EC 228 01, Fall 2008: Answers to Problem Set 1

September 24, 2008

Problem A.7.

(i) By exponentiation left and right parts of equation we get $Salary = e^{10.6+.027exper}$. Therefore, for $exper = 0$ we have $Salary_1 = e^{10.6} \approx 40134.84$, and for $exper = 5$ $Salary_2 = e^{10.6+.027*5} = e^{10.735} \approx 45935.80$

(ii) $\frac{Salary_2 - Salary_1}{Salary_1} \approx \ln Salary_2 - \ln Salary_1 = .027 * 5 = .135$

(iii) $\frac{Salary_2 - Salary_1}{Salary_1} = \frac{45935.80-40134.84}{40134.80} \approx .144 > .135$

Problem B.2.

(i) P$(X \leq 6) = P[(X-5)/2 \leq (6-5)/2] = P(Z \leq 0.5) \approx 0.692.$, where Z denotes a Normal(0,1) random variable. [We obtain $P(Z) \leq 0.5$ from Table G.1]

(ii) $P(X > 4) = P[(X-5)/2 > (4-5)/2] = P(Z > -0.5) = P(Z \leq 0.5) \approx 0.692$

(iii) $P(|X-5| > 1) = P(X-5 > 1) + P(X-5 < -1) = P(X > 6) + P(X < 4) \approx (1-0.692) + (1-0.692) = 0.616$, where we have used answers from parts (i) and (ii).

Problem B.5.

(i) As stated in the hint, if X is the number of jurors convinced of Simpson's innocence, then $X \sim$ Binomial(12,20). We want $P(X \geq 1) = 1 - P(X = 0) = 1 - (.8)^{12} \approx .931$.

(ii) Above, we computed $P(X = 0)$ as about .069. We need P(X=1), which we obtain from (B.14) with $n = 12, \theta = .2, and x = 1$: $P(X = 1) = 12 \cdot (.2)(.8)^{11} \approx .206$. Therefore, $P(X \geq 2) \approx 1 - (.069 + .206) = .725$, so there is almost a three in four chance that the jury had at least two members convinced of Simpson's innocence prior to the trial.

Problem C.1.

(i) This is just a special case of what we covered in the text, with $n = 4$: $E(\bar{Y}) = \mu$ and $Var(\bar{Y}) = \sigma^2/4$.

(ii)$E(W) = E(Y_1)/8 + E(Y_2)/8 + E(Y_3)/4 + E(Y_4)/2 = \mu[(1/8) + (1/8) + (1/4) + (1/2)] = \mu(1 + 1 + 2 + 4)/8 = \mu$, which shows that W is unbiased. Because the Y_i are independent,

$$\begin{aligned} Var(W) &= Var(Y_1)/64 + Var(Y_2)/64 + Var(Y_3)/16 + Var(Y_4)/4 = \\ &= \sigma^2((1/64) + (1/64) + (4/64) + (16/64)) = \sigma^2(22/64) = \sigma(11/32). \end{aligned} \tag{1}$$

(iii) Because $11/32 > 8/32 = 1/4, Var(W) > Var(\bar{Y})$ for any $\sigma^2 > 0$, so $\bar{Y}$ is preferred to W because each is unbiased.

Problem C.6.

(i) $H_0 : \mu = 0$
(ii)$H_1 : \mu < 0$
(iii)The standard error of $\bar{y}$ is $s/\sqrt{n} = 466.4/30 \approx 15.55$. Therefore, the t statistic for testing $H_0 : \mu = 0$ is $t = \bar{y}/se(\bar{y}) = -32.8/15.55 \approx -2.11$. We obtain the p-value as $P(Z \leq -2.11)$, where $Z \sim$ Normal(0,1). These probabilities are in Table G.1: p-value=.0174. Because the p-value is below .05, we reject H_0 against the one-sided alternative at the 5
(iv) The estimated reduction, about 33 ounces, does not seem large for an entire year's consumption. If the alcohol is beer, 33 ounces is less than three 12-ounces cans of beer. Even if this is hard liquor, the reduction seems small. (On the other hand, when aggregated across the entire population, alcohol distributors might not think the effect is so small.)
(v) The implicit assumption is that other factors that affect liquor consumption - such as income, or changes in price due to transportation costs, are constant over the two years.

1 C1.1

(i)

```
. summ educ
```

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	526	12.56274	2.769022	0	18

(ii)

```
. summ wage
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	526	5.896103	3.693086	.53	24.98

(iii) CPI for 2003 is 184.0 and for 1976 it is 56.9, respectively.

(iv) Therefore, mean salary in 2003 dollars is $5.896 \cdot \frac{184.0}{56.9} = 19.066$

(v)

```
. tab female
```

female	Freq.	Percent	Cum.
0	274	52.09	52.09
1	252	47.91	100.00
Total	526	100.00	

2 C1.2

(i)

```
. gen cigdummy=(cigs>0)
. tab cigdummy male, col
```

Key
frequency
column percentage

cigdummy	male 0	1	Total
0	275 79.94	299 85.43	574 82.71
1	69 20.06	51 14.57	120 17.29
Total	344 100.00	350 100.00	694 100.00

There are 344 women in the sample and approximately 20 percentages among them smoke.

(ii)

```
. bysort male: summ cigs
```

-> male = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
cigs	344	2.69186	6.59829	0	40

-> male = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
cigs	350	2.137143	6.268907	0	50

Mean number of cigarette is obviously not a very good measure. Clearly it is very unlikely that woman can smoke 2.69 (2 or 3) cigarettes.

(iii)

```
. bysort male cigdummy: summ cigs

-> male = 0, cigdummy = 1
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        cigs |        69    13.42029     8.57391          1         40
```

So, among smoking women mean is 13.42 and it approximately exceeds average number of cigarettes by five times! That happens that in the whole sample 80 percent smoke 0 cigarettes per day.

(iv)

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    fatheduc |       589    13.20204    2.650554          3         18
```

There are just 589 observations instead of 694 as not everyone answered question and there are missing values.

(v)

```
. summ faminc

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      faminc |       694    29.03602     18.5336         .5         65
```

3 C1.3

(i)

```
. summ math4

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       math4 |       729    72.88683    19.75993          0        100
```

This range tells us that in the best school all students fourth year passed math and in the worst school nobody was able to pass. Probably, both cases may occur for very small classes.

(ii)

```
. tab math4 if math4==100
 % students |
satisfactor |
     y, 4th |
 grade math |      Freq.     Percent        Cum.
------------+-----------------------------------
        100 |         19      100.00      100.00
------------+-----------------------------------
```

```
      Total |         19      100.00
```

```
. disp _N 729
. disp 19/729*100
2.60631
```

Only 19 out of 729 or 2.6 percent of the total sample had perfect pass rete.

(iii)

```
. tab math4 if math4==50

 % students |
satisfactor |
     y, 4th |
 grade math |      Freq.     Percent        Cum.
------------+-----------------------------------
         50 |          6      100.00      100.00
------------+-----------------------------------
      Total |          6      100.00
```

(iv)

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       math4 |       729    72.88683    19.75993          0        100
       read4 |       729    60.80905    19.15681        7.7        100
```

It can be seen from the table that pass rate for mathematics is higher, but no significantly (less than one standard deviation). Anyway, it is likely that reading test is harder to pass.

(v)

```
. correlate math4 read4 (obs=729)

             |    math4    read4
-------------+------------------
       math4 |   1.0000
       read4 |   0.8686   1.0000
```

High positive correlation means that school that performs better in mathematics usually (in 86.86 percentage) also performs better in reading.

(vi)

```
. summ expp

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       exppp |       729    5167.997     1057.09   1548.689   11957.64
```

```
. disp 1057.09/5167.997
.2045454
```

No. As standard deviation is about 20 percent of the mean expenditures, it means only 3 percent of leading schools (above two standard deviations) spend just more than 40 percent above than average level.

(vii) A's spending exceeds B's by $frac6000 - 55005500100 \approx 9.09$ percents. As $100 \cdot [log(6000) - log(5500)] \approx 8.70$. Approximation is slightly lower as its second order member in Taylor row is negative.

4 C1.4

(i)
```
. tab train

      =1 if |
assigned to |
        job |
   training |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        260       58.43       58.43
          1 |        185       41.57      100.00
------------+-----------------------------------
      Total |        445      100.00
```

(ii)
```
. bysort train: summ re78

------------------------------------------------------------------------------------------
-> train = 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        re78 |       260    4.554802    5.483837          0    39.4835

------------------------------------------------------------------------------------------
-> train = 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        re78 |       185    6.349145    7.867405          0    60.3079
. disp (6.349-4.555)/4.555*100
39.385291
```

So job training give 40 percent premium to real earnings. That difference is really big and comparable e.g. to premium for the higher education.

(iii) `. tab unem78 train, column`

```
     =1 if | =1 if assigned to job
 unem. all |       training
   of 1978 |         0          1 |     Total
-----------+----------------------+----------
         0 |       168        140 |       308
           |     64.62      75.68 |     69.21
-----------+----------------------+----------
         1 |        92         45 |       137
           |     35.38      24.32 |     30.79
-----------+----------------------+----------
     Total |       260        185 |       445
           |    100.00     100.00 |    100.00
```

It can be seen from the table that for those who didn't receive job training unemployment level is above 35 percent, while for trained workers it is below 25 percent, which is statistically and economically significant.

(iv) No, all differences may occur due to unobservable characteristics: e.g. men with higher abilities and productivity were trained and that sample selection determines observed differences in earnings and unemployment. Maybe it is better to compare real earnings after training with their earnings before, the same can be done for unemployment.