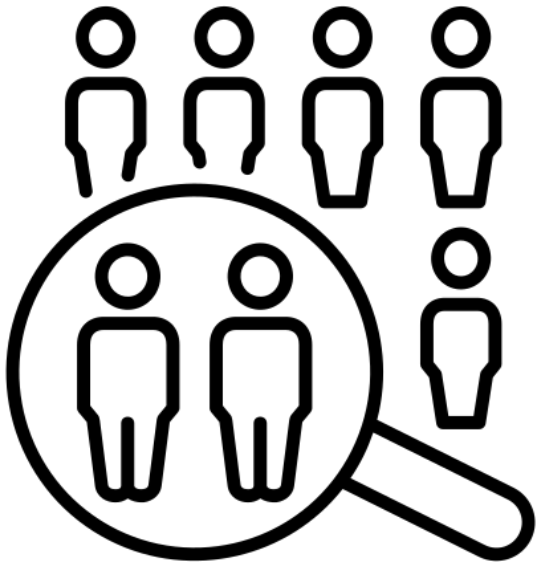


Menaksir Angka Anggota Keluarga

Sampling juniarto-bECj - Continuation Batch 9

Introduction

Introduction



- Topik yang akan kita bahas adalah menaksir anggota keluarga dari populasi menggunakan sampel. Populasi yang digunakan adalah Pegawai Pemerintahan Kota Pontianak.
 -
- Berdasarkan laporan Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017, rata-rata jumlah anak ideal baik dari segi ekonomi maupun kesehatan ada pada rentang 2,7 s.d 2,9 anak sehingga anggota keluarga ideal yang dimiliki adalah 3 orang. Hal ini akan menjadi petunjuk untuk melihat gambaran dari pemerintah kota dalam hal ini pegawai pemerintah kota pontianak yang sudah memiliki anggota keluarga.

Rancangan Sampling

Rancangan Sampling

- **Populasi** : Populasi yang digunakan yaitu seluruh pegawai daerah kota Pontianak baik kota maupun kelurahan.
- **Populasi Target** : Populasi target yang digunakan yaitu pegawai yang memiliki anggota keluarga (telah berkeluarga).
- **Kerangka Sampling**: Sampling diperoleh dari open data pemerintah kota Pontianak.
- **Unit Sampling**: Unit sampling pertama adalah kelurahan di bawah kota Pontianak. Unit sampling kedua adalah individu pegawai dari kelurahan tersebut.
- **Unit Observasi**: Unit observasi yang diteliti adalah pada sampling pegawai yang mempunyai anggota keluarga
- **Karakteristik yang diteliti** : Berapa jumlah anggota keluarga yang dimiliki.

Metode Sampling

Metode Sampling.

- Secara garis besar kerangka yang akan digunakan dalam pengambilan sampel adalah:



- Populasi yang di gunakan adalah data seluruh pegawai kelurahan yang berada pada area Pemerintah Kota Pontianak. Dengan pertimbangan cakupan yang tidak begitu luas maka kita membuat cluster dengan karakteristik homogen tingkat pertama yaitu sebanyak 22 kelurahan. Unit observasi yang berada pada cluster yaitu kelurahan terdiri dari karakteristik yang berbeda (heterogen) yaitu pegawai pns, honorer, pegawai tetap non pns.

Langkah Pengambilan Sampling

Pra Survey Untuk Mendapat jumlah Cluster

- Langkah pertama yang dilakukan adalah menentukan ukuran sampel melalui pra survey, survey dilakukan kepada 5 kelurahan dengan hasil sebagai berikut:

```
#Load Dataset Pra Sample
data = pd.read_excel(r'C:\Users\jkurn\Documents\Sampling\pra_sample.xlsx')
data.head()
```

	no	nama	status	jumlah	tgl	kelurahan	kecamatan
0	1	RIZKI FEBRIYANTI, S.STP, M.AP	ASN	5	2020-03-13 00:00:00	KELURAHAN DARAT SEKIP	PONTIANAK KOTA
1	2	ITA ROSMAYANTY, S.E, M.M	ASN	3	2020-03-16 00:00:00	KELURAHAN DARAT SEKIP	PONTIANAK KOTA
2	3	SETIA BUDI, S.Sos	ASN	5	2020-02-28 00:00:00	KELURAHAN DARAT SEKIP	PONTIANAK KOTA
3	4	ELITHA RIA PANGGABEAN	ASN	6	2020-03-18 00:00:00	KELURAHAN DARAT SEKIP	PONTIANAK KOTA
4	5	ARIF WIBISONO, A.Md	ASN	4	2020-03-17 00:00:00	KELURAHAN DARAT SEKIP	PONTIANAK KOTA

- Kemudian kita hitung jumlah sample dengan formula sbb

$$n = \frac{N\sigma_c^2}{N.D + \sigma_c^2}$$

Pra Survey Untuk Mendapat jumlah Cluster

- Langkah selanjutnya buat tabulasi per clusternya untuk menghitung M_i dan y_i .

```
#Buat Hasilnya per Cluster untuk menghitung jumlah unit tiap kelurahan dan jumlah hasil observasi
jumlah = data.groupby('kelurahan')['nama'].count().reset_index(name='total_unit(Mi)')
total = data.groupby('kelurahan')['jumlah'].sum().reset_index(name='jml_data(yi)')
pra_sample = pd.merge(jumlah, total, on=['kelurahan'])
pra_sample
```

	kelurahan	total_unit(Mi)	jml_data(yi)
0	KELURAHAN BANGKA BELITUNG DARAT	12	54
1	KELURAHAN BANSIR LAUT	8	36
2	KELURAHAN DARAT SEKIP	10	45
3	KELURAHAN SAIGON	11	45
4	KELURAHAN SUNGAI JAWI DALAM	7	31

- Hitung Nilai D dengan Bound Of Error yang telah ditentukan.

$$D = \frac{B^2}{4N^2}$$

```
#Hitung Nilai D dengan bound of error sebesar 7 dari perhitungan
B = 7
N = 22 # Total sample

# Maka diperoleh nilai D sebagai berikut
D = B**2/(4*(N**2))
D
```

0.0253099173553719

Pra Survey Untuk Mendapat jumlah Cluster

- Hitung nilai μ dengan rumus

$$\hat{\mu} = \frac{\hat{\tau}}{M}$$

$$\hat{\tau} = \frac{N}{n} \sum_{i=1}^n y_i = N\bar{y}$$

```
# cari r atau mu
# y_i adalah jumlah anggota keluarga dari kelurahan yang disample
y_total_1 = np.sum(pra_sample["jml_data(yi)"])

# M_i adalah jumlah orang/unit dari kelurahan yang disample
M1 = np.sum(pra_sample["total_unit(Mi)"])

mu = y_total_1/M1
print(f"rata-rata anggota keluarga yang diestimasi : {mu:.0f} orang")

rata-rata anggota keluarga yang diestimasi : 4 orang
```

Pra Survey Untuk Mendapat jumlah Cluster

- Setelah didapat nilai μ kemudian cari varians berdasarkan tabel tabulasi per kelurahan(cluster) di atas.

$$\sigma_c^2 = \frac{\sum_{i=1}^N (y_i - \hat{\mu}M_i)^2}{N - 1}$$

```
pra_sample['muMi'] = pra_sample['total_unit(Mi)'] * mu
pra_sample['yi-muMi'] = pra_sample['jml_data(yi)'] - pra_sample['muMi']
pra_sample['yi-muMi^2'] = pra_sample['yi-muMi'] ** 2
pra_sample
```

	kelurahan	total_unit(Mi)	jml_data(yi)	muMi	yi-muMi	yi-muMi^2
0	KELURAHAN BANGKA BELITUNG DARAT	12	54	52.750000	1.250000	1.562500
1	KELURAHAN BANSIR LAUT	8	36	35.166667	0.833333	0.694444
2	KELURAHAN DARAT SEKIP	10	45	43.958333	1.041667	1.085069
3	KELURAHAN SAIGON	11	45	48.354167	-3.354167	11.250434
4	KELURAHAN SUNGAI JAWI DALAM	7	31	30.770833	0.229167	0.052517

```
# Cari Varians
y_kuadrat = np.sum(pra_sample["yi-muMi^2"])
varians = y_kuadrat / (N-1)
varians
```

0.6973792989417988

Pra Survey Untuk Mendapat jumlah Cluster

- Setelah didapat semua variabelnya maka masukkan ke formula perhitungan :

$$n = \frac{N\sigma_c^2}{N.D + \sigma_c^2}$$

```
# hitung n
n = N*varians/((N*D)+varians)
n = math.ceil(n)
print(f"Maka Sampel yang harus di ambil adalah sebanyak : {n} cluster")
```

Maka Sampel yang harus di ambil adalah sebanyak : 13 cluster

Pengumpulan Data

Pengumpulan Data



- Setelah diperoleh sebanyak 13 cluster yang dijadikan sampel, maka selanjutnya adalah mengacak dengan cara random sampling 13 cluster dari 22 cluster yang tersedia.
- Observasi data dilakukan dengan mengambil data pada situs open data Pemkot Pontianak (<https://data.pontianak.go.id/>) sehingga jenis data adalah data sekunder.

Penaksiran Parameter

Penaksiran Parameter

- Data Hasil Sampling ke 13 CLuster

```
#Load Data Hasil Sampling ke 13 Cluster  
data_net = pd.read_excel(r'C:\Users\jkurn\Documents\Sampling\net_sample1.xlsx')  
data_net.head()
```

	no	nama	status	jumlah	tgl	kelurahan	kecamatan
0	1	Dra.DWI SURYANTI	ASN	7	2020-03-28 00:00:00	KELURAHAN BANJAR SERASAN	PONTIANAK TIMUR
1	2	NORANI,SH,ME	ASN	4	2020-03-04 00:00:00	KELURAHAN BANJAR SERASAN	PONTIANAK TIMUR
2	3	SYARIF YUSUF	ASN	6	2020-03-05 00:00:00	KELURAHAN BANJAR SERASAN	PONTIANAK TIMUR
3	4	PAKU ERNANTO DJATI	ASN	6	2020-03-05 00:00:00	KELURAHAN BANJAR SERASAN	PONTIANAK TIMUR
4	5	SYF.ROSMAWATI,A.Md	ASN	5	2020-03-03 00:00:00	KELURAHAN BANJAR SERASAN	PONTIANAK TIMUR

Penaksiran Parameter

- Langkah pertama buat tabulasi per clusternya untuk menghitung M_i dan y_i .

```
#Buat Hasilnya per Cluster untuk menghitung jumlah unit tiap kelurahan dan jumlah hasil observasi
jumlah1 = data_net.groupby('kelurahan')['nama'].count().reset_index(name='total_unit(Mi)')
total1 = data_net.groupby('kelurahan')['jumlah'].sum().reset_index(name='jml_data(yi)')
net_sample = pd.merge(jumlah1, total1, on=['kelurahan'])
net_sample
```

	kelurahan	total_unit(Mi)	jml_data(yi)
0	KELURAHAN BENUA MELAYU DARAT	10	48
1	KELURAHAN BANGKA BELITUNG DARAT	12	54
2	KELURAHAN BANGKA BELITUNG LAUT	11	48
3	KELURAHAN BANJAR SERASAN	10	44
4	KELURAHAN BANSIR DARAT	8	30
5	KELURAHAN BANSIR LAUT	8	36
6	KELURAHAN DALAM BUGIS	10	39
7	KELURAHAN SAIGON	11	45
8	KELURAHAN SIANTAN TENGAH	11	43
9	KELURAHAN SUNGAI JAWI DALAM	7	31
10	KELURAHAN TAMBELAN SAMPIT	6	27
11	KELURAHAN TANJUNG HILIR	8	38
12	KELURAHAN TANJUNG HULU	11	56

Penaksiran Rata-rata

- Karena ukuran unit tiap cluster berbeda maka menggunakan pendekatan ratio estimation r .
- Kemudian Hitung nilai rata-ratanya:

$$\hat{\mu}_r = r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n M_i}$$

Penaksiran Rata-rata

Karena ukuran unit tiap cluster berbeda maka menggunakan pendekatan ratio estimation r

```
# y_i adalah jumlah anggota keluarga dari kelurahan yang disample
y_total = np.sum(net_sample["jml_data(yi)"])

# M_i adalah jumlah orang/unit dari kelurahan yang disample
M = np.sum(net_sample["total_unit(Mi)"])

# r or mu_r adalah sebagai berikut
mu_r = y_total / M
mu_r

print(f"jumlah anggota keluarga dari kelurahan yang disample      : {y_total:.0f} orang")
print(f"jumlah orang/unit dari kelurahan yang disample             : {M} orang")
print(f"rata-rata anggota keluarga yang diestimasi                   : {mu_r:.0f} orang")

jumlah anggota keluarga dari kelurahan yang disample      : 539 orang
jumlah orang/unit dari kelurahan yang disample             : 123 orang
rata-rata anggota keluarga yang diestimasi                 : 4 orang
```

Penaksiran Estimator dengan 95% Confidence Interval

- Lakukan Perhitungan Variance Sample

$$s_c^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - rM_i)^2$$

```
# Lakukan Perhitungan varians tiap cluster sample
# Hitung nilai yi-rMi2
net_sample['yi_rMi2'] = (net_sample['jml_data(yi)'] - (mu_r*net_sample['total_unit(Mi)']))**2
net_sample
```

	kelurahan	total_unit(Mi)	jml_data(yi)	yi_rMi2
0	KELURAHAN BENUA MELAYU DARAT	10	48	17.462886
1	KELURAHAN BANGKA BELITUNG DARAT	12	54	2.001190
2	KELURAHAN BANGKA BELITUNG LAUT	11	48	0.041311
3	KELURAHAN BANJAR SERASAN	10	44	0.031992
4	KELURAHAN BANSIR DARAT	8	30	25.572345
5	KELURAHAN BANSIR LAUT	8	36	0.889418
6	KELURAHAN DALAM BUGIS	10	39	23.243374
7	KELURAHAN SAIGON	11	45	10.260824
8	KELURAHAN SIANTAN TENGAH	11	43	27.073832
9	KELURAHAN SUNGAI JAWI DALAM	7	31	0.105757
10	KELURAHAN TAMBELAN SAMPIT	6	27	0.500297
11	KELURAHAN TANJUNG HILIR	8	38	8.661775
12	KELURAHAN TANJUNG HULU	11	56	60.789279

```
sc_square = np.sum(net_sample["yi_rMi2"]) / (n-1)
print(f"Perhitungan varians tiap kelurahan sample : ({np.sqrt(sc_square):.0f})^2")

# Next we calculate the variance of total population estimator
var_y_total_est = N*(N-n)*(sc_square/n)
print(f"Varians dari total populasi. : ({np.sqrt(var_y_total_est):.0f})^2")

Perhitungan varians tiap kelurahan sample : (4)^2
Varians dari total populasi. : (15)^2
```

Penaksiran Estimator dengan 95% Confidence Interval

- Kemudian Hitung Variance dari Rata-rata

$$\widehat{\text{var}}(\hat{\mu}_r) = \frac{1}{(NM)^2} \widehat{\text{var}}(\hat{\tau}_r)$$

```
# hitung rata-rata unit dari cluster yang di sample
M_bar = np.mean(net_sample['total_unit(Mi)'])
print(f"rata-rata pegawai dari kelurahan yang di sample      : {M_bar:.0f}")

# Hitung Total dari populasi berdasarkan rata-rata
M_tot_est = N * M_bar
print(f"Estimasi total pegawai dari kelurahan yang di sample : {M_tot_est:.0f}")

# Hitung Varians dari rata-rata
var_y_mean_est = var_y_total_est / (M_tot_est**2)
print(f"Varians dari rata-rata                                : ({np.sqrt(var_y_mean_est):.2f})^2")

rata-rata pegawai dari kelurahan yang di sample      : 9
Estimasi total pegawai dari kelurahan yang di sample : 208
Varians dari rata-rata                                : (0.07)^2
```


Penaksiran Estimator dengan 95% Confidence Interval

- Setelah didapat Variance dari rata-rata, maka dapat dihitung nilai d untug melihat Confidence Interval nya.

```
# Hitung z_stat @ 95% Confidence Interval
alpha = 0.05
z_stat = stats.norm.ppf(1 - alpha/2.)
print(f"z_stat          : {z_stat:.2f}")
```

```
# Hitung Margin of error (d)
d = z_stat * np.sqrt(var_y_mean_est)
print(f"margin of error (d) : {d:.2f}")
```

```
z_stat          : 1.96
margin of error (d) : 0.14
```

```
# Calculate the CI
lb = mu_r - d
ub = mu_r + d

print(f"Dengan Confidence Interval 95% rata-rata jumlah anggota")
print(f"keluarga pada pegawai kelurahan kota pontianak adalah {lb:.2f} orang and {ub:.2f} orang")
```

```
Dengan Confidence Interval 95% rata-rata jumlah anggota
keluarga pada pegawai kelurahan kota pontianak adalah 4.24 orang and 4.52 orang
```

Konklusi

Konklusi

- Berdasarkan perhitungan teknik sampling survey, maka dari 22 cluster kelurahan yang ada di kota pontianak, sebanyak 13 cluster digunakan sebagai sampling untuk mendapat hasil penaksiran yang baik.
- Dari hasil penaksiran dengan Confidence Interval sebesar 95% maka rata-rata anggota keluarga berada pada rentang 4,2 s.d 4,5, lebih tinggi dari anggota keluarga ideal yaitu 2,7 s.d 2,9. Kedepannya perlu dilakukan awareness terkait pentingnya mempunyai keluarga yang ideal secara ekonomi dan kesehatan.
- Sampling dengan cluster dari pegawai kelurahan jika diambil data secara visit ke lapangan, dapat menghemat biaya serta efisiensi karena pegawai berkantor di satu gedung yang sama, sehingga lebih mudah dalam observasinya.
- Hal yang perlu diperhatikan adalah, perlu dilakukan kajian hasil antara Cluster Sampling dengan Simple Random Sampling. Untuk data yang diobservasi juga perlu data tambahan misalkan status anggota keluarga agar hasil dapat lebih mendetail.

Reference

- data.pontianak.go.id
- the best sampling method | Data Science and Machine Learning | Kaggle
- Laporan Survey Demografi dan Kesehatan Indonesia SDKI 2017.

Thank You
