Adriana Scanteianu

Basic Demographic Methods 6090

```
options(digits=20)
```

Problem Set 1

1. Calculate the exact number of person-years lived in the population for each of the five calendar years (2005, 2006,…, 2009), with the assumption that the population changed linearly during each of the three intercensal periods (1/1/2005-1/1/2006; 1/1/2006-1/1/2008; 1/1/2008-1/1/2010).

For linear growth rate:

$$PY[0,t] = \frac{N(t) - N(0)}{2} * t$$

$PY[2005, 2006] = \frac{15641+16144}{2} = 15892.5$

```
(15641+16144)/2
```

```
## [1] 15892.5
```

$PY[2006, 2007] = \frac{16144+\frac{16144+16752}{2}}{2} = 16296$

```
(16144+((16144+16752)/2))/2
```

```
## [1] 16296
```

$PY[2007, 2008] = \frac{16448+16752}{2} = 16600$

```
(((16144+16752)/2)+16752)/2
```

```
## [1] 16600
```

$PY[2008, 2009] = \frac{16752+\frac{16752+17936}{2}}{2} = 17048$

```
(16752+((16752+17936)/2))/2
```

```
## [1] 17048
```

$PY[2009, 2010] = \frac{17344+17936}{2} = 17640$

```
(((16752+17936)/2)+17936)/2
```

```
## [1] 17640
```

2. Calculate the Crude Growth Rate (CGR) in this population for the period 1/1/2005-1/1/2010, using your person-years estimates from (1). Give at least 5 decimals in your result.

$$CGR[0,t] = \frac{N(t) - N(0)}{PY[0,t]}$$

$PY[2006, 2008] = \frac{16144+16752}{2} * 2 = 32896$

$PY[2008, 2010] = \frac{16752+17936}{2} * 2 = 34688$

$CGR[2005, 2010] = \frac{17936-15641}{15892.5+32896+34688} = 0.02749$

```
(17936-15641)/(15892.5 + 32896 + 34688)
```

```
## [1] 0.027492767425562881362
```

3. Calculate the mean annualized growth rate for each of the three intercensal periods (1/1/2005-1/1/2006; 1/1/2006-1/1/2008; 1/1/2008-1/1/2010). Also give at least 5 decimals in your results.

$$\bar{r}[0,t] = \frac{ln[\frac{N(t)}{N(0)}]}{t}$$

$\bar{r}[2005, 2006] = \frac{ln[\frac{16144}{15641}]}{1} = 0.03165$

```
log(16144/15641)
```

```
## [1] 0.031652791919148680344
```

$\bar{r}[2006, 2008] = \frac{ln[\frac{16752}{16144}]}{2} = 0.01848$

```
(log(16752/16144))/2
```

```
## [1] 0.018484595258463902556
```

$\bar{r}[2008, 2010] = \frac{ln[\frac{17936}{16752}]}{2} = 0.03415$

```
(log(17936/16752))/2
```

```
## [1] 0.034146106100811493556
```

4. In view of your results in (3), do you expect the mean annualized growth rate for the period 1/1/2005-1/1/2010 to be equal to the CGR for the same period? Why, or why not? You don’t need to make additional calculations to answer this question.

No, since the mean annualized growth rate $\bar{r}[2005, 2010]$ doesn't take into account the different rates of growth in each intercensal period.

Indeed, $\bar{r}[2005, 2010] = \frac{ln[\frac{17936}{15641}]}{2} = 0.06846$

```
(log(17936/15641))/2
```

```
## [1] 0.068457097318849774448
```

5. Calculate the Crude Birth Rate (CBR) for the following time periods, using the same assumption as in (1). Give at least 5 decimals in your results.
   1. 1/1/2006-1/1/2007
   2. 1/1/2006-1/1/2009
   3. 1/1/2005-1/1/2010

$$CBR = \frac{births[0,t]}{PY[0,t]}$$

$CBR[2006, 2007] = \frac{521}{\frac{16144+\frac{16144+16752}{2}}{2}} = 0.03197$

```
521/((16144+((16144+16752)/2))/2)
```

```
## [1] 0.031971035837015215897
```

$CBR[2006, 2009] = \frac{521+465+530}{32896+\frac{16752+\frac{17936+16752}{2}}{2}} = 0.03035$

```
(521+465+530)/(32896+((16752+((16752+17936)/2))/2))
```

```
## [1] 0.030353996476053180659
```

$CBR[2005, 2010] = \frac{556+521+465+530+512}{15892.5+32896+34688} = 0.03095$

```
(556+521+465+530+512)/(15892.5+32896+34688)
```

```
## [1] 0.030954819619893023802
```

B. The population size of the United States on April 1, 2010 was, according to the Census of that date, 308,745,538 persons. The census of April 1, 2020 recorded 331,449,281 persons.

1. Suppose that the instantaneous growth rate was constant over the interval.

$$r(t) = \frac{d(ln(N(t)))}{dt} \rightarrow N(t) = N(0)e^{\int_0^t r(t)dt} \rightarrow ln(\frac{N(t)}{N(0)}) = r^* * t$$

```
1.  What was its value?
```

$\frac{331449281}{308745538} = e^{\int_0^t r(t)dt}$

$ln(\frac{331449281}{308745538}) = (r(t) * t)|_0^t = r^* * t = 0.07096$

$r^* = 0.007095736$

```
(log(331449281/308745538))/10
```

```
## [1] 0.0070957363412853542012
```

```
2.  At this rate, how long would it take for the population to triple in size?
```

$N(t) = 3N(0)$

$3N(0) = N(0)e^{r^* * t}$

$t = \frac{ln3}{r^*} = \frac{1.098612}{0.007095736} = 154.82710$

```
log(3)/0.0070957363412853538542
```

```
## [1] 154.82710120949928978
```

```
3.  At this rate, what would be the population size on January 1, 2050?
```

$N(2050) = N(2020)e^{\int_{2020}^{2050} 0.007095736dt}$

$N(2050) = 331449281e^{0.007095736*29.75} = 409350990.5$

```
331449281*exp(0.0070957363412853538542*29.75)
```

```
## [1] 409350990.52976280451
```

```
4.  How many person-years were lived in the U.S. over the intercensal period?
```

$PY[0,t] = \frac{N(t)-N(0)}{ln(\frac{N(t)}{N(0)})} * t \; PY[2010, 2020] = \frac{331449281-308745538}{ln\frac{331449281}{308745538}} * 10 = 3199631709.5$

```
((331449281-308745538)/log(331449281/308745538))*10
```

```
## [1] 3199631709.5242214203
```

```
5.  What was the population size at the middle of the intercensal period?
```

$N(2015) = 308745538e^{0.007095736(5)} = 319896055.9$

```
308745538*exp(0.0070957363412853538542*5)
```

```
## [1] 319896055.90262937546
```

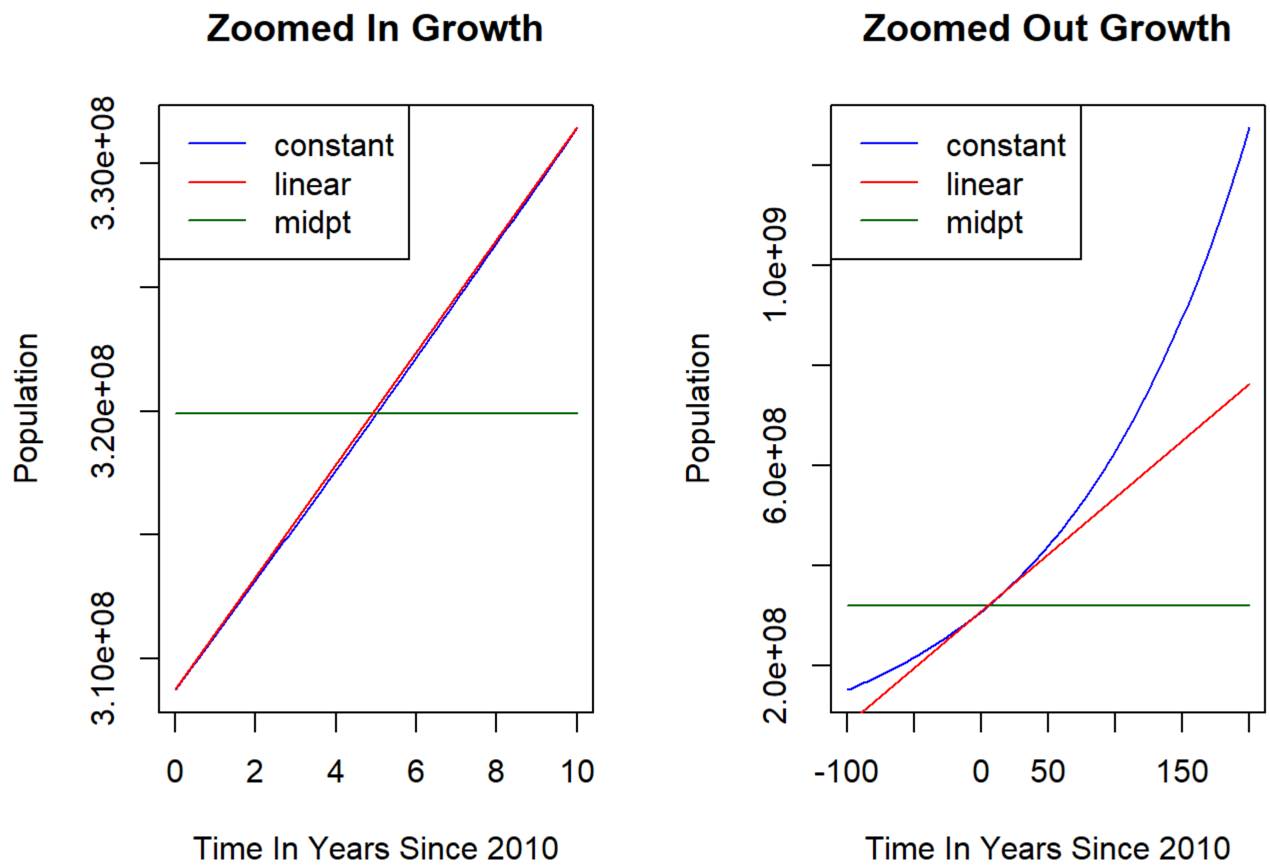
```
6.  Suppose the only population information you had was your answer to (1e). What would your est
imate of person-years lived over the period be? How would this estimate differ from the "true" p
erson-years lived?
```

Using only the information in part (1e), here written as 1.5., the estimate for person-years would be as follows.

$$PY[2010, 2020] = 319896055.35675 * 10 = 3198960553.5675$$

This is an underestimate, as seen by the following depiction of the difference between midpoint and constant estimates of growth rates. Note, a linear approximation of the person-years would lead to an overestimate.

```r
equation1=function(x){308745538*exp(0.0070957363412853538542*(x))}
equation2=function(x){319896055.90262937546}+0*x
equation3=function(x)(308745538+((331449281-308745538)/10)*x)

par(mfrow = c(1, 2))
curve(equation1, from=0, to=10, xlab="Time In Years Since 2010", ylab="Population",
      col="blue", lwd=1, main="Zoomed In Growth" )
curve(equation2, from=0, to=10, n=3000, add=TRUE, xlab="X", ylab="Y", col="darkgreen",lwd=1)

curve(equation3, from=0, to=10, n=3000, add=TRUE, xlab="X", ylab="Y", col="red",lwd=1)

legend(x="topleft", legend=c("constant", "linear", "midpt"), lty = c(1,1,1), col=c("blue","red",
"darkgreen"))

curve(equation1, from=-100, to=200, xlab="Time In Years Since 2010", ylab="Population",
      col="blue", lwd=1, main="Zoomed Out Growth" )

curve(equation2, from=-100, to=200, n=3000, add=TRUE, xlab="X", ylab="Y", col="darkgreen",lwd=1)

curve(equation3, from=-100, to=200, n=3000, add=TRUE, xlab="X", ylab="Y", col="red",lwd=1)

legend(x="topleft", legend=c("constant", "linear", "midpt"), lty = c(1,1,1), col=c("blue","red",
"darkgreen"))
```



Calculating these integrals, we see the midpoint method yields an underestimate, which according to Preston (15) will always be true when comparing midpoint person-years estimates to constant-rate estimates.

```
"constant"
```

```
## [1] "constant"
```

```
integrate(equation1, 0, 10)
```

```
## 3199631709.5242176056 with absolute error < 3.6e-05
```

```
"linear"
```

```
## [1] "linear"
```

```
integrate(equation3, 0, 10)
```

```
## 3200974095.0000004768 with absolute error < 3.6e-05
```

```
"midpoint"
```

```
## [1] "midpoint"
```

```
integrate(equation2, 0, 10)
```

```
## 3198960559.0262937546 with absolute error < 3.6e-05
```

2. Suppose that the instantaneous growth rate varied during the interval. What was its mean value (i.e., the mean annualized growth rate)?

$$\bar{r} = \frac{ln(\frac{N(t))}{N(0))})}{t} = 0.007095736$$

```
(log((331449281)/308745538))/10
```

```
## [1] 0.0070957363412853542012
```

3. Suppose that the two censuses recorded the same number of people but the second census was taken on February 1, 2010 instead of April 1, 2010.

    1. What would be your answer to (1a)?

$$N(t) = N(0)e^{(r^{*}*t)}|_{2009\frac{5}{6}}^{2020} \; r^{*} = \frac{ln(\frac{(331449281)}{308745538}}{10\frac{1}{6}} \; r^{*} = 0.00698$$

```
(log(331449281/308745538))/(10+1/6)
```

```
## [1] 0.0069794127947069061271
```

```
2.  What would be your answer to (1e)? Compare your answer with the answer to question (1e).
```

$$N(2015\frac{1}{12}) = 308745538 * e^{0.00698*5\frac{1}{6}} = 319896055.9$$

```
308745538*exp(0.0069794127947069061271*(5+1/12))
```

```
## [1] 319896055.90262937546
```

The estimate is the same because the population estimate at the middle of the intercensal period is independent of time regardless of when the first census was taken as long as the population at the start and end of the period we are investigating is the same (see below, t in exponent cancels out).

$$N(MidIntercensal) = N(Initial)e^{r^{*}*t}$$

$$N(MidIntercensal) = N(Initial)e^{\frac{ln(\frac{N(t)}{N(0)})}{t}*t}$$

4. Calculate the annual geometric growth rate, using the following compounding periods: annually,monthly, daily, continuously (give at least 7 decimals).

*skip

C. The mean annualized growth rate for the population of Ukraine is estimated to -0.91% for the year 2021. What is the halving time for Ukraine?

$\frac{ln(0.5)}{-0.0091} = 76.17002$ years

```
(log(0.5))/(-0.0091)
```

```
## [1] 76.170019841752221623
```

D. Suppose that in a country, urban and rural populations are growing at constant – but distinct – rates between two dates. Demonstrate that the rate of growth in the ratio of urban to rural populations (Urban Pop / Rural Pop) between these two dates is equal to the difference between urban and rural growth rates.

Consider:

$N_{u/r}(t) = \frac{N_u(t)}{N_r(t)} = \frac{N_u(0)*e^{\int_0^t r_u(t)dt}}{N_r(0)*e^{\int_0^t r_r(t)dt}}$, and assuming r(t) is constant $= \frac{N_u(0)*e^{r_u^* * t}}{N_r(0)*e^{r_r^* * t}} = N_{u/r}(0) * e^{r_{u/r}^* * t}$

Taking ln and rearranging:

$$r_{u/r}^* * t = ln(N_u(t)) - ln(N_r(t)) - ln(N_u(0)) - ln(N_r(0)) =$$

$$= ln(N_u(t)) - ln(N_u(0)) - ln(N_r(t)) - ln(N_r(0)) = ln(\frac{\frac{N_u(t)}{N_u(0)}}{\frac{N_r(t)}{N_r(0)}}) = ln(\frac{N_u(t)}{N_u(0)}) - ln(\frac{N_r(t)}{N_r(0)}) = t(r_u^* - r_r^*)$$

Dividing through by t, $r_u^* - r_r^* = r_{u/r}^*$.