**BUAN 6356.002**

Problem Set 2

Huyen Nguyen (Htn180001)

**Question 1:**

1. Holding other factors fixed

ΔvoteA = β1Δlog(ExpendA)

= (β1 /100) [100 \*Δlog(ExpendA)]

≈ (β1 /100) [% Δlog(ExpendA)]

When campaign expenditure by Candidate A increases/ decreases by 1%, percentage of vote received will increase/ decrease by β1/100

2. H0: β2 = - β1

H1: β2 ≠ - β1

3. voteA = 45.08788 + 6.08136\*ln(expendA) - 6.61563\*ln(expendB) + 0.15201\*prystrA

R-sq = 0.7925, df = 169

The coefficient of ln(expendA) is very significant (t value = 15.915) and 1% increase in candidate A expense increases the percentage of vote received by 0.0608136

The coefficient of ln(expendB) is -17.461 and 1% increase in spending of candidate B reduces voteA by 0.0661563

The signs coefficient on ln(expendA) and ln(expendB) are opposite, which aligns with the hypothesis in part 2. Therefore, we can use these results to test the hypothesis in part 2.

4. We can write the hypothesis in part 2 as:

θ1 = β1 + β2

or β1 = θ1 - β2

-> voteA = β0 + θ1 ln(expendA) + β2 [ln(expendB) - ln(expendA)] + β3 prystrA

Estimate this model, we have θ1-hat ≈ -0.53427, se(θ1-hat) ≈ 0.53311

5. t-statistic on log(expendA) = -1.002

Fail to reject null hypothesis: H0: β2 = - β1

-> 1% increase in A’s expenditures is not offset by a 1% increase in B’s expenditures.

**Question 2:**

1. H0: β5 = 0

Estimate this model, we have:

ln(salary) = 8.3432262 + 0.0046965\*LSAT + 0.2475239\*GPA + 0.0949932\*log(libvol) + 0.0375539\*log(cost) - 0.0033246\*rank

β5-hat = -0.0033246

t value = -9.541

-> Reject null hypothesis. The rank of law schools has effect on median starting salary.

If rank decreases by 10 (which is an improvement of the school), median starting salary is predicted to increase by about 3.3%.

2. LSAT t value = 1.171 -> not significant

GPA t value = 2.749 -> signicant

R-square unrestricted regression: 0.8417

R-square restricted regression: 0.8174 (with missing values from GPA and LSAT dropped)

F-statistic = [(0.8417-0.8174)/2] / [(1-0.8417)/130] ≈ 9.95 (with 2 and 130 df)

LSAT and GPA are jointly very significant

3. Add clsize and faculty to the regression

R-square unrestricted regression: 0.844

R-square restricted regression: 0.8416 (with missing values from GPA and LSAT dropped)

F-statistic = [(0.844-0.8416)/2]/[(1-0.8416)/12 ≈ 0.95 (with 2 and 123 df)

-> These two variables are not jointly significant, no need to be added to this equation jointly.

4. Other factors might influence the rank of the law school: acceptance rate, replacement rate, faculty quality (measured by publications), student body (gender, race composition),...

**Question 3:**

1. ln(price) = 4.766 + 0.0003794\*sqrft + 0.02888\*bdrms

θ1-hat = 150\*0.0003794 + 0.02888 = 0.08579

-> An additional 150 square foot bedroom increases the predicted price by about 8.6%

2. β2 = θ1 - 150β1

ln(price) = β0 + β1(sqrft -150bdrms) + θ1bdrms + u

3. se(θ1 -hat) = 2.677e-02

Confident interval is 0.0325803714 to 0.1390223618

**Question 4:**

1. Ho: β2 = β3

2. Set θ2 = β2 - β3

-> ln(wage) = β0 + β1\*educ + θ2\*exper + β3\*(exper + tenure) + u

exper confident interval: -0.007355358 to 0.01126271

-> θ2 is not statistically different from zero at the 5% level, we fail to reject Ho: β2 = β3 at the 5% level.

**Question 5:**

1. There are 2,017 single-person households in the sample

2. nettfa = -43.03981 + 0.79932\*inc + 0.84266\*age

R-sq = 0.1193, df = 2017

Holding age fixed, $1 increase in income leads to about $0.8 increase in predicted nettfa

Holding income fixed, if a person gets one year older, his/her nettfa is expected to increase by around $843

This is not surprising because usually, income and age affect net total financial assets.

3. The intercept in part 2 doesn't have interesting meaning as it gives the predicted nettfa for inc = 0 and age = 0. However, there is no one with these values in the relevant population.

4. t statistic = (0.84266 - 1)/ 0.09202 = -1.709846

Against the one-sided alternative H1: β2 < 1, the p-value is about 0.044

We cannot reject Ho: β2 = 1 at the 1% significance level

5. nettfa = -10.5709 + 0.8207\*inc

The obtained slope coefficient on inc is not very different from the 0.79932 obtained in part 2. Correlation between inc and age in the sample of single people is 0.03905864, which helps explain why the simple and multiple regression estimates are not very different

**Question 6:**

1. As the presence of the incinerator depresses housing prices, we expect β1 >= 0 (all other relevant factors equal)

It is better to have a home farther away from the incinerator

ln(price) = 8.04716 + 0.36488 ln(dist)

R-sq = 0.1803, df = 140

-> 1% increase in distance from the incinerator is associated with a predicted price that is about 0.36488% higher

2. Add the variables ln[intst], ln[area], ln[land], rooms, bath, and age, we have coefficient on log(dist) = 0.055389 (se = 0.057621)

-> The distance effect is much smaller now and statistically insignificant

1 and 2 give conflicting result because some other factors that determine the house's quality as well as its location have been controlled. This is consistent with the hypothesis that the incinerator was located near less desirable homes

3. Add ln[intst]^2

log(price) = -3.318025 + 0.185256 log(dist) + 2.072959 log(intst) + 0.359352 log(area) + 0.091386 log(land) + 0.038106 rooms + 0.149533 baths - 0.002927 age - 0.119329 log(intst)^2

R-sq = 0.7775, df = 142

The coefficient on log(dist), log(inst) and [log(inst)]^2 are now very statistically significant, with t statistic log(dist) = 2.971

t statistic of log(inst) = 4.138

t statistic of [log(inst)]^2 = -4.236

Adding [log(inst)]2 has a very considerable effect. This means that distance from the incinerator and distance from the interstate are correlated in some nonlinear way that also affects housing price.

4. Coefficient on [log(dist)]^2 when added to the model estimated in part 3 is -0.036418 with t statistic = -0.331

-> It is not necessary to add this complication.

**Question 7:**

1. ln(wage) = 0.1263226 + 0.0906207 educ + 0.0409731 exper - 0.0007121 (exper^2)

R-sq = 0.3003, df = 522

2. t statistic on exper^2 = -6.141 -> p-value ≈ 0

-> exper is significant at the 1% level

3. %Δwage (5th year) = 100\*(0.0409731 - 2\*0.0007121\*4) = 3.52763%

%Δwage (20th year) = 100\*(0.0409731 - 2\*0.0007121\*19) = 1.39133%

4. Value of exper that additional experience actually lower predicted ln[wage]

0.0409731/(2\*0.0007121) = 28.7692 years

There are 121 people with at least 29 years of experiencein the sample.

**Question 8:**

1. Holding experience fixed, we have

Δlog(wage) = β1Δeduc + β3 Δeduc exper = Δeduc (β1 + β3 exper)

->Δlog(wage)/ Δeduc = β1 + β3 exper

2. Ho: β3 = 0

The appropriate alternative is β3 > 0 as normally we see that people with more experience work more effectively given another year of education.

3. log(wage) = 5.949455 + 0.044050 educ - 0.021496 exper + 0.003203 educ exper

R-sq = 0.1349, df = 931

t statistic on educ\*exper = 2.095

With p-value < 0.02 against H1: > 0. Therefore, we reject H0: = 0 against H1: > 0 at the 2% level.

4. β1 = θ1 - 10β3

log(wage) = β0 + θ1educ + β2exper + β3educ(exper-10) + u

θ1-hat = 0.076080, se(θ1-hat) = 0.006615

The 95% confident interval for θ1 is 0.0630973587 to 0.089061718

**Question 9:**

1. sat = 997.981 + 19.814 hsize - 2.131 hsize^2

R-sq = 0.00765, df = 4134

The quadratic term is very statistically significant, with t statistic = -3.881

2. “Optimal” high school size: This is the turning point in the parabola, where sat-hat reaches its maximum

19.81/(2\*2.13) = 4.650235

-> 465 students is the “optimal” class size

However, the very small R-squared shows that class size explains only a tiny amount of the variation in SAT score.

3. This analysis is not representative of the academic performance of all high school seniors as only students who take the SAT exam are in the sample.

4. ln(sat) = 6.8960291 + 0.0196029 hsize - 0.0020872 hsize^2

R-sq = 0.007773, df = 4134

0.0196029/(2\*0.0020872) = 4.69598

-> Now 469 students is the “optimal” class size, which is not much different from what you obtained in part 2

**Question 10:**

1. ln(price) = -1.29704 + 0.16797 ln(lotsize) +0.70023 ln(sqrft) + 0.03696 bdrms

R-sq = 0.643, df = 84

2. With lotsize = 20,000, sqrft = 2,500, and bdrms = 4, we have

lprice = -1.29704 + 0.16797\*log(20000) + 0.70023\*log(2500) + 0.03696\*4 = 5.992921

-> predicted price = 400.583

3. price = -2.177e+01 + 2.068e-03 lotsize + 1.228e-01 sqrft + 1.385e+01 bdrms

R-sq = 0.6724, df = 84

R-sq of the model in part 3 is higher. Therefore, for predicting price, the model in part 3 is better.