

Fakultas Ilmu Komputer Universitas Indonesia  
CSGE602013 - Statistika & Probabilitas, Semester Genap 2016/2017

**Ujian Tengah Semester**

Senin, 3 April 2017

Waktu: 150 menit

**Petunjuk**

1. Ujian bersifat "*open notes*", mahasiswa diperbolehkan membuka catatan berupa **satu lembar** kertas A4 (bolak-balik).
2. Tidak dibagikan kertas buram. Mahasiswa dipersilakan mencoret-coret pada halaman kosong di set soal ini.
3. Diperbolehkan menggunakan kalkulator (bukan alat komunikasi seperti handphone, pad, dll)
4. Berkas soal terdiri dari **7** halaman yang terdiri 10 soal pilihan ganda dan 5 soal essay  
Mahasiswa wajib memeriksa kelengkapan berkas soal sebelum mulai mengerjakan ujian.
5. Tuliskan nama Anda pada setiap lembar berkas soal untuk mengantisipasi kertas tercecer.

Soal	Bobot	Nilai
Pilihan Ganda	25	
Essay 1	5	
Essay 2	15	
Essay 3	17	
Essay 4	20	
Essay 5	18	
<b>Total</b>	<b>100</b>	

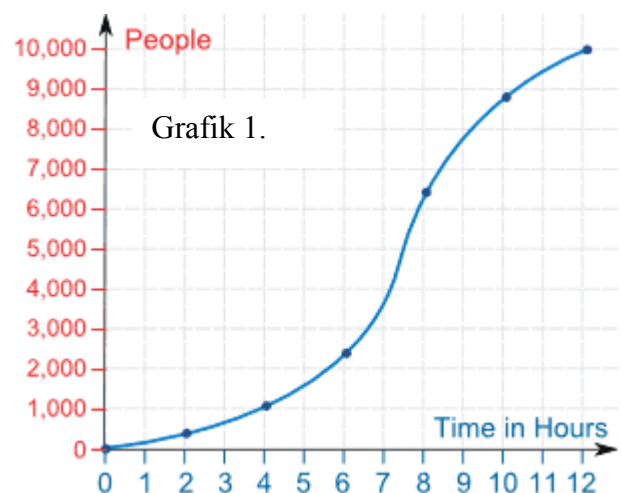
**SOAL PILIHAN GANDA (Bobot 20)**

Pilihlah jawaban yang benar untuk pertanyaan yang diberikan. Untuk setiap nomor, jawaban yang benar **hanya satu jawaban**.

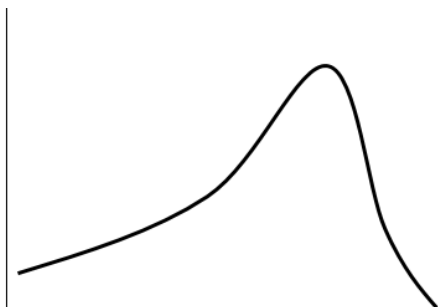
- Berdasarkan data-data sebelumnya, skor IQ score mahasiswa Fasilkom UI adalah variable acak dengan rata-rata 115 dan standard deviasi 5. Manakan berikut pernyataan yang benar?
  - Probabilitas seorang mahasiswa baru mempunyai skor IQ diatas 125 minimal 0.95
  - Probabilitas seorang mahasiswa baru mempunya skor IQ antara 105-1025 minimal 0.5
  - Jumlah sampel mahasiswa baru yang harus diambil adalah 25 sehingga probabilitas nya paling tidak 99% untuk mendapatkan rata-rata skor IQ antara 110-120
  - Semakin banyak sample mahasiswa yang diambil, semakin kecil probabilitas untuk mendapatkan rata-rata sample mendekati rata-rata variable acak nya.

- Grafik 1 menyatakan jumlah kumulatif orang yang menghabiskan sekian waktu di depan komputer dalam sehari. Pernyataan yang **benar** terkait Grafik 1 adalah

- Percentile rank dari 4 adalah 7
- Percentile 30 terletak di 6.5 jam
- Median data adalah 6
- Ada lebih dari 1000 orang yang menghabiskan waktu 2-4 jam di depan komputer



- Pernyataan yang **benar** tentang gambar berikut adalah:



- Menggambarkan distribusi frekuensi yang *positively skewed*
- Menggambarkan CDF dari suatu data
- Median lebih kecil daripada modus
- Mean data lebih besar daripada modus

- Diketahui data pada tabel berikut.

Maka dapat disimpulkan bahwa:

- Jika Budi mendapat nilai 60 di Kuis 2 maupun Kuis 3, maka nilai z-score nya di Kuis 2 lebih tinggi daripada nilai z-scorenya di Kuis 3
- Jika nilai ingin distandarkan sehingga reratanya menjadi 50 dan deviasi 10, maka nilai 70 Ani di Kuis 3 berubah menjadi 80
- Nilai 70 di Kuis 1 setara dengan nilai 55 di Kuis 3
- Jika Akbar mendapat nilai 70 di ketiga kuis, sebenarnya Akbar mendapat nilai terbaik di Kuis 2 jika dibandingkan dengan teman-temannya.

	Rerata	Standar Deviasi
Kuis 1	60	10
Kuis 2	50	10
Kuis 3	50	5

5. Pernyataan yang **salah** tentang peluang adalah
- $P(A|B) = 0$  jika A dan B *mutually exclusive*
  - $P(AB|CD) = P(A|BCD) \cdot P(BCD) / P(CD)$
  - $P(A) = P(A \cup B) + P(B) - P(AB)$**
  - $P(ABCD) = P(A) \cdot P(B|A) \cdot P(C|AB) \cdot P(D|ABC)$
6. Pernyataan yang **salah** tentang korelasi adalah
- Korelasi hanya dapat diterapkan pada data ordinal, interval atau rasio.
  - Korelasi merupakan nilai kovarian yang dinormalisasi sehingga nilai mutlaknya maksimal 1.
  - Nilai Pearson-r dan Spearman-rho antara dua buah data selalu sama asalkan keduanya berskala rasio.**
  - Nilai korelasi 0 berarti tidak ada hubungan linier antara kedua data.
7. Pernyataan yang **salah** tentang ekspektasi, variansi dan kovariansi adalah
- $E[2X + 3] = 2 \cdot E[X] + 3$
  - $\text{Var}[2X + 3] = 4 \text{Var}[X] + 3$**
  - $\text{Var}(2X) = 2 \cdot \text{Var}(X) + 2 \cdot \text{Cov}(X, X)$
  - $\text{Cov}(X, 2X) = 2 \cdot \text{Var}(X)$
8. Syarat suatu fungsi  $f_X(x)$  dapat menjadi *probability density function* (pdf) adalah ...
- $\int_{-\infty}^{\infty} x f_X(x) dx = 1$
  - $x$  boleh negatif, tapi  $f_X(x)$  tidak boleh negatif**
  - Monoton
  - Nilai di setiap titik  $x$  tidak boleh lebih dari 1
9. Seorang peneliti ingin melakukan survey terhadap 500 mahasiswa yang berbaris membentuk 10 barisan berdasarkan kisaran IPK. Manakah dari pernyataan berikut ini yang **benar**?
- Jika dia memilih mahasiswa ke 5, 10, 15, dst di setiap barisan, berarti dia menggunakan teknik *stratified sampling*
  - Jika dia memilih semua mahasiswa di barisan ke 3 dan 5 saja, berarti dia melakukan teknik *systematic sampling*
  - Jika dia memilih 5 orang secara acak di setiap barisan, berarti dia menggunakan teknik *random sampling*
  - Jika dia menggunakan *random number generator* untuk memilih tiga angka  $\{x, y, z\}$  antara 1-10 lalu memilih semua mahasiswa di barisan ke- $\{x, y, z\}$ , berarti dia melakukan teknik *cluster sampling***
10. Pernyataan yang **benar** tentang Markov dan Chebyshev *inequality* adalah:
- Markov *inequality* dapat digunakan untuk mencari batas peluang jika mean dan variansi diketahui
  - Chebyshev *inequality* tidak ada hubungannya dengan *weak law of large number*
  - Chebyshev *inequality* menyatakan bahwa peluang nilai suatu variabel acak berada pada kisaran  $\text{mean} \pm 2$  standar deviasi maksimal adalah  $1/4$
  - Chebyshev *inequality* menyatakan bahwa peluang nilai suatu variabel acak berada di luar kisaran  $\text{mean} \pm 3$  standar deviasi maksimal adalah  $1/9$**

**Soal Essay (80 poin)**

1. [5] Jelaskan perbedaan antara statistik deskriptif dan statistik inferensial

Statistik deskriptif: Cara mendeskripsikan dan menarik kesimpulan dari dataset (biasanya data hasil survey atau sampling), tidak menggunakan teori probabilitas.

Statistik inferensial: Cara mendeduksi parameter populasi dari data/statistik sampel (menggunakan teori probabilitas).

2. [15] Tini mempunyai 6 pasang sepatu (jadi total ada 12 buah sepatu). Tini akan mengambil 6 buah sepatu secara acak. Berapakah probabilitasnya dari ke 6 sepatu yang diambil:

- a. tidak ada sepatu yang berpasangan.

**Jawab:** event A = tidak ada sepatu yang berpasangan.

Cara 1:

$$P(A) = \frac{2 * C(6,6) + 2 * C(6,5) * C(1,1) + 2 * C(6,4) * C(2,2) + C(6,3) * C(3,3)}{C(12,6)}$$

$$= \frac{2 + 12 + 30 + 20}{924} = \frac{64}{924} = 0.069264$$

Cara 2:

$$P(A) = \frac{C(6,0) * C(6,6) * 2^6}{C(12,6)} = \frac{64}{924} = 0.069264$$

$C(6,0)$  = cara memilih pasangan sepatu terambil, dalam hal ini 0 karena tidak ada yang terambil

$C(6,6)$  = cara memilih jenis sepatu yang tidak berpasangan

$2^6$  yaitu cara untuk memilih dari sepatu kiri atau kanan, dari yang tidak berpasangan

$C(12,6)$  = sample space

- b. ada tepat satu pasang sepatu yg merupakan pasangan.

**Jawab:** event B= tepat satu sepatu yang berpasangan

Cara 1: Pilih 1 pasang dari 6 pasang sepatu x (pilih 4ka/ki-1ki/ka + 3ka/ki-2ki/ka + 3ka-3ki)

$$P(B) = \frac{C(6,1) * (2 * C(5,4) + 2 * C(5,3) * C(2,1) + C(5,2) * C(3,2))}{C(12,6)} = \frac{6(10 + 40 + 30)}{924}$$

$$= \frac{480}{924} = 0.51948$$

Cara 2: pilih 1 pasang dari 6 pasang x pilih 4 sepatu beda pasang dari 5 pasang tersisa x tiap pasang pilih ki/ka

$$P(B) = \frac{C(6,1) * C(5,4) * 2^4}{C(12,6)} = \frac{480}{924} = 0.51948$$

$C(6,1)$  = cara memilih pasangan sepatu terambil, pilih 1 dari 6 pasang sepatu

$C(5,4)$  = 4 sepatu yang lain tidak boleh perpasangan, jadi pilih 4 dari 5 pasangan tak terambil

$2^4$  yaitu cara untuk memilih dari sepatu kiri atau kanan, dari yang tidak berpasangan

- c. berapakah angka harapan (expected value) jumlah pasang sepatu yang terambil.

**Jawab:**

Menggunakan *indicator variable*  $X_i$ , untuk  $i=1,2,3,\dots,6$ , dimana

$$X_i = \begin{cases} 1 & \text{pasangan sepatu ke - i terambil} \\ 0 & \text{selainnya} \end{cases}$$

ditanyakan  $E[X_1+X_2+X_3+X_4+X_5+X_6]=?$

$$P(X_i = 1) = \frac{C(2,2) \times C(10,4)}{C(12,6)} = \frac{210}{924} = 0.2272727273$$

$$E[X_i] = P(X_i = 1) = 0.227$$

$$E\left[\sum_{i=1}^6 X_i\right] = 6xE[X_i] = 6 \times 0.2272727273 = 1.36363636$$

Jadi dari 6 sepatu yang terambil, angka harapan jumlah pasang sepatu adalah 1.3636

Atau bisa juga dengan cara seperti di soal nomor a,b:

$X$  = jumlah pasang sepatu yang terambil

$$P(X = 0) = 64/924 \quad (64 = C(6,0) \times C(6,6) \times 2^6, \text{ jawaban soal a})$$

$$P(X = 1) = 480/924 \quad (480 = C(6,1) \times C(5,4) \times 2^4, \text{ jawaban soal b})$$

$$P(X = 2) = 360/924 \quad (360 = C(6,2) \times C(4,2) \times 2^2)$$

$$P(X = 3) = 20/924 \quad (20 = C(6,3) \times C(3,0) \times 2^0)$$

$$E[X] = (0 \times 64 + 1 \times 480 + 2 \times 360 + 3 \times 20)/924 = 1.363636$$

3. [17] Suatu alat pendeteksi virus flu-kucing sebelum dijual secara komersil diuji coba terhadap sejumlah sukarelawan. Dari 25 orang yang benar-benar terinfeksi, 24 orang di antaranya berhasil dideteksi positif oleh alat tersebut, satu lainnya negatif. Alat itu juga diujicoba terhadap 50 orang yang benar-benar sehat dan ternyata alat tersebut hanya gagal terhadap satu orang saja (ia dideteksi positif). Alat itu kemudian dipakai di suatu populasi yang menurut suatu studi terdapat 5 orang diantara 1000 yang terinfeksi virus tsb.

- a. a Si A diperiksa menggunakan alat tersebut. Jika alat tsb menyatakan si A positif terinfeksi virus, berapakah probabilitasnya si A benar-benar terinfeksi?

**Jawaban:** Event S: terinfeksi virus, Event P: hasil pengujian positif.

Diketahui  $P(P|S) = 24/25$ ,  $P(P'|S') = 49/50$ ,  $P(S) = 5/1000$ .

Ditanyakan  $P(S|P)=?$

$$P(P|S) = 24/25 \rightarrow P(P'|S) = 1/25.$$

$$P(P'|S') = 49/50 \rightarrow P(P|S') = 1/50.$$

$$P(S) = 5/1000 \rightarrow P(S') = 995/1000.$$

$$\text{Total Probability: } P(P) = P(P|S).P(S) + P(P|S').P(S') = (24/25) \times (5/1000) + (1/50) \times (995/1000) = 0.0247$$

$$\text{Teorema Bayes: } P(S|P) = P(P|S).P(S)/P(P) = (24/25) \times (5/1000) / 0.0247 = 0.1943.$$

- b. Si B diperiksa menggunakan alat tersebut. Jika alat tsb menyatakan si B negatif tidak terinfeksi virus, berapakah probabilitasnya si B sebenarnya terinfeksi?

**Jawaban:**

Ditanyakan  $P(S|P')=?$

$$\text{Teorema Bayes: } P(S|P') = P(P'|S).P(S)/P(P') = (1-P(P|S)).P(S)/(1-P(P))$$

$$= (1-24/25) \times (5/1000) / (1-0.0247) = 0.000205$$

- c. Berapa tingkat kesalahannya alat tersebut (sehat dideteksi positif, sakit dideteksi negatif) saat digunakan disitu?

**Jawaban:**  $\text{Error} = (S \cap P') \cup (S' \cap P)$

Sehingga  $P(\text{error}) = P(S \cap P') + P(S' \cap P)$  atau  $P(\text{error}) = P(SP') + P(S'P)$

$P(SP') = P(P' | S) \cdot P(S) = (1/25) \cdot (5/1000) = 1/5000 = 0.0002$  atau

$P(S'P) = P(P | S') \cdot P(S') = (1/50) \cdot (995/1000) = 0.0199$

$P(SP') + P(S'P) = 0.0002 + 0.0199 = 0.0201 \rightarrow$  tingkat kesalahan 2%.

Banyak mhs yang menjawab  $P(\text{error}) = P(P' | S) + P(P | S') = 6\%$  dimana pada jawaban ini, tidak diperhatikan komposisi populasi yang terdiri dari 5% terinfeksi virus.

Seharusnya komposisi populasi diperhitungkan, karena di soal ada kata “di situ” 😊

4. [20] Ada dua program pembangkit angka random yang dijalankan; program pertama membangkitkan angka random 1 atau 2, sedangkan program kedua membangkitkan angka random 1,2 atau 3. Variabel acak X adalah selisih kedua angka yg dihasilkan dalam kedua eksekusi program tersebut, sedangkan Variabel Acak Y adalah jumlah kedua angka yg dihasilkan.

- a. Carilah joint PMF  $p_{xy}(x,y)$

**Jawab :**

$X=\{0,1,2\}$

$X=0 \rightarrow \{(1,1),(2,2)\}$ ,  $X=1 \rightarrow \{(1,2),(2,1),(2,3)\}$ ,  $X=2 \rightarrow \{(1,3)\}$

$Y=\{2,3,4,5\}$

$Y=2 \rightarrow \{(1,1)\}$ ,  $Y=3 \rightarrow \{(1,2),(2,1)\}$ ,  $Y=4 \rightarrow \{(1,3),(2,2)\}$ ,  $Y=5 \rightarrow \{(2,3)\}$

$X \backslash Y$	2	3	4	5	$P(X)$
0	1/6	0	1/6	0	2/6
1	0	2/6	0	1/6	3/6
2	0	0	1/6	0	1/6
$P(Y)$	1/6	2/6	2/6	1/6	

- b. Hitung probabilitas maksimal selisih nya 1 dan maksimal jumlahnya 4

**Jawab :**  $F_{xy}(1,4) = p_{xy}(0,2) + p_{xy}(0,4) + p_{xy}(1,3) = 1/6 + 1/6 + 2/6 = 4/6 = 0.667$

- c. Hitung probabilitas selisihnya adalah 2 jika diketahui jumlahnya adalah 4

**Jawab :**  $p_{x|y}(2 | 4) = p_{xy}(2,4) / p_y(4) = (1/6) / (2/6) = 1/2 = 0.5$

- d. Hitung COV(X,Y)

**Jawab :**

$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$

$E[X] = 0 \times 2/6 + 1 \times 3/6 + 2 \times 1/6 = 5/6 = 0.833$

$E[Y] = 2 \times 1/6 + 3 \times 2/6 + 4 \times 2/6 + 5 \times 1/6 = 21/6 = 3.5$

$E[XY] = 3 \times 2/6 + 5 \times 1/5 + 8 \times 1/6 = 19/6 = 3.167$

$\text{Cov}(X,Y) = 19/6 - 5/6 \times 21/6 = 9/6 = 1.5$

5. [20] Ali dan Badu janjian untuk ketemu di perpustakaan sekitar jam 12.30. Pola waktu kedatangan Ali dan Badu tidak bergantung satu sama lain. Ali menggunakan sepeda pergi ke perpustakaan dengan pola waktu kedatangan variable acak  $X$  dengan pdf waktu kedatangan  $=1/60$  antara jam 12.00-13.00. Sedangkan Badu menggunakan sepeda motor pergi ke perpustakaan dengan pola waktu kedatangan variable acak  $Y$  dengan pdf waktu kedatangan  $=1/30$  antara jam 12.15-12.45.

a. Tentukan joint pdf waktu kedatangan Ali dan Badu di perpustakaan.

**Jawab:**

$$f_{xy}(x, y) = ?$$

$$f_x(x) = \begin{cases} \frac{1}{60} & , 0 \leq x \leq 60 \\ 0 & , \text{selainnya} \end{cases}$$

$$f_y(y) = \begin{cases} \frac{1}{30} & , 15 \leq y \leq 45 \\ 0 & , \text{selainnya} \end{cases}$$

$$f_{xy}(x, y) = f_x(x)f_y(y), \text{ karena } X \text{ dan } Y \text{ tidak bergantung satu sama lain (idenpendent)}$$

$$f_{xy}(x, y) = \begin{cases} f_x(x)f_y(y) = \frac{1}{60} \frac{1}{30} = \frac{1}{1800} & , 0 \leq x \leq 60 \text{ dan } 15 \leq y \leq 45 \\ 0 & , \text{selainnya} \end{cases}$$

b. Berapakah probabilitas Badu datang duluan di perpustakaan.

Ditanyakan:  $P(Y < X) = ?$

Jawab :

Mencari range  $X$  dan  $Y$  dengan constraint :

$$0 \leq x \leq 60, 15 \leq y \leq 45 \text{ dan } y < x$$

dan didapat range nya :  $15 \leq y \leq 45$  dan  $y \leq x \leq 60$

Maka,

$$\begin{aligned} P(Y < X) &= \int_{y=15}^{45} \int_{x=y}^{60} \frac{1}{1800} dx dy = \int_{y=15}^{45} \frac{60-y}{1800} dy = \frac{60(45-15) - 0.5(45^2 - 15^2)}{1800} \\ &= \frac{1800 - 900}{1800} = \frac{900}{1800} = 0.5 \end{aligned}$$

c. Berapakah probabilitas yang datang duluan akan menunggu yang lain kurang dari 5 menit ?

Ditanyakan:  $P(|X - Y| < 5) = ?$

Jawab :

Mencari range  $X$  dan  $Y$  dengan constraint :

$$0 \leq x \leq 60, 15 \leq y \leq 45 \text{ dan } |x - y| < 5 \Leftrightarrow 0 \leq x \leq 60, 15 \leq y \leq 45, x - y < 5 \text{ dan } x - y > -5$$

dan didapat range nya :  $15 \leq y \leq 45$  dan  $y - 5 \leq x \leq y + 5$

Maka,

$$P(|X - Y| < 5) = \int_{y=15}^{45} \int_{x=y-5}^{y+5} \frac{1}{1800} dx dy = \int_{y=15}^{45} \frac{(y+5) - (y-5)}{1800} dy = \int_{y=15}^{45} \frac{1}{180} dy = \frac{(45-15)}{180} = \frac{30}{180} = \frac{1}{6}$$