 77 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 1.954e+05 on 2 and 77 DF,  p-value: < 2.2e-16
```

```r
h1 <- female$h
h2 <- h1^2
```

```r
quadratic2.model <- lm(female$w ~ h1 + h2)
summary(quadratic2.model)
```

```
##
## Call:
## lm(formula = female$w ~ h1 + h2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40854 -0.08172 -0.01320  0.10256  0.39396
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.504e+02  2.733e+00   55.03   <2e-16 ***
## h1          -1.938e+00  3.314e-02  -58.47   <2e-16 ***
## h2           8.504e-03  9.989e-05   85.14   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1502 on 75 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 2.734e+05 on 2 and 75 DF,  p-value: < 2.2e-16
```

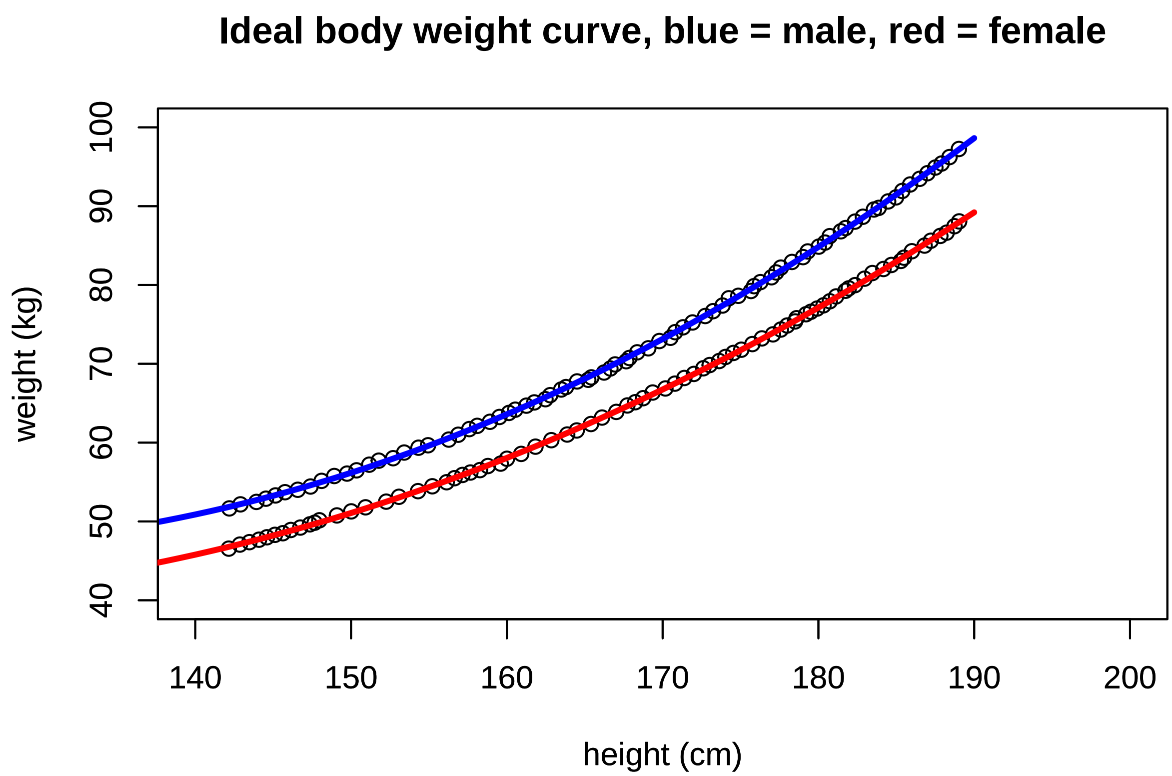
For both linear regressions, we have a multiple r-squared of 0.9998 and 0.9999 for males and females, indicating that our model is very accurate. We also fitted a cubic equation, but there was basically no difference.

```r
height_values_male <- seq(130, 190, 1)
predicted_weight_male <- predict(quadratic1.model,list(height1=height_values_male, height2=height_values

height_values_female <- seq(130, 190, 1)
predicted_weight_female <- predict(quadratic2.model,list(h1=height_values_female, h2=height_values_femal

plot(male, main="Ideal body weight curve, blue = male, red = female", xlab="height (cm)", ylab="weight
lines(height_values_male, predicted_weight_male, col = "blue", lwd = 3)

par(new=TRUE) #plots second graph on top of first
plot(female, xlab="height (cm)", ylab="weight (kg)", xlim=c(140, 200), ylim=c(40, 100))
lines(height_values_female, predicted_weight_female, col = "red", lwd = 3)
```



Let IW denote the ideal weight equation. We establish that the two ideal-weight formula used in Feldschuh's paper is the following:

$$IW_{male} = 200.3 - 2.557h + 0.01064h^2.$$

$$IW_{female} = 150.4 - 1.938h + 0.008504h^2$$

$$\text{multiple r-squared} = 0.9998 \text{ and } 0.9999$$

Model for Feldschuh's patients

```r
patient <- read.csv("patient_data.csv")
div <- patient$div
div2 <- div^2
div3 <- div^3
div4 <- div^4

cubic.model <- lm(patient$bv ~ div + div2 + div3) #cubic curve
summary(cubic.model)
```

```
##
## Call:
## lm(formula = patient$bv ~ div + div2 + div3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6084  -4.1431  -0.3651   4.0943  17.8979
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  7.026e+01  5.672e-01 123.878  < 2e-16 ***
## div         -5.212e-01  1.857e-02 -28.063  < 2e-16 ***
## div2         3.941e-03  3.820e-04  10.315  < 2e-16 ***
## div3        -9.873e-06  1.707e-06  -5.785 3.06e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.69 on 184 degrees of freedom
## Multiple R-squared:  0.8786, Adjusted R-squared:  0.8766
## F-statistic:   444 on 3 and 184 DF,  p-value: < 2.2e-16
```

```r
four.model <- lm(patient$bv ~ div+div2+div3+div4)
summary(four.model)
```

```
##
## Call:
## lm(formula = patient$bv ~ div + div2 + div3 + div4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.7008  -4.2906  -0.4514   4.1599  17.8642
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.974e+01  6.968e-01 100.082  < 2e-16 ***
## div         -5.167e-01  1.887e-02 -27.385  < 2e-16 ***
## div2         4.706e-03  7.114e-04   6.616 3.95e-10 ***
## div3        -2.111e-05  8.973e-06  -2.352   0.0197 *
## div4         3.696e-08  2.898e-08   1.275   0.2039
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.68 on 183 degrees of freedom
## Multiple R-squared:  0.8797, Adjusted R-squared:  0.8771
## F-statistic: 334.5 on 4 and 183 DF,  p-value: < 2.2e-16
```

```r
exponential.model <- lm(log(patient$bv) ~ div)
summary(exponential.model)
```

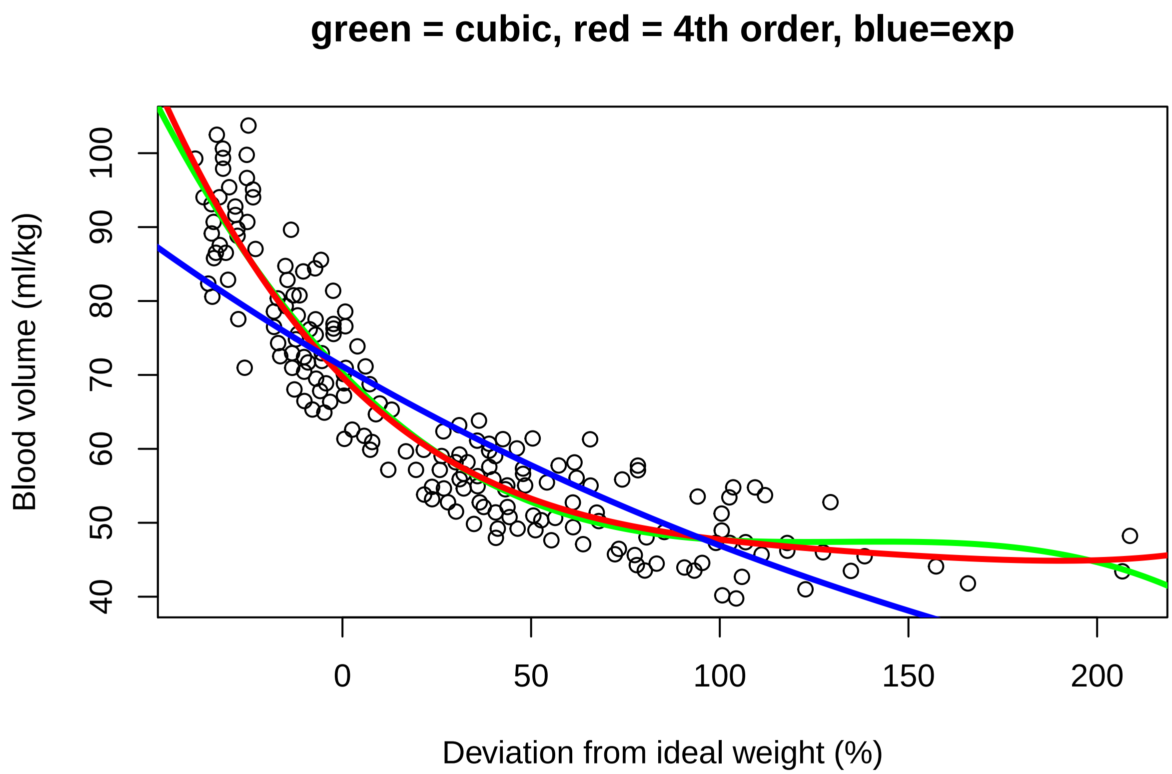
```
##
## Call:
## lm(formula = log(patient$bv) ~ div)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22530 -0.10357 -0.01262  0.07976  0.47996
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.2648908  0.0101864  418.69   <2e-16 ***
## div         -0.0041632  0.0001736  -23.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1209 on 186 degrees of freedom
```

```
## Multiple R-squared:  0.7556, Adjusted R-squared:  0.7543
## F-statistic: 575.1 on 1 and 186 DF,  p-value: < 2.2e-16
```

```
deviation_values <- seq(-50, 220, 1)
predicted_bv_cube <- predict(cubic.model,list(div=deviation_values, div2=deviation_values^2, div3=devia

predicted_bv_four <- predict(four.model, list(div=deviation_values,div2=deviation_values^2, div3=deviat
bv.exponential2 <- exp(predict(exponential.model,list(div=deviation_values)))

plot(patient, main="green = cubic, red = 4th order, blue=exp", xlab="Deviation from ideal weight (%)", y
lines(deviation_values, predicted_bv_cube, col = "green", lwd = 3)
lines(deviation_values, predicted_bv_four, col = "red", lwd = 3)
lines(deviation_values, bv.exponential2, lwd=3, col = "blue")
```



Here we chose a cubic polynomial to approximate the dataset, because Nadler's formula[1] was also a cubic function. On the other hand, there's no reason to stop there, so we fitted a 4th, 5th, . . . , 9th order polynomial too. Turns out that increasing the order beyond 4 worsens the fit, but 4th order is clearly better than 3rd order. We arrive at the following formula:

$$BV_{cubic} = 70.26 - 0.5212d + 0.003941d^2 - 0.000009873d^3$$

$$BV_{quartic} = 69.74 - 0.5167d + 0.0004706d^2 - 0.00002111d^3 + 0.0000000386d^3$$

$$\text{multiple r-squared} = 0.8786 \text{ and } 0.8797$$

References

1. Nadler, Samuel B., John U. Hidalgo, and Ted Bloch. "Prediction of blood volume in normal human adults." Surgery Vol 51 Issue 2, (1962), pg 224 ~ 232.
2. Feldschuh J, Enson Y. Prediction of the normal blood volume. Relation of blood volume to body habitus. Circulation. 1977 Oct;56(4 Pt 1):605-12.