4-C-Skip-Test-EX-SUS

October 31, 2017

In [3]:
```
from utility import db_connect, query2csv
from settings import  DBNAME, DBPASS, DBUSER, DBHOST
```

In [4]:
```
qsql="""
with d as (
  select generate_series(0,6) as dayofweek
),
m as (
  select generate_series(1,12) as month
),
-- v_ijmy:Compute an average by day of week for each month.
v_ijmy as (
  select
      baadv.analysis_area_id,
      to_char(baadv.date, 'YYYY') as year,
      avg(baadv.volume)::bigint as volume_i,
      avg(baadv.volume) as volume,
      d.dayofweek,
      m.month
  from
      baa_ex_sus.analysis_areas_daily_volume as baadv,
      d,
      m
  where
      extract(dow from baadv.date) in (d.dayofweek)
      AND date_part('month', baadv.date) = m.month
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
),
-- madt: average volume each month, each year for sites
madt as (
  select
      analysis_area_id,
      month,
      year,
      avg(volume)::bigint as volume_i,
      avg(volume) as volume
  from
```

```
      v_ijmy
      group by analysis_area_id, year, month
      having count(dayofweek)=7 -- having 7 days of data each week
),
AADT as (
select
  analysis_area_id,
  year,
  avg(volume)::bigint as AADT_i,
  round(avg(volume), 2) as AADT
from madt
  group by analysis_area_id, year
  having count(month) = 12 -- having 12 months of data
),
-- daily_exclude_holiday: daily counts for sites excluding holidays
daily_exclude_holiday as (
select
 baaad.analysis_area_id,
 baaad.date,
 baaad.volume,
 date_part('month', baaad.date) as month,
 date_part('dow', baaad.date) as dow
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  left join baa.holidays as baahd on baaad.date::date = baahd.holiday_date
where
  baahd.holiday_id is null
  group by 1,2,3
),
V_jmyl_exclude_holiday as (
  select
      baadv.analysis_area_id,
      to_char(baadv.date, 'YYYY') as year,
      avg(baadv.volume) as volume,
      d.dayofweek,
      m.month
  from
      daily_exclude_holiday as baadv,
      d,
      m
  where
      extract(dow from baadv.date) in (d.dayofweek)
      AND date_part('month', baadv.date) = m.month
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
),
-- 84 factors volume count should exclude holiday weeks
factor84 as (
select
```

```
  v_jmyl_nh.analysis_area_id,
  v_jmyl_nh.volume as v_jmyl,
  AADT.aadt as aadt,
  round(v_jmyl_nh.volume/aadt::numeric, 2) as f_jmys,
  v_jmyl_nh.dayofweek,
  v_jmyl_nh.month,
  v_jmyl_nh.year
from
  V_jmyl_exclude_holiday as v_jmyl_nh inner join AADT using(analysis_area_i
where
  AADT.AADT <> 0
),
-- Calculating 84 factors, first calculate V_jmyl excluding holiday weeks
wkstart as (
select
  baaad.analysis_area_id,
  date_trunc('week', baaad.date) AS week_start,
  to_char(baaad.date, 'YYYY') as year
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  left join baa.holidays as baahd on baaad.date::date = baahd.holiday_date
where
  baahd.holiday_id is null
  group by 1,2,3
  having count(baaad.date)=7
),
-- daily_no_holiday: daily counts for sites excluding holiday weeks
daily_no_holiday as (
select
 baaad.analysis_area_id,
 baaad.date,
 baaad.volume,
 date_part('month', baaad.date) as month,
 se.year,
 date_part('dow', baaad.date) as dow,
 date_part('doy', baaad.date) as doy
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  inner join wkstart as se using (analysis_area_id)
where
  baaad.date <= se.week_start + interval '6' day
  and baaad.date >=se.week_start
),
V_jmyl_no_holiday as (
  select
      aa_dnh.analysis_area_id,
      aa_dnh.year,
      avg(aa_dnh.volume)::bigint as volume,
```

```
        d.dayofweek,
        m.month
    from
        daily_no_holiday as aa_dnh,
        d,
        m
    where
        extract(dow from aa_dnh.date) in (d.dayofweek)
        AND date_part('month', aa_dnh.date) = m.month
        group by 1,2,4,5
),
-- average 84 factors for group of sites that exclude the test site
f84_est as (
select
    fg.city,
    fg.weekly_group,
    fg.mode,
    fg.analysis_area_test_id,
    fg.trial_analysis_area_id_list,
    f84.dayofweek,
    f84.month,
    f84.year,
    round(avg(f84.f_jmys), 2) as f_jmys_avg
  from
    factor84 as f84 inner join baa_ex_sus.factor_group_skip_test as fg
    on f84.analysis_area_id = Any(fg.trial_analysis_area_id_list::int[])
    group by
    fg.city,
    fg.weekly_group,
    fg.mode,
    fg.analysis_area_test_id,
    fg.trial_analysis_area_id_list,
    f84.dayofweek,
    f84.month,
    f84.year
    order by fg.city,
    fg.weekly_group,
    fg.mode,f84.year, f84.month, f84.dayofweek
 )
select
  aa_dnh.analysis_area_id, -- this is the test site id excluded from trial
  f84_est.city,
  f84_est.weekly_group,
  f84_est.mode,
  aa_dnh.year,
  aa_dnh.month,
  aa_dnh.date,
  aa_dnh.dow,
```

```
        aa_dnh.doy,
        f84_est.trial_analysis_area_id_list as trial_factor_group,
        array_append(f84_est.trial_analysis_area_id_list, aa_dnh.analysis_area_id
        aa_dnh.volume,
        f84_est.f_jmys_avg as trial_factor,
        round(aa_dnh.volume/f84_est.f_jmys_avg, 2) as aadb_est,
        aadt.aadt as aadb
      from
        daily_no_holiday as aa_dnh, f84_est, aadt
      where
        f84_est.analysis_area_test_id = aa_dnh.analysis_area_id
        and f84_est.dayofweek = aa_dnh.dow
        and f84_est.month = aa_dnh.month
        and f84_est.year = aa_dnh.year
        and aadt.analysis_area_id=aa_dnh.analysis_area_id
        and aadt.year = aa_dnh.year
        and f84_est.f_jmys_avg <> 0
        order by aa_dnh.analysis_area_id, aa_dnh.date, aa_dnh.dow, aa_dnh.month,
      """
      csvfile='4-C-Skip-test_EX_SUS.csv'
      query2csv(qsql,csvfile)
```

```
<IPython.core.display.HTML object>
```

```
In [5]: qsql="""
        with d as (
          select generate_series(0,6) as dayofweek
        ),
        m as (
          select generate_series(1,12) as month
        ),
        -- v_ijmy:Compute an average by day of week for each month.
        v_ijmy as (
          select
              baadv.analysis_area_id,
              to_char(baadv.date, 'YYYY') as year,
              avg(baadv.volume)::bigint as volume_i,
              avg(baadv.volume) as volume,
              d.dayofweek,
              m.month
          from
              baa_ex_sus.analysis_areas_daily_volume as baadv,
              d,
              m
          where
              extract(dow from baadv.date) in (d.dayofweek)
              AND date_part('month', baadv.date) = m.month
```

```
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
),
-- madt: average volume each month, each year for sites
madt as (
  select
      analysis_area_id,
      month,
      year,
      avg(volume)::bigint as volume_i,
      avg(volume) as volume
  from
      v_ijmy
      group by analysis_area_id, year, month
      having count(dayofweek)=7 -- having 7 days of data each week
),
AADT as (
select
  analysis_area_id,
  year,
  avg(volume)::bigint as AADT_i,
  round(avg(volume), 2) as AADT
from madt
  group by analysis_area_id, year
  having count(month) = 12 -- having 12 months of data
),
-- daily_exclude_holiday: daily counts for sites excluding holidays
daily_exclude_holiday as (
select
 baaad.analysis_area_id,
 baaad.date,
 baaad.volume,
 date_part('month', baaad.date) as month,
 date_part('dow', baaad.date) as dow
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  left join baa.holidays as baahd on baaad.date::date = baahd.holiday_date
where
  baahd.holiday_id is null
  group by 1,2,3
),
V_jmyl_exclude_holiday as (
  select
      baadv.analysis_area_id,
      to_char(baadv.date, 'YYYY') as year,
      avg(baadv.volume) as volume,
      d.dayofweek,
      m.month
  from
```

```
      daily_exclude_holiday as baadv,
      d,
      m
  where
      extract(dow from baadv.date) in (d.dayofweek)
      AND date_part('month', baadv.date) = m.month
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
),
-- 84 factors volume count should exclude holiday weeks
factor84 as (
select
  v_jmyl_nh.analysis_area_id,
  v_jmyl_nh.volume as v_jmyl,
  AADT.aadt as aadt,
  round(v_jmyl_nh.volume/aadt::numeric, 2) as f_jmys,
  v_jmyl_nh.dayofweek,
  v_jmyl_nh.month,
  v_jmyl_nh.year
from
  V_jmyl_exclude_holiday as v_jmyl_nh inner join AADT using(analysis_area_i
where
  AADT.AADT <> 0
),
-- Calculating 84 factors, first calculate V_jmyl excluding holiday weeks
wkstart as (
select
  baaad.analysis_area_id,
  date_trunc('week', baaad.date) AS week_start,
  to_char(baaad.date, 'YYYY') as year
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  left join baa.holidays as baahd on baaad.date::date = baahd.holiday_date
where
  baahd.holiday_id is null
  group by 1,2,3
  having count(baaad.date)=7
),
-- daily_no_holiday: daily counts for sites excluding holiday weeks
daily_no_holiday as (
select
 baaad.analysis_area_id,
 baaad.date,
 baaad.volume,
 date_part('month', baaad.date) as month,
 se.year,
 date_part('dow', baaad.date) as dow,
 date_part('doy', baaad.date) as doy
from
```

```
  baa_ex_sus.analysis_areas_daily_volume as baaad
  inner join wkstart as se using (analysis_area_id)
where
  baaad.date <= se.week_start + interval '6' day
  and baaad.date >=se.week_start
),
V_jmyl_no_holiday as (
  select
      aa_dnh.analysis_area_id,
      aa_dnh.year,
      avg(aa_dnh.volume)::bigint as volume,
      d.dayofweek,
      m.month
  from
      daily_no_holiday as aa_dnh,
      d,
      m
  where
      extract(dow from aa_dnh.date) in (d.dayofweek)
      AND date_part('month', aa_dnh.date) = m.month
      group by 1,2,4,5
),
-- average 84 factors for group of sites that exclude the test site
f84_est as (
select
    fg.city,
    fg.weekly_group,
    fg.mode,
    fg.analysis_area_test_id,
    fg.trial_analysis_area_id_list,
    f84.dayofweek,
    f84.month,
    f84.year,
    round(avg(f84.f_jmys), 2) as f_jmys_avg
  from
    factor84 as f84 inner join baa_ex_sus.factor_group_skip_test as fg
    on f84.analysis_area_id = Any(fg.trial_analysis_area_id_list::int[])
    group by
    fg.city,
    fg.weekly_group,
    fg.mode,
    fg.analysis_area_test_id,
    fg.trial_analysis_area_id_list,
    f84.dayofweek,
    f84.month,
    f84.year
    order by fg.city,
    fg.weekly_group,
```

```
    fg.mode,f84.year, f84.month, f84.dayofweek
 ),
-- aadb estimate for test sites
aadb_est as (
select
  aa_dnh.analysis_area_id, -- this is the test site id excluded from trial
  f84_est.city,
  f84_est.weekly_group,
  f84_est.mode,
  aa_dnh.year,
  aa_dnh.month,
  aa_dnh.date,
  aa_dnh.dow,
  aa_dnh.doy,
  f84_est.trial_analysis_area_id_list as trial_factor_group,
  array_append(f84_est.trial_analysis_area_id_list, aa_dnh.analysis_area_id
  aa_dnh.volume,
  f84_est.f_jmys_avg as trial_factor,
  round(aa_dnh.volume/f84_est.f_jmys_avg, 2) as aadb_est,
  aadt.aadt as aadb
from
  daily_no_holiday as aa_dnh, f84_est, aadt
where
  f84_est.analysis_area_test_id = aa_dnh.analysis_area_id
  and f84_est.dayofweek = aa_dnh.dow
  and f84_est.month = aa_dnh.month
  and f84_est.year = aa_dnh.year
  and aadt.analysis_area_id=aa_dnh.analysis_area_id
  and aadt.year = aa_dnh.year
  and f84_est.f_jmys_avg <> 0
  order by aa_dnh.analysis_area_id, aa_dnh.date, aa_dnh.dow, aa_dnh.month,
),
-- aadb estimate for sites and their actual aadt (same as aadb)
aadb_est_daily as (
select
  aadb_est.*,
  aadt.aadt
from
  aadb_est inner join aadt using(analysis_area_id, year)
),
-- Aggregate aadt estimate into weekly estimate
aadt_est_weekly as (
select
  analysis_area_id,
  mode,
  year,
  date_part('month', date_trunc('week', date)) as month,
  date_trunc('week', date) AS monday,
```

```
          factor_group,
          trial_factor_group,
          aadt,
          round(avg(aadb_est),2) as aadt_est_weekly
        from
          aadb_est_daily
          group by 1,2,3,4,5,6,7,8
        )
        select
          *,
          round((aadt_est_weekly-aadt)/aadt::numeric,2) as error
        from aadt_est_weekly
        where aadt <> 0
        order by analysis_area_id, mode, year, month, monday
        """
        csvfile='4-C-test_error_EX_SUS.csv'
        query2csv(qsql,csvfile)
```

```
<IPython.core.display.HTML object>
```