INSH5301 Intro Computational Statistics

Ali Banijamali

03/23/2020

PROBLEM 1: College Distance (Revisited)

For this problem we are going to explore the effect of distance from college on educational attainment. The dependent variable (Y) is years of completed education ed. Run any regression using the lm command and display regression tables using the stargazer command.

```r
# Required Libraries:
# install.packages('stargazer')
# install.packages('car')
library('stargazer')
library('ggplot2')
library('car')

# Reading the data:
col.dist <- read.csv(
  paste('C:/Users/alibs/Google Drive/Courses/INSH5301 Intro Computational Statistics/',
  'Module 9 - Multiple Regression 2/collegeDistance/CollegeDistance.csv', sep=''),
  stringsAsFactors = F)
# In my previous HW you mentioned you are ok with paths being out of border,
# but I left them like this any way, it is tidier this way :)
```

1.1. Test the hypothesis that college distance is related with educational attainment using a bivariate regression model.

ANS.

```r
ed.vs.colDist <- lm(ed ~ dist, data=col.dist)

stargazer(ed.vs.colDist, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment (Years) vs. College Distance",
          dep.var.labels=c("Educational Attainment (Years)"),
          covariate.labels=c("College Distance"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # TO hold the table at it's place (Not floating)
          # Floting tables get out of order in the printed output
         )
```

Table 1: Educational Attainment (Years) vs. College Distance

	Dependent variable:
	Educational Attainment (Years)
College Distance	−0.073***
	(0.014)
Constant	13.956***
	(0.038)
Observations	3,796
R^2	0.007
Adjusted R^2	0.007
Note:	*p<0.1; **p<0.05; ***p<0.01

As can be seen from the summary of the results, the slope of regression is -0.073. There is a negative relationship, meaning that as the college distance increases, students attain less years of education. the $R^2/AdjustedR^2$ are 0.007 howerver, meaning that this parameter (College Distance) only explains %7 of the variation in educational attainments. We can improve the model by adding more useful parameters.

1.2. A regression will suffer from omitted variable bias when two conditions hold. What are these two conditions? Do these conditions seem to hold in this case?

ANS.
First let's see what is Omitted Variable Bias (OVB):
In statistics, omitted-variable bias (OVB) occurs when a statistical model leaves out one or more relevant variables. The bias results in the model attributing the effect of the missing variables to the estimated effects of the included variables.

For this condition to exist in a linear regression, 2 conditions must hold true:
- The omitted variable must be a determinant of the dependent variable (i.e., its true regression coefficient must not be zero); and
- The omitted variable must be correlated with an independent variable specified in the regression (i.e., cov(z,x) must not equal zero).

Both of these conditions hold true here as we found out in the previous HW, however we can check these conditions again:

```
# 1. Checking the correlation between parameters:
cor(col.dist)
```

```
##                 female       black    hispanic      bytest     dadcoll
## female    1.000000000  0.04478205 -0.03302302 -0.07289937 -0.03854340
## black     0.044782050  1.00000000 -0.20506904 -0.31467768 -0.11264367
## hispanic -0.033023018 -0.20506904  1.00000000 -0.12989571 -0.06794587
## bytest   -0.072899369 -0.31467768 -0.12989571  1.00000000  0.25586014
## dadcoll  -0.038543401 -0.11264367 -0.06794587  0.25586014  1.00000000
## momcoll  -0.017521065 -0.03640195 -0.07094757  0.19748799  0.44540301
## ownhome  -0.046632926 -0.14045532 -0.07704685  0.12874932  0.07776654
## urban    -0.004875662  0.18928751  0.15327707 -0.10373517 -0.07195792
## cue80     0.027641792 -0.06554238  0.09065139 -0.01840956 -0.08830895
## stwmfg80 -0.031614444 -0.16203034 -0.03376490  0.11265098  0.02433183
## dist     -0.003368955 -0.09989332  0.02114091 -0.06153932 -0.11325766
## tuition  -0.001989497 -0.02073031 -0.26642256  0.17180732  0.07223171
## ed       -0.002228095 -0.10395177 -0.01744247  0.47600412  0.29297352
## incomehi -0.048863884 -0.12018661 -0.09947671  0.17593158  0.36045622
##               momcoll      ownhome        urban        cue80     stwmfg80
## female   -0.017521065 -0.046632926 -0.004875662  0.027641792 -0.031614444
## black    -0.036401953 -0.140455316  0.189287507 -0.065542381 -0.162030343
## hispanic -0.070947568 -0.077046850  0.153277074  0.090651394 -0.033764900
## bytest    0.197487989  0.128749318 -0.103735173 -0.018409562  0.112650978
## dadcoll   0.445403013  0.077766543 -0.071957922 -0.088308947  0.024331835
## momcoll   1.000000000  0.066421483 -0.044360318 -0.079993210 -0.001643723
## ownhome   0.066421483  1.000000000 -0.125392632  0.015110814  0.078427486
## urban    -0.044360318 -0.125392632  1.000000000 -0.036417213 -0.039086366
## cue80    -0.079993210  0.015110814 -0.036417213  1.000000000  0.311321564
## stwmfg80 -0.001643723  0.078427486 -0.039086366  0.311321564  1.000000000
## dist     -0.079642396  0.050378121 -0.295551459  0.249481380 -0.008638147
## tuition   0.041328936 -0.001027916 -0.048455948  0.199838909  0.289825469
## ed        0.233311833  0.092818819 -0.018914574 -0.009724351  0.021936505
## incomehi  0.253276558  0.135453753 -0.084352593 -0.065536951  0.084828250
##                  dist      tuition           ed    incomehi
## female   -0.003368955 -0.001989497 -0.002228095 -0.04886388
## black    -0.099893318 -0.020730311 -0.103951768 -0.12018661
## hispanic  0.021140913 -0.266422562 -0.017442465 -0.09947671
## bytest   -0.061539318  0.171807322  0.476004125  0.17593158
## dadcoll  -0.113257661  0.072231707  0.292973520  0.36045622
## momcoll  -0.079642396  0.041328936  0.233311833  0.25327656
## ownhome   0.050378121 -0.001027916  0.092818819  0.13545375
## urban    -0.295551459 -0.048455948 -0.018914574 -0.08435259
## cue80     0.249481380  0.199838909 -0.009724351 -0.06553695
## stwmfg80 -0.008638147  0.289825469  0.021936505  0.08482825
## dist      1.000000000 -0.191838458 -0.086310913 -0.08453223
## tuition  -0.191838458  1.000000000  0.057107818  0.10515168
## ed       -0.086310913  0.057107818  1.000000000  0.21673400
## incomehi -0.084532226  0.105151685  0.216734001  1.00000000
```

By looking at the dist column, we can see it has non-zero correlations with these variables:
dadcoll (-0.11): 1 = Father is a College Graduate/ 0 = Father is not a College Graduate,
urban (-0.29): 1 = School in Urban Area / = School not in Urban Area,
cue80 (0.25): County Unemployment rate in 1980,
tuition (-0.19): Avg. State 4yr College Tuition in $1000's.

So, one of the conditions hold true for these variabes so far. Now let's look at the regression coefficient of these parameters:

```
# 2. Checking the regression parameter:
# 2.a. dadcol:
ed.vs.dadcol <- lm(ed ~ dadcoll, data=col.dist)
ed.vs.dadcol$coefficients
```

```
## (Intercept)     dadcoll
##   13.561902    1.323366
```

```
# 2.b. urban:
ed.vs.urban <- lm(ed ~ urban, data=col.dist)
ed.vs.urban$coefficients
```

```
## (Intercept)       urban
##   13.848780   -0.079882
```

```r
# 2.c. cue80:
ed.vs.cue80 <- lm(ed ~ cue80, data=col.dist)
ed.vs.cue80$coefficients
```

```
##  (Intercept)        cue80
## 13.876412025 -0.006155298
```

```r
# 2.d. tuition:
ed.vs.tuition <- lm(ed ~ tuition, data=col.dist)
ed.vs.tuition$coefficients
```

```
## (Intercept)     tuition
##  13.4957215   0.3653029
```

As can be seen above, the slope for 2 of these variables (urban & cue80) is zero. However, for dadcoll and tuition, the slope is not zero and therefore at least these two variables have the criteria described above and must be included in the model.

1.3. Consider the various control variables in the dataset. Which do you think should be included in the regression? (Explain)

ANS.
We can run a linear regression with all of these parameters in the model and then decide to include which one of the parameters. However, as discussed in the lecture notes, we should think carefully how independent parameters can affect the dependent parameters and whether they are meaningful or not. Let's first include all of the parameters in the model:

```r
ed.vs.all <- lm(ed ~ female + black + hispanic + bytest + dadcoll + momcoll +
                  ownhome + urban + cue80 + stwmfg80 + dist + tuition + incomehi,
                data=col.dist)

stargazer(ed.vs.all, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment (Years) vs. All Parameters",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 2: Educational Attainment (Years) vs. All Parameters

	Dependent variable:
	Educational Attainment (Years)
female	0.144^{***}
	(0.050)
black	0.338^{***}
	(0.072)
hispanic	0.349^{***}
	(0.078)
bytest	0.093^{***}
	(0.003)
dadcoll	0.574^{***}
	(0.074)
momcoll	0.379^{***}
	(0.082)
ownhome	0.143^{**}
	(0.067)
urban	0.065
	(0.064)
cue80	0.028^{***}
	(0.010)
stwmfg80	-0.043^{**}
	(0.020)
dist	-0.033^{**}
	(0.013)
tuition	-0.185^{*}
	(0.101)
incomehi	0.374^{***}
	(0.061)
Constant	8.894^{***}
	(0.253)
Observations	3,796
R^2	0.284
Adjusted R^2	0.281
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

I think the most important parameter is incomehi which is the family income. This parameter includes some other parameters, i.e. if a family have higher income, there is a higher chance for them to own a home (ownhome), or if they have a higher income, they can afford higher tuition fees (tuition). The effect of dad's education looks stronger than mom's education. Also, the ethnicity information like hispanic and black variables have high p-value and small errors. The last variable is the gender information which seems to play an important role too.

1.4. Estimate a set of multivariate regression models. Add one additional variable at a time – that is, each new regression should have at most one additional variable than the previous one –. Display all regressions (including (1.1)) in a single table using stargazer. At this stage only consider OVB when deciding to add more variables. Compare the results of the multivariate regressions with the bivariate model.

ANS.

```
# I am going based on correlations between all of the predictors and education:
cor(col.dist[,-13], col.dist$ed) # column 13 is ed which is removed
```

```
             [,1]
```

female -0.002228095 black -0.103951768 hispanic -0.017442465 bytest 0.476004125 dadcoll 0.292973520 momcoll 0.233311833 ownhome 0.092818819 urban -0.018914574 cue80 -0.009724351 stwmfg80 0.021936505 dist -0.086310913 tuition 0.057107818 incomehi 0.216734001

```
# Now having dist in the model we start adding the highest cor predictors:
# 1. bytest:
col.m1 <- lm(ed ~ dist + bytest, data=col.dist)
stargazer(col.m1, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment Model",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 3: Educational Attainment Model

	Dependent variable:
	Educational Attainment (Years)
dist	−0.049***
	(0.012)
bytest	0.097***
	(0.003)
Constant	8.957***
	(0.155)
Observations	3,796
R^2	0.230
Adjusted R^2	0.229
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary: All of the variables are significant, *adjusted* $R^2 = 0.229$. Let's add more parameters:

```
# 2. dadcoll:
col.m2 <- lm(ed ~ dist + bytest + dadcoll, data=col.dist)
stargazer(col.m2, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment Model",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 4: Educational Attainment Model

	Dependent variable:
	Educational Attainment (Years)
dist	-0.034^{***}
	(0.012)
bytest	0.088^{***}
	(0.003)
dadcoll	0.809^{***}
	(0.066)
Constant	9.237^{***}
	(0.153)
Observations	3,796
R^2	0.259
Adjusted R^2	0.259
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary: All of the variables are significant, *adjusted* $R^2 = 0.259$. We'll keep these, let's add more parameters:

```r
# 3. dadcoll:
col.m3 <- lm(ed ~ dist + bytest + dadcoll + momcoll, data=col.dist)
stargazer(col.m3, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment Model",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 5: Educational Attainment Model

	Dependent variable:
	Educational Attainment (Years)
dist	-0.032^{***}
	(0.012)
bytest	0.086^{***}
	(0.003)
dadcoll	0.651^{***}
	(0.072)
momcoll	0.435^{***}
	(0.082)
Constant	9.282^{***}
	(0.153)
Observations	3,796
R^2	0.265
Adjusted R^2	0.264
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary: All of the variables are significant, *adjusted* $R^2 = 0.264$. We'll keep these, let's add more parameters:

```
# 4. incomehi:
col.m4 <- lm(ed ~ dist + bytest + dadcoll + momcoll + incomehi, data=col.dist)
stargazer(col.m4, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment Model",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 6: Educational Attainment Model

	Dependent variable:
	Educational Attainment (Years)
dist	−0.029**
	(0.012)
bytest	0.085***
	(0.003)
dadcoll	0.549***
	(0.074)
momcoll	0.392***
	(0.082)
incomehi	0.314***
	(0.060)
Constant	9.279***
	(0.153)
Observations	3,796
R^2	0.270
Adjusted R^2	0.269
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary: All of the variables are significant, *adjusted* $R^2 = 0.269$. dist is starting to get less significant but the adjusted R-squared is still increasing. We'll keep these, let's add more parameters:

```
# 5. black:
col.m5 <- lm(ed ~ dist + bytest + dadcoll + momcoll + incomehi + black, data=col.dist)
stargazer(col.m5, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment Model",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 7: Educational Attainment Model

	Dependent variable:
	Educational Attainment (Years)
dist	-0.023^{*}
	(0.012)
bytest	0.089^{***}
	(0.003)
dadcoll	0.561^{***}
	(0.074)
momcoll	0.376^{***}
	(0.082)
incomehi	0.330^{***}
	(0.060)
black	0.255^{***}
	(0.068)
Constant	9.035^{***}
	(0.166)
Observations	3,796
R^2	0.273
Adjusted R^2	0.272
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary: All of the variables are significant, *adjusted* $R^2 = 0.272$. At this point dist is no longer significant. We might want to remove it from the model. In the next step, let's add cue80:

```r
# 6. cue80:
col.m6 <- lm(ed ~ dist + bytest + dadcoll + momcoll + incomehi + black + cue80, data=col.dist)
stargazer(col.m6, align=TRUE, no.space=TRUE, header=FALSE,
          title="Educational Attainment Model",
          dep.var.labels=c("Educational Attainment (Years)"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 8: Educational Attainment Model

	Dependent variable:
	Educational Attainment (Years)
dist	-0.030^{**}
	(0.012)
bytest	0.089^{***}
	(0.003)
dadcoll	0.567^{***}
	(0.074)
momcoll	0.383^{***}
	(0.082)
incomehi	0.334^{***}
	(0.060)
black	0.263^{***}
	(0.068)
cue80	0.021^{**}
	(0.009)
Constant	8.882^{***}
	(0.178)
Observations	3,796
R^2	0.274
Adjusted R^2	0.273
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary: All of the variables are significant, $adjusted\ R^2 = 0.273$. The in improvement in adjusted R-squared is very minor. We might want to stop at this poin. dist variable is significant again but both dist and cue80 are not very impressive variables.

1.5. Check if the estimated model sufferes from multicollinearity.

ANS.
(NOTE!! I first did problem 2, therefore I explained multi-collinearity in more detail there.)
I am checking multi-collinearity w/ vif function from car package. The smallest possible value of VIF is 1 (absence of multicollinearity). As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity:

```
vif(lm(ed ~ dist + bytest + dadcoll + momcoll + incomehi + black + cue80, data=col.dist))
```

```
##     dist   bytest  dadcoll  momcoll incomehi    black    cue80
## 1.094049 1.196728 1.405117 1.278342 1.180795 1.141242 1.075611
```

As can be seen, there is no multi-collinearity in our model.

PROBLEM 2: Blood Pessure

For this problem we are going to explore the effect of weight on blood pressure. The dependent variable (Y) is blood pressure bp. Run any regression using the lm command and display regression tables using the stargazer command.

```
# Reading the data:
blood.press <- read.csv(
  paste('C:/Users/alibs/Google Drive/Courses/INSH5301 Intro Computational Statistics/',
  'Module 9 - Multiple Regression 2/bloodPressure/bloodpress.csv', sep=''),
  stringsAsFactors = F)
# paste is for breaking the path into multiple lines
```

2.1. Test the hypothesis that weight is related with blood presure using a bivariate model.

ANS.

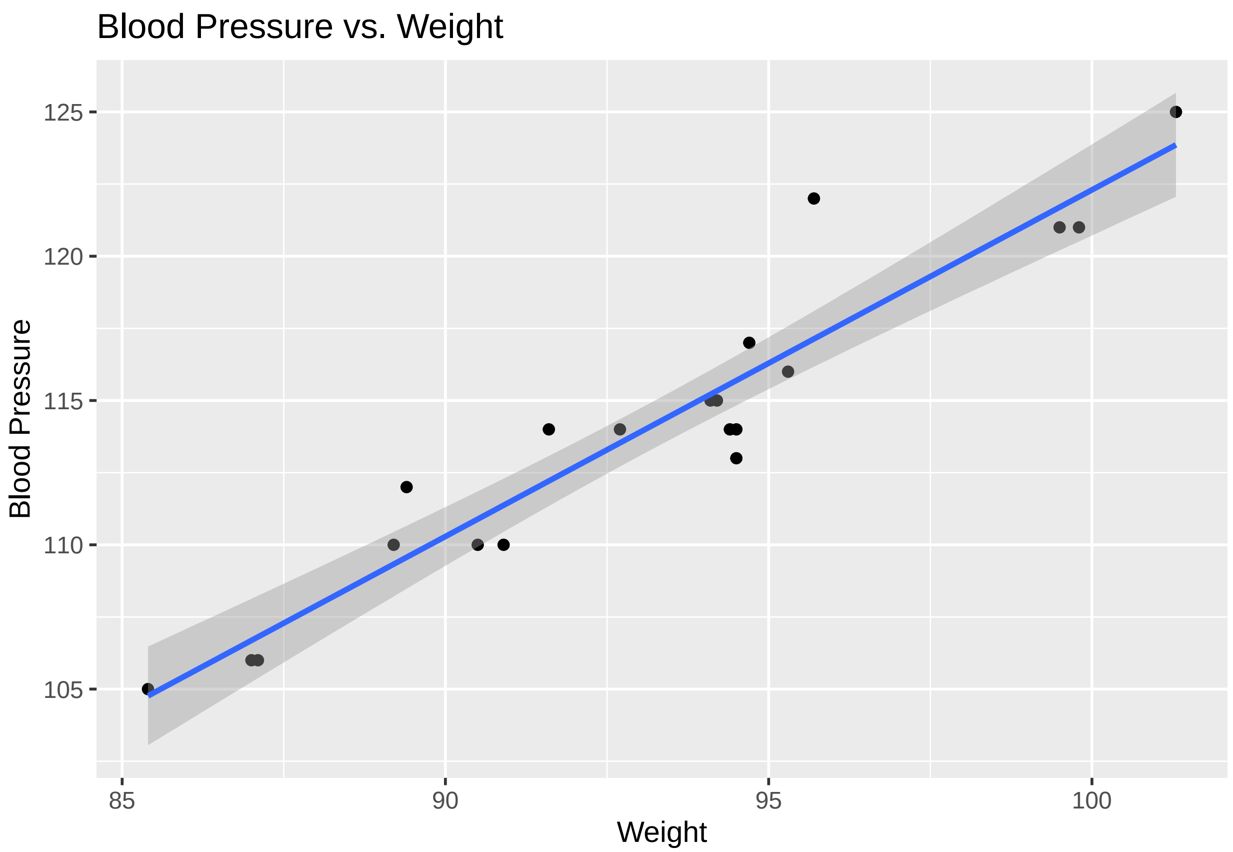
```
bp.vs.weight <- lm(BP ~ Weight, data=blood.press)

stargazer(bp.vs.weight, align=TRUE, no.space=TRUE, header=FALSE,
          title="Blood Pressure vs. Weight",
          dep.var.labels=c("Blood Pressure"),
          covariate.labels=c("Weight"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 9: Blood Pressure vs. Weight

	Dependent variable:
	Blood Pressure
Weight	1.201***
	(0.093)
Constant	2.205
	(8.663)
Observations	20
R^2	0.903
Adjusted R^2	0.897
Note:	*p<0.1; **p<0.05; ***p<0.01

The results look promising. The variables are positively related, i.e. As weight increases, the blood pressure goes up. R^2 is high (0.90) and the *adjusted* R^2 is also high (0.897). Which means about %90 of the variations in blood pressure, is explained by weight. Let's also plot these two variables against each other to see how they are related:

```
ggplot(data=blood.press, aes(x=Weight, y=BP))+
  geom_point()+
  geom_smooth(method=lm)+
  xlab('Weight')+ylab('Blood Pressure')+ggtitle('Blood Pressure vs. Weight')
```



The plot looks alright too and agrees with the model.

2.2. Use the other variables in the model to correct for OVB.

ANS.
Let's do what we did in the previous problem. First let's look at the correlations between weight and other variables:

```
cor(blood.press)
```

```
##                 Pt         BP        Age     Weight         BSA       Dur
## Pt      1.00000000 0.03113499 0.04269354 0.02485650 -0.03128800 0.1762455
## BP      0.03113499 1.00000000 0.65909298 0.95006765  0.86587887 0.2928336
## Age     0.04269354 0.65909298 1.00000000 0.40734926  0.37845460 0.3437921
## Weight  0.02485650 0.95006765 0.40734926 1.00000000  0.87530481 0.2006496
## BSA    -0.03128800 0.86587887 0.37845460 0.87530481  1.00000000 0.1305400
## Dur     0.17624551 0.29283363 0.34379206 0.20064959  0.13054001 1.0000000
## Pulse   0.11228508 0.72141316 0.61876426 0.65933987  0.46481881 0.4015144
## Stress  0.34315169 0.16390139 0.36822369 0.03435475  0.01844634 0.3116398
##            Pulse     Stress
## Pt     0.1122851 0.34315169
## BP     0.7214132 0.16390139
## Age    0.6187643 0.36822369
## Weight 0.6593399 0.03435475
## BSA    0.4648188 0.01844634
## Dur    0.4015144 0.31163982
## Pulse  1.0000000 0.50631008
## Stress 0.5063101 1.00000000
```

As can be seen above, weight has high correlations with Age(0.41), BSA (Body Surface Area)(0.87), Pulse (0.66) and some correlation with Dur (Duration of hypertension)(0.21). Let's examine these parameters:

```
# Age:
bp.w_age <- lm(BP ~ Weight + Age, data=blood.press)
bp.w_age$coefficients
```

```
## (Intercept)      Weight         Age
## -16.5793694   1.0329611   0.7082515
```

```
# BSA:
bp.w_bsa <- lm(BP ~ Weight + BSA, data=blood.press)
bp.w_bsa$coefficients
```

```
## (Intercept)      Weight         BSA
##    5.653398    1.038734    5.831250
```

```
# Pulse:
bp.w_pulse <- lm(BP ~ Weight + Pulse, data=blood.press)
bp.w_pulse$coefficients
```

```
## (Intercept)      Weight       Pulse
##  -1.4534677   1.0608694   0.2399014
```

```
# Duration of Hypertension:
bp.w_dur <- lm(BP ~ Weight + Dur, data=blood.press)
bp.w_dur$coefficients
```

```
## (Intercept)      Weight         Dur
##   2.9867668   1.1739221   0.2694908
```

2.3. Test for multicollinearity.

ANS.
We've already ran the test for multi-collinearity. It is simply to look at the correlations between independent variables. Multi-collinearity happens when there is a high correlation between two or more variables, meaning that some independent variables can predict the others. Calculating the correlation coeff between all of the variables shows this:

```
all.cors <- data.frame(cor(blood.press))

# I consider the correlation above 0.8 high, to make it easier to find the correlations
# above 0.8, I set all the cors<0.8 to zero:
all.cors[all.cors<0.8] <- 0
all.cors
```

```
##        Pt        BP Age    Weight       BSA Dur Pulse Stress
## Pt      1 0.0000000   0 0.0000000 0.0000000   0     0      0
## BP      0 1.0000000   0 0.9500677 0.8658789   0     0      0
## Age     0 0.0000000   1 0.0000000 0.0000000   0     0      0
## Weight  0 0.9500677   0 1.0000000 0.8753048   0     0      0
## BSA     0 0.8658789   0 0.8753048 1.0000000   0     0      0
## Dur     0 0.0000000   0 0.0000000 0.0000000   1     0      0
## Pulse   0 0.0000000   0 0.0000000 0.0000000   0     1      0
## Stress  0 0.0000000   0 0.0000000 0.0000000   0     0      1
```

Now we can quickly recognize the multi-collinearity between weight and BSA. The relationship between these two is obvious! If one has a higher weight, they must have a higher body surface area. Therefore using these two variable as predictor is redundat.

We could also use VIF (Variance Inflation Factors) function. The smallest possible value of VIF is 1 (absence of multicollinearity). As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity.

```r
vif(lm(BP ~ Pt + Weight + Age + BSA + Dur + Pulse + Stress, data=blood.press))
```

```
##       Pt   Weight      Age      BSA      Dur    Pulse   Stress 
## 1.205226 8.741853 1.766424 5.463757 1.253932 4.546765 2.116450
```

Using this method again we can see that Weight/BSA are problematic. Pulse is also on the border.

2.4. Choose a final functional form and interpret the results.

ANS.
Based on all of the discussions in the previous parts, I am going with the following model. I included stress and duration in the model initially but they have extremely bad p-values (0.573304 and 0.545485 respectively). The best model is with the following variables. It has an *adjusted* R^2 of 0.9904 which is pretty good.

```r
final.bp.model <- lm(BP ~ Weight + Age, data=blood.press)

stargazer(final.bp.model, align=TRUE, no.space=TRUE, header=FALSE,
          title="Blood Pressure",
          dep.var.labels=c("Blood Pressure"),
          covariate.labels=c("Weight", "Age"),
          omit.stat=c("LL","ser","f"),
          table.placement = "H" # Hold the position of the table
         )
```

Table 10: Blood Pressure

	Dependent variable:
	Blood Pressure
Weight	1.033***
	(0.031)
Age	0.708***
	(0.054)
Constant	−16.579***
	(3.007)
Observations	20
R^2	0.991
Adjusted R^2	0.990
Note:	*p<0.1; **p<0.05; ***p<0.01