

Laporan Proyek MATH 1042 – Peluang dan Statistika

Proyek 2A: Survei Kepuasan Bandara



Kelompok 1:

Charis

Grace Ping

Isabella

Mega

Moody

Yehezkiel

Calvin Institute of Technology

Jakarta

2021

I. PENGANTAR

A. Latar Belakang

Sebuah bandara internasional di Texas, yaitu Bandara Austin Bergstrom meminta tim manajemen untuk mencari tahu bagaimana penilaian pengunjung bandara terhadap 33 layanan di bandara tersebut agar nantinya pihak bandara bisa mengetahui layanan mana yang perlu ditingkatkan. Tentunya mewawancarai satu per satu pengunjung bandara adalah hal yang tidak efektif, maka tim manajemen memutuskan untuk melakukan survei acak terhadap 3501 pengunjung bandara. Nantinya dari penilaian 3501 sampel ini, tim statistik dapat menarik kesimpulan untuk populasi seluruh pengunjung bandara. Data sampel tersebut dapat diakses dari <https://query.data.world/s/itm46bpjsz5gwvfxegils6fz3pjt7c>.

Dalam data tersebut dapat dilihat adanya kolom quarter atau kuartal yang dituliskan dalam bentuk XQYY. Di mana $X=\{1, 2, 3, 4\}$ yang adalah waktu kuartal survei tersebut dilakukan dengan rincian sebagai berikut:

- Kuartal 1: Januari-Maret
- Kuartal 2: April-Juni
- Kuartal 3: Juli-September
- Kuartal 4: Oktober-Desember

Dan YY menunjukkan tahun survei tersebut dilakukan. Data sampel yang diambil berisi data survei dari kuartal pertama tahun 2015 hingga kuartal kedua tahun 2017.

B. Tujuan

- Menyimpulkan rata-rata dan standar deviasi dari populasi skor kepuasan dari 3 kategori layanan dengan skor terburuk dengan mengolah 3501 sampel.
- Menentukan skor minimum agar bisa dengan yakin mengatakan 95% pengunjung memberikan skor kepuasan dengan minimum skor tersebut.
- Membuat inferensi statistik untuk proporsi pengunjung yang tergolong puas.
- Menentukan layanan dan kuartal paling seragam dan variatif.
- Merekomendasikan satu layanan yang perlu ditingkatkan.

II. RANGKUMAN HASIL TUGAS PEMROGRAMAN DAN INTERPRETASI

A. Memilih tiga layanan yang akan dianalisis

Dengan menggunakan fungsi `choose_feature`, dataset (seluruh layanan) yang ada akan diurutkan berdasarkan rata-ratanya dan diambil 3 layanan dengan rata-rata paling rendah.

```
In [ ]: def choose_feature(dataset):
        chosen_feature = set()
        for key, val in dataset.mean().sort_values()[0:3].items():
            chosen_feature.add(dataset.columns.get_loc(key))
        return chosen_feature
```

```
In [36]: feature_index = choose_feature(dataset)
        for index in feature_index:
            print(dataset.columns[index])

Business/executive lounges
Ease of making connections
Availability of banks/ATM/money changing
```

Melengkapi dataset

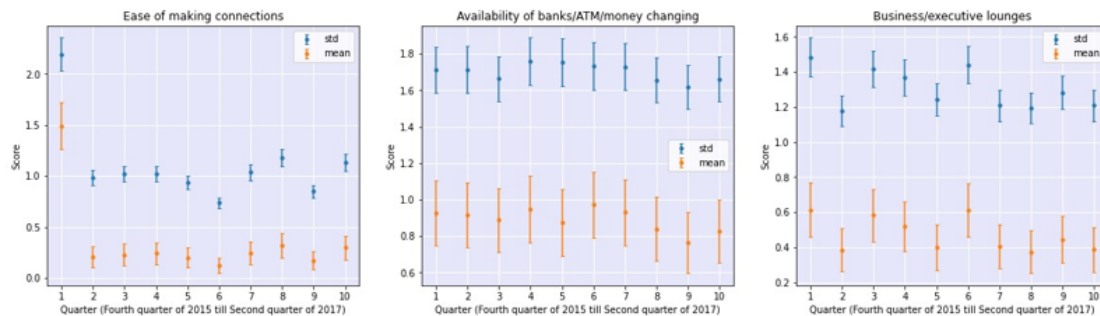
```
In [38]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3501 entries, 0 to 3500
Data columns (total 4 columns):
 #   Column                                                                 Non-Null Count  Dtype
---  -
 0   Quarter                                                                3501 non-null  object
 1   Ease of making connections                                              3418 non-null  float64
 2   Availability of banks/ATM/money changing                               3460 non-null  float64
 3   Business/executive lounges                                             3410 non-null  float64
dtypes: float64(3), object(1)
memory usage: 109.5+ KB
```

Dapat dilihat bahwa pada dataframe yang telah difilter masih terdapat beberapa sel yang kosong. Oleh karena itu, sel kosong tersebut akan diisi dengan nilai rata-rata dari kolom tempat sel tersebut berada. Kami menggunakan nilai rata-rata karena nantinya yang akan kami cari adalah estimasi rata-rata populasi, sehingga sebisa mungkin nilai rata-rata sampel tidak berubah.

B. Nilai rata-rata dan standar deviasi skor kepuasan populasi pengunjung dan visualisasinya

Dengan menggunakan rumus dan algoritma T-Student untuk mencari estimasi rata-rata dan chi square untuk mencari estimasi standar deviasi, maka ditemukan hasil perhitungan sebagai berikut:



Kategori Layanan	Estimasi Rata-Rata	Estimasi Std
Ease of Making Connections	(0.31, 0.39)	(1.2, 1.25)
Availability of Banks/ATM/Money Changing	(0.83, 0.94)	(1.65, 1.73)
Business/Executive Lounges	(0.43, 0.51)	(1.27, 1.33)

Beberapa hasil interpretasi kami terhadap perhitungan estimasi rata-rata dan standar deviasi adalah sebagai berikut:

- Layanan "Ease of Making Connection": pada kuartal pertama, nilai rata-rata dan standar deviasinya lebih tinggi daripada kuartal lainnya menunjukkan bahwa pada kuartal tersebut ada lebih banyak penerbangan ke bandara yang beragam, sehingga beberapa pengunjung memberikan nilai yang tidak terlalu buruk. Namun, pada kuartal-kuartal berikutnya, bandara Austin Bergstrom tidak banyak terhubung dengan bandara lainnya.
- Layanan "Availability of Banks/ATM/Money Changing": layanan ini cenderung mendapatkan penilaian yang stabil pada setiap kuartal karena memang jumlah fasilitas keuangan di bandara ini tidak ditambah ataupun dikurangi. Hal ini menunjukkan bahwa fasilitas keuangannya hanya itu-itu saja dan sudah terkenal kurang memuaskan bagi pengunjung.

- Layanan "Business/Executive Lounges": skor layanan ini tidak begitu stabil dan cenderung dinamis. Hal ini mungkin terkait dengan tingkat kebersihan lounges yang berbeda di setiap kuartalnya, tergantung pada pegawai yang ada.

III. PERTANYAAN DISKUSI

A. Menentukan batas bawah skor kepuasan s , sehingga kita dapat cukup yakin bahwa 95% total pengunjung bandara pasti memberikan penilaian lebih besar atau sama dengan s

Fungsi yang digunakan dalam bagian ini sama dengan fungsi yang digunakan dalam mencari nilai standar deviasi dengan tingkat kepercayaan 95%. Di mana dalam fungsi tersebut, akan ditemukan nilai batas bawah dan batas atas untuk nilai tingkat kepercayaan yang berpusat di tengah. Sedangkan, dalam kasus ini, yang diinginkan hanya nilai batas bawah agar 95% penilaian lainnya pasti lebih besar dari batas bawah tersebut atau dengan kata lain, batas atasnya adalah infiniti. Karena itu kami memasukkan tingkat kepercayaan 90% dengan fungsi standar deviasi ini agar nilai tengahnya adalah 90% sedangkan kedua sisi sisanya bernilai sama, yaitu 5%. Dengan demikian, batas bawah untuk tingkat kepercayaan 90% di tengah adalah sama dengan batas bawah untuk menyatakan 95% nilai lainnya pasti lebih besar dari batas bawah tersebut. Hasil akhirnya adalah sebagai berikut:

```
[272] #Ease of making connections
      mean_population_interval(df[df.columns[1]], 0.90)[0]
```

```
0.31753798444928527
```

```
[273] #Availability of banks/ATM/money changing
      mean_population_interval(df[df.columns[2]], 0.90)[0]
```

```
0.8416239111709057
```

```
[274] #Business/executive lounges
      mean_population_interval(df[df.columns[3]], 0.90)[0]
```

```
0.4353779150727187
```

B. Inferensi statistika untuk proporsi banyaknya responden yang tergolong 'puas' terhadap layanan apabila diasumsikan seseorang tergolong 'puas' terhadap suatu layanan jika ia mengisi skor lebih besar atau sama dengan 4

Untuk mencari proporsinya sama dengan mencari peluang distribusi normalnya. Dalam hal ini, yang diinginkan adalah skor yang lebih besar atau sama dengan 4 sehingga nilai mean populasi yang dimasukkan adalah 4. Dengan begini, kita akan mengetahui proporsi golongan puas secara populasi. Untuk ketiga layanan dengan skor terburuk, proporsi golongan puasnya mendekati 0.

```
In [52]: def proportion(dataset):  
         num_sample = len(dataset)  
         dof = num_sample-1  
         sample_mean = dataset.mean()  
         sample_std = dataset.std()  
         t_value = (-4 + sample_mean)/(sample_std/math.sqrt(num_sample))  
         p = t.cdf(t_value, dof)  
         return p
```

```
In [53]: #Ease of making connections  
         proportion(df[df.columns[1]])
```

```
Out[53]: 0.0
```

```
In [54]: #Availability of banks/ATM/money changing  
         proportion(df[df.columns[2]])
```

```
Out[54]: 0.0
```

```
In [55]: #Business/executive lounges  
         proportion(df[df.columns[3]])
```

```
Out[55]: 0.0
```

C. Menentukan layanan dan kuartal dengan tingkat kepuasan yang cenderung seragam serta cenderung bervariasi

Sebenarnya dengan melihat grafik standar deviasi saja sudah terlihat layanan dan kuartal mana yang paling seragam dan yang paling variatif. Layanan dan kuartal dengan standar deviasi terendah adalah yang paling seragam, yaitu layanan "Ease of Making

Connection" kuartal ke-6. Sedangkan, layanan dan kuartal dengan standar deviasi tertinggi adalah yang paling variatif, yaitu layanan "Ease of Making Connection" kuartal ke-1.

```
[40] def most_uniform(dataset):
    quarters = dataset['Quarter'].unique()
    serv = ""
    q = ""
    max = (-np.inf, -np.inf)
    for quarter in quarters:
        for service in dataset:
            if (service) != "Quarter":
                df = dataset[dataset['Quarter']==quarter][service]
                std = std_population_interval(df, 0.95)
                if std[1] > max[1]:
                    max = std
                    serv = service
                    q = quarter
    return q, serv, max

def most_varied(dataset):
    quarters = dataset['Quarter'].unique()
    serv = ""
    q = ""
    min = (np.inf, np.inf)
    for quarter in quarters:
        for service in dataset:
            if (service) != "Quarter":
                df = dataset[dataset['Quarter']==quarter][service]
                std = std_population_interval(df, 0.95)
                if std[0] < min[0]:
                    min = std
                    serv = service
                    q = quarter
    return q, serv, min
```

```
[41] most_varied(df)
```

```
('2Q16',
'Ease of making connections',
(0.6814138426076525, 0.7905669188011353))
```

```
[42] most_uniform(df)
```

```
('1Q15', 'Ease of making connections', (2.030108495638066, 2.3553038078692095))
```

D. Merekomendasikan satu layanan yang paling perlu ditingkatkan

Untuk menentukan layanan mana yang paling perlu ditingkatkan, kami hanya melihatnya dari rata-rata skor yang diberikan pengunjung dan menemukan bahwa layanan dengan rata-rata paling rendah adalah layanan "Ease of Making Connection".

```
[304] df.mean().sort_values()[0:1]
```

```
Ease of making connections    0.351614  
dtype: float64
```

IV. KESIMPULAN & SARAN

A. Kesimpulan

- Tiga layanan dengan skor terburuk adalah layanan kemudahan terhubung dengan bandara lain, ketersediaan layanan keuangan, dan kualitas lounge untuk kelas bisnis dan eksekutif. Ketiganya memiliki rata-rata dan standar deviasi penilaian di antara 0 dan 1.
- Dengan tingkat kepercayaan 95%, kita bisa yakin bahwa pengunjung akan memberikan skor 0 untuk ketiga layanan terburuk tersebut.
- Secara populasi, hampir tidak ada pengunjung yang bisa dikatakan puas untuk ketiga layanan tersebut. Dengan kata lain, proporsi puasnya mendekati 0.
- Dengan melihat standar deviasinya, dapat terlihat layanan dan kuartal yang paling seragam dan yang paling variatif.
- Satu layanan yang paling perlu untuk ditingkatkan adalah layanan terhubung dengan bandara lain.

B. Saran

- Jika tidak ada batasan jumlah layanan yang dapat ditingkatkan, maka semua layanan dengan tingkat keyakinan 95% pengunjung memberikan skor minimum di bawah 4 harus ditingkatkan.

- Cara filter data dapat disesuaikan dengan kepentingan penggunaan. Dalam kasus ini, yang ingin didapatkan adalah layanan yang paling perlu ditingkatkan sehingga filter yang digunakan adalah untuk mendapatkan layanan dengan skor terburuk.