Homework 10 Solutions
Statistics 200B
Due April 25

1. Derive the expression for the variance of $\hat{\beta}$ in simple linear regression, given in equation (13.11) in Wasserman, using the multivariate normal distribution for $\hat{\beta}$ we found in class (page 163 of the notes).

   Proof:

   From page 163 of the notes, we have

   $$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}),$$

   where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ and $\beta = (\beta_0, \beta_1)^T$.

   In simple linear regression,

   $$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

   So

   $$X'X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}.$$

   $$\begin{aligned} \Rightarrow (X'X)^{-1} &= \frac{1}{n \sum_{i=1}^n (X_i - \bar{X}_n)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \\ &= \frac{1}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{X}_n \\ -\bar{X}_n & 1 \end{bmatrix} \\ &= \frac{1}{n s_X^2} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{X}_n \\ -\bar{X}_n & 1 \end{bmatrix}, \end{aligned}$$

   where $s_X^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

   Therefore,

   $$V[\hat{\beta}] = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{n s_X^2} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{X}_n \\ -\bar{X}_n & 1 \end{bmatrix}.$$

2. Assume a multiple linear regression model with normal errors. Take σ to be known. Show that the model with the highest AIC is the model with the lowest Mallows C_p statistic.

   Proof:

   The multiple linear regression model is

   $$Y = X\beta + \epsilon,$$

   where $\epsilon \sim N(0, \sigma^2 I_n)$.

   Given a model S, and define $|S|$ as the number of parameters in S, the AIC is

   $$\begin{aligned} AIC(S) &= \ell(\hat{\beta}) - |S| \\ &= -n\log(2\pi) - \frac{n}{2}\log\sigma - \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{2\sigma^2} - |S| \\ &= \text{constant} - \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{2\sigma^2} - |S|, \end{aligned}$$

   where $\hat{\beta}$ is the MLE.

   The Mallows C_p statistic is

   $$\begin{aligned} \hat{R}(S) &= \hat{R}_{tr}(S) + 2|S|\sigma^2 \\ &= \sum_{i=1}^{n}\left(\hat{Y}_i(S) - Y_i\right)^2 + 2|S|\sigma^2 \\ &= \sum_{i=1}^{n}\left(X_i\hat{\beta} - Y_i\right)^2 + 2|S|\sigma^2 \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + 2|S|\sigma^2 \\ &= -2\sigma^2 AIC(S). \end{aligned}$$

   Therefore, maximizing the AIC is equivalent to minimizing the Mallows C_p statistic.

3. Consider the simple linear regression model

   $$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

   with $E[\epsilon_i] = 0$ and $V[\epsilon_i] = \sigma^2$. Under the following two cases, find what happens to the estimates, standard errors, and Wald test statistics for β_0 and β_1.

(a) Instead of X, we regress Y on a new variable $Q = aX + b$.

(b) Instead of Y, we construct a new variable $R = aY + b$ and regress this on X.

Solution:

Under the original model, the estimates are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2},$$
$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n.$$

The standard errors are

$$se(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2}},$$
$$se(\hat{\beta}_0) = \sigma \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X}_n)^2}}.$$

The Wald test statistics are

$$W(\hat{\beta}_1) = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)},$$
$$W(\hat{\beta}_0) = \frac{\hat{\beta}_0}{\hat{se}(\hat{\beta}_0)},$$

where $\hat{se}$ denotes substituting $\hat{\sigma}^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 / n$ for σ^2.

(a) $X = \frac{Q-b}{a}$. The model becomes

$$Y_i = \beta_0 + \beta_1 \frac{Q_i - b}{a} + \epsilon_i = (\beta_0 - \frac{\beta_1 b}{a}) + \frac{\beta_1}{a} Q_i + \epsilon_i.$$

Define $\alpha_0 = \beta_0 - \frac{\beta_1 b}{a}$ and $\alpha_1 = \frac{\beta_1}{a}$. Then $\beta_1 = a\alpha_1$ and $\beta_0 = \alpha_0 + b\alpha_1$. The model is

$$Y_i = \alpha_0 + \alpha_1 Q_i + \epsilon_i.$$

The estimates of α_1, α_0 are

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^n (Q_i - \bar{Q}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (Q_i - \bar{Q}_n)^2} = \frac{\sum_{i=1}^n a(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n a^2(X_i - \bar{X}_n)^2} = \frac{1}{a}\hat{\beta}_1,$$
$$\hat{\alpha}_0 = \bar{Y}_n - \hat{\alpha}_1 \bar{Q}_n = \bar{Y}_n - \frac{1}{a}\hat{\beta}_1(a\bar{X}_n + b) = \bar{Y}_n - \hat{\beta}_1\bar{X}_n - \frac{b}{a}\hat{\beta}_1 = \hat{\beta}_0 - \frac{b}{a}\hat{\beta}_1.$$

The standard errors of $\hat{\alpha}_1$, $\hat{\alpha}_0$ and their covariance are

$$se(\hat{\alpha}_1) = \frac{1}{a} se(\hat{\beta}_1),$$
$$se(\hat{\alpha}_0) = \sqrt{V(\hat{\alpha}_0)} = \sqrt{V(\hat{\beta}_0) + \frac{b^2}{a^2}V(\hat{\beta}_1) - 2\frac{b}{a}Cov(\hat{\beta}_0, \hat{\beta}_1)}.$$

The Wald test statistics are

$$W(\hat{\alpha}_1) = \frac{\hat{\alpha}_1}{\hat{se}(\hat{\alpha}_1)} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)},$$
$$W(\hat{\alpha}_0) = \frac{\hat{\alpha}_0}{\hat{se}(\hat{\alpha}_0)} \neq \frac{\hat{\beta}_0}{\hat{se}(\hat{\beta}_0)}.$$

By comparing $\hat{\alpha}_1$, $\hat{\alpha}_0$, $se(\hat{\alpha}_1)$, $se(\hat{\alpha}_0)$, $W(\hat{\alpha}_1)$ and $W(\hat{\alpha}_0)$ to $\hat{\beta}_1$, $\hat{\beta}_0$, $se(\hat{\beta}_1)$, $se(\hat{\beta}_0)$, $W(\hat{\beta}_1)$ and $W(\hat{\beta}_0)$, we can see that the slope and intercept estimates and their standard errors are different. However, the Wald test statistic of the slope remains the same, while the one for the intercept is changed.

(b) $Y = \frac{R-b}{a}$. The model becomes

$$\frac{R_i - b}{a} = \beta_0 + \beta_1 X_i + \epsilon_i.$$

$$\Rightarrow R_i = (a\beta_0 + b) + a\beta_1 X_i + a\epsilon_i.$$

Define $\gamma_0 = a\beta_0 + b$ and $\gamma_1 = a\beta_1$. Then $\beta_1 = \frac{\gamma_1}{a}$ and $\beta_0 = \frac{\gamma_0 - b}{a}$. Define $\delta_i = a\epsilon_i$. The model is

$$R_i = \gamma_0 + \gamma_1 X_i + \delta_i,$$

with $E[\delta_i] = 0$ and $V[\delta_i] = a^2\sigma^2$.

The estimates of γ_1, γ_0 are

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)(R_i - \bar{R}_n)}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} = \frac{a\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} = a\hat{\beta}_1,$$
$$\hat{\gamma}_0 = \bar{R}_n - \hat{\gamma}_1\bar{X}_n = a\bar{Y}_n + b - a\hat{\beta}_1\bar{X}_n = a(\bar{Y}_n - \hat{\beta}_1\bar{X}_n) + b = a\hat{\beta}_0 + b.$$

The standard errors of $\hat{\gamma}_1$, $\hat{\gamma}_0$ and their covariance are

$$se(\hat{\gamma}_1) = a \cdot se(\hat{\beta}_1),$$
$$se(\hat{\gamma}_0) = a \cdot se(\hat{\beta}_0).$$

The Wald test statistics are

$$W(\hat{\gamma}_1) = \frac{\hat{\gamma}_1}{\hat{se}(\hat{\gamma}_1)} = \frac{a\hat{\beta}_1}{a\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)},$$
$$W(\hat{\gamma}_0) = \frac{\hat{\gamma}_0}{\hat{se}(\hat{\gamma}_0)} \neq \frac{\hat{\beta}_0}{\hat{se}(\hat{\beta}_0)}.$$

The results can be explained as follows. First, the Wald test for the slope is to check if the covariate has significant effect on the response variable. It is reasonable that the test result remains the same after the linear transformation of the covariate or the response. Second, the Wald test for the intercept is to check if the regression line goes through the origin, and the result will be changed after linear transformation of the covariate or the response.

4. Consider the multiple regression model with k possible predictors. For a particular model, let $S \subseteq \{1, \ldots, k\}$ denote the indices of the included regressors. Prove that
$$R^2(S) = \frac{\sum_{i=1}^{n}(\hat{Y}_i(S) - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$
can not decrease by adding an additional term to S.

Proof:

First, we can show that

$$R^2(S) = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

Note that you may use the fact that $TSS = ESS + RSS$ from last homework to show the above, or to prove it for the multiple regression case, please see below.

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}\left[(Y_i - \hat{Y}_i(S)) + (\hat{Y}_i(S) - \bar{Y})\right]^2 \\ &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))^2 + \sum_{i=1}^{n}(\hat{Y}_i(S) - \bar{Y})^2 + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))(\hat{Y}_i(S) - \bar{Y}),\end{aligned}$$

and

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))(\hat{Y}_i(S) - \bar{Y}) = \sum_{i=1}^{n}\hat{\epsilon}_i(X_i\hat{\beta} - \bar{Y}) = (X\hat{\beta})^T\hat{\epsilon} - \bar{Y}\sum_{i=1}^{n}\hat{\epsilon}_i = \hat{\beta}^T X^T\hat{\epsilon} - \bar{Y}\sum_{i=1}^{n}\hat{\epsilon}_i,$$

where $\hat{\beta}$ is the least squares estimate, which minimizes

$$f(\beta) = \sum_{i=1}^{n}(Y_i - X_i\beta)^2 = (Y - X\beta)^T(Y - X\beta).$$

So

$$\bigtriangledown f(\hat{\beta}) = -2X^T(Y - X\hat{\beta}) = -2X^T\hat{\epsilon} = 0.$$

Since

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}_{n\times k} = \begin{bmatrix} 1 & X_{12}\cdots & X_{1k} \\ \vdots & \vdots\cdots & \vdots \\ 1 & X_{n2}\cdots & X_{nk} \end{bmatrix},$$

$X^T\hat{\epsilon} = 0$ implies that $\sum_{i=1}^{n}\hat{\epsilon}_i = 0$.

Hence,

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))(\hat{Y}_i(S) - \bar{Y}) = 0,$$

and

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))^2 + \sum_{i=1}^{n}(\hat{Y}_i(S) - \bar{Y})^2.$$

So

$$R^2(S) = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

When an additional term is added to S, there are $k+1$ regressors on the model. Define the least squares estimate for the new model S^* as $\hat{\beta}^*$, and the least squares estimate for the old model as $\hat{\beta}$. We can write $\hat{\beta}^*$ and $\hat{\beta}$ as

$$\hat{\beta}^* = \begin{pmatrix} \hat{\beta}_1^* \\ \vdots \\ \hat{\beta}_k^* \\ \hat{\beta}_{k+1}^* \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \\ 0 \end{pmatrix},$$

because in the old model, the $(k+1)$th regressor is not included, its coefficient estimate is 0.

Hence, $\hat{\beta}$ can be considered as a parameter estimate candidate for the new model. Its fitted values are $\hat{Y}_i(S)$.

Because $\hat{\beta}^*$ is the least squares estimate for the new model S^*, it minimizes

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S^*))^2.$$

So

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i(S^*))^2 \leq \sum_{i=1}^{n}(Y_i - \hat{Y}_i(S))^2.$$

Therefore,

$$R^2(S^*) \leq R^2(S).$$

We have shown that $R^2(S)$ can not decrease by adding an additional term to S.

5. Use `load("hitters.RData")` to load the data frame `hitters` into `R`. This data frame contains information about baseball player salaries in the 1987 season, as well as a number of possible predictors of salary, described below.

   (a) Using the `stepAIC` function in `R` (you will first need to call `library(MASS` to load the package with this function), use backward stepwise selection to choose two models for predicting the log of the 1987 salary, one based on AIC and one based on BIC. (*Hint: look at the help page for* `stepAIC` *to see how to "trick" stepAIC into using BIC instead.*) What variables are included in the chosen models for each? How do they compare?

(b) Write a function in `R` to compute the leave-one-out cross-validation risk estimator for a fitted model `mod` obtained using `lm`. *Hint:* `lm.influence(mod)$hat` *returns the diagonal elements of the matrix U in (13.30) of Wasserman.* Use your function to compute the risk estimates for the models selected via AIC and BIC from part (a). Which model is preferred by this criterion?

(c) Look at the p-values for the individual t-tests being computed for the full model (with all possible regressors) to the model selected by BIC. You should see that many of the individual t-tests are significant relative to the smaller model, but are not significant in the larger model. Why do you think this occurs?

Description of the baseball data: The Statistical Graphics and Statistical Computing Sections of the American Statistical Association sponsor a bi-annual "Data Exposition" (`http://stat-computing.org/dataexpo`), for which this data appeared in 1988. The data as I present it here incorporates some corrections and variable transformations as described in "Applied Regression Analysis and Generalized Linear Models" by John Fox.

In addition to salary (`salary1987`), the data set contains the following variables for the 1986 season: at bats (`AB86`), hits (`H86`), home runs (`HR86`), runs scored (`R86`), runs batted in (`RBI86`), walks (`W86`), put-outs (`PO86`), assists (`A86`), errors (`E86`), batting average (`AVG86`), and on-base percentage (`OBP86`). The next eight variables in the data frame are equivalent to the first eight 1986 variables, but over the course of the player's professional career. Next we have number of years in the major league (`years`), and the same eight variables but on a per-year basis (each career variable divided by (`years`). The next four variables are indicators for position: middle infielders (`MI`), catchers ((`C`), center fielders (`CF`), and designated hitters (`DH`). (Note that this dataset does not contain pitchers.) After 3 years in the major leagues, players are eligible for salary arbitration, and after 6 years they are eligible for free agency (can negotiate a contract with any team). The last two variables (`Y35` and `YG6`) are dummy variables for these two categories.

Solutions:

First, read in the data.

```
> load("hitters.Rdata")
> head(hitters)
```

(a) • Backward stepwise selection via AIC:

```
> library(MASS)
> full <- lm(log(salary87)~., data=hitters)
> backwardAIC <- stepAIC(full, direction = "backward")
> backwardAIC$call
lm(formula = log(salary87) ~ AB86 + H86 + RBI86 + W86 + A86 +
    OBP86 + careerAB + careerR + careerW + careerAVG + careerOBP +
    pyAB + pyH + pyRBI + MI + C + CF + DH + Y35 + YG6, data = hitters)
```

Variable included in the chosen model via AIC can be seen above.

- Backward stepwise selection via BIC:

```
> backwardBIC <- stepAIC(full, direction = "backward",
    k=log(nrow(hitters)))
> backwardBIC$call
lm(formula = log(salary87) ~ H86 + W86 + A86 + careerAVG + pyAB +
    MI + CF + Y35 + YG6, data = hitters)
```

Variable included in the chosen model via BIC can be seen above. By comparing the selected models via AIC and BIC respectively, we can see that BIC selects a smaller model, whose regressors are a subset of the regressors selected via AIC. This is what we expected, as BIC has a heavier penalty on the number of regressors in the model.

(b) The leave-one-out cross validation is written as below.

```
LOOCV <- function(mod) {
sum((mod$residuals/(1-lm.influence(mod)$hat))^2)
}
```

The leave-one-out cross validation risk estimator for the models selected via AIC and BIC can be calculated as

```
> LOOCV(backwardAIC)
[1] 41.10894
> LOOCV(backwardBIC)
[1] 41.04930
```

Therefore, the model selected by BIC is slightly preferred by this criterion.

(c) The p-vaues for the individual t-tests for the full model and the model selected by BIC can be seen below.

```
> summary(full)

Call:
```

```
lm(formula = log(salary87) ~ ., data = hitters)

Residuals:
Min        1Q    Median        3Q       Max
-1.378680 -0.243997 -0.006161  0.279648  0.828883

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.427e+00  9.554e-01   2.541   0.0117 *
AB86        -4.143e-03  2.203e-03  -1.880   0.0613 .
H86          1.650e-02  7.707e-03   2.141   0.0333 *
HR86        -1.602e-03  9.797e-03  -0.163   0.8703
R86          8.389e-04  4.450e-03   0.189   0.8506
RBI86       -3.557e-03  3.761e-03  -0.946   0.3453
W86          1.090e-02  6.287e-03   1.734   0.0843 .
PO86         2.921e-05  1.084e-04   0.269   0.7879
A86         -6.313e-04  4.130e-04  -1.529   0.1277
E86         -2.119e-04  5.796e-03  -0.037   0.9709
AVG86       -1.295e+00  4.607e+00  -0.281   0.7788
OBP86       -4.573e-02  4.483e-02  -1.020   0.3088
careerAB    -3.263e-04  2.702e-04  -1.208   0.2284
careerH      5.618e-04  1.357e-03   0.414   0.6793
careerHR    -5.377e-04  3.099e-03  -0.174   0.8624
careerR      1.152e-03  1.424e-03   0.809   0.4195
careerRBI    5.108e-04  1.455e-03   0.351   0.7258
careerW     -6.471e-04  5.987e-04  -1.081   0.2809
careerAVG    7.458e+00  5.046e+00   1.478   0.1408
careerOBP    3.358e-02  4.570e-02   0.735   0.4632
years        7.704e-03  2.425e-02   0.318   0.7510
pyAB         7.309e-03  3.429e-03   2.131   0.0341 *
pyH         -2.157e-02  1.417e-02  -1.522   0.1294
pyHR         6.193e-03  2.616e-02   0.237   0.8131
pyR          6.249e-04  1.182e-02   0.053   0.9579
pyRBI        5.100e-03  1.213e-02   0.420   0.6746
pyW         -4.059e-05  8.818e-03  -0.005   0.9963
MITRUE       2.421e-01  1.142e-01   2.120   0.0351 *
CTRUE        1.751e-01  9.222e-02   1.898   0.0589 .
CFTRUE      -2.097e-01  1.021e-01  -2.055   0.0411 *
```

```
DHTRUE      -1.720e-01  1.402e-01  -1.227   0.2210
Y35TRUE      9.254e-01  9.070e-02  10.202   <2e-16 ***
YG6TRUE      1.544e+00  1.183e-01  13.044   <2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0

Residual standard error: 0.3864 on 229 degrees of freedom
Multiple R-squared: 0.8324,Adjusted R-squared: 0.809
F-statistic: 35.55 on 32 and 229 DF,  p-value: < 2.2e-16

> summary(backwardBIC)

Call:
lm(formula = log(salary87) ~ H86 + W86 + A86 + careerAVG + pyAB +
MI + CF + Y35 + YG6, data = hitters)

Residuals:
Min       1Q   Median       3Q      Max
-1.46564 -0.26579  0.01443  0.28319  0.86108

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.1188462  0.3357682   6.310 1.24e-09 ***
H86          0.0023662  0.0009519   2.486  0.01358 *
W86          0.0067132  0.0014217   4.722 3.88e-06 ***
A86         -0.0010304  0.0003109  -3.315  0.00105 **
careerAVG    5.7935964  1.4516964   3.991 8.64e-05 ***
pyAB         0.0021764  0.0002712   8.025 3.86e-14 ***
MITRUE       0.2679487  0.1037334   2.583  0.01036 *
CFTRUE      -0.2313267  0.0896821  -2.579  0.01046 *
Y35TRUE      0.9040871  0.0803966  11.245  < 2e-16 ***
YG6TRUE      1.4886726  0.0745177  19.977  < 2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0

Residual standard error: 0.3878 on 252 degrees of freedom
Multiple R-squared: 0.8143,Adjusted R-squared: 0.8077
F-statistic: 122.8 on 9 and 252 DF,  p-value: < 2.2e-16
```

We can see that individual t-tests for `H86`, `W86`, `A86`, `careerAVG` and `pyAB` are more significant in the smaller model than in the larger model. The reason is that, in the larger model, there are other variables who are possibly linearly dependent of them, and hence their effects are to some extent masked by those variables.