Statistics 312/612, fall 2016
Homework # 4
Due: Monday 3 October

[1] The catheter data set is taken from a well known text book. If you happen to know the book please do not just repeat the analysis it presents.

```
cath <- read.table("catheter.txt",header=T)
outHW <- lm(distance ~ height + weight,cath)
outH <- lm(distance ~ height, cath)
outW <-lm(distance ~ weight, cath)
cath[1:3,]

##   height weight distance
## 1   42.8   40.0     37.0
## 2   63.5   93.5     49.5
## 3   37.5   35.5     34.5
```

The summary information for each fit (`outHW`, `outH`, `outW`) seems to suggest that height by itself is a good predictor of distance, that weight by itself is a good predictor of distance, but when both predictors are used then neither is particularly useful. (The stars in the summary table suggest 'significance'.)

(i) (15 points) Add more variables to the cath data.frame: height and weight centered to zero means (call them hcen and wcen); and things like `wres <-lm(wcen~ hcen, cath)$res`. Explain why some of the coeffs and std. errors are the same and some are different for the models:

```
distance ~ height
distance ~ hcen
distance ~ hcen + wres
distance ~ height+wres
distance ~ hcen+wcen
distance ~ height+weight
```

(ii) (5 points) Explains why `summary(outHW)` is misleading regarding the value of `height` and `weight` as predictors.

(iii) (5 points) Explain how the output from `cor(cath)` is relevant to the problem.

[2] The handout two_factors.pdf showed how to calculate several least squares fits using the Box-Cox data:

```
BC <- read.table("../Handouts/boxcox.data", header=T,sep="\t")
BC$rate <- 1/BC$time # transformation suggested by BHH page 235
BC$Htreatment <- C(BC$treatment,helmert)
out5 <- lm(rate ~ -1 + treatment,BC)
out7 <- lm(rate ~ Htreatment,BC)
```

In class I showed (page 9 of the handout) how to transform results from one parametrization into results for a different parametrization, using `out5` and `out6` as an example. For this homework problem I want you to recreate the shortened summary

```
## lm(formula = rate ~ -1 + treatment, data = BC)
##           tA    tB    tC    tD
## Est    3.519 1.862 2.947 2.161
## StdErr 0.292 0.292 0.292 0.292
```

using only the information contained in `out7`, which is essentially the same as the `out9` generated by:

```
C7 <- contrasts(BC$Htreatment)
dummyT <- outer(BC$treat,levels(BC$treat),"==")+0
X7 <- cbind(1, dummyT %*% C7)
out9 <- lm(BC$rate ~ -1+X7)
```

Display all the **R** code that you use.

(i) (5 points) Show that `X7` is equal to `dummyT %*% K7` where `K7 <- cbind(1,C7)`.

(ii) (10 points) If $\widehat{g}$ is the vector of coefficients from `out7` and $\widehat{b}$ is the vector of coefficients from `out5`, show that $\widehat{b} = K_7\widehat{g}$.

(iii) (10 points) Use (ii) and `out7` to recreate the shortened summary for `out5`.