ps6

Sijun Zhang 89934761

2019/12/8

In Problem Set 6, all confidence intervals are calculated at the level 0.05 and the lwr indicates the 2.5% lower bound and the upr indicates the 97.5% upper bound for each confidence interval.

Question 1

Firstly, the csv data file generated by ps2_q3.R is imported to the .sas script as the data table for manuplation. The confidence interval is calculated at level 0.05

```
/* data library for reading/writing data: ---------------------------------- */
libname mylib '/folders/myfolders/Problem Set 6/';

/* import delimited data with proc import: --------------------------------- */
proc import
 datafile='/folders/myfolders/Problem Set 6/mousetrap_data.csv'
 out=mylib.ps2q2d
 replace;
run;

/* transform the data into log form and set auc < 1 as 1 */
data mylib.ps6_q1;
 set mylib.ps2q2d;
 AUC=max(AUC, 1);
 tot_dist = log(tot_dist);
 max_abs_dev = log(max_abs_dev);
 avg_abs_dev = log(avg_abs_dev);
 AUC = log(AUC);
 by subject_nr count_trial;
run;

proc format library=mylib.ps6_q1;
 value Condition
  1="Typical"
  2="Atypical";
run;

/* Use Linear Mixed model fit tot_dist */
ods output
 SolutionF = tot_dist_re;

proc mixed data=mylib.ps6_q1;
 class Condition Exemplar;
 model tot_dist = Condition / cl;
 random intercept / subject=Exemplar;
 random intercept / subject=subject_nr;
run;

/* Use Linear Mixed model fit max_abs_dev */
ods output
 SolutionF = max_abs_dev_re;

proc mixed data=mylib.ps6_q1;
 class Condition Exemplar;
 model max_abs_dev = Condition / cl;
 random intercept / subject=Exemplar;
 random intercept / subject=subject_nr;
run;

/* Use Linear Mixed model fit avg_abs_dev */
ods output
 SolutionF = avg_abs_dev_re;

proc mixed data=mylib.ps6_q1;
 class Condition Exemplar;
 model avg_abs_dev = Condition / cl;
 random intercept / subject=Exemplar;
 random intercept / subject=subject_nr;
run;

/* Use Linear Mixed model fit AUC */
ods output
 SolutionF = AUC_re;

proc mixed data=mylib.ps6_q1;
 class Condition Exemplar;
 model AUC = Condition / cl;
 random intercept / subject=Exemplar;
 random intercept / subject=subject_nr;
run;

/* add label name */
data name;
```

```
 input label$;
 cards;
  tot_list
  max_dev
  avg_dev
  AUC;
run;

/* collect the result for output */
data est_result;
 set tot_dist_re max_abs_dev_re avg_abs_dev_re AUC_re;
 if (_n_=2)|(_n_=5)|(_n_=8)|(_n_=11);
 keep Estimate Lower Upper;
run;

data final_result;
 merge name est_result;
 Relative_effect_est = exp(Estimate);
 Relative_effect_lwr = exp(Lower);
 Relative_effect_upr = exp(Upper);
run;

/* Saving result into file PS6_q1.csv, and create the .lst file */
proc export data=final_result
 outfile='/folders/myfolders/Problem Set 6/ps6_q1.csv' replace;
run;

ods listing;
proc printto print='/folders/myfolders/Problem Set 6/ps6_q1.lst'
new;
run;

proc print data=final_result;
 title 'Final result of Question 1';
run;

proc printto;
run;
ods listing close;
```

```
ps6_q1 = read.csv("ps6_q1.csv")
knitr::kable(ps6_q1)
```

label	Estimate	Lower	Upper	Relative_effect_est	Relative_effect_lwr	Relative_effect_upr
tot_list	0.1613	0.08516	0.2375	1.175095	1.088895	1.268119
max_dev	0.5101	0.27850	0.7417	1.665423	1.321146	2.099415
avg_dev	0.6511	0.38500	0.9172	1.917738	1.469685	2.502387
AUC	0.3503	0.14160	0.5590	1.419492	1.152080	1.748973

From the above table, we find the coefficient of Condition on “the average absolute deviation of the observed trajectory from the direct path” is the maximum. Moreover, the expectation of average absolute deviation is also the smallest, which means the unit change in Condition variable is more influential in changing the distribution of average absolute deviation. Considering the two facts in average absolute deviation, we can conclude that the Condition has largest effect on the average absolute deviation. Moreover, the 95% confidence intervals are obtained in the above table.

Question 2

In this Part we revisit problem set 2, question 1 parts a, c-d. For this problem set.

Part a

```
/* a. First, use proc transpose to reshape the repliacte weights to a longer format.  */
/* Save these to disk as brrwt_long.sas7bdat using a two part filename. */
proc transpose data=mylib.recs
               out=mylib.brrwt_long;
               var BRRWT1-BRRWT96 ;
               by DOEID;
run;

data mylib.brrwt_long;
 set mylib.brrwt_long;
 repl = _NAME_;
 wt = COL1;
 keep DOEID repl wt;
run;
```

Part b

```
data temp;
 set mylib.recs;
 if (TEMPNITE ~= -2)&&(HEATHOME = 1);
 keep DOEID TEMPNITE NWEIGHT;
run;

/* need to eliminate the NA data repl */
data brrwt_long_b;
 set mylib.recs;
 if (TEMPNITE ~= -2)&&(HEATHOME = 1);
 keep DOEID BRRWT1-BRRWT96;
run;

proc transpose data=brrwt_long_b
               out=brrwt_long_b;
               var BRRWT1-BRRWT96 ;
               by DOEID;
run;

data brrwt_long_b;
 set brrwt_long_b;
 repl = _NAME_;
 wt = COL1;
 keep DOEID repl wt;
run;

/* Join home type to weights: ------------------------------------------------ */
data temp;
 merge temp brrwt_long_b;
 by DOEID;
 temp_est=TEMPNITE*NWEIGHT;
 temp_repl=TEMPNITE*wt;
run;

/* use the summary to calc the sum of variable, need to drop ovrall summary*/
proc summary data=temp;
 class repl;
 output out=temp_sum
 sum(temp_est) = sum_temp_est
 sum(wt) = sum_wt_repl
 sum(temp_repl) = sum_temp_repl
 sum(NWEIGHT) = sum_nwt;
run;

/* calc estimation and sse */
data temp;
 set temp_sum;
 if (_TYPE_ ~= 0);
 temp_est = sum_temp_est/sum_nwt;
 temp_repl = sum_temp_repl/sum_wt_repl;
 std_err = (4*(temp_est-temp_repl)**2);
run;

/* average the error */
proc summary data=temp;
 output out=temp
 mean(temp_est) = temp_est
 mean(std_err) = std_err;
run;

/* calc CI */
data temp;
 set temp;
 CI_lwr = temp_est + quantile('NORMAL',0.025)*std_err**0.5;
 CI_upr = temp_est + quantile('NORMAL',0.975)*std_err**0.5;
 if (_TYPE_ ~= .);
 drop _TYPE_ _FREQ_ std_err;
run;

/* output the result */
proc print data=temp;
 title 'national average home temperature at night';
run;

proc export data=temp
 outfile='/folders/myfolders/Problem Set 6/ps6_q2b.csv' replace;
run;
```

```
ps6_q2b = read.csv("ps6_q2b.csv")
knitr::kable(ps6_q2b)
```

temp_est	CI_lwr	CI_upr
68.10518	67.9265	68.28386

The above is the national average home temperature at night, among homes that use space heating.

Part c

```
data temp_c;
 set mylib.recs;
 if (TEMPNITE ~= -2)&&(TEMPHOME ~= -2)&&(TEMPGONE ~= -2);
 keep DOEID DIVISION TEMPNITE TEMPHOME TEMPGONE NWEIGHT;
run;

/* need to eliminate the NA data repl */
data brrwt_long_c;
 set mylib.recs;
 if (TEMPNITE ~= -2)&&(TEMPHOME ~= -2)&&(TEMPGONE ~= -2);
 keep DOEID DIVISION BRRWT1-BRRWT96;
run;

proc transpose data=brrwt_long_c
               out=brrwt_long_c;
               var BRRWT1-BRRWT96 ;
               by DOEID DIVISION;
run;

data brrwt_long_c;
 set brrwt_long_c;
 repl = _NAME_;
 wt = COL1;
 keep DOEID DIVISION repl wt;
run;

/* Join home type to weights: ------------------------------------------------ */
data temp_c;
 merge temp_c brrwt_long_c;
 by DOEID;
 night_temp_est=TEMPNITE*NWEIGHT;
 night_temp_repl=TEMPNITE*wt;
 home_temp_est=TEMPHOME*NWEIGHT;
 home_temp_repl=TEMPHOME*wt;
 gone_temp_est=TEMPGONE*NWEIGHT;
 gone_temp_repl=TEMPGONE*wt;
run;

/* use the summary to calc the sum of variable, need to drop ovrall summary*/
proc summary data=temp_c;
 class DIVISION repl;
 output out=temp_sum_c
 sum(night_temp_est) = sum_night_temp_est
 sum(wt) = sum_wt_repl
 sum(night_temp_repl) = sum_night_temp_repl
 sum(NWEIGHT) = sum_nwt
 sum(home_temp_est) = sum_home_temp_est
 sum(home_temp_repl) = sum_home_temp_repl
 sum(gone_temp_est) = sum_gone_temp_est
 sum(gone_temp_repl) = sum_gone_temp_repl;
run;

/* calc estimation and sse */
data temp_c;
 set temp_sum_c;
 if (_TYPE_ = 3);
 night_temp_est=sum_night_temp_est/sum_nwt;
 night_temp_repl=sum_night_temp_repl/sum_wt_repl;
 home_temp_est=sum_home_temp_est/sum_nwt;
 home_temp_repl=sum_home_temp_repl/sum_wt_repl;
 gone_temp_est=sum_gone_temp_est/sum_nwt;
 gone_temp_repl=sum_gone_temp_repl/sum_wt_repl;
 std_err_night = (4*(night_temp_est-night_temp_repl)**2);
 std_err_home = (4*(home_temp_est-home_temp_repl)**2);
 std_err_gone = (4*(gone_temp_est-gone_temp_repl)**2);
run;

/* average the error */
proc summary data=temp_c;
 class DIVISION;
 output out=temp_c
 mean(night_temp_est) = night_temp_est
 mean(std_err_night) = std_err_night
 mean(home_temp_est) = home_temp_est
 mean(std_err_home) = std_err_home
 mean(gone_temp_est) = gone_temp_est
 mean(std_err_gone) = std_err_gone;
run;

/* calc CI */
data temp_c;
 set temp_c;
 night_lwr = night_temp_est + quantile('NORMAL',0.025)*std_err_night**0.5;
 night_upr = night_temp_est + quantile('NORMAL',0.975)*std_err_night**0.5;
 home_lwr = home_temp_est + quantile('NORMAL',0.025)*std_err_home**0.5;
```

```
 home_upr = home_temp_est + quantile('NORMAL',0.975)*std_err_home**0.5;
 gone_lwr = gone_temp_est + quantile('NORMAL',0.025)*std_err_gone**0.5;
 gone_upr = gone_temp_est + quantile('NORMAL',0.975)*std_err_gone**0.5;
 if (_TYPE_ ~= 0);
 drop _TYPE_ _FREQ_;
run;

/* decoding the division code */
data names;
 input DIVISION names$;
 cards;
  1 New_England
  2 Middle_Atlantic
  3 East_North_Central
  4 West_North_Central
  5 South_Atlantic
  6 East_South_Central
  7 West_South_Central
  8 Mountain_North
  9 Mountain_South
  10 Pacific
;

data temp_c;
 merge names temp_c;
 by DIVISION;
 drop DIVISION std_err_home std_err_night std_err_gone;
 rename names=DIVISION;
run;

/* output the result */
proc print data=temp_c;
 title 'national average home temperature by census division';
run;

proc export data=temp_c
 outfile='/folders/myfolders/Problem Set 6/ps6_q2c.csv' replace;
run;
```

```
ps6_q2c = read.csv("ps6_q2c.csv")
knitr::kable(ps6_q2c)
```

DIVISION	night_temp_est	home_temp_est	gone_temp_est	night_lwr	night_upr	home_lwr	home_upr	gone_lwr	gone_upr
New_Engl	65.67196	68.18011	64.61019	64.88227	66.46164	67.30774	69.05247	64.01626	65.20412
Middle_A	67.30342	69.20760	66.48801	66.83123	67.77562	68.82265	69.59255	66.02259	66.95344
East_Nor	67.93268	69.70346	66.83603	67.55995	68.30540	69.35766	70.04927	66.54731	67.12475
West_Nor	68.06797	69.63445	67.17892	67.58524	68.55071	69.22405	70.04486	66.60541	67.75243
South_At	69.19868	70.64661	68.54698	68.77466	69.62270	70.35649	70.93672	68.15903	68.93492
East_Sou	69.30772	70.62840	68.13987	68.75929	69.85616	70.18734	71.06946	67.44907	68.83067
West_Sou	69.86110	71.29428	68.26726	69.50399	70.21821	70.88730	71.70127	67.42280	69.11172
Mountain	66.80776	69.24982	65.93051	65.72636	67.88916	68.59538	69.90426	65.35916	66.50185
Mountain	68.88158	70.82877	67.41744	66.74826	71.01490	69.89128	71.76626	65.78798	69.04691
Pacific	66.54376	69.75933	65.34039	66.04240	67.04512	69.32516	70.19351	64.98769	65.69310

The above is the average winter home temperatures at night, during the day with someone home, and during the day with no one home (when applicable), by census division.

Question 3

Repeat the second two parts (b and c) of question 2 above, using proc SQL.

Part b

```
libname mylib '/folders/myfolders/Problem Set 6/';

/* import delimited data with proc import: ---------------------------------- */
proc import
    datafile='/folders/myfolders/Problem Set 6/recs2015_public_v4.csv'
    out=mylib.recs
    replace;
run;

/* b. Estiamte the national average home temperature at night, among homes */
/* that use space heating. */

proc sql;
 create table temp as select DOEID,TEMPNITE,HEATHOME,NWEIGHT
 from mylib.recs;
 create table bt as select *
 from mylib.brrwt_long;
 /*  diff from in q2, we can first merge the whole set and then drop the uncessary point */
 create table temp_b as
 select temp.DOEID,temp.TEMPNITE,temp.HEATHOME,temp.NWEIGHT,bt.repl,bt.wt from bt left join
  (select * from temp)
  on temp.DOEID = bt.DOEID
 where HEATHOME = 1 and not TEMPNITE = -2;
/*   calc the est and stderr */
  create table temp_b_re as
  select mean(temp_est) as temp_est,
         (mean(4*(temp_est - temp_repl)**2))**0.5 as std_err
  from (select sum(TEMPNITE*NWEIGHT)/sum(NWEIGHT) as temp_est,
               sum(TEMPNITE*wt)/sum(wt) as temp_repl
        from temp_b
        group by repl
  );
/*   calc CI */
  create table temp_b_re as
    select temp_est, temp_est+quantile('normal',0.025)*std_err as CI_lower,
                  temp_est+quantile('normal',0.975)*std_err as CI_upper
    from temp_b_re;

quit;
run;

/* output the result */
proc print data=temp_b_re;
 title 'national average home temperature at night';
run;

proc export data=temp_b_re
 outfile='/folders/myfolders/Problem Set 6/ps6_q3b.csv' replace;
run;
```

```
ps6_q3b = read.csv("ps6_q3b.csv")
knitr::kable(ps6_q3b)
```

temp_est	CI_lower	CI_upper
68.10518	67.9265	68.28386

The above is the national average home temperature at night, among homes that use space heating.

Part C

```
/* c. Next, by census division, estimate the average winter home
temperatures at night, during the day with someone home, and during
the day with no one home (when applicable) */
proc sql;
 create table temp_c_ini as select DOEID,TEMPNITE,TEMPHOME,TEMPGONE,NWEIGHT,DIVISION
 from mylib.recs;
 create table bt as select *
 from mylib.brrwt_long;
 /*  diff from in q2, we can first merge the whole set and then drop the uncessary point */
 create table temp_c as
 select tc.DIVISION,tc.DOEID,tc.TEMPNITE,tc.TEMPHOME,tc.TEMPGONE,tc.NWEIGHT,bt.repl,bt.wt from bt left join
  (select * from temp_c_ini tc)
  on tc.DOEID = bt.DOEID
  where not TEMPNITE = -2 and not TEMPHOME = -2 and not TEMPGONE = -2;

  /*   calc the est and stderr */
  create table temp_c_re as
  select DIVISION,
           mean(night_temp_est) as night_temp_est,
         (mean(4*(night_temp_est - night_temp_repl)**2))**0.5 as night_std_err,
         mean(home_temp_est) as home_temp_est,
         (mean(4*(home_temp_est - home_temp_repl)**2))**0.5 as home_std_err,
         mean(gone_temp_est) as gone_temp_est,
         (mean(4*(gone_temp_est - gone_temp_repl)**2))**0.5 as gone_std_err
  from (select sum(TEMPNITE*NWEIGHT)/sum(NWEIGHT) as night_temp_est,
                sum(TEMPNITE*wt)/sum(wt) as night_temp_repl,
                sum(TEMPHOME*NWEIGHT)/sum(NWEIGHT) as home_temp_est,
                sum(TEMPHOME*wt)/sum(wt) as home_temp_repl,
                sum(TEMPGONE*NWEIGHT)/sum(NWEIGHT) as gone_temp_est,
                sum(TEMPGONE*wt)/sum(wt) as gone_temp_repl,
                DIVISION
        from temp_c
        group by DIVISION,repl
  )
  group by DIVISION;
/*   calc CI */
  create table temp_c_re as
    select night_temp_est, night_temp_est+quantile('normal',0.025)*night_std_err as night_lower,
         night_temp_est+quantile('normal',0.975)*night_std_err as night_upper,
         home_temp_est, home_temp_est+quantile('normal',0.025)*home_std_err as home_lower,
         home_temp_est+quantile('normal',0.975)*home_std_err as home_upper,
         gone_temp_est, gone_temp_est+quantile('normal',0.025)*gone_std_err as gone_lower,
         gone_temp_est+quantile('normal',0.975)*gone_std_err as gone_upper,
         DIVISION
    from temp_c_re;
/* send decoding information  */
  alter table temp_c_re add division_name char(20);
  update temp_c_re set division_name = 'New_England' WHERE DIVISION = 1;
  update temp_c_re set division_name = 'Middle_Atlantic' WHERE DIVISION = 2;
  update temp_c_re set division_name = 'East_North_Central' WHERE DIVISION = 3;
  update temp_c_re set division_name = 'West_North_Central' WHERE DIVISION = 4;
  update temp_c_re set division_name = 'South_Atlantic' WHERE DIVISION = 5;
  update temp_c_re set division_name = 'East_South_Central' WHERE DIVISION = 6;
  update temp_c_re set division_name = 'West_South_Central' WHERE DIVISION = 7;
  update temp_c_re set division_name = 'Mountain_North' WHERE DIVISION = 8;
  update temp_c_re set division_name = 'Mountain_South' WHERE DIVISION = 9;
  update temp_c_re set division_name = 'Pacific' WHERE DIVISION = 10;
  alter table temp_c_re drop division;

quit;
run;

/* output the result */
proc print data=temp_c_re;
 title 'national average home temperature by census division';
run;

proc export data=temp_c_re
 outfile='/folders/myfolders/Problem Set 6/ps6_q3c.csv' replace;
run;
```

```
ps6_q3c = read.csv("ps6_q3c.csv")
knitr::kable(ps6_q3c)
```

night_temp_est	night_lower	night_upper	home_temp_est	home_lower	home_upper	gone_temp_est	gone_lower	gone_upper	division_
65.67196	64.88227	66.46164	68.18011	67.30774	69.05247	64.61019	64.01626	65.20412	New_Engl
67.30342	66.83123	67.77562	69.20760	68.82265	69.59255	66.48801	66.02259	66.95344	Middle_At
67.93268	67.55995	68.30540	69.70346	69.35766	70.04927	66.83603	66.54731	67.12475	East_Nort
68.06797	67.58524	68.55071	69.63445	69.22405	70.04486	67.17892	66.60541	67.75243	West_Nor
69.19868	68.77466	69.62270	70.64661	70.35649	70.93672	68.54698	68.15903	68.93492	South_Atl
69.30772	68.75929	69.85616	70.62840	70.18734	71.06946	68.13987	67.44907	68.83067	East_Sou

night_temp_est	night_lower	night_upper	home_temp_est	home_lower	home_upper	gone_temp_est	gone_lower	gone_upper	division_
69.86110	69.50399	70.21821	71.29428	70.88730	71.70127	68.26726	67.42280	69.11172	West_Sou
66.80776	65.72636	67.88916	69.24982	68.59538	69.90426	65.93051	65.35916	66.50185	Mountain_
68.88158	66.74826	71.01490	70.82877	69.89128	71.76626	67.41744	65.78798	69.04691	Mountain_
66.54376	66.04240	67.04512	69.75933	69.32516	70.19351	65.34039	64.98769	65.69310	Pacific

The above is the average winter home temperatures at night, during the day with someone home, and during the day with no one home (when applicable), by census division.