

**Pengantar Analisis Data:  
Studi kasus dengan data  
iSIKHNAS**

**Juni 2014**

**Brendan Cowled, Angus Cameron, dan Etih Sudarnika**

# Isi

<b>Kata Pengantar .....</b>	<b>4</b>
Prasyarat.....	4
Tujuan pelatihan .....	4
Pendekatan pembelajaran .....	4
<b>Ikhtisar analisis data .....</b>	<b>6</b>
Relevansi analisis data dengan kebijakan kesehatan hewan .....	6
Langkah-langkah analisis data.....	7
1. Tujuan .....	7
2. Manajemen data .....	7
3. Deskripsi data.....	7
4. Pengujian hipotesis .....	7
Catatan: Konsep kunci mengenai penarikan sampel.....	8
<b>Studi kasus 1: Pengukuran kinerja pelayanan veteriner .....</b>	<b>12</b>
Pengantar ke studi kasus.....	12
Keahlian yang dikembangkan dalam studi kasus ini .....	12
Daftar file yang digunakan untuk studi kasus.....	13
Langkah-langkah dalam analisis data studi kasus 1 (latihan).....	14
Langkah 1: Tujuan.....	14
Langkah 2: Manajemen data .....	17
Catatan: Konsep kunci mengenai kesalahan data .....	20
Langkah 3: Deskripsi data .....	26
Catatan: Konsep kunci mengenai statistika deskriptif.....	26
Langkah 4: Pengujian hipotesis statistika .....	32
Rangkuman studi kasus 1.....	32
<b>Studi Kasus 2: Prevalensi musiman diare pada sapi .....</b>	<b>34</b>
Pengantar ke studi kasus.....	34
Keahlian yang dikembangkan dalam studi kasus ini .....	34
Daftar file yang digunakan untuk studi kasus.....	34
Langkah-langkah dalam analisis data studi kasus 2 (latihan).....	35
Langkah 1: Tujuan.....	35
Langkah 2: Manajemen data .....	35
Langkah 3: Deskripsi data.....	39

Catatan: Konsep kunci mengenai pengukuran penyakit dan asosiasi .....	44
Langkah 4: Pengujian hipotesis statistika .....	48
Rangkuman studi kasus 2.....	50
<b>Studi kasus 3: Swasembada daging sapi (berdasarkan statistika pemotongan sapi) .....</b>	<b>51</b>
Pengantar ke studi kasus.....	51
Keahlian yang dikembangkan dalam studi kasus ini .....	51
Daftar file yang digunakan untuk studi kasus.....	51
Catatan: Konsep kunci mengenai pengujian hipotesis dan studi kasus 3 .....	52
Langkah-langkah analisis data studi kasus 3 (Latihan).....	58
Langkah 1: Tujuan.....	58
Langkah 2: Manajemen data .....	59
Langkah 3: Deskripsi data .....	59
Langkah 4: Pengujian hipotesis statistika .....	60
Rangkuman studi kasus 3.....	62
<b>Catatan Penutup: .....</b>	<b>63</b>
<b>Lampiran 1: Latihan lanjutan menggunakan R. ....</b>	<b>64</b>
Pengantar .....	Error! Bookmark not defined.
Perangkat lunak yang perlu diunduh .....	64
Latihan .....	65
Pengembangan lanjutan R dan keterampilan statistika.....	65
<b>Lampiran 2: Kunci jawaban latihan66</b>	
<b>Daftar Pustaka .....</b>	<b>Error! Bookmark not defined.</b>

# Kata Pengantar

## Prasyarat

Pelatihan ini adalah untuk para dokter hewan yang bekerja dalam sistem kesehatan hewan di Indonesia. Para peserta diasumsikan telah menyelesaikan modul-modul pelatihan Excel, epidemiologi lapangan, dan surveilans sebelum mengikuti pelatihan ini. Meskipun demikian, staf lain yang relevan juga dapat mengikuti pelatihan ini. Misalnya, jika Anda telah mengetahui tentang epidemiologi dan cara menggunakan Excel, maka prasyarat tadi tidaklah diperlukan.

## Tujuan pelatihan

Tujuan yang lebih luas dari pelatihan ini adalah untuk membantu para peserta didalam membuat kebijakan kesehatan hewan yang berdasarkan bukti. Hal ini dapat membantu mereka didalam meningkatkan produksi serta kesehatan ternak di Indonesia.

Untuk dapat melakukan hal itu, para peserta harus dapat mengakses, memahami, serta menganalisis informasi mengenai kesehatan hewan di Indonesia. Untunglah pada saat ini di Indonesia telah tersedia suatu sistem informasi kesehatan hewan yang dapat diakses oleh staf kesehatan hewan di Indonesia, yaitu iSIKHNAS. Sistem informasi kesehatan hewan ini saat ini merupakan yang terbaik di dunia. Sistem ini memberikan sejumlah besar informasi berkualitas tinggi (data<sup>1</sup>) yang dapat digunakan untuk pengambilan keputusan yang baik dalam kesehatan hewan. Oleh karena itu, maka tujuan pelatihan ini adalah memberikan pengetahuan kepada para peserta mengenai cara mengunduh, memahami, mengevaluasi, menganalisis, dan menginterpretasikan data iSIKHNAS.

## Pendekatan pembelajaran

Di Ddlam pelatihan ini, para pesera akan diajarkan metode analisis data dengan cara menganalisis data iSIKHNAS yang sesungguhnya dalam bentuk studi kasus. Ada tiga studi kasus yang disajikan, dan terdapat latihan-latihan yang harus dikerjakan oleh peserta. Pada setiap pertanyaan latihan, diberikan ruang kosong untuk para peserta menuliskan jawabannya. Kunci jawaban disediakan pada Lampiran 2. Akan tetapi sebaiknya Anda menjawab terlebih dahulu pertanyaan-pertanyaan pada latihan tersebut sebelum membaca jawabannya.

Terkadang akan disajikan catatan atau konsep sebelum atau pada saat studi kasus, guna mendukung proses pembelajaran. Catatan ini akan diberi latar belakang warna abu-abu agar Anda dapat membedakan bagian catatan dengan latihan.

---

<sup>1</sup>Definisi data dalam kamus Oxford adalah: 'Fakta serta statistik yang dikumpulkan bagi tujuan referensi atau analisis'

Pelatihan ini sangat relevan dan dapat diterapkan bagi staf kesehatan hewan di Indonesia. Tiga studi kasus difokuskan pada: 1) kajian mengenai layanan veteriner (kinerja staf), 2) manajemen penyakit (diare pada sapi), dan produksi ternak (swasembada daging sapi).

Data yang digunakan didalam pelatihan ini diunduh pada awal 2014, ketika iSIKHNAS baru beroperasi selama kira-kira satu tahun di sebagian kecil wilayah di Indonesia. Data awal ini digunakan agar kami dapat menyediakan kunci jawaban latihan. Pada saat pelatihan berlangsung, Anda mungkin ingin mengunduh data yang lebih baru serta lebih lengkap untuk tujuan analisis. Untuk itu mohon diperhatikan bahwa jika Anda menggunakan data yang berbeda, tentu hasil perhitungan Anda akan berbeda dengan kunci jawaban yang tersedia.

Sifat alamiah data yang senantiasa berubah seiring waktu menyebabkan tidak ada kesimpulan nyata yang dapat ditarik dari hasil analisis data yang dilakukan selama pelatihan ini. Analisis data yang dilakukan di dalam pelatihan ini hanya mengajarkan dan memperagakan cara melakukan analisis data dan menarik kesimpulan, untuk nanti diterapkan pada data sesungguhnya yang akan dianalisis. . Sejalan dengan waktu, akan tersedia data yang lebih lengkap. Pada saat itulah Indonesia akan dapat melakukan analisis yang lebih lengkap serta lebih akurat.

Pelatihan ini diberikan dalam bentuk Excel. Bentuk ini dipilih karena ekonomis, tersedia secara luas dan dikuasai oleh kebanyakan staf pelayanan veteriner di Indonesia.. Analisis dalam bentuk Excel akan mendukung staf untuk menarik kesimpulan yang bermakna mengenai data iSIKHNAS.

Jika Anda bermaksud melakukan banyak kerja statistik yang penting, Anda akan perlu belajar menggunakan paket statistik yang lengkap disamping program Excel.Oleh karena itu, kami juga menyertakan soal ekstra bagi para peserta yang ingin memperluas pengetahuan mereka di luar Excel. Soal ini disajikan dalam Lampiran 1, yang memberikan pengenalan mengenai R, suatu paket statistik *online* gratis. R adalah salah satu paket perangkat lunak yang paling bermanfaat di dunia. Program ini gratis dan dapat diunduh dari internet. Lampiran 1 mengulangi penjelasan studi kasus 1 dengan menggunakan program R.

Program analisis data lainnya yang diperkenalkan di dalam panduan ini adalah EpiTools, yaitu suatu program yang sangat membantu di dalam analisis data epidemiologi yang dikembangkan oleh AusVet Animal Health Services. Untuk masuk ke situs EpiTools, Anda dapat mengaksesnya melalui <http://epitools.ausvet.com.au/>.

Selama pelatihan berlangsung, akan digunakan serangkaian video pendek untuk membantu Anda memahami pengerjaan latihan-latihan yang disediakan. Video-video ini dapat ditayangkan dengan menggunakan berbagai perangkat lunak, termasuk Windows Media Player.

## **Ikhtisar analisis data**

### **Relevansi analisis data dengan kebijakan kesehatan hewan**

Agar dapat menyusun kebijakan kesehatan hewan yang baik, seorang dokter hewan perlu memahami situasi kesehatan hewan di tempat mereka bekerja. Misalnya, Apa saja penyakit yang ada di wilayah tersebut? Atau, apa yang menyebabkan suatu penyakit dan apakah berbagai upaya intervensi yang dilakukan berjalan dengan baik di lapangan?

Untuk memperoleh pemahaman ini, berbagai upaya dapat dilakukan oleh seorang dokter hewan. Misalnya dengan memperkirakan situasi yang ada atau menyusun asumsi berdasarkan pengalaman mereka sendiri. Pada umumnya ini cara seperti ini tidak menghasilkan keputusan yang baik. Keputusan yang diambil dengan cara mengira-ngira, tentu saja dibuat tidak berdasarkan informasi faktual. Keputusan yang diambil berdasarkan pengalaman mereka sendiri mungkin bermanfaat, namun secara umum keputusan-keputusan tersebut juga berdasarkan kepada pengalaman yang terbatas. Misalnya keputusan yang diambil berdasarkan kepada pengalaman dari satu dokter hewan saja, walaupun dokter hewan tersebut mungkin sangat berpengalaman.

Suatu cara yang lebih baik bagi dokter hewan untuk mengambil keputusan adalah berdasarkan informasi yang mencerminkan situasi kesehatan hewan yang lebih luas. Informasi ini dapat diperoleh melalui beberapa cara, seperti melalui data kesehatan hewan, publikasi, buku teks, serta berbagai laporan. Pada saat ini di Indonesia telah dikembangkan suatu sistem informasi yang untuk mengumpulkan informasi (data) kesehatan hewan dari seluruh Indonesia. Data ini disimpan dalam iSIKHNAS dan dapat digunakan untuk membantu dalam pengambilan keputusan. Pengambilan keputusan yang baik tentunya hanya dapat dilakukan jika data iSIKHNAS tersebut telah dianalisis serta ditafsirkan dengan baik. Oleh karena itu, maka pelatihan ini bertujuan untuk membantu Anda didalam melakukan analisis data serta menarik kesimpulan dari data iSIKHNAS.

Setelah mengikuti pelatihan ini, Anda diharapkan untu segera mengunduh dan menganalisis data iSIKHNAS anda sendiri.. Dengan demikian makapemahaman Anda akan semakin mantap, keterampilan Anda akan semakin meningkat, serta kemampuan Anda di dalam pengambilan keputusan berdasarkan bukti data juga akan diperkuat. Karena itu, sisihkanlah satu hari di minggu depan untuk melakukan analisis data Anda sendiri, dengan menggunakan keterampilan yang telah Anda pelajari. Kemudian, lakukan analisis data iSIKHNAS secara berkala. Sejalan dengan waktu, keterampilan Anda di dalam analisis data akan semakin meningkat.

## **Langkah-langkah analisis data.**

Terdapat beberapa langkah untuk melakukan analisis dan interpretasi data, yaitu (1) menentukan tujuan analisis data, (2) manajemen data, (3) deskripsi data, dan (4) pengujian hipotesis. Setiap langkah ini akan dibahas secara singkat dalam panduan ini. Selanjutnya, keempat langkah tersebut akan digunakan pada studi kasus-studi kasus yang dipelajari.

### **Langkah dasar didalam analisis data:**

#### **1. Tujuan**

Tujuan harus dinyatakan yang jelas serta ringkas. Misalnya, berapa prevalensi diare pada sapi pada tahun 2014? Tujuan akan memfokuskan pekerjaan Anda dan membantu Anda mencari sumber data guna menjawab tujuan tersebut. Kemudian, tujuan tersebut diterjemahkan ke dalam bentuk hipotesis untuk selanjutnya diuji.

#### **2. Manajemen data**

Penting bagi seorang dokter hewan untuk mengetahui cara mengakses serta menggunakan data iSIKHNAS. Kami akan membantu Anda mengunduh data iSIKHNAS. Kami juga akan membantu Anda menyimpan, memeriksa kesalahan, membuat, serta mengevaluasi data iSIKHNAS.

#### **3. Deskripsi data**

Langkah berikutnya adalah mendeskripsikan data. Tujuan deskripsi data adalah: (1) memeriksa data untuk melihat kesalahan yang mungkin ada, (2) memahami struktur serta sifat alamiah data, dan (3) memahami hubungan antara berbagai bagian data. Dalam langkah ini, digunakan rangkuman variabel tunggal, rangkuman hubungan antara variabel, serta plot. Deskripsi data membantu Anda didalam memulai pengujian hipotesis (langkah 4).

#### **4. Pengujian hipotesis**

Untuk menjawab tujuan analisis data, maka disusunlah pertanyaan (atau hipotesis) yang dapat diuji, yang diturunkan dari tujuan Anda. Kemudian hipotesis ini diuji menggunakan uji statistik yang sesuai, dan setelah itu Anda dapat memutuskan apakah data mendukung gagasan Anda atau tidak.

## Catatan: Konsep kunci mengenai penarikan sampel

### 1. Penarikan sampel

Misalnya kita ingin mengetahui proporsi sapi yang sero-positif terhadap bruselosis di suatu wilayah. Untuk tujuan tersebut, kita dapat menguji semua sapi yang ada di wilayah tersebut terhadap bruselosis. Jika kita melakukan pengukuran pada semua hewan yang ada di dalam suatu populasi, maka hal ini disebut sebagai *sensus*. Biasanya terlalu sulit untuk melakukan suatu sensus karena terlalu luas dan jumlahnya besar. Cara lain yang dapat dilakukan adalah dengan cara menguji sejumlah kecil hewan dari populasi, yang disebut sebagai sampel.

Jika sampel tersebut mewakili populasi, hasil yang kita peroleh dari sampel dapat diterapkan bagi populasi. Misalnya jika sampel memiliki *seroprevalence* sebesar 20%, maka kita dapat menyimpulkan bahwa 20% dari keseluruhan populasi akan memberikan hasil positif.

Seberapa baik hasil sampel mencerminkan keadaan populasi yang sebenarnya, bergantung pada bias dan variabilitas penarikan sampel.

Istilah *bias* digunakan untuk menjelaskan suatu situasi ketika hasil sampel tidak mencerminkan apa yang terjadi didalam populasi. Hal ini sering diakibatkan oleh pemilihan individu di dalam penarikan sampel yang kurang baik. Dengan demikian, kita harus mempertimbangkan dengan baik cara pemilihan sampel guna memastikan bahwa hewan yang dipilih sebagai sampel dapat mewakili populasi target.

Misalnya didalam penarikan sampel, hewan yang anda pilih untuk pengujian adalah hanya hewan yang terbesar, sehat, dan muda. Jika hasil uji menunjukkan bahwa semua hewan ini negatif terhadap bruselosis, dapatkah Anda berasumsi bahwa prevalensi brucellosis adalah 0%? Jawabannya adalah tidak. Mungkin terdapat hewan lain dalam populasi (hewan yang lebih tua dan sakit) yang telah terinfeksi, dan hasil dari sampel hewan yang sehat serta muda tidak mewakili keseluruhan populasi. Pemilihan sampel yang buruk ini dapat menyebabkan hasil studi yang *bias*. Pengambilan sampel harus dilakukan secara representatif. Ini berarti bahwa baik hewan yang tua maupun yang muda disertakan dalam sampel tersebut.

**Variabilitas pengambilan sampel** berarti bahwa setiap kali sampel diambil dari populasi, hasil dari sampel akan berbeda. Misalnya kita mengambil lima sampel yang terpisah, yaitu pengambilan 20 sapi dari tiap wilayah. Juga Diasumsikan bahwa *seroprevalence* bruselosis yang sebenarnya, namun belum diketahuialah 20%. Kita akan memperoleh nilai dugaan *seroprevalence* yang berbeda dari lima sampel yang kita ambil, misalnya 21%, 18%, 22%, 25%, dan 14%. Semua nilai penduga tersebut tidak ada yang benar-benar tepat, namun semuanya mendekati 20%. Melalui penggunaan statistika, kita dapat meningkatkan estimasi kita dari hanya satu sampel dengan menunjukkan seberapa yakin kita terhadap nilai estimasi tersebut.

Jika Anda menggunakan penarikan sampel yang representatif dan memperhitungkan variabilitas pengambilan sampel, maka penarikan sampel dapat memberikan informasi yang nyaris sama dengan yang akan Anda peroleh dari sensus, dengan biaya yang jauh lebih rendah. Misalnya dari suatu penarikan contoh yang representatif dan perhitungan statistika yang benar,

diperoleh *seroprevalence* adalah 19%, dan kita sangat percaya bahwa nilai yang sebenarnya dari *seroprevalence* terletak pada selang antara 17% dan 23%. Jika suatu estimasi yang memiliki variasi sebesar 3% dapat diterima (baik 3% lebih besar atau lebih kecil), maka Anda dapat menghemat banyak uang melalui penarikan sampel daripada melakukan sensus. Dalam hal ini, penggunaan statistika sangatlah penting. Dalam pelatihan ini, kami akan memperlihatkan cara menggunakan statistik semacam ini dengan benar.

## 2. Gugus data dan variabel

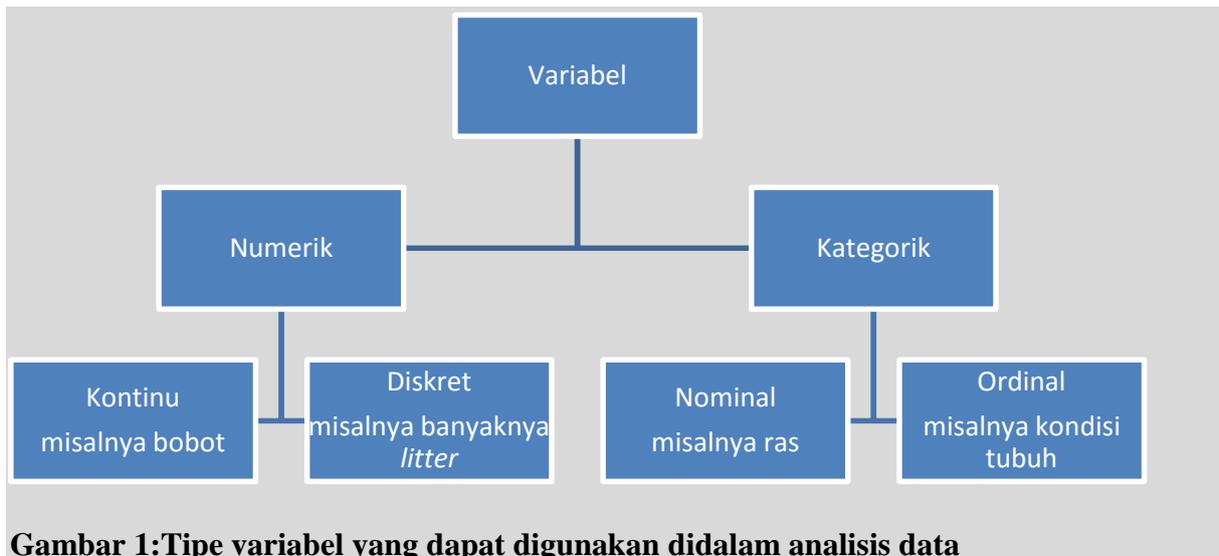
Misalnya kita ingin mengukur bobot hidup (kilogram) sejumlah besar sapi bali dewasa. Hasil pengukuran bobot hidup masing-masing sapi tersebut kemudian dimasukkan ke dalam kolom Excel, dengan masing bobot hidup per hewandicatat per baris dalam satu kolom dengan judul kolom ***bobot***. Dalam epidemiologi dan statistika, kita dapat menyebut kolom pengukuran bobot ini sebagai suatu ***variabel*** dengan nama variabel ***bobot***.

Gugus data yang sama dapat mengikutsertakan variabel tambahan yang diukur pada hewan yang sama. Misalnya, kita dapat mencatat nomor peneng kuping atau bentuk identifikasi hewan lainnya (***id hewan***); jenis kelamin hewan dapat dicatat sebagai pejantan, jantan muda, betina (***jenis kelamin***); sementara umur hewan dapat dicatat dalam kategori umum seperti pedet, sapi lepas-sapih, dewasa (***kelas umur***). Beberapa dari variabel ini dapat berisi data numerik (nomor peneng kuping dan bobot hidup), sementara yang lainnya dapat berisi catatan berupa teks. Tiap baris akan berisi catatan yang dari satu hewan yang sama. Secara bersama-sama, berbagai variabel ini membentuk satu gugus data.

Variabel yang berbeda (misalnya kelas umur, bobot, jenis kelamin, ID hewan) dapat diklasifikasikan agar dapat lebih mudah dipahami. Variabel dapat bersifat numerik atau kategorik.

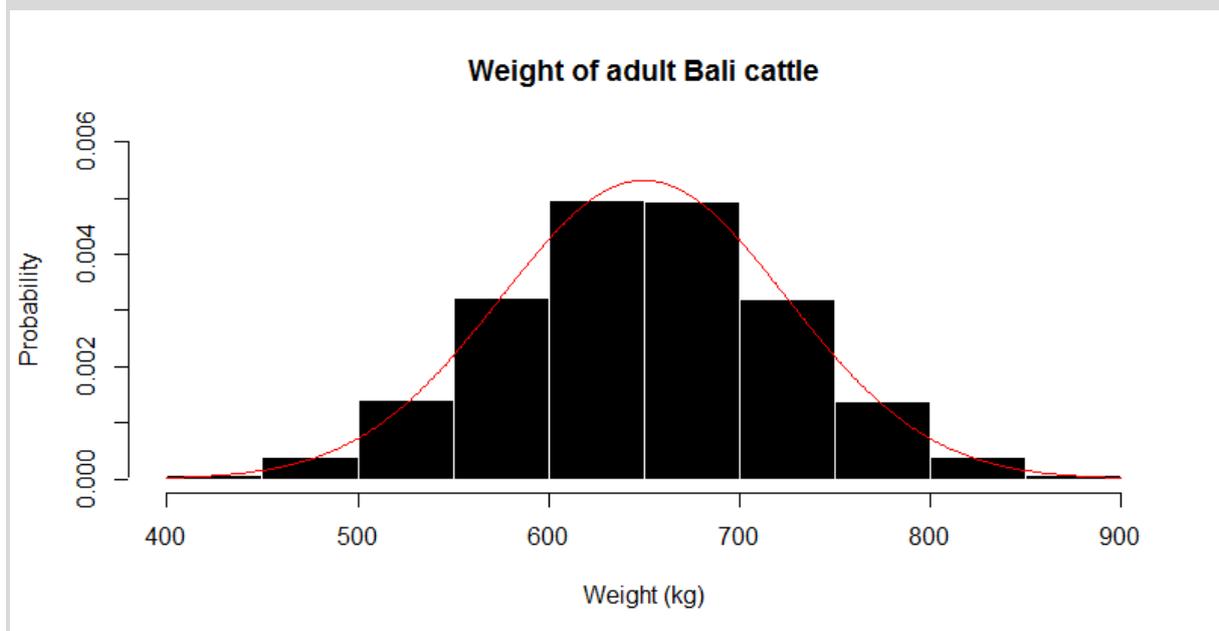
Variabel numerik diukur menggunakan bilangan nyata dan dapat bersifat kontinu atau diskret. Variabel kontinu dapat berupa suatu angka apa pun dari suatu gugus bilangan nyata, misalnya bobot. Variabel diskret hanya dapat berupa suatu angka tertentu (biasanya bilangan cacah), misalnya banyaknya *litter*.

Variabel kategorik dapat berupa data nominal atau ordinal. Nilai Suatu variabel nominal hanya 'nama', tidak menggambarkan besaran nilai pengukuran tertentu. Misalnya untuk variabel jenis kelamin, kita memberi kode 1=jantan, 2=betina. Pada variabel tersebut tidak menggambarkan bahwa betina adalah 2 kali jantan. Variabel ordinal hampir sama dengan variabel nominal, tetapi variabel ordinal memiliki suatu urutan tertentu, meskipun intervalnya tidak konsisten. Misalnya umur diklasifikasikan menjadi 1= pedet, 2= anak sapi lepas-sapih, dan 3= dewasa.. Penjelasan jenis-jenis variabel ini secara ringkas dapat dilihat pada Gambar 1.



### 3. Sebaran

Misalnya kita memiliki Gugus data yang berisi sejumlah baris yang berisi data bobot hidup (kg) sapi Bali dewasa. Untuk memplot data seperti itu, biasanya digunakan histogram, seperti yang ditunjukkan pada gambar 2.



**Gambar 2: Fungsi sebaran peluang (garis merah) dan histogram (batang hitam) bobot hidup sapi bali dewasa**

Histogram ini memberikan nilai peluang<sup>2</sup> pada sumbu vertikal. Data diurutkan dan diklasifikasikan dalam bentuk batang- batang dalam sumbu horisontal. Dalam hal ini, setiap batang memiliki selang 50 kg (400-450, 451-400, 501-550, dan seterusnya).

<sup>2</sup>Peluang adalah suatu ukuran yang menggambarkan kemungkinan munculnya suatu kejadian. Peluang diberi nilai antara 0 (tidak akan muncul) dan 1 (pasti akan terjadi). Jadi jika peluang (p) = 0,5, maka berarti suatu

Histogram di atas memperlihatkan peluang seekor hewan memiliki bobot hidup pada setiap selang tersebut. Histogram tersebut memperlihatkan pola penyebaran bobot hidup sapi bali dalam sampel kita, atau *sebaran* dari sampel kita. Dari histogram tersebut kita dapat melihat bahwa bobot terkecil adalah antara 400-450 kg, bobot yang paling umum adalah di antara 600-700 kg, dan bobot yang tertinggi adalah antara 850-900 kg. Juga dapat dilihat bahwa bobot sapi yang lebih rendah dan lebih tinggi jarang ditemukan.

Histogram di atas menunjukkan bentuk yang simetris (*bell-shape*). Bentuk lain yang mungkin ditunjukkan oleh sebaran data adalah menjulur ke kanan (menjulur positif) jika ekor kanan lebih panjang dari ekor kiri, atau menjulur ke kiri (menjulur negatif) jika terjadi sebaliknya. Pada sebaran yang simetris, nilai tengah dan mediannya terletak pada posisi yang sama. Jika menjulur ke kanan, nilai tengah akan lebih besar dari pada mediannya. Artinya, sebagian besar data memiliki nilai-nilai yang kecil di ekor kiri. Keadaan sebaliknya terjadi jika sebaran data menjulur ke kiri.

Pada histogram di atas, garis merah memperlihatkan sebaran normal teoritis. Sebaran normal adalah bentuk sebaran yang umum terjadi di alam. Sebaran ini memiliki satu puncak pusat dengan bentuk simetris. Sebaran normal sering disebut sebagai kurva lonceng (*bell curve*) karena bentuknya seperti lonceng. Histogram yang kita peroleh di atas menunjukkan bahwa data ini menyebar normal, karena polanya tampak seperti kurva normal berbentuk lonceng.

Dalam statistika, kita dapat menggunakan dua jenis pengukuran ringkasan data sederhana untuk menggambarkan suatu sebaran normal, yaitu nilai rata-rata (pusat dari sebaran), dan ragam atau simpangan baku (pengukuran penyebaran data di sekitar nilai rata-rata).

Banyak uji statistika yang didasarkan pada sebaran normal. Jika Anda mengambil sampel cukup besar dan Anda melakukan pendugaan suatu statistik yang umum, maka terdapat banyak pendekatan statistika yang dapat Anda gunakan, karena Anda mengsumsikan bahwa statistik yang Anda perhitungkan ini menyebar normal. Kita akan menggunakan asumsi ini kemudian.

Hal penting yang harus diketahui adalah bahwa terdapat banyak tipe sebaran data selain sebaran normal. Beberapa sebaran data yang umum digunakan didalam analisis data kesehatan hewan adalah sebaran binomial, sebaran Poisson, sebaran khi-kuadrat, dan sebaran logistik.

---

kejadian memiliki kemungkinan sama untuk muncul dan tidak muncul. Jika  $p = 0,999$  ini berarti suatu kejadian memiliki kemungkinan besar akan muncul. Jika  $p = 0,001$  ini berarti suatu kejadian memiliki kemungkinan kecil untuk muncul.

# **Studi kasus 1: Pengukuran kinerja pelayanan veteriner**

## **Pengantar ke studi kasus**

Dalam studi kasus 1 ini, akan dijelaskan mengenai analisis data numerik. Dalam hal ini, variabel yang akan dianalisis adalah variabel waktu tunggu antara saat pelaporan dan saat petugas melakukan investigasi, yang diukur dalam satuan jam. Variabel ini merupakan variabel yang bersifat kontinu.

Dokter hewan senior di tingkat lokal memiliki kepentingan untuk dapat mengkaji kinerja pelayanan veteriner setempat. Suatu tugas kunci pelayanan veteriner di daerah adalah menyelidiki sindrom penyakit ternak yang dapat mengindikasikan munculnya wabah penyakit yang penting.

Salah satu cara untuk melakukan tinjauan kinerja terhadap layanan veteriner setempat adalah dengan melihat kecepatan dilakukannya investigasi terhadap sindrom penyakit ternak yang penting. Kita akan menggunakan contoh ini untuk studi kasus 1.

Data iSIKHNAS yang akan kita gunakan mencatat kemunculan sejumlah sindrom prioritas, seperti:

- Keguguran dan pembengkakan sendi
- Menggigit dan perubahan perilaku
- Demam pada babi
- Pincang, berliur, dan lesi
- Kematian unggas
- Kematian mendadak

Semua ini dirancang untuk mendukung deteksi awal dari penyakit penting yaitu rabies, penyakit-penyakit vesikuler seperti PMK (Penyakit Mulut dan Kuku), HPAI, brucellosis, dan antraks.

Data yang digunakan dalam pelatihan ini diunduh pada awal 2014, yaitu ketika iSIKHNAS telah beroperasi selama satu tahun dengan wilayah cakupan yang masih terbatas. Dengan tidak lengkapnya data yang kita gunakan dalam latihan ini, berarti tidak ada kesimpulan nyata yang dapat ditarik terkait hasil analisis data yang dilaksanakan pada pelatihan ini. Namun, seiring waktu data yang lebih lengkap akan tersedia sehingga petugas kesehatan hewan di seluruh Indonesia dapat melakukan analisis yang lebih baik.

## **Keahlian yang dikembangkan dalam studi kasus ini**

- Menyusun tujuan analisis data
- Mengunduh data iSIKHNAS
- Menyimpan dan menjaga data asli

- Mengevaluasi data (memahami struktur gugus data, mengidentifikasi kesalahan, memahami bias)
- Membuat data baru
- Deskripsi data numerik (ukuran pemusatan, ukuran penyebaran, pembuatan plot)

### **Daftar file yang digunakan untuk studi kasus**

#### *Data:*

'isikhnas\_priority disease syndromes\_March\_2013.csv'

#### *Video:*

- Case study 1\_data download.avi
- Case study 1\_exercise 3\_missing data.avi
- Case study 1\_exercise 5\_creating new data.avi
- Case study 1\_exercise 6\_descriptive statistics.avi
- Case study 1\_exercise 6b\_descriptive statistics without tool pack.avi
- Case study 1\_exercise 7\_plotting histograms.avi

## Langkah-langkah dalam analisis data studi kasus 1 (latihan)

Kita akan mengerjakan langkah-langkah analisis data untuk studi kasus ini.

### Langkah 1: Tujuan

Sebelum bekerja menggunakan data, kita harus memiliki gambaran mengenai pertanyaan yang ingin kita jawab—inilah yang kita sebut sebagai tujuan. Dalam studi kasus ini, tujuan umum kita adalah untuk melakukan kajian terhadap pelayanan veteriner dalam hal kecepatan dalam menginvestigasi laporan-laporan sindrom prioritas. Setelah itu kita akan mempertajam tujuan umum ini dalam latihan (latihan 2).

Buka data iSIKHNAS (isikhnas\_priority disease syndromes\_March\_2013.csv) dan teliti data tersebut. Gambar 3 merupakan tangkapan layar dari lembar kerja sindrom prioritas, dan ini serupa dengan apa yang akan Anda lihat pada layar komputer Anda. Latihan 1 difokuskan pada pemahaman data. Latihan 2 difokuskan pada penyusunan suatu tujuan yang lebih berguna. Selesaikan latihan 1 terlebih dahulu dan kemudian lanjutkan dengan latihan 2.

#### Latihan 1: Soal (menelaah data)

Bentuk kelompok yang terdiri atas 2-4 orang dan lakukanlah latihan berikut ini. Tuliskan jawaban Anda dalam kotak di bawah ini.

**Teliti dengan baik kolom *worksheet* dan data yang terdapat didalam setiap kolom.**

Apa arti dari setiap kolom?

Berapa banyak laporan sindrom prioritas yang ada dalam *worksheet*?

**Pikirkan tujuan dari analisis yang ingin kita lakukan.**

Kolom data mana yang menurut Anda paling berguna untuk memenuhi tujuan kita? Sebagai petunjuk, dua kolom penting diberi warna berbeda dalam Gambar 3.

**Presentasikan hasil temuan Anda ke kelompok besar.**

#### Latihan 1: Jawaban (menelaah data)

Apa arti dari tiap kolom?

Berapa banyak laporan sindrom prioritas yang ada dalam lembar kerja?

Kolom data mana yang Anda anggap paling bermanfaat untuk menjawab tujuan kita? Sebagai petunjuk, dua kolom penting diberi warna berbeda dalam Gambar 3.

**Kunci jawaban disediakan dalam Lampiran 2. Silakan tuliskan jawaban Anda sebelum memeriksa jawaban di Lampiran 2. Ini akan membantu proses Anda belajar.**

Setelah Anda meneliti data dan memahaminya, selanjutnya kita perlu menyusun tujuan yang lebih terperinci serta bermanfaat bagi analisis kita. Inilah tujuan Latihan 2.

**Latihan 2: Soal (menguraikan tujuan secara terperinci)**

**Tentukan suatu tujuan analisis yang lebih terperinci.**

Petunjuk: apa perbedaan yang diperlihatkan di antara dua kolom tanggal yang disorot dalam Gambar 3? Gunakan ini untuk menyusun tujuan spesifik dari analisis data.

**Tiap Kelompok memaparkan kepada seluruh kelas mengenai tujuan analisis yang telah dipertajam tersebut.**

**Latihan 2: Jawaban (suatu tujuan spesifik)**

Tuliskan apa yang menurut Anda merupakan tujuan yang lebih spesifik dari analisis data Anda.

isikhnas\_priority disease syndromes\_March\_2013.csv - Microsoft Excel

ID	Tanggal laporan	Jumlah kasus	Syndroma	Nama pengirim	Provinsi	Kabupaten	Kecematar	Desa	Tanggal diinvestigasi	Staf teknis dinas	Tipe investigasi	Diagnosa sementara
28241	10/03/2014	1	Biting and behaviour change	Jaswarmen	Sumatera Barat				10/03/2014	Jaswarmen	Kunjungan	Rabies
28233	10/03/2014	1	Biting and behaviour change	Syafri	Sumatera Barat				10/03/2014	Gusmedi	Kunjungan	Rabies
28225	10/03/2014	1	Biting and behaviour change	Syafri	Sumatera Barat				10/03/2014	Dedrinaldi	Kunjungan	Rabies
28217	5/03/2014	1	Limping, drooling and vesicles	Setiawan Destika	Sumatera Barat							
28209	5/03/2014	1	Limping, drooling and vesicles	Setiawan Destika	Sumatera Barat							
28175	3/03/2014	1	Biting and behaviour change	Anton Firdaus	Sumatera Barat							
28167	3/03/2014	1	Biting and behaviour change	Lila Susanti	Sumatera Barat				3/03/2014	Lila Susanti	Kunjungan	Rabies
28159	2/03/2014	1	Biting and behaviour change	Riswarman	Sumatera Barat				2/03/2014	Riswarman	Kunjungan	Rabies
28142	28/02/2014	1	Biting and behaviour change	Wellya Roza	Sumatera Barat				28/02/2014	Wellya Roza	Kunjungan	Rabies
28134	28/02/2014	5	Abortion or swollen joints	Rohmat Jumhana	Jawa Barat			Sukaratu	28/02/2014	Rohmat Jumhana	Kunjungan	Brucellosis
28126	28/02/2014	5	Abortion or swollen joints	Rohmat Jumhana	Jawa Barat			Sukaratu	28/02/2014	Rohmat Jumhana	Kunjungan	Brucellosis
28118	27/02/2014	28	Poultry mortality	Rohmat Jumhana	Jawa Barat				27/02/2014	Rohmat Jumhana	Kunjungan	Avian Influenza (Highly pathogenic)
28100	27/02/2014	28	Poultry mortality	Rohmat Jumhana	Jawa Barat			Sukaratu	27/02/2014	Rohmat Jumhana	Kunjungan	Avian Influenza (Highly pathogenic)
28092	25/02/2014	1	Biting and behaviour change	Muhammad Rijal	Sumatera Barat				26/02/2014	Sri Hilmayeni Tri Putri	Telepon	Rabies
28076	24/02/2014	1	Biting and behaviour change	Gusmedi	Sumatera Barat				24/02/2014	Gusmedi	Kunjungan	Rabies
28050	23/02/2014	10	Poultry mortality	ARIFIN	Sulawesi Barat							
28027	22/02/2014	64	Poultry mortality	Iwan Setiawan	Jawa Barat				22/02/2014	Iwan Setiawan	Kunjungan	Avian Influenza (Highly pathogenic)
28027	22/02/2014	64	Poultry mortality	Iwan Setiawan	Jawa Barat				22/02/2014	Iwan Setiawan	Kunjungan	Avian Influenza (Highly pathogenic)
28019	20/02/2014	2	Abortion or swollen joints	HASANUDDIN	Sulawesi Selatan							
27995	13/02/2014	32	Poultry mortality	Heru Mulyadi	Jawa Barat				13/02/2014	Hendra Komara	Kunjungan	Avian Influenza (Highly pathogenic), Newcastle Disease
27987	11/02/2014	42	Poultry mortality	Heru Mulyadi	Jawa Barat				11/02/2014	Hendra Komara	Kunjungan	Avian Influenza (Highly pathogenic), Newcastle Disease
27979	11/02/2014	42	Poultry mortality	Heru Mulyadi	Jawa Barat							
27961	5/02/2014	1	Abortion or swollen joints	Sulaeman	Sulawesi Selatan							
27953	20/01/2014	1	Sudden death	AKBAR	Sulawesi Selatan				20/01/2014	AKBAR	Kunjungan	Accident
27920	8/01/2014	12	Abortion or swollen joints	tine purwanti	Jawa Barat				8/01/2014	tine purwanti	Kunjungan	Brucellosis
27904	8/01/2014	2	Biting and behaviour change	Sirajuddin	Sulawesi Selatan				8/01/2014	Sirajuddin	Kunjungan	Rabies
27888	8/01/2014	1	Abortion or swollen joints	MUH. LISMI	Sulawesi Barat							
27870	4/01/2014	1	Biting and behaviour change	Hasrun	Sulawesi Selatan				4/01/2014	Hasrun	Kunjungan	Rabies
27854	31/12/2013	1	Abortion or swollen joints	HERNAWATI	Sulawesi Barat				31/12/2013	HERNAWATI	Kunjungan	Brucellosis
27847	31/12/2013	3	Poultry mortality	Iwan Setiawan	Jawa Barat				31/12/2013	Iwan Setiawan	Kunjungan	Avian Influenza (Highly pathogenic)
27839	31/12/2013	3	Poultry mortality	Iwan Setiawan	Jawa Barat				31/12/2013	Iwan Setiawan	Kunjungan	Avian Influenza (Highly pathogenic)
27839	30/12/2013	110	Poultry mortality	dini dinarwati	Jawa Barat				31/12/2013	dini dinarwati	Kunjungan	Avian Influenza (Highly pathogenic)
27839	30/12/2013	110	Poultry mortality	dini dinarwati	Jawa Barat				30/12/2013	dini dinarwati	Kunjungan	Avian Influenza (Highly pathogenic), Newcastle Disease
27805	20/12/2013	8	Poultry mortality	Sony Sonjaya	Sulawesi Barat				20/12/2013	Sony Sonjaya	Kunjungan	Avian Influenza (Highly pathogenic)
27763	13/12/2013	1	Sudden death	Makmur	Sulawesi Selatan				13/12/2013	Makmur	Kunjungan	Strangles, Bloat/Ruminal tympany, Anthrax
25189	9/12/2013	2	Poultry mortality	saleng	Sulawesi Selatan				9/12/2013	Abdul Hakim	Kunjungan	Newcastle Disease
25171	8/12/2013	1	Biting and behaviour change	Isnaniah Bagenda	Sulawesi Barat				8/12/2013	Isnaniah Bagenda	Kunjungan	Rabies
25148	30/11/2013	1	Biting and behaviour change	Makmur	Sulawesi Selatan				30/11/2013	Makmur	Kunjungan	Rabies
25130	30/11/2013	1	Biting and behaviour change	RUSLAN	Sulawesi Barat							
25122	30/11/2013	1	Biting and behaviour change	MUH. LISMI	Sulawesi Barat							

Gambar 3: Lembar kerja sindrom prioritas yang diunduh dari iSIKHNAS

## Langkah 2: Manajemen data

Dalam bagian manajemen data ini kita akan:

- Mengunduh data iSIKHNAS
- Menyimpan dan menjaga data asli
- Mengevaluasi data
- Membuat data baru.

### a. Mengunduh data iSIKHNAS

Berikut ini adalah latihan untuk mengunduh data iSIKHNAS. Anda dipersilakan untuk menyaksikan tayangan video 'Case study 1\_data download.avi' untuk memperoleh peragaan mengenai cara mengunduh data dari iSIKHNAS.

Langkah-langkah berikut adalah mengenai cara mengunduh data iSIKHNAS.

- i. Buka situs web ([www.isikhnas.com](http://www.isikhnas.com))
- ii. Masuk dengan nama pengguna (nomor ponsel atau email) serta kata sandi
- iii. Pilih gugus data (laporan, penyakit, sindrom prioritas, laporan penyakit prioritas)
- iv. Pilih kisaran data yang Anda minati
- v. Pilih area geografis yang ingin Anda periksa datanya
- vi. Pilih jalankan laporan
- vii. Gulir ke bawah dan klik unduhan, dan suatu lembar berformat csv akan terunduh
- viii. Simpan dan beri nama sesuai kebutuhan.

Sebelum melanjutkan pelatihan, berhenti dan praktikkan cara mengunduh data iSIKHNAS yang relevan bagi Anda.

Kami telah mengunduh data yang Anda butuhkan untuk latihan ini. Hal ini kami lakukan agar kami dapat menyediakan kunci jawaban latihan.

### b. Menyimpan dan menjaga data

Setiap kali Anda mengunduh berkas data, Anda harus menyimpan dua versi terpisah dari file yang sama.

Satu salinan harus disimpan dalam keadaan tidak berubah sebagai salinan asli dari file data. Anda akan selalu memiliki salinan file data yang asli dan tidak diubah, yang sama dengan saat ketika Anda mengunduhnya.

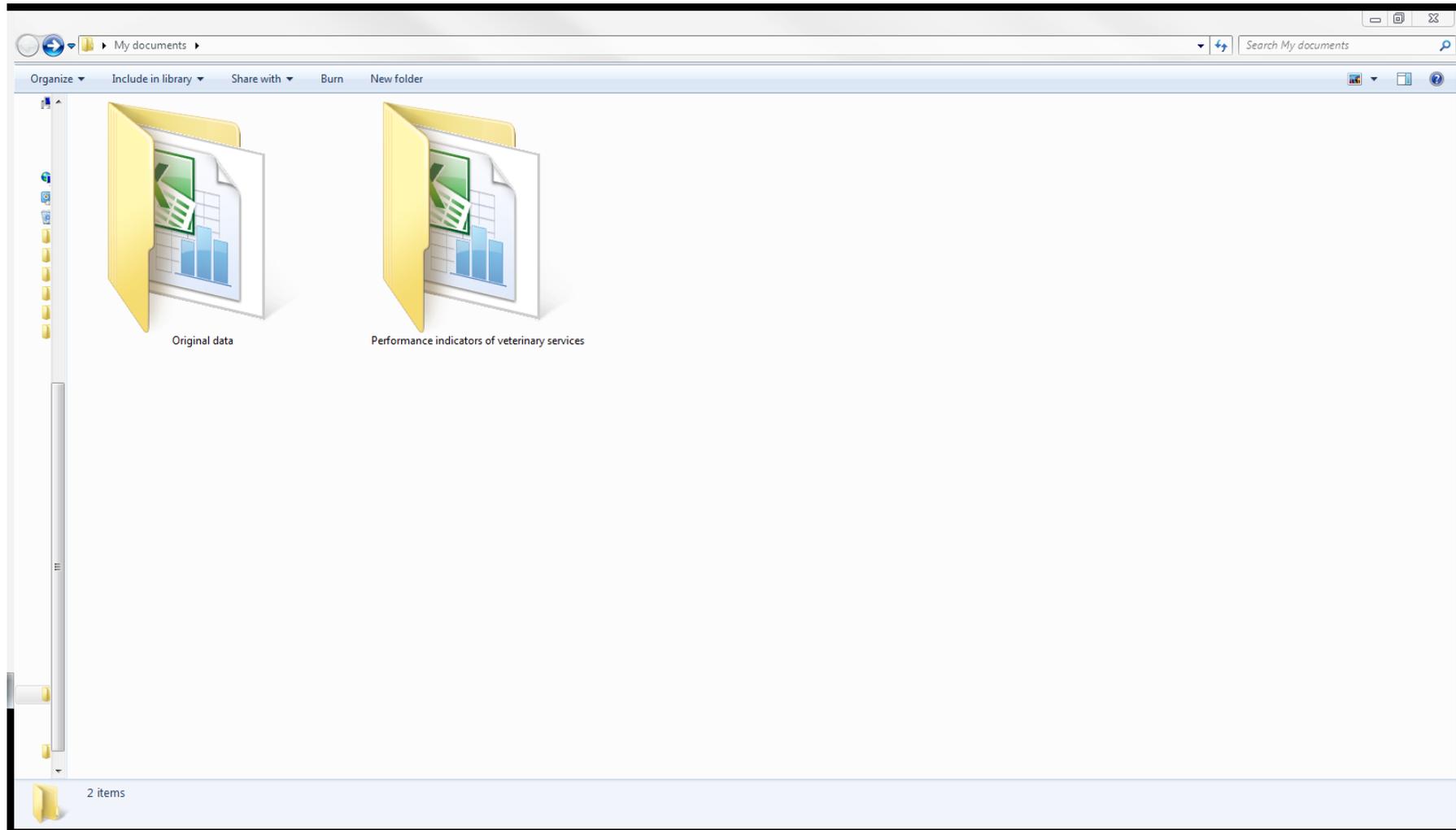
Salinan yang lain dapat digunakan sebagai file kerja untuk analisis data Anda. Selama proses analisis, Anda mungkin membuat berbagai perubahan terhadap data, termasuk menghapus atau menambah data serta variabel.

Cara yang baik untuk melakukan langkah-langkah tersebut adalah dengan mengunduh data asli ke suatu tempat khusus untuk data-data asli (misalnya dalam *folder* 'Data Asli' dalam 'My Documents'). Cara lain adalah dengan menyimpan salinan asli dengan penambahan kata *\_ASLI* pada nama filenya, sehingga Anda dapat dengan mudah mengidentifikasinya. Jangan pernah bekerja dengan menggunakan file data asli.

Namai file-file data unduhan dari iSIKHNAS dengan menggunakan nama yang memiliki makna dan aturan atau tata cara penamaan yang baik, agar Anda dapat membedakan berbagai file tersebut. Pada latihan ini, kita menamai data kita sebagai 'isikhnas\_priority disease syndromes\_March\_2013.csv'.

Dengan memiliki dua file data terpisah (file data kerja dan file data asli), Anda dapat selalu kembali membuka file data asli untuk memulai analisis baru, atau memeriksa perubahan yang telah dilakukan dalam salinan atau file data kerja. Jika Anda kembali membuka file data asli untuk memulai analisis yang baru, kembali buatlah salinan atau file data kerja baru dari file data asli tersebut.

Dalam Gambar 4 kami memperlihatkan tangkapan layar dari suatu kemungkinan struktur folder untuk data asli, yang menyimpan berkas Excel hasil unduhan, serta folder terpisah untuk analisis data, yang menyimpan salinan terpisah dari data yang diunduh.



**Gambar 4: Kemungkinan struktur folder untuk memastikan bahwa data asli yang diunduh tetap terjaga, dan agar analisis dilakukan dalam folder terpisah dengan menggunakan salinan data**

## Catatan: Konsep kunci mengenai kesalahan data

Langkah berikutnya adalah manajemen data. Salah satu bagian dari langkah ini adalah mengevaluasi data untuk memeriksa kesalahan serta kelayakannya. Untuk membantu Anda, kami telah menyediakan beberapa catatan mengenai kesalahan data.

Sebelum dianalisis, data harus diperiksa terlebih dahulu untuk mengetahui apakah ada kesalahan, dan juga untuk memikirkan potensi masalah lain yang mungkin ada dalam data tersebut. Setelah itu, Anda baru dapat menilai manfaat kesimpulan yang dapat ditarik dari data tersebut. Terdapat dua jenis kesalahan yang utama, yaitu data hilang, dan data yang tidak dicatat dengan benar.

### Data hilang

Data hilang dapat terjadi melalui dua cara, yaitu. terdapat sel-sel kosong dalam suatu baris data—dan ini dapat dengan mudah dilihat—atau, ada pengamatan (baris) yang hilang, dan ini sulit terdeteksi.

#### **a. Sel kosong dalam data**

Dalam lembar kerja yang Anda hadapi saat ini, sel kosong akan muncul dalam kejadian penyelidikan sindrom prioritas (suatu sel kosong dalam baris data).

Ini mungkin terjadi karena ada yang lupa mencatat data, atau tak punya waktu untuk melakukannya.

Dalam lembar data `isikhnas_priority disease syndromes_March_2013.csv`, contoh data yang hilang adalah nilai-nilai yang tidak ada dalam kolom Tanggal diinvestigasi. Ini terjadi karena beberapa investigator tidak mencatat tanggal investigasi laporan sindrom prioritas. Kemungkinan lainnya adalah investigasi memang tidak dilakukan. Dalam latihan di bawah ini, kita akan menganggap investigasi telah dilakukan dan data tidak dicatat. Cara utama bagi seorang dokter hewan untuk mengoreksi data yang tidak ada ini adalah dengan menghubungi pencatat data dan melengkapi data tersebut.

#### **b. Tidak ada Observasi**

Mungkin saja seluruh data dari suatu investigasi sindrom prioritas tidak pernah dicatat, sehingga data tersebut tidak tercantum di file data, atau dengan kata lain dikatakan data hilang (misalnya yang terjadi pada baris-baris yang hilang secara keseluruhan pada data dalam file `isikhnas_priority disease syndromes_March_2013.csv`).

Sayangnya, sangat sulit untuk mendeteksi masalah ini karena data memang tidak ada. Jika terdapat suatu pola dari baris-baris yang hilang ini, data yang tidak ada tersebut dapat menyebabkan masalah karena dapat menimbulkan *bias seleksi*. Bias seleksi terjadi dalam hal pengumpulan data yang tidak menyerupai populasi yang dipelajari.

Dari contoh yang telah dijelaskan sebelumnya mengenai pendugaan proporsi sapi yang positif brusellosis, dan peneliti hanya memilih sapi yang sehat dan muda sebagai sampel. Dari ilustrasi tersebut maka tidak ada sapi tua yang diambil sebagai sampel, dan diperoleh hasil

seroprevalensi brucellosis adalah sebesar 0%. Sebetulnya sapi-sapi yang lebih tua di kabupaten tersebut adalah positif brucellosis, namun sapi-sapi tersebut tidak terpilih sebagai sampel. Jika hal ini terjadi maka data sapi-sapi yang lebih tua tersebut tidak tercantum di dalam gugus data kita karena tidak terpilih sebagai sampel, dan akan menghasilkan baris-baris yang kosong di dalam gugus data.

Kita harus sungguh-sungguh memperhatikan metode pengumpulan data serta memperkirakan apakah ada bias seleksi, karena hal ini akan memengaruhi tingkat kepercayaan dalam hasil yang kita peroleh.

### **Data yang salah (galat)**

#### **a. Galat acak**

Kesalahan pencatatan data bisa terjadi dalam sebagian besar gugus data. Untungnya, kemungkinan kesalahan pencatatan tersebut tidak terlalu banyak dalam data iSIKHNAS dibandingkan dengan gugus data lain. Hal ini adalah karena data ekstrem yang tidak masuk akal tidak dapat dimasukkan ke dalam iSIKHNAS. Metode pemasukan data akan mencegah terjadinya hal ini. Sebagai contoh, jika bobot seekor sapi dicatat sebagai 5000 kg ini berarti kesalahan, karena angka tersebut 10 kali lipat dari nilai yang kita perkirakan ada untuk seekor sapi dewasa. Kesalahan ekstrem yang demikian secara umum dicegah pada tahap pemasukan data ke iSIKHNAS.

Namun, dapat terjadi tercatatnya data yang salah tapi masih berada dalam kisaran normal. Ini sangat sulit dideteksi karena data tampak normal. Data seperti ini cenderung ‘menggangu’ gugus data, dan memperkecil kemungkinan pendeteksian pengaruh saat Anda melakukan analisis. Sebagai contoh, jika berat sebenarnya dari seekor sapi adalah 450 kg namun ada kesalahan pencatatan menjadi 490 kg (karena kesalahan pengetikan), hal ini tidak akan terdeteksi oleh iSIKHNAS karena merupakan nilai yang masuk akal. Kesalahan ini akan dimasukkan ke dalam gugus data.

#### **b. Galat sistematis**

Kesalahan lain yang mungkin terjadi pada data adalah terjadinya kesalahan sistematis dalam cara pengukuran. Dalam hal ini, nilai sebenarnya adalah salah tetapi dicatat sebagai data yang “benar” didalam iSIKHNAS.

Sebagai contoh, seorang dokter hewan di area tertentu mungkin tidak cukup terlatih dalam pendeteksian kasus rabies. Hal tersebut mengakibatkan banyaknya investigasi Gilgalak yang akan memiliki diagnosis sementara yang salah. Misalnya kasus rabies tersebut pada diagnosis sementara ditulis ‘keracunan’, padahal anjing tersebut menggigit dan memiliki perubahan perilaku akibat rabies. Kesalahan-kesalahan ini kemudian dicatat dalam iSIKHNAS, dan merupakan catatan yang salah karena sang dokter hewan melakukan kesalahan. Kesalahan di dalam uji klinis juga dapat terjadi dengan cara yang sama.

Bias seperti ini dapat memberi dampak serius dalam penafsiran Anda terhadap data. Galat seperti ini juga sangat sulit dideteksi dan tidak dapat dikoreksi dengan mudah (kecuali jika tingkat kesalahan diketahui, misalnya kesalahan yang terjadi pada sensitifitas dan spesifisitas

tertentu dari suatu uji diagnostik). Galat seperti ini menimbulkan bias yang dikenal sebagai *bias informasi*.

Catatan ini menyimpulkan pembahasan mengenai berbagai kesalahan dalam data, dan sekarang saatnya untuk melanjutkan studi kasus dengan berbagai latihan dalam hal galat.

### c. Mengevaluasi data

**Mengidentifikasi berbagai kesalahan dan masalah-masalah yang mungkin ada dalam data**

#### **Latihan 3: Soal (menghitung data entri yang tidak ada dengan menggunakan Excel)**

Hitung jumlah poin data yang hilang dalam kolom Tanggaldinvestigasi dari file isikhnas\_priority disease syndromes\_March\_2013.csv. Petunjuk: gunakan fungsi filter dalam Excel untuk memilih hanya sel yang kosong (data yang tidak ada), kemudian hitung, atau gunakan fungsi countif dalam Excel.

**Petunjuk:** Lihat video untuk memperoleh bantuan (Case study 1\_exercise 3\_missing data.avi).

#### **Latihan 3: Jawaban (menghitung data entri yang tidak ada dengan menggunakan Excel)**

Berapa banyak nilai yang tidak ada, yang berhasil Anda hitung menggunakan fungsi filter?

Berapa banyak nilai yang tidak ada, yang berhasil Anda hitung menggunakan formula 'countif'?

#### **Latihan 4: Soal (bias seleksi dan bias informasi).**

Dalam kelompok Anda, diskusikan bentuk bias seleksi dan bias informasi yang mungkin ada dalam data iSIKHNAS.

#### **Petunjuk:**

Menurut Anda, investigasi sindromik apa yang mungkin secara rutin tidak ada dalam gugus data iSIKHNAS? Apa dampaknya terhadap kesimpulan yang dapat Anda tarik berdasarkan data yang Anda miliki? Hal ini merupakan bias seleksi.

Dapatkah Anda membayangkan kesalahan yang sering muncul ketika Anda menginvestigasi suatu kejadian sindromik dan membuat diagnosis sementara? Hal ini merupakan bias informasi.

**Latihan 4: Jawaban (bias seleksi dan bias informasi).**

Catat hasil diskusi kelompok mengenai bagaimana bias seleksi dan bias informasi dapat terjadi.

***Membuat data baru***

Penting bagi Anda untuk dapat membuat kolom data baru (atau variabel) dengan menggunakan kolom yang sudah ada. Hal ini akan membantu Anda dalam membuat analisis yang lebih terperinci terhadap data iSIKHNAS. Misalnya anda ingin menganalisis Anda akan perlu melakukan hal tersebut ketika ingin menentukan waktu antara tanggal laporan dan tanggal investigasi untuk mengetahui kecepatan petugas dalam merespon laporan masyarakat. Untuk keperluan tersebut maka Anda memerlukan satu variabel baru yang merupakan selisih antara tanggal investigasi dan tanggal laporan. Dalam latihan berikut ini, Anda harus membuat kolom baru dengan nama 'WaktuTunggu' dan mengisinya dengan data selisih jam antara pelaporan dan investigasi yang dalam hal ini disebut jumlah jam sebelum investigasi dilakukan.

**Latihan 5: Soal (membuat data baru: jumlah jam sebelum investigasi dilakukan)**

Buat kolom baru yang mengukur selisih waktu antara investigasi dan pelaporan.

**Langkah-langkah (petunjuk):**

Ubah format kolom Tanggalaporan dan Tanggaldiinvestigasi dalam lembar kerja isikhnas\_priority disease syndromes\_March\_2013.csv menjadi format tanggal (jika perlu).

Membuat kolom baru berjudul 'Masatunggu'.

Membuat formula dalam kolom tersebut, berupa Tanggalaporan dikurangi Tanggaldiinvestigasi untuk memperoleh jumlah hari di antara kedua tanggal tersebut.

Kalikan dengan 24 untuk menentukan jumlah jam antara pelaporan dan investigasi (WaktuTunggu).

Konversi jawaban ke kolom nilai baru menggunakan perintah paste special (values).

Hapus kolom tambahan, dan biarkan kolom WaktuTunggu.

Hapus pesan kesalahan #NUM! dengan menggunakan fungsi filter. Hal ini merupakan nilai-nilai yang tidak ada/tidak tercatat.

Sekarang Anda telah membuat data baru yang dapat digunakan untuk mengetahui berapa lama waktu yang dibutuhkan sebelum laporan sindrom prioritas sdiinvestigasi oleh petugas. Hal ini akan membantu Anda untuk menilai kinerja pelayanan veteriner.

**Petunjuk:** Lihat video untuk peragaan (Case study 1\_exercise 5\_creating new data).

**Latihan 5: Jawaban (membuat data baru: jumlah jam sebelum investigasi dilakukan)**

Buat catatan singkat mengenai kesulitan atau kemudahan yang Anda temui dalam Latihan 5.

Latihan mengakhiri langkah 2: manajemen data. Kita telah mempertimbangkan bias seleksi dan bias informasi, kita telah menelaah kesalahan data, dan kita telah membuat data baru. Berikutnya, kita akan mendeskripsikan data.

### Langkah 3: Deskripsi data

Tahap berikutnya dalam analisis data ini adalah menelaah kolom WaktuTunggu dan mendeskripsikannya. WaktuTunggu adalah variabel numerik. Untuk deskripsi data tersebut, pertama-tama kita akan mendeskripsikan kolom WaktuTunggu dengan menggunakan statistika deskriptif, dan kemudian memplotnya dalam suatu grafik. Namun sebelum melakukan ini, kami akan memberikan catatan mengenai statistika deskriptif.

#### Catatan: Konsep kunci mengenai statistika deskriptif

Suatu tahap penting dalam analisis data adalah mendeskripsikan data. Suatu variabel adalah karakteristik, angka, atau kuantitas yang dapat diukur atau dihitung. WaktuTunggu adalah suatu variabel kontinu.

#### Ukuran pemusatan (rataan dan median)

Langkah pertama dalam mendeskripsikan variabel WaktuTunggu adalah menduga nilai yang paling umum. Terdapat beberapa cara untuk melakukan ini, namun cara terbaik yang biasa dilakukan adalah dengan mengukur rataan dan median.

#### Rataan

Rataan berguna untuk menduga pusat dari semua nilai yang kita miliki untuk WaktuTunggu. Anda dapat menduga nilai rataan dengan menjumlahkan semua nilai dari WaktuTunggu dan membagi jumlah itu dengan jumlah nilai yang kita miliki.

Rumus untuk rataan adalah:

$$\bar{x} = \frac{\sum x}{n}$$

Dengan  $\bar{x}$  = rataan sampel,  $\sum$  = simbol untuk jumlah,  $x$  = variabel yang tengah kita telaah (dalam hal ini adalah WaktuTunggu) dan  $n$  = jumlah pengamatan.

Sebagai contoh:

Kita memiliki 5 pengamatan untuk variabel WaktuTunggu (24, 0, 48, 78, dan 96 jam).

Maka nilai rataannya adalah:

$$\text{Rataan (WaktuTunggu)} = \bar{x} = \frac{24 + 0 + 48 + 78 + 96}{5} = 49,2 \text{ jam}$$

Kita akan dapat memperkirakan bahwa diperlukan waktu 49,2 jam untuk melakukan investigasi suatu laporan sindrom prioritas.

#### Median

Median mengukur nilai bagian tengah. Dibandingkan dengan rata-rata, median tidak terlalu dipengaruhi oleh pencilan (nilai ekstrem yang berbeda dengan observasi lainnya dalam suatu variabel). Untuk menghitung median, kita mengurutkan semua pengukuran dari yang terendah hingga tertinggi, kemudian mengambil nilai yang berada di tengah sebagai mediannya. Dari data sebelumnya, maka daftar yang diurutkan adalah:

0, 24, 48, 78, 96 jam.

Dari data yang sudah diurutkan di atas, nilai yang ada di tengah adalah 48 jam. Dalam kasus ini, median hampir sama nilainya dengan rata-ratanya. Rataan dan median memiliki nilai yang mirip jika data menyebar normal atau jika ukuran sampelnya besar.

Berikutnya kita akan menghitung rata-rata dan median dari data yang sebenarnya dengan menggunakan Excel.

### **Ukuran penyebaran (wilayah (*range*), wilayah antarkuartil, simpangan baku, dan selang kepercayaan)**

Selain nilai tengah, hal lain yang penting untuk diketahui adalah seberapa jauh nilai-nilai suatu variabel tersebar di sekitar rata-rata atau median. Suatu ukuran penyebaran membantu Anda memahami seberapa baik sampel Anda dalam menduga rata-rata atau median. Jika sampel besar atau variabilitas dalam observasi rendah, maka nilai ukuran penyebaran akan lebih kecil. Nilai ukuran penyebaran yang lebih kecil ini menunjukkan bahwa Anda memiliki keyakinan yang lebih tinggi bahwa rata-rata atau median yang Anda duga mendekati nilai yang sebenarnya.

Ukuran penyebaran yang umum digunakan adalah simpangan baku atau selang kepercayaan untuk rata-rata, dan jarak antarkuartil untuk median. Adapun wilayah (*range*) dapat digunakan baik untuk rata-rata maupun median.

#### **Wilayah**

Wilayah adalah selisih antara observasi tertinggi dan terendah. Misalnya jika nilai variabel WaktuTunggu terlama adalah 96 jam dan terendah adalah 0 jam, maka wilayah adalah  $96-0=96$  jam).

#### **Wilayah antarkuartil**

Wilayah antarkuartil merupakan selisih antara kuartil atas (Q3) dan kuartil bawah (Q1). Kuartil atas (Q3) adalah nilai yang ada di tengah di antara median dan nilai tertinggi. Kuartil bawah (Q1) adalah nilai yang ada di tengah di antara median dan nilai terendah. Dengan demikian, nilai kuartil membagi data menjadi 25% di bagian atas dan bawah, sehingga 50% nilai tercakup dalam wilayah antarkuartil.

Misalnya untuk variabel WaktuTunggu diperoleh nilai minimum, Q1, median, Q3 dan nilai maksimum sebagai berikut:

Pengukuran (jam)	Parameter
------------------	-----------

0	Pengukuran terendah
24	Q1 (kuartil bawah)
48	Median (pengukuran tengah)
78	Q3 (kuartil atas)
96	Pengukuran tertinggi

Maka wilayah antarkuartilnya adalah  $Q3-Q1 = 78-24 = 54$  jam. Dengan demikian, kita bisa menafsirkan bahwa 50% sebaran waktu tunggu diagnosis adalah sekitar nilai 54 jam, yaitu di antara 24 jam sampai 78 jam.

### Simpangan baku

Simpangan baku (SB) mengukur rata-rata perbedaan antara nilai setiap pengukuran dengan nilai rata-rata. Berikut ini perhitungan SB untuk WaktuTunggu:

$$\begin{aligned}
 SB &= \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \\
 &= \sqrt{\frac{(0 - 49,2)^2 + (24 - 49,2)^2 + (48 - 49,2)^2 + (78 - 49,2)^2 + (96 - 49,2)^2}{5}} \\
 &= \sqrt{\frac{6077}{5}} = \sqrt{1215} = 35
 \end{aligned}$$

Dengan  $x = 0, 24, 48, 78, 96$  jam;  $n = 5$ ;  $\bar{x} = 49,2$  jam; dan  $SB =$  simpangan baku.

Hasil ini bisa kita tafsirkan sebagai berikut: rata-rata perbedaan waktu tunggu diagnosis terhadap nilai rata-ratanya (49,2 jam) adalah 35 jam. Dari nilai simpangan baku tersebut kita dapat mengetahui bahwa data sangatlah beragam. Karena nilai simpangan baku yang besar itu, maka tidak benar-benar yakin bahwa nilai rata-rata kita yang sebesar 49,2 jam tersebut memiliki nilai yang dekat dengan setiap pengukuran.

### Selang kepercayaan

Selang kepercayaan (SK) adalah suatu selang yang masuk akal dari suatu nilai dugaan. Suatu selang kepercayaan sebesar 95% dapat dihitung sebagai berikut:

$$95\% SK = \bar{x} \pm t_{\alpha/2, \nu} SB / \sqrt{n}$$

dengan  $t_{\alpha/2, \nu}$  = nilai t pada  $\alpha/2$ , derajat bebas  $\nu = n-1$ , SB = simpangan baku,  $n$  = ukuran sampel, dan  $\bar{x}$  adalah penduga rata-rata. (Tabel T terlampir)

Jika ukuran sampel ( $n$ )  $\geq 30$ , maka selang kepercayaan 95% dapat dihitung sebagai berikut:

$$95\% SK = \bar{x} \pm Z SB / \sqrt{n}$$

dengan  $Z = 1,96$ ;  $SB$  = simpangan baku;  $n$  = ukuran sampel; dan  $\bar{x}$  adalah penduga rata-rata.

$$95\% SK = \bar{x} \pm 1,96 SB = 49,2 \pm 1,96(35) = 49,2 \pm 69 = -20 \text{ hingga } 118 \text{ jam}$$

Dengan demikian, selang kepercayaan kita untuk variabel WaktuTunggu adalah dari -20 hingga 118 jam. Nilai -20 bukanlah nilai yang mungkin (kita tidak dapat melakukan investigasi sebelum menerima laporan!), jadi kita ambil nilai 0 jam sebagai gantinya.

Kita dapat menafsirkan hasil ini sebagai berikut:

Kita yakin 95% bahwa rata-rata populasi yang sebenarnya untuk variabel WaktuTunggu terletak di antara -20 (0) dan 118 jam. Artinya, pengambilan sampel kita tidak cukup baik untuk memperkirakan rata-rata dengan akurat karena selang kepercayaannya cukup lebar.

Penafsiran yang lebih formal adalah, apabila kita mengulangi pengambilan sampel dari populasi yang sama dengan cara yang sama berkali-kali, maka 95% dari penghitungan rata-rata yang kita peroleh akan menghasilkan nilai di antara -20 (0) dan 118 jam.

Selang kepercayaan dapat dihitung untuk berbagai jenis ukuran lainnya, bukan hanya untuk rata-rata.

#### a. Mendeskripsikan variabel tunggal

#### **Latihan 6: Soal (menghitung ukuran pemusatan dan penyebaran untuk variabel WaktuTunggu)**

Sebelumnya kita telah menggunakan 5 angka simulasi untuk memperagakan perhitungan ukuran pemusatan dan penyebaran. Sekarang kita akan melakukan penghitungan ukuran yang sama dengan menggunakan data yang sebenarnya dari iSIKHNAS.

Hitunglah rata-rata, median, simpangan baku, selang kepercayaan, wilayah, dan wilayah antarkuartil untuk variabel WaktuTunggu dalam lembar kerja Sindrom prioritas.

#### **Petunjuk:**

Gunakan Excel, terutama gunakan *descriptive analysis tool* dalam Excel.

**Petunjuk:** Lihat video peragaan (Case study 1\_exercise 6\_descriptive statistics.avi). Video ini menguraikan cara untuk melakukan analisis dengan menggunakan Analysis ToolPak dalam Excel.

Jika Anda tidak memiliki akses ke Analysis ToolPak, Anda dapat melihat video berikut ini tentang cara menghitung nilai-nilai deskripsi statistika secara manual dengan menggunakan formula Excel (Case study 1\_exercise 6b\_descriptive statistics without tool pack.avi).

**Latihan 6: Jawaban (menghitung ukuran pemusatan dan penyebaran untuk variabel WaktuTunggu)**

Tuliskan jawaban Anda bagi tiap parameter dalam tabel berikut ini:

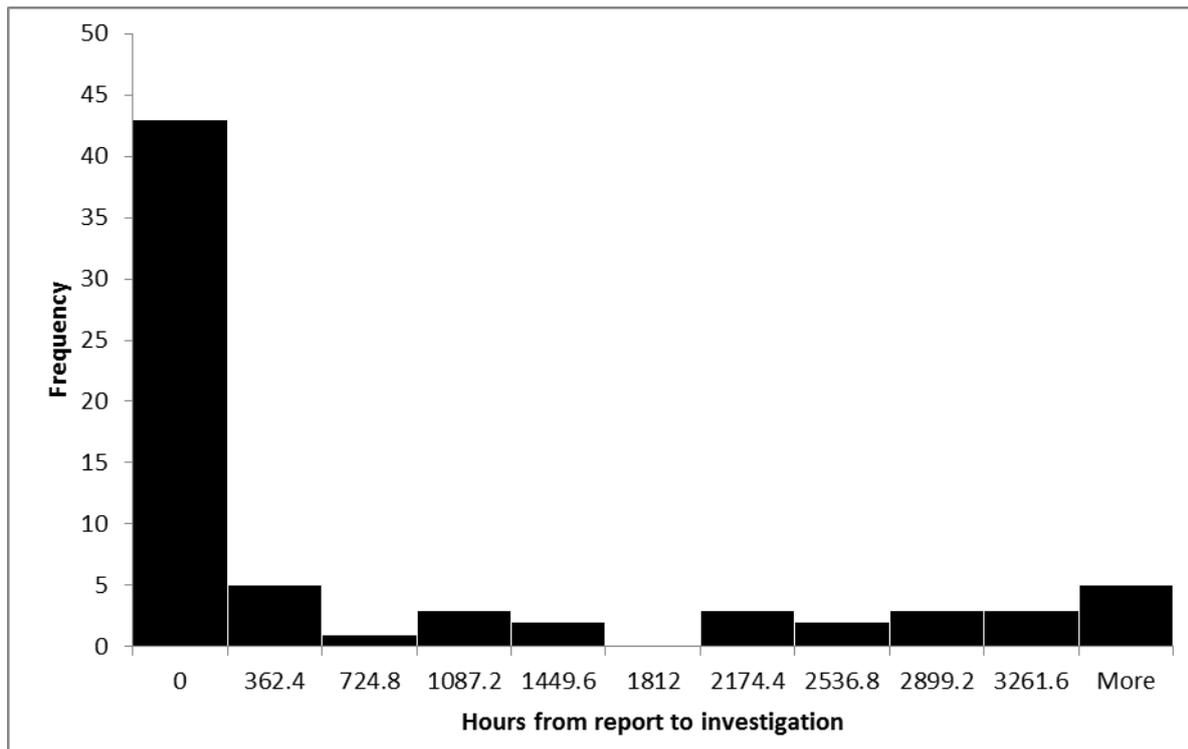
<b>Parameter</b>	<b>Penduga</b>
Rataan	
median	
simpangan baku	
selang kepercayaan	
wilayah	
Wilayah antarkuartil	

***b. Plotting***

Sekarang kita telah selesai menghitung ukuran pemusatan dan penyebaran untuk WaktuTunggu.

Terdapat indikasi bahwa data tidak menyebar normal, karena nilai-nilai rataaan dan median yang sangat berbeda, yaitu rataanya adalah 750 jam dan mediannya adalah 0 jam. Pola sebaran data penting diperhatikan agar kita dapat memilih nilai penduga yang paling sesuai. Misalnya pada kasus data ini apakah kita akan memfokuskan perhatian pada rataaan, atau pada median?

Untuk mempertimbangkan hal ini, kita akan membuat histogram untuk variabel WaktuTunggu, yang akan memberi gambaran kepada kita mengenai sebaran data. Kita akan membuat histogram dengan menggunakan Excel dalam latihan berikutnya, namun sebelum itu kami akan membahas dan mendiskusikan tentang histogram tersebut terlebih dahulu. Berikut ini adalah histogram untuk variabel WaktuTunggu (Gambar 5).



**Gambar 5:Histogram waktu tunggu antara laporan dan investigasi penyakit prioritas (jam)**

Dengan menelaah histogram ini, kita dapat melihat bahwa untuk data variabel WaktuTunggu:

- Memiliki selang kelas dari 0 jam hingga lebih besar dari 3261,6 jam.
- Terdapat 43 sindrom prioritas yang diinvestigasi dengan segera (dalam waktu 0 jam). Ini merupakan indikasi bahwa banyak nilai variabel WaktuTunggu yang rendah. Hal ini merupakan indikasi pelayanan veteriner yang baik.
- Histogram ini memiliki ekor (*tail*) panjang yang memperlihatkan bahwa ada masa jeda yang panjang sebelum beberapa laporan penyakit prioritas diinvestigasi.
- Data tidak menyebar normal. Dalam hal ini, median dan wilayah antarkuartil merupakan ukuran deskriptif yang jauh lebih baik dibandingkan rata-rata dan simpangan baku atau selang kepercayaan.

Kita perlu memperhatikan mengapa beberapa laporan membutuhkan waktu begitu panjang untuk diinvestigasi. Masa jeda yang panjang ini umumnya terlihat pada tahun 2013, saat iSIKHNAS masih baru dan mungkin mengindikasikan masalah pencatatan data yang muncul di awal implementasi, atau bisa juga prosedur investigasi telah membaik sejak saat pencatatan data baru dimulai. Bagaimana pun, dibutuhkan upaya lebih lanjut untuk mengetahui alasan di balik lamanya waktu yang dibutuhkan sebelum investigasi dilakukan.

**Latihan 7: Soal (histogram variabel WaktuTunggu dalam Excel)**

Buat histogram WaktuTunggu.

**Petunjuk:** Lihat video yang disediakan untuk membantu Anda (Case study 1\_exercise 7\_plotting histograms).

### **Latihan 7: Jawaban (histogram)**

Di bagian kosong di bawah ini, tempatkan histogram yang Anda buat untuk data lamanya waktu tunggu antara laporan dan investigasi sindrom prioritas. Anda dapat menggambar dengan tangan, atau menyalinnya (*copy-paste*) secara elektronik dari lembar kerja Excel ke dalam dokumen ini, atau mencetak dan kemudian menempelkannya di bagian kosong ini.

### **Langkah 4: Pengujian hipotesis statistika**

Langkah keempat dan terakhir dalam analisis data adalah untuk melakukan pengujian hipotesis. Akan tetapi, hal ini tidak relevan bagi studi kasus ini, karena di sini kita hanya ingin memeriksa waktu yang dibutuhkan untuk antara laporan dan investigasi suatu sindrom prioritas. Kita akan melakukan pengujian hipotesis dalam dua studi kasus berikutnya.

### **Rangkuman studi kasus 1**

Dalam studi kasus ini, pertama-tama Anda telah belajar untuk mengunduh data iSIKHNAS.

Anda telah melakukan beberapa langkah penting dalam analisis data:

1. Menetapkan suatu tujuan
2. Melakukan manajemen data (dengan membuat salinan data asli, meneliti jika terdapat kesalahan pada data, dan membuat data baru).
3. Mendeskripsikan data numerik.

Tiga langkah umum ini harus selalu dilakukan, tanpa memandang jenis data yang Anda miliki.

Yang penting diingat, terkadang cara Anda mendeskripsikan data (langkah 3) akan berbeda-beda. Sebagai contoh, di sini Anda menggunakan ukuran pemusatan (misalnya median), dan ukuran penyebaran (Q1:Q3) untuk mendeskripsikan data. Deskripsi yang demikian hanya bermanfaat bagi variabel kontinu. Tipe data lain akan membutuhkan pendekatan yang berbeda dalam pendeskripsiannya (lihat Studi kasus 2, yaitu data kategorik dideskripsikan dengan tabel kontingensi). Anda harus selalu mencoba mendeskripsikan data dengan memplotnya pada grafik, tanpa memandang jenis data Anda.

Dalam hal studi kasus ini, langkah ke-4 (pengujian hipotesis) tidak perlu dilakukan, karena kita hanya ingin menelaah waktu yang dibutuhkan untuk menyelidiki suatu laporan penyakit prioritas.

**Ini adalah akhir studi kasus 1.**

## **Studi Kasus 2: Prevalensi musiman diare pada sapi**

### **Pengantar studi kasus**

Pada studi kasus 2 ini akan dijelaskan mengenai analisis data kategorik, yang banyak dijumpai pada studi epidemiologi.

Diare merupakan tanda klinis dari berbagai penyakit produksi yang serius, misalnya akibat adanya parasit atau kecacingan. Pada studi kasus ini kita akan melakukan pendugaan prevalensi diare pada sapi selama setahun untuk menentukan apakah terdapat puncak musiman dalam jumlah kasus diare. Hal ini berguna dalam penyusunan rencanapenanganan diare. Misalnya apakah kita dapat mengidentifikasi kasus diare terbanyak pada setiap tahunnya? jika dapat diidentifikasi, kapankah terjadinya puncak kasus tersebut? Jika informasi tersebut diketahui, maka Anda dapat meminta pasokan anthelmintik sebelum puncak kasus terjadi, atau menerapkan tindakan pencegahan sebelum puncak musiman tersebut. Di dalam latihan ini disediakan data 'Cattle clinical signs.xls' untuk contoh kasus seperti itu.

Seperti telah dijelaskan sebelumnya bahwa data yang disediakan pada pelatihan ini diunduh pada awal 2014, pada saat iSIKHNAS beroperasi selama satu tahun dengan wilayah cakupan yang terbatas. Ketidaklengkapan data ini berarti bahwa kesimpulan yang ditarik dari pelatihan ini tidak mencerminkan data yang sebenarnya pada saat ini. Seiring berjalannya waktu, maka data akan tersedia lebih lengkap, dan Anda dapat melakukan analisis data yang lebih baik.

### **Keahlian yang dikembangkan dalam studi kasus ini**

- Evaluasi data
- Pembuatan data baru
- Mendeskripsikan data kategorik dengan menggunakan tabel kontingensi serta plot
- Pengujian hipotesis menggunakan risiko relatif, odds ratio, dan selangl kepercayaan

### **Daftar file yang digunakan untuk studi kasus**

#### *Data*

Cattle clinical signs.xls

#### *Video*

- Case study 2\_exercise 10.avi
- 'Case study 2\_Exercise 11.avi
- Case study 2\_Exercise 12.avi
- Case study 2\_Exercise 13.avi
- Case study 2\_Exercise 14.avi

## Langkah-langkah dalam analisis data studi kasus 2 (latihan)

### Langkah 1: Tujuan

Tujuan dari studi kasus ini adalah untuk menentukan apakah ada perbedaan prevalensi diare pada sapi saat musim kemarau dan musim penghujan. Kita membutuhkan informasi ini agar dapat meningkatkan pengelolaan kesehatan hewan.

Kegunaan dari hasil analisis data ini adalah kita mungkin ingin memesan obat tepat pada waktunya untuk menghadapi puncak diare musiman tersebut. Kita juga mungkin ingin menerapkan program pengelolaan pada saat yang tepat untuk mengurangi kejadian diare pada tahun tersebut. Untuk dapat melakukan hal ini, tentu kita harus mengetahui apakah ada puncak musiman untuk kejadian diare, dan jika ada, kapan puncak musiman itu terjadi.

Di dalam studi kasus ini, kami tidak memberi latihan dalam hal penentuan tujuan.

### Langkah 2: Manajemen Data

Gugus data yang digunakan dalam studi kasus 2 ini adalah “Cattle clinical signs.xlsx”.

Seperti pada studi kasus 1, langkah pertama yang dilakukan adalah menyimpan dan memelihara data asli. Buat salinan untuk kerja dan arsipkan berkas asli.

Selanjutnya, kita akan mengevaluasi data. Kita akan memutuskan data mana yang paling relevan, dan kita akan memeriksa apakah ada masalah dalam data tersebut (galat dan bias). Kemudian kita akan memutuskan apakah kita perlu membuat variabel baru untuk dapat menjawab tujuan yang telah kita tentukan.

#### a. Data apakah yang penting?

Terdapat banyak kolom data (variabel) di gugus data yang kita miliki. Kita akan memulai analisis data kita dengan mempertimbangkan kolom data mana yang penting dalam “Cattle clinical signs.csv”. Terdapat dua jenis data yang kita butuhkan:

1. Data yang memungkinkan kita melakukan pencacahan jumlah penyakit. Artinya, kita akan membutuhkan baik data numerator (jumlah kasus mencret pada sapi) dan juga data denominator (jumlah sapi yang berisiko terkena penyakit)
2. Perubahan penyakit sejalan dengan waktu. Kita perlu mengetahui kapan observasi terhadap sapi yang sakit diketahui

**Latihan 8: Soal (Menentukan variabel mana yang penting untuk menjawab tujuan yang telah kita tentukan, dalam file “Cattle clinical signs.xlsx”).**

Petunjuk:

Teliti kolom yang mewakili “Tanda yang dilaporkan” serta “Diagnosamentara”. Variabel mana yang menurut Anda paling tepat dalam memberikan indikasi kemungkinan diagnosis

diare pada sapi? Apakah ada data yang menunjukkan kapan infeksi terjadi? Data apa yang dapat digunakan sebagai denominator?

**Latihan 8: Jawaban (Menentukan variabel mana yang penting untuk menjawab tujuan yang telah kita tentukan, dalam “Cattle clinical signs.xlsx”).**

Berdasarkan diskusi kelompok Anda, tuliskan variabel yang menurut Anda merupakan kolom data terpenting beserta alasannya?

Setelah Anda membicarakan masalah ini bersama kelompok, lihat jawaban yang telah disediakan untuk membantu Anda menjawab pertanyaan ini. Tuliskan jawabannya di sini.

**Mengidentifikasi berbagai kesalahan dan masalah-masalah yang mungkin ada dalam data**

Seperti pada studi kasus sebelumnya, sebelum melakukan analisis data, kita harus menentukan apakah ada data yang hilang, dan apakah ada bias (bias seleksi maupun bias informasi) atau kesalahan dalam data. Kita akan melakukan ini dalam latihan berikut ini.

**Latihan 9: Soal (memeriksa kesalahan dan kemungkinan bias)**

**Petunjuk:**

Gunakan fungsi filter dalam Excel untuk memilih sel-sel yang kosong (data yang hilang) dan kemudian hitung untuk mengetahui jumlah pengamatan yang tidak ada. Untuk keperluan

tersebut, dapat juga digunakan fungsi countif dalam Excel. Lakukan ini bagi dua variabel yang sedang kita pelajari.

Dengan menggunakan fungsi filter, periksa juga semua kategori dalam dua kolom variabel yang kita pelajari tersebut, apakah terdapat kategori yang janggal atau berpotensi salah?

Apa saran Anda mengenai kesalahan tersebut (jika ada)?

Apakah terdapat kemungkinan bias dalam data?

### **Latihan 9: Jawaban (memeriksa kesalahan dan kemungkinan bias)**

Apakah Anda mendeteksi adanya kesalahan (ya/tidak)?

Tuliskan catatan hasil diskusi kelompok mengenai bias yang ada.

### **b. Membuat data baru untuk memfasilitasi analisis selanjutnya**

Untuk melakukan analisis data, kita harus membuat beberapa data tambahan. Lakukan latihan 10 untuk membuat data ini.

### **Latihan 10: Soal (membuat data baru)**

Kita perlu membuat beberapa data baru untuk dapat menganalisis data secara optimal. Pertimbangkanlah data apa saja yang Anda perlukan.

Petunjuk: Kita perlu mengetahui apakah kasus diare terjadi saat musim hujan atau kemarau, dan apakah tanda-tanda klinis yang dicatat adalah diare atau mungkin ada tanda klinis lain. Kita memerlukan penggunaan formula guna mengisi tiga kolom baru tersebut (variabel), yaitu nama bulan, musim, dan diare/tanda klinis lain. Bagaimana cara Anda melakukan ini dalam Excel, dan variabel mana di dalam gugus data yang anda perlukan untuk membuat data baru tersebut?

Setelah Anda mendiskusikan ini, lihat dokumen kunci jawaban untuk mengetahui tentang pendekatan umum (serta rumus), dan tonton video untuk latihan 10 (Case study 2\_exercise 10.avi).

**Latihan 10: Jawaban (membuat data baru)**

Tuliskan catatan mengenai video latihan, jika perlu.

### Langkah 3: Deskripsi Data

Tahap ini penting untuk memeriksa lebih lanjut apakah ada kesalahan, dan untuk memahami mengenai hubungan antarvariabel. Pemahaman tersebut akan membantu kita dalam menjawab pertanyaan yang kita ajukan.

#### a. Deskripsi data

Pertama-tama kita akan menelaah dan meringkas data dari tiap kolom yang diamati. Untuk melakukan hal ini, kita membutuhkan *pivot table* dalam Excel. Kita akan melakukan hal ini dalam latihan berikut ini.

#### **Latihan 11: Soal [ringkasan data tiap kolom yang tengah kita amati (bulan, musim, dan kasus diare?)]**

Buatlah 3 tabel kontingensi, yaitu untuk mencacah jumlah tanda klinis (diare atau yang lainnya), untuk mencacah musim (kemarau atau hujan), dan untuk mencacah nama bulan terjadinya diare.

#### **Petunjuk:**

Tiap kolom penting berisi data yang sangat sederhana, yaitu nilai 1 sampai 12 untuk kolom Integer bulan, hujan atau kemarau untuk kolom Musim, dan diare atau lainnya untuk kolom Tanda klinis. *Data analysis tool* untuk menghitung ukuran pemusatan dan penyebaran di dalam Excel yang sebelumnya digunakan dalam studi kasus 1, tidak relevan digunakan di dalam studi kasus ini.

Untuk data seperti ini, *pivot table* akan memungkinkan Anda meringkas data tiap kategori. *Pivot table* akan mempermudah Anda dalam mencacah jumlah observasi untuk setiap kategori. Misalnya untuk peubah Musim, akan ditampilkan berapa jumlah kasus diare yang terjadi pada musim hujan dan musim kemarau.

Lihat video 'Case study 2\_Exercise 11.avi'.

#### **Latihan 11: Jawaban [Ringkasan data tiap kolom yang yang dianalisis (bulan, musim dan diare)]**

Masukkan hasil *pivot table* ke dalam bagian kosong di bawah ini.

## b. Hubungan antarvariabel

### Tabel Kontingensi

Langkah berikutnya adalah menyelidiki hubungan antara berbagai variabel (misalnya: gejala klinis terhadap musim). Hal ini dilakukan sebelum pengujian hipotesis secara resmi. Dalam studi kasus ini, data yang kita miliki adalah data kategorik, yaitu musim hujan/kemarau dan diare/lainnya). Untuk tujuan tersebut maka kita kembali menggunakan tabel kontingensi. Data subtotal untuk setiap kategori yang diperoleh dari tabel kontingensi akan kita gunakan dalam pengukuran asosiasi. Hal ini akan kita bahas kemudian.

#### Latihan 12: Soal (tabel kontigensi)

Buatlah dua tabel kontingensi (dengan menggunakan *pivot table*) dalam Excel dan perhatikan data yang ada.

Petunjuk: Tabel Kontingensi yang dihasilkan akan terlihat seperti berikut ini.

*Diare menurut musim:*

		Musim	
		Kemarau	Penghujan
Gejala klinis	Diare		
	Lainnya		
	Total sapi yang dilaporkan		
	Prevalensi diare pada sapi yang sakit (%)		

*Diare menurut bulan:*

		Bulan												
		1	2	3	4	5	6	7	8	9	10	11	12	
Gejala klinis	Lainnya													
	Diare													
	Total sapi													
	Prevalensi diare pada sapi yang sakit (%)													

**Petunjuk:** Lihatlah tayangan video untuk menyelesaikan latihan dan mengisi tabel kontingensi (Case study 2\_exercise 12.avi).

Apakah terlihat adanya hubungan antara musim atau bulan terhadap diare?

**Latihan 12: Jawaban (tabel kontigensi)**

Salin kedua tabel kontingensi yang telah anda buat pada tempat yang tersedia di bawah ini dan perhatikan hubungan antara waktu dan kasus diare.

**Tabel kontigensi yang telah diisi lengkap sebagai bahan diskusi**

Berikut ini tersedia tabel kontingensi yang telah terisi lengkap.

Tampaknya ada sedikit peningkatan pada proporsi sapi sakit yang menderita diare pada musim penghujan (27% dibanding 22%). Selain itu, terlihat juga bahwa prevalensi sapi sakit yang menderita diare meningkat sepanjang tahun itu, dan pada bulan-bulan di tengah tahun, yaitu saat musim kemarau, memiliki prevalensi diare yang relatif rendah.

Akan tetapi, hasil yang ada sangat beragam. Nanti kita akan menggunakan pengujian hipotesis statistika untuk menarik kesimpulan yang lebih baik.

*Diare menurut musim:*

	<b>Musim Kemarau</b>	<b>Musim Hujan</b>
<b>Diare</b>	246	544
<b>Lainnya</b>	875	1437
<b>Total sapi</b>	<b>1121</b>	<b>1981</b>
<b>Prevalensi diare pada sapi sakit (%)</b>	22	27

*Diare menurut bulan:*

		<b>Bulan</b>											
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b>Gejala klinis</b>	<b>Diare</b>	106	57	63	17	30	20	69	39	44	44	92	209
	<b>Lainnya</b>	282	193	224	97	118	115	260	164	114	104	249	392
	<b>Total sapi</b>	<b>388</b>	<b>250</b>	<b>287</b>	<b>114</b>	<b>148</b>	<b>135</b>	<b>329</b>	<b>203</b>	<b>158</b>	<b>148</b>	<b>341</b>	<b>601</b>
	<b>Prevalensi diare pada sapi sakit (%)</b>	27	23	22	15	20	15	21	19	28	30	27	35

**c. Plotting**

Sangat penting untuk menggambarkan data dalam bentuk plot karena plot memberikan banyak informasi kepada pembaca.

Untuk keperluan pembuatan plot tersebut, kita harus membuat tabel kontingensi bulan terhadap kasus diare. Kita akan meneliti data ini dengan menggunakan diagram pencar (*scatter plot*) dan

*line of best fit* dalam Excel. Diagram pencar digunakan untuk melakukan plot data yang terdiri dari dua variabel yang berhubungan pada setiap observasi. Diagram pencar biasanya digunakan untuk variabel yang berjenis ordinal atau kontinu.

Pada studi kasus ini, prevalensi diare tiap bulan pada seluruh sapi yang sakit merupakan variabel kontinu (yaitu suatu nilai persentase yang nilainya 0-100). Adapun bulan merupakan variabel ordinal. Dengan demikian maka variabel kita adalah variabel ordinal dan kontinu sehingga kita dapat membuat diagram pencar dengan bulan sebagai sumbu x dan prevalensi diare pada sapi sakit sebagai sumbu y. Dari diagram pencar tersebut kita dapat menelaah data secara visual dan dapat melihat apakah prevalensi diare pada sapi berubah sepanjang tahun.

**Latihan 13: Soal (plot prevalensi diare berdasarkan bulan)**

Buat diagram pencar variabel bulan terhadap prevalensi diare pada sapi sakit.

**Petunjuk:**

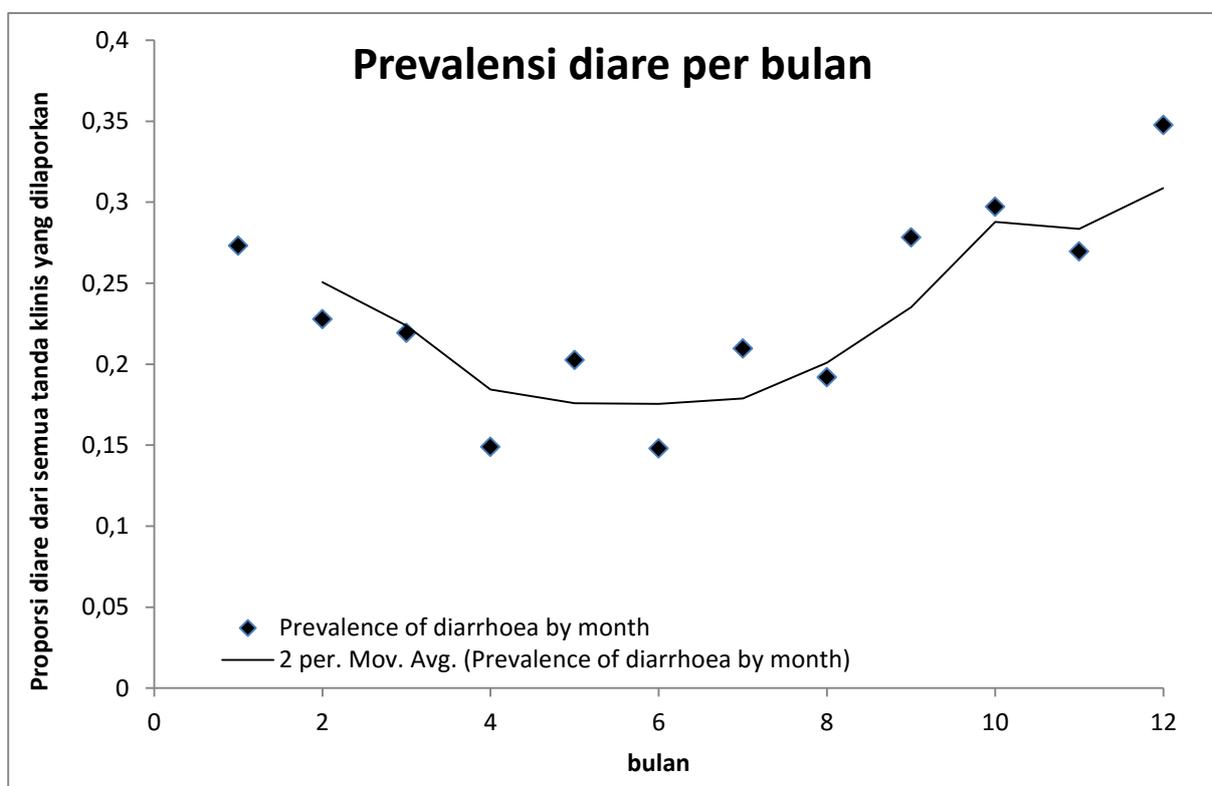
Untuk membuat diagram pencar tersebut, gunakan tabel kontingensi bulan terhadap prevalensi diare yang sudah kita buat di atas. Lihatlah tayangan video ('Case study 2\_exercise 13.avi') untuk menuntun anda menyelesaikan soal ini.

**Latihan 13: Jawaban (plot prevalensi diare berdasarkan bulan)**

Gambarlah diagram pencar bulan terhadap prevalensi diare pada tempat yang disediakan di bawah ini.

Hasil plot yang telah Anda buat tersebut dapat dilihat pada Gambar 6. Ada beberapa hal yang harus diingat pada pembuatan diagram pencar, yaitu variabel bebas adalah sebagai sumbu

horizontal dan variabel tak bebas sebagai sumbu vertikal. Di dalam latihan kita, variabel bulan dianggap memengaruhi variabel prevalensi diare, sehingga variabel bulan adalah variabel bebas dan variabel prevalensi diare adalah variabel tak bebas. Jangan lupa untuk memberi judul pada setiap sumbu. Pada plot di bawah ini, disertakan juga garis suai terbaik (*line of best fit*) yang diperoleh dengan menghitung rata-rata bergerak (*moving average*). Plot ini membantu kita dalam memberikan informasi mengenai hubungan antara musim dengan prevalensi diare pada sapi. Dari plot tersebut dapat dilihat bahwa garis suai terbaik menurun di musim kemarau yang menandakan adanya hubungan antara diare dan bulan. Prevalensi terendah tampak terjadi di musim kemarau, yaitu sekitar bulan ke-4 hingga bulan ke-8, dan terjadi peningkatan prevalensi di musim penghujan.



**Gambar 6** Diagram pencar dan garis suai terbaik untuk data asosiasi antara bulan dan prevalensi diare.

**Catatan: Konsep kunci mengenai pengukuran penyakit dan asosiasi**

Mari kita mengingat kembali informasi mengenai ukuran frekuensi penyakit dan asosiasi yang telah dipelajari saat pelatihan epidemiologi lapangan. Parameter-parameter ini akan membantu Anda saat melakukan langkah 4 (pengujian hipotesis).

**Ukuran frekuensi penyakit**

Penting sekali bagi kita untuk dapat menggambarkan besaran penyakit (atau suatu ukuran lain) yang ada pada suatu populasi. Salah satu ukuran tersebut adalah prevalensi.

Prevalensi mengukur besaran penyakit pada satu titik waktu (atau pada suatu selang waktu tertentu).

Misalnya jika Anda memiliki kawanan yang terdiri dari 100 ekor sapi dan Anda mengambil sampel dari semua sapi lalu memperoleh hasil 50 ekor sapi memiliki antibodi terhadap *Brucella*, maka prevalensi antibodi *Brucella* adalah sebesar 0,50 (50%). Kesimpulan Anda adalah bahwa 50% sapi telah terpapar *Brucella* sebelumnya. Namun Anda tidak yakin apakah hasil positif tersebut adalah karena infeksi alami yang baru terjadi, infeksi alami di masa lampau, ataukah karena vaksinasi *Brucella*.

$$\text{Prevalensi antibodi Brucella} = \frac{\text{numerator}}{\text{denominator}} = \frac{50 \text{ antibodi positif}}{100 \text{ ekor dalam kawanan}} = 0,50 \\ = 50\%$$

### Ukuran asosiasi penyakit

Ukuran asosiasi merupakan ukuran yang penting dalam membantu kita menelaah suatu 'penyebab' penyakit. Sebenarnya yang diukur adalah asosiasi, bukan penyebab. Kita akan membahas perbedaan antara keduanya dalam pembahasan di bawah ini. Ukuran asosiasi ini mengukur asosiasi antara faktor risiko yang dihipotesiskan terhadap besaran penyakit. Berikut ini kita akan mempelajari dua jenis ukuran asosiasi.

#### *Risiko relatif (RR)*

Risiko relatif adalah perbandingan antara risiko terjadinya penyakit pada kelompok yang terpapar faktor risiko dengan risiko terjadinya penyakit pada kelompok yang tidak terpapar faktor risiko. Risiko ( $p$ ) merupakan peluang sakit, yaitu dihitung dengan:

$$p = \frac{\text{Jumlah hewan sakit}}{\text{Jumlah hewan yang (tidak) terpapar faktor}}$$

$$RR = \frac{p_1}{p_2}$$

RR = risiko relatif,  $p_1$  = peluang sakit pada kelompok yang terpapar faktor risiko, dan  $p_2$  = peluang sakit pada kelompok yang tidak terpapar faktor risiko

Jika faktor risiko tidak mempengaruhi kejadian penyakit, besar risiko pada kelompok yang terpapar faktor dan tidak terpapar faktor adalah sama, sehingga  $RR = 1$ . Dengan demikian, jika dari hasil pendugaan nilai RR diperoleh selang kepercayaan yang mencakup nilai 1, maka dapat dikatakan bahwa faktor risiko tidak berpengaruh terhadap kejadian penyakit.

#### *Odds Ratio*

*Odds* adalah perbandingan antara peluang terjadinya suatu kejadian dengan peluang tidak terjadinya kejadian tersebut. Dalam konteks penyakit, *odds* adalah perbandingan antara peluang sakit dan peluang tidak sakit atau secara matematis ditulis:

$$odds = \frac{p}{1 - p}$$

*Odds ratio* (OR) adalah perbandingan antara *odds* terjadinya penyakit pada kelompok yang terpapar faktor risiko dengan *odds* terjadinya penyakit pada kelompok yang tidak terpapar faktor risiko, yaitu dihitung dengan:

$$OR = \frac{Odds1}{Odds2}$$

OR = *odds ratio*, Odds1 = *Odds* terjadinya sakit pada kelompok yang terpapar faktor risiko, dan Odds2 = *Odds* terjadinya sakit pada kelompok yang tidak terpapar faktor risiko.

Seperti halnya RR, jika faktor risiko tidak mempengaruhi kejadian penyakit, maka nilai *odds* pada kelompok yang terpapar dan tidak terpapar faktor risiko adalah sama, sehingga OR = 1. Dengan demikian, jika dari hasil pendugaan nilai OR diperoleh selang kepercayaan yang mencakup nilai 1, maka dapat dikatakan bahwa faktor risiko tidak berpengaruh terhadap kejadian penyakit.

Misalnya di dalam suatu kawanan terdiri dari 100 ekor sapi yang telah kita bahas di atas, dan terdapat 50 ekor sapi positif *Brucella* dalam kawanan itu. Misalnya kita ingin mengetahui apakah umur merupakan faktor yang mempengaruhi terjadinya *Brucella* positif pada sapi, dan kita dapat mengetahui apakah sapi muda atau tua. Untuk tujuan tersebut, maka kita dapat menghitung *odds ratio* yang memungkinkan kita menjawab pertanyaan apakah umur memiliki hubungan dengan prevalensi *Brucella*.

Misalnya terdapat 50 sapi muda dan 15 ekor diantaranya positif *Brucella*. Dengan demikian maka 50 ekor sapi lainnya adalah sapi tua dan 35 ekor diantaranya adalah positif *Brucella*.

Dari data tersebut kita dapat membuat tabel kontingensi sebagai berikut.

	Tua	Muda	
Positif <i>Brucella</i>	35	15	<b>50</b>
Negatif <i>Brucella</i>	15	35	<b>50</b>
	<b>50</b>	<b>50</b>	<b>100</b>

Perhitungan nilai RR:

Peluang bruselosis pada sapi tua dan muda adalah:

$$P(\text{bruselosis pada sapi tua}) = \frac{\text{jumlah kasus}}{\text{jumlah sapi tua}} = \frac{35}{50} = 0,7$$

$$P(\text{bruselosis pada sapi muda}) = \frac{\text{jumlah kasus}}{\text{jumlah sapi muda}} = \frac{15}{50} = 0,3$$

Risiko relatif (RR) adalah rasio dari dua peluang:

$$RR = \frac{P(\text{bruselosis pada sapi tua})}{P(\text{bruselosis pada sapi muda})} = \frac{0,7}{0,3} = 2,33$$

Nilai RR = 2,33 berarti bahwa risiko terjadinya antibodi bruselosis positif pada sapi tua adalah 2,33 kali lebih besar dibandingkan dengan pada sapi muda. Untuk mengetahui apakah nilai ini bermakna dalam menunjukkan adanya asosiasi antara umur sapi terhadap kejadian bruselosis, kita harus menghitung selang kepercayaannya. Dengan menggunakan bantuan EpiTools (demo penggunaannya ditunjukkan di video), diperoleh nilai untuk selang kepercayaan 95% adalah 1.47 sampai 3.7. Nilai 1 tidak tercakup pada selang tersebut sehingga dapat disimpulkan bahwa terdapat asosiasi antara umur sapi dengan antibodi bruselosis positif.

Perhitungan nilai OR:

Kita kemudian dapat menghitung *odds* *Brucella* baik pada sapi tua maupun muda.

$$\text{odds bruselosis pada sapi tua} = \frac{\text{jumlah kasus}}{\text{jumlah bukan kasus}} = \frac{35}{15} = 2,33$$

$$\text{odds bruselosis pada sapi muda} = \frac{\text{jumlah kasus}}{\text{jumlah non kasus}} = \frac{15}{35} = 0,43$$

Secara sederhana, *Odds ratio* adalah rasio dari kedua *odds* ini:

$$\text{odds ratio} = \frac{\text{odds bruselosis pada sapi tua}}{\text{odds bruselosis pada sapi muda}} = \frac{2,33}{0,43} = 5,44$$

Hal ini ditafsirkan sebagai *odds Brucella* pada sapi tua dibandingkan dengan sapi muda adalah 5,44. Artinya adalah *odds* terjadinya positif *Brucella* lebih tinggi (544% lebih tinggi) pada sapi tua. Tentu saja penghitungan selang kepercayaan untuk *odds ratio* penting dalam melakukan penafsiran. Nilai selang kepercayaan 95%-nya adalah 2.31 sampai 12.81, tidak melewati 1. Ini menandakan bahwa pertambahan umur berasosiasi dengan peningkatan *odds* terjadinya antibodi positif *Brucella*. Jika *odds ratio* mendekati 1 (dan selang kepercayaan melewati 1), kita akan mengasumsikan bahwa umur tidak berdampak kepada positif *Brucella*.

### Interpretasi terhadap ukuran asosiasi penyakit

Ukuran asosiasi yang signifikan (yaitu ukuran asosiasi yang tidak sama dengan 1 dan selang kepercayaan tidak melewati 1) menandakan adanya asosiasi antara faktor risiko yang dicurigai terhadap keberadaan suatu penyakit.

Hal itu tidak membuktikan bahwa faktor risikolah yang menyebabkan penyakit tersebut. Untuk menunjukkan bahwa memang faktor risiko itulah yang menyebabkan penyakit, maka kita

memerlukan sejumlah kriteria lainnya, seperti masuk akal secara biologis dan apakah paparan terjadi sebelum penyakit berkembang, dan sebagainya. Namun, suatu asosiasi statistika yang dikombinasikan dengan pemahaman terhadap sistem dan penalaran biologis dapat mengidentifikasi penyebab penyakit atau faktor risiko penyakit.

#### Langkah 4: Pengujian hipotesis statistika

##### a. Ukuran asosiasi dan pengujian hipotesis

Terdapat dua cara umum dalam pengujian hipotesis dengan menggunakan metode standar dalam epidemiologi veteriner, yaitu dengan menggunakan ukuran asosiasi atau pengujian hipotesis statistika. Pada bagian ini kita akan menggunakan ukuran asosiasi untuk menjawab pertanyaan yang kita miliki. Kita akan menggunakan pendekatan pengujian hipotesis statistika yang lebih tradisional di dalam studi kasus berikutnya (studi kasus 3). Dalam studi kasus 3 juga akan disediakan satu set informasi latar belakang yang lengkap untuk memperkenalkan prinsip-prinsip pengujian hipotesis statistika.

Dalam studi kasus ini kita akan membuat sebuah hipotesis, menghitung ukuran asosiasi (dan selang kepercayaan), serta menafsirkan hasilnya.

Pertanyaan kita adalah apakah terdapat perbedaan prevalensi diare di saat musim kemarau dan penghujan. Pertama kali kita harus menentukan hipotesis nol. Hipotesis nol selalu menyatakan tidak ada pengaruh. Hipotesis nol ditolak atau diterima berdasarkan nilai ukuran asosiasi (dan selang kepercayaannya).

*Hipotesis nol =  $H_0$  = Tidak ada perbedaan antara odds diare pada musim kemarau dan penghujan.*

Hipotesis alternatif pada umumnya adalah hipotesis yang bertentangan dengan hipotesis nol. Hipotesis alternatif diterima jika Anda memiliki cukup bukti secara statistika untuk menolak hipotesis nol. Hipotesis alternatif pada studi kasus ini adalah:

*$H_A$  = Terdapat perbedaan antara odds diare pada musim kemarau dan penghujan.*

#### **Latihan 14: Pengukuran asosiasi antara diare dan musim.**

Hitung ukuran asosiasi dan selang kepercayaannya antara diare dan musim.

##### **Petunjuk:**

Hitung risiko relatif dan *odds ratio* menggunakan data tabel kontingensi dua arah tentang musim dan mencret dari Latihan 12.

Kita akan menggunakan Epitools untuk latihan ini (<http://epitools.ausvet.com.au/content.php?page=home>). Silakan buka laman Epitools. Juga perhatikan video tayangan Epitools secara terperinci dan bagaimana menggunakan Epitools untuk menghitung ukuran asosiasi.

**Latihan 14: Ukuran asosiasi antara diare dan musim.**

Tuliskan hasil yang diperoleh dari epitools pada tempat yang disediakan berikut ini.

Berikut ini kita akan menghitung risiko relatif dan *odds ratio* secara manual untuk membantu Anda memahami latihan ini. Di bawah ini adalah tabel kontingensinya.

	<b>Musim Kemarau</b>	<b>Musim Penghujan</b>
<b>Diare</b>	246	544
<b>Lainnya</b>	875	1437
<b>Total sapi</b>	<b>1121</b>	<b>1981</b>
<b>Prevalensi mencret pada sapi sakit (%)</b>	22	27

P(diare di musim penghujan):

$$544/1981 = 0,27$$

P(diare di musim kemarau):

$$246/1121 = 0,22$$

Risiko relatif:

$$0,27/0,22 = 1,25$$

*Odds* diare di musim penghujan:

$$544/1437 = 0,3786$$

*Odds* diare di musim kemarau:

$$246/875 = 0,2811$$

*Odds ratio*

$$0,3786/0,2811 = 1,35$$

Dari hasil analisis data dengan epitools diperoleh bahwa selang kepercayaan 95% untuk RR adalah 1,10 sampai 1,43, dan selang kepercayaan 95% untuk OR adalah 1,13 sampai 1,60. Ini

berarti bahwa ada hubungan antara musim dan diare. Dengan demikian maka kita menolak hipotesis nol dan menerima hipotesis alternatif, yaitu musim mempengaruhi proporsi diare pada sapi yang disurvei secara pasif. Namun perlu dicatat bahwa hal ini hanyalah hubungan secara statistika antara musim penghujan dan diare, dan tidak membuktikan bahwa musim penghujanlah yang menyebabkan diare, Hal ini hanya menunjukkan bahwa ada kemungkinan bahwa musim penghujan dapat menyebabkan diare. Diperlukan investigasi dan kajian lebih lanjut untuk menentukan faktor-faktor apa saja yang sebenarnya menyebabkan penyakit.

## Rangkuman studi kasus 2

Pada studi kasus ini Anda telah melakukan beberapa langkah penting dalam analisis data, yaitu:

1. Tujuan analisis data telah disajikan untuk Anda
2. Anda melakukan manajemen data (menyimpan salinan data asli, menelaah apakah terdapat kesalahan pada data, dan membuat data baru).
3. Anda mendeskripsikan data
4. Anda melakukan pengujian hipotesis menggunakan risiko relatif, *odds ratio* dan selang kepercayaan.

Pada studi kasus ini Anda mendeskripsikan data dengan membuat tabel kontingensi (yaitu *pivot table* dalam Excel) dan dengan membuat *diagram pencar* (*scatterplot*) dan garis suai terbaik (*lines of best fit*). Langkah-langkah ini hampir cukup untuk menjawab pertanyaan penelitian Anda mengenai apakah terdapat perbedaan kejadian diare pada musim hujan dan kemarau. Dari plot yang telah dibuat, tampak terdapat penurunan diare di musim kemarau. Kemudian kita melakukan pengujian hipotesis menggunakan ukuran asosiasi dan selang kepercayaannya untuk menghitung pengaruh keragaman sampel, sehingga kita yakin bahwa hasil yang ditunjukkan oleh tabel dan plot kemungkinan besar merupakan pengaruh yang sebenarnya.

**Ini adalah akhir studi kasus 2**

## **Studi kasus 3: Swasembada daging sapi (berdasarkan statistika pemotongan sapi)**

### **Pengantar ke studi kasus**

Jumlah, jenis, dan tanggal pemotongan ternak di RPH di Indonesia telah dicatat dalam iSIKHNAS. Kita akan menggunakan data ini dalam analisis studi kasus yang akan kita pelajari ini.

Dalam studi kasus ini kita akan menentukan apakah ada perubahan jumlah pada berbagai jenis sapi yang dipotong dalam masa pengumpulan data iSIKHNAS (misalnya antara sapi betina produktif dan sapi impor Australia) . Kita dapat membuat kesimpulan sementara mengenai upaya pemerintah Indonesia dalam mencapai swasembada daging sapi berdasarkan data perubahan jumlah penyembelihan berbagai kategori sapi ini.

Seperti telah dijelaskan pada dua studi kasus sebelumnya bahwa bahwa data yang digunakan dalam pelatihan ini berasal dari data iSIKHNAS pada awal 2014. Data ini dikumpulkan ketika iSIKHNAS baru beroperasi selama sekitar satu tahun dengan area cakupan terbatas di Indonesia. Dengan ketidaklengkapan data ini, berarti belum ada kesimpulan nyata yang dapat ditarik mengenai hasil analisis data yang dilaksanakan pada pelatihan ini. Namun, seiring waktu, akan tersedia data lebih lengkap agar Indonesia dapat melakukan analisis yang lebih baik.

### **Keahlian yang dikembangkan dalam studi kasus ini**

- Manajemen data.
- Meringkas data dengan menggunakan plot
- Memahami hipotesis nol dan hipotesis alternatif
- Pengujian hipotesis menggunakan uji statistika

### **Daftar file yang digunakan untuk studi kasus**

#### *Data*

Abattoir\_April\_2014.xlsx

#### *Video*

Case study 3\_Exercise 16.avi

Case study 3\_Exercise 17.avi

Case study 3\_Exercise 18.avi

### Catatan: Konsep kunci mengenai pengujian hipotesis dan studi kasus 3

Pengujian hipotesis statistika merupakan bagian penting dalam analisis data. Dalam studi kasus ini kita akan memperkenalkan pendekatan uji hipotesis formal dengan menggunakan contoh data statistika pemotongan sapi di RPH di Indonesia.

Langkah-langkah uji hipotesis adalah:

1. Menyusun pertanyaan atau tujuan dari analisis data
2. Membuat hipotesis nol dan hipotesis alternatif
3. Menentukan uji statistika yang sesuai
4. Menghitung nilai statistik uji
5. Menentukan wilayah penolakan atau penerimaan hipotesis nol.
6. Menentukan nilai peluang statistik uji yang diperoleh kemungkinan jika hipotesis nol benar
7. Menolak atau menerima hipotesis nol
8. Membuat kesimpulan

Penjelasan untuk setiap langkah dalam pengujian hipotesis tersebut adalah sebagai berikut.

#### Menyusun pertanyaan (tujuan)

Salah satu cara untuk mengetahui apakah produksi sapi dalam rangka swasembada daging sapi menunjukkan peningkatan adalah dengan melihat data statistika penyembelihan di RPH. Anda dapat meneliti perubahan jumlah pemotongan pada berbagai kategori sapi pada suatu kisaran waktu tertentu. Jika jumlah pemotongan pada beberapa kategori seperti sapi betina reproduktif atau sapi impor Australia menurun, maka hal ini bisa menunjukkan bahwa kita telah semakin dekat ke arah swasembada. Diperlukan pendekatan uji hipotesis formal untuk menelaah permasalahan ini.

Dengan asumsi bahwa perubahan pola statistika pemotongan dapat menunjukkan ke arah keberhasilan swasembada, maka pertanyaan penelitiannya adalah: *apakah jumlah sapi yang dipotong untuk masing-masing kategori sapi adalah tetap dari waktu ke waktu?*

#### Membuat hipotesis nol dan hipotesis alternatif

Tahap awal pengujian hipotesis adalah menentukan hipotesis nol. Hipotesis nol adalah hipotesis yang menyatakan tidak ada perbedaan pengaruh antarperlakuan. Penolakan atau penerimaan hipotesis nol adalah berdasarkan nilai statistik uji dan peluangnya (atau ukuran asosiasi dan selang kepercayaan).

*Hipotesis nol = Tidak ada asosiasi antara waktu (bulan) terhadap jumlah pemotongan berbagai kategori sapi.*

Hipotesis alternatif adalah hipotesis yang bertentangan dengan hipotesis nol. Hipotesis alternatif diterima jika Anda memiliki cukup bukti secara statistika untuk menolak hipotesis

nol. Sebaliknya, jika tidak terdapat bukti yang cukup untuk menolak hipotesis nol, maka hipotesis nol diterima. Hipotesis alternatif pada studi kasus ini adalah:

$H_A = \text{Terdapat pengaruh waktu (bulan) terhadap jumlah pemotongan berbagai kategori sapi.}$

### Menentukan uji statistika yang sesuai

Terdapat banyak uji statistika yang dapat digunakan untuk menganalisis data yang kita miliki, tetapi hanya ada uji yang terbaik. Untuk memberikan gambaran kepada Anda tentang bagaimana memilih uji statistika, Anda dapat menyimak file presentasi PowerPoint ‘uji mana yang dipakai.ppt’ yang ditulis oleh Dr Brendan Cowled dan Dr Nigel Perkins dari AusVet. Presentasi tersebut akan memberi Anda banyak informasi tentang hal-hal yang perlu dipertimbangkan ketika memilih uji statistika yang sesuai.

Jenis data yang akan kita analisis pada studi kasus ini adalah data kategorik. Dalam hal ini, sapi-sapi yang dipotong dikelompokkan menurut jenis sapi, sehingga merupakan data kategorik. Sapi-sapi tersebut juga dikelompokkan menurut bulan pemotongan, dan juga merupakan data kategorik. Karena kedua kelompok data tersebut berjenis kategorik, maka uji statistika yang cocok adalah uji khi-kuadrat.

### Menghitung statistik uji

Pengujian secara statistika tergantung pada penghitungan suatu statistik uji. Kami akan menunjukkan cara penghitungan suatu statistik uji khi-kuadrat. Dalam penghitungan statistik ini kita hanya akan menggunakan sebagian dari keseluruhan data dari studi kasus ini, yaitu hanya data dari dua bulan dan dua jenis sapi. Kemudian pada saat latihan Anda akan menggunakan gugus data lengkap untuk menghitung statistik uji.

- a. Langkah pertama adalah membuat sebuah tabel kontingensi yang membagi data menjadi beberapa kelompok menurut jenisnya masing-masing.

Berikut ini adalah tabel yang berisi data iSIKHNAS. Tabel ini adalah suatu tabel kontingensi 2 x 2 untuk data jumlah sapi Australia atau sapi Indonesia produktif yang disembelih pada November dan Desember 2013.

*Data hasil pengamatan*

Jenis sapi	Bulan	
	11-2013	12-2013
Sapi Australia	3962	6988
Sapi betina produktif Indonesia	654	958

Kemudian pada tabel kontingensi tersebut, tambahkan baris dan kolom untuk jumlah marjinal baris/kolom, serta hitunglah total keseluruhan

Jenis sapi	Bulan		Total Baris
	11-2013	12-2013	
Sapi Australia	3962	6988	10950

<b>Sapi betina produktif Indonesia</b>	654	958	1612
<b>Total kolom</b>	4616	7946	Total keseluruhan= 12562

b. Langkah berikutnya adalah menghitung nilai harapan.

Nilai harapan adalah nilai-nilai yang kita harapkan jika tidak ada asosiasi antara peubah yang kita teliti, atau dengan kata lain jika data yang diamati menyebar menurut sebaran khi-kuadrat. Jika tidak ada hubungan antara peubah-peubah yang diteliti, maka perbedaan antara nilai harapan dan nilai observasi akan kecil. Namun, jika nilai harapan sangat berbeda dengan nilai observasi, maka ini menunjukkan bahwa ada hubungan antara kedua variabel yang kita teliti.

Nilai harapan dihitung dengan rumus sebagai berikut:

$$\text{Nilai harapan suatu sel} = \frac{\text{total baris} * \text{total kolom}}{\text{total keseluruhan}}$$

Rumus ini dihitung pada setiap sel data dalam tabel kontingensi, bukan pada total marjinal atau total keseluruhan.

*Nilai harapan*

	Bulan		Total baris
Jenis sapi	11-2013	12-2013	
<b>Sapi Australia</b>	$\frac{10950 * 4616}{12562} = 4024$	$\frac{10950 * 7946}{12562} = 6926$	10950
<b>Sapi betina produktif lokal</b>	$\frac{1612 * 4616}{12562} = 592$	$\frac{1612 * 7946}{12562} = 1020$	1612
<b>Total kolom</b>	4616	7946	Total keseluruhan = 12562

c. Kemudian kita menghitung statistik khi-kuadrat ( $\chi^2$ )

Khi-kuadrat dihitung dengan rumus sebagai berikut:

$$\chi^2 = \sum \frac{(\text{pengamatan} - \text{harapan})^2}{\text{harapan}}$$

Fokus rumus ini adalah menghitung seberapa jauh perbedaan antara nilai pengamatan dan nilai harapan dalam setiap sel.

Misalnya untuk data pada sel pertama (sapi dari Australia yang disembelih pada November 2013), menggunakan rumus sebagai berikut:

Nilai observasi = 3962, nilai harapan = 4024

Nilai  $\chi^2$  untuk sel ini adalah:  $\frac{(\text{pengamatan} - \text{harapan})^2}{\text{harapan}} = \frac{(3962 - 4024)^2}{4024} = 0,955$

Perhitungan ini dilakukan untuk tiap data di seluruh sel yang ada, kecuali data total marginal dan total keseluruhan, dan kemudian seluruh hasilnya dijumlahkan. Lambang  $\Sigma$  berarti penjumlahan.

Maka total  $\chi^2$  adalah:  $0,955+0,555+6,493+1,939=9,942$

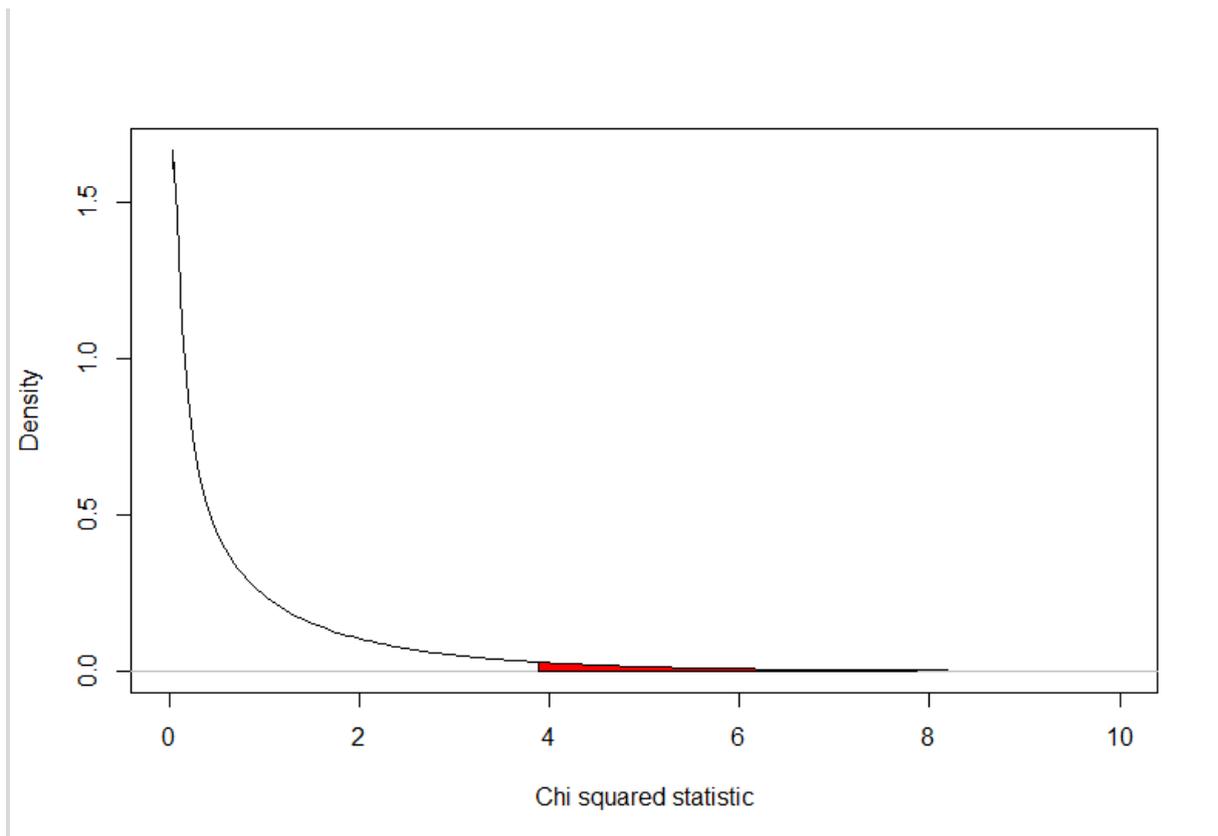
### **Menentukan wilayah penolakan atau penerimaan hipotesis nol**

Andaikan Anda melakukan pengambilan data berulang-ulang (beberapa kali *sampling*) dengan jumlah pengulangan yang cukup besar pada suatu populasi dimana tidak ada hubungan antara dua variabel yang diteliti (hipotesis nol benar). Kemudian setiap statistik uji yang dihitung dari tiap sampel diplot dalam sebuah fungsi sebaran peluang. Adanya keragaman pengambilan sampel menyebabkan semua hasil uji tidak sama dan membentuk suatu sebaran.

Gambar 7 menyajikan distribusi khi-kuadrat yang disertai dengan nilai titik kritik. Sebaran ini adalah sebaran khi-kuadrat untuk ilustrasi data yang kita gunakan jika hipotesis nol adalah benar.

Berdasarkan sebaran ini, maka kita dapat menetapkan nilai kritik yang membagi sebaran data menjadi dua bagian. Nilai kritik ini ditentukan sehingga 95% dari nilai statistik uji adalah lebih rendah daripada nilai kritik tersebut jika hipotesis nol benar. Dengan demikian, hanya tersisa 5% nilai dari sebaran tersebut yang melebihi nilai kritik jika hipotesis nol benar.

**Gambar 7: Fungsi sebaran peluang khi-kuadrat untuk 1 derajat bebas. Daerah yang diarsir berwarna merah adalah wilayah penolakan hipotesis nol (nilai khi-kuadrat lebih besar dari 3,84). Nilai 3,84 adalah nilai kritik.**



**Gambar 8: Fungsi sebaran peluang khi-kuadrat untuk 1 derajat bebas. Daerah yang diarsir berwarna merah adalah wilayah penolakan hipotesis nol (nilai khi-kuadrat lebih besar dari 3,84). Nilai 3,84 adalah nilai kritis.**

Nilai kritik ini biasanya dicantumkan dalam buku teks atau dihitung menggunakan perangkat lunak dan dapat dicari atau ditafsirkan secara otomatis untuk Anda oleh perangkat lunak tersebut. Tabel nilai kritik untuk sebaran  $\chi^2$  terlampir.

Kita dapat memutuskan menolak atau menerima hipotesis nol dengan membandingkan  $\chi^2_{hitung}$  dengan nilai kritisnya ( $\chi^2_{tabel}$ ). Jika  $\chi^2_{hitung} > \chi^2_{tabel}$ , diputuskan untuk menolak hipotesis nol yang berarti ada asosiasi antara kedua variabel, dan sebaliknya jika  $\chi^2_{hitung} < \chi^2_{tabel}$ , diputuskan untuk menerima hipotesis nol yang berarti tidak ada asosiasi antara kedua variabel.

Dalam latihan ini titik kritiknya adalah  $\chi^2=3,84$ .

### **Menentukan nilai peluang dari suatu statistik uji jika hipotesis nol benar**

Kemudian kita akan membandingkan nilai dugaan kita ( $\chi^2=9,9$ ) dan melihat apakah nilai tersebut lebih besar dari nilai kritik khi-kuadrat, yaitu sebesar 3,84. Ternyata memang nilai dugaan yang kita peroleh adalah lebih besar dari titik kritik. Pada kenyataannya bahwa nilai peluang nilai khi-kuadrat observasi sama besar dengan nilai khi-kuadrat jika hipotesis nol benar adalah sangat kecil (mendekati 0). Secara umum, kita selalu menghitung nilai p berdasarkan nilai statistik uji. Jika p kurang dari 0,05, maka kita akan menolak hipotesis nol.

### **Menolak atau menerima hipotesis nol**

Dari hasil yang kita peroleh, kita menolak hipotesis nol dan menerima hipotesis alternatif. Dengan demikian dapat disimpulkan bahwa terdapat hubungan antara jenis sapi yang disembelih dan bulan penyembelihan. (**Membuat inferensia mengenai populasi yang diteliti**)

Dari pengujian hipotesis yang telah dilakukan, kita menolak hipotesis nol dan memilih hipotesis alternatif. Dengan demikian kita menyimpulkan bahwa seiring waktu terdapat efek atau perbedaan dalam proporsi jenis sapi yang dipotong.

Kita dapat menguji perbedaan antara nilai-nilai observasi dan nilai harapan untuk menentukan hubungan alamiah dari variabel tersebut dan membuat kesimpulan lebih lanjut.

Dari pengujian hipotesis yang telah dilakukan, Tampak bahwa pada November lebih sedikit sapi Australia dan lebih banyak sapi betina produktif yang dipotong, dan kebalikannya terjadi pada Desember. Hal ini mungkin disebabkan adanya pembukaan kembali impor ternak sapi Australia ke Indonesia. Artinya, adanya sapi impor memberikan lebih sedikit tekanan untuk menyembelih sapi betina produktif. Alternatif lain adalah mungkin kesimpulan kita tentang alasan mengapa terdapat perbedaan yang signifikan tersebut adalah salah karena kita tidak memahami sistem produksi dengan sempurna. Mungkin saja perbedaan tersebut terjadi karena perubahan musiman atau tahunan.

Hal ini menunjukkan bahwa akal sehat dan pemahaman tentang sistem yang diteliti adalah penting dalam menginterpretasikan hasil-hasil pengujian statistika.

### **Kesalahan dalam penolakan/penerimaan hipotesis**

Kesalahan dalam penolakan/penerimaan hipotesis mungkin saja terjadi.

Kesalahan jenis pertama yang mungkin terjadi adalah ketika kita menolak hipotesis nol padahal kenyataannya hipotesis nol tersebut benar (kesalahan  $\alpha$ ).

Seperti pada latihan sebelumnya, kita telah menentukan titik kritik penolakan hipotesis nol pada sebaran khi-kuadrat bagian atas (yaitu pada titik di mana 95% nilai-nilai khi-kuadrat berada di bawah titik kritik ketika hipotesis nol benar). Namun, kita mengetahui bahwa pada sebaran khi-kuadrat tersebut, 5% dari nilai  $\chi^2$  akan melebihi titik kritik walaupun hipotesis nol benar. Jika nilai statistik uji yang kita peroleh lebih besar dari titik kritik, maka kita akan menolak hipotesis nol dengan peluang 95% (misalnya 95 kali dari 100 kali pengulangan pengambilan sampel yang sama). Namun mungkin saja sampel yang kita peroleh adalah termasuk ke dalam peluang 5% nilai hipotesis nol ada pada nilai-nilai tersebut, sehingga sebenarnya hipotesis nol benar. Dengan demikian, kita mungkin salah dalam menolak hipotesis nol dengan peluang 5% (misalnya 5 kali dari 100 kali pengulangan pengambilan sampel yang sama). Jadi, kita yakin dengan tingkat kepercayaan 95% mengenai kesimpulan-kesimpulan yang kita ambil dalam kaitannya dengan penolakan hipotesis nol.

Jenis kesalahan kedua adalah ketika kita menerima hipotesis nol padahal sebenarnya hipotesis nol salah (kesalahan  $\beta$ ). Cara untuk mengurangi jenis kesalahan ini adalah dengan meningkatkan besaran sampel sebanyak mungkin. Kuasa (kehandalan) suatu studi adalah seberapa baik studi tersebut dapat menghindari kesalahan-kesalahan seperti itu, atau dengan

kata lain seberapa baik studi tersebut mendeteksi adanya suatu pengaruh/asosiasi jika memang terdapat pengaruh/asosiasi antar variabel yang diteliti.

Berikut ini adalah latihan untuk menunjukkan bagaimana cara melakukan pengujian hipotesis statistika. Dalam latihan ini kita akan menggunakan seluruh data pemotongan di RPH, bukan hanya sebagian data seperti yang telah kita lakukan sebelumnya.

## Langkah-langkah analisis data Studi Kasus 3 (Latihan)

### Langkah 1: Tujuan

#### Latihan 15:

Bentuk kelompok kecil dan bahas tujuan dari studi kasus kita. Apakah hipotesis yang telah kita buat mendukung tujuan analisis?

Sampaikan kesimpulan kelompok Anda kepada seluruh kelas.

#### Petunjuk:

Fokus yang lebih luas dari analisis data dalam studi kasus ini adalah menjawab hal penting bagi Indonesia, yaitu pencapaian swasembada daging sapi. Salah satu cara untuk mengetahui apakah swasembada daging sapi meningkat adalah dengan meneliti statistika penyembelihan.

Hipotesisnya adalah:

Hipotesis nol: Tidak ada hubungan antara bulan dan jenis sapi yang dipotong. Dengan kata lain, tidak ada pengaruh bulan terhadap jumlah total pemotongan pada berbagai jenis sapi yang dipotong.

Hipotesis alternatif: Terdapat hubungan antara bulan dan jenis sapi yang dipotong. Dengan kata lain, terdapat pengaruh bulan terhadap jumlah total pemotongan pada berbagai jenis sapi yang dipotong. Apakah hipotesis-hipotesis tersebut memadai?

#### Latihan 15: Apakah hipotesis mendukung tujuan analisis data?

Tuliskan catatan singkat dari diskusi kelompok pada topik ini.

## Langkah 2: Manajemen data

Simpan salinan data Anda sebelum diubah!

### Latihan 16: Soal (buat *pivot table* yang berisi statistika pemotongan berdasarkan bulan).

#### Petunjuk:

Pilih seluruh data sapi dari worksheet Excel berjudul "Abattoir 2014.xls" dan simpan di dalam worksheet baru. Buka data pemotongan sapi, lakukan pemeriksaan kesalahan, dan buat sebuah tabel kontingensi (*pivot table* dalam Excel).

Lihatlah video (Studi Kasus 3, Latihan 16). Anda telah melakukan beberapa hal ini dalam latihan sebelumnya.

### Latihan 16: Jawaban (buat *pivot table* yang berisi statistika pemotongan berdasarkan bulan).

Tuliskan *pivot table* yang Anda telah buat di pada tempat yang tersedia di bawah ini.

## Langkah 3: Deskripsi data

### Latihan 17: Soal (plot ringkasan data)

Buatlah beberapa plot ringkasan data untuk tabel kontingensi. Tujuannya adalah untuk memahami data sebelum pengujian hipotesis dilakukan.

**Petunjuk:** Saksikan tayangan video (Studi Kasus 3, Latihan 17)

**Latihan 17: Jawaban (plot ringkasan data)**

Tulislah plot ringkasan data yang telah Anda buat pada tempat yang tersedia di bawah ini.

**Langkah 4: Pengujian hipotesis statistika**

**Latihan 18: Soal (analisis khi-kuadrat untuk data pemotongan sapi)**

Lakukan uji statistika khi-kuadrat pada *pivot table* yang telah dibuat.

**Petunjuk:** Saksikan tayangan video (Studi Kasus 3, Latihan 18) dan video penggunaan Epitools untuk melakukan uji khi-kuadrat.

**Latihan 18: Jawaban (analisis khi-kuadrat untuk data pemotongan sapi)**

Sertakan hasil uji khi-kuadrat Anda di pada tempat yang tersedia di bawah ini.

**Latihan 19: Soal (inferensia)**

Bentuk kelompok kecil dan bahas hasil uji khi-kuadrat yang signifikan. Apa artinya? Buatlah beberapa inferensia.

**Petunjuk:**

Lihat hasil pengujian statistika, dengan cara melihat plot ringkasan pada latihan 17 dan bandingkan nilai observasi dan nilai harapan pada latihan 18. Pada sel mana pada tabel kontingensi tersebut yang memiliki perbedaan antara nilai harapan dan nilai observasi yang besar sehingga memberikan kontribusi terbesar terhadap nilai khi-kuadrat? Itulah area yang memiliki pengaruh terbesar.

**Latihan 19: Jawaban (inferensia)**

Tuliskan catatan singkat hasil diskusi kelompok Anda mengenai inferensia hasil uji khi-kuadrat pada data pemotongan sapi.

### **Rangkuman studi kasus 3**

Dalam studi kasus ini, pertama-tama Anda diberikan informasi mengenai pendekatan pengujian hipotesis statistika. Hal ini mungkin agak rumit, tetapi pendekatan ini adalah pendekatan yang konsisten digunakan dalam berbagai studi. Setelah itu, kemudian Anda kembali melakukan beberapa langkah penting dalam analisis data, yaitu:

1. Memeriksa hipotesis untuk melihat apakah hipotesis tersebut mendukung tujuan penelitian
2. Manajemen data, yaitu dengan membuat salinan data asli, memeriksa jika terdapat kesalahan pada data, dan membuat data baru.
3. Deskripsi data dengan menggunakan tabel kontingensi dan plot
4. Melakukan pengujian hipotesis

Pada studi kasus ini Anda melakukan pengujian hipotesis dengan menggunakan pendekatan statistika formal yaitu uji khi-kuadrat. Langkah-langkah utama dalam pendekatan uji hipotesis ini dapat digunakan untuk hampir semua gugus data. Anda sekarang telah memiliki keterampilan dasar dalam melakukan pengujian hipotesis statistika ini. Seiring dengan berjalannya waktu dan bertambahnya pengalaman, maka pengetahuan Anda mengenai pemilihan uji statistik mana yang sebaiknya digunakan dan bagaimana menerapkannya akan semakin meningkat.

**Ini adalah akhir studi kasus 3.**

## Catatan kesimpulan

Analisis data adalah sangat penting dalam manajemen kesehatan hewan dan penyusunan kebijakan. Indonesia beruntung memiliki iSIKHNAS, sistem informasi kesehatan hewan yang baik dan mutakhir. Namun, sistem informasi terbaik di dunia ini tidak akan berguna jika tidak ada yang menggunakan data itu. Karena itulah, maka fokus pelatihan ini adalah untuk membantu Anda dalam memulai menggunakan data iSIKHNAS. Pendekatan ini dapat membantu Anda dalam peningkatan proses pengambilan keputusan kesehatan hewan.

Terdapat beberapa langkah standar dalam analisis data, yaitu menyusun pertanyaan penelitian (tujuan), manajemen data, deskripsi data, dan pengujian hipotesis. Semua langkah ini diajarkan dalam pelatihan ini. Untuk memantapkan keterampilan baru Anda ini, sebaiknya Anda mengaplikasikan langkah-langkah tersebut pada persoalan-persoalan yang tengah Anda teliti segera setelah Anda kembali ke tempat kerja Anda. Jika misalnya saja Anda menghabiskan satu atau dua hari per bulan selama beberapa bulan ke depan, maka keterampilan Anda dalam analisis data ini akan semakin meningkat.

Dalam pelatihan ini, kita hanya membahas sebagian kecil dari sejumlah pendekatan statistik yang tersedia (misalnya ukuran asosiasi dan uji khi-kuadrat). Anda dapat memperluas pengetahuan Anda di luar apa yang telah kita pelajari dalam pelatihan ini. Ada beberapa buku teks berguna yang dapat membantu Anda mengembangkan keterampilan statistika Anda. *Veterinary Epidemiologic Research* (Dohoo et al. 2009) adalah satu buku teks yang dapat anda gunakan yang mencakup epidemiologi veteriner dan menyajikan beberapa analisis statistika. *Statistics for Veterinary and Animal Science* (Petrie and Watson 2006) adalah buku teks yang menggunakan statistika secara lebih umum dalam ilmu veteriner. Jika Anda ingin mempelajari pendekatan R serta memperluas pengetahuan statistika Anda, Anda perlu membaca bagian lampiran mengenai R (Lampiran 1). *Introductory Statistics with R* (Dalgaard 2008) merupakan buku teks yang sangat berguna yang memperkenalkan baik ilmu statistik maupun R. Jika Anda mulai menggunakan R, maka dunia statistik akan mudah Anda genggam!

## Lampiran 1: Latihan lanjutan menggunakan R.

### Pendahuluan

Lampiran ini memperkenalkan pembaca kepada R, dan ditujukan kepada peserta yang kerap atau ingin melakukan sejumlah analisis yang substansial, dan ingin meningkatkan kemampuan mereka lebih daripada menggunakan Excel. Mungkin tidak terdapat cukup waktu menyelesaikan lampiran ini selama pelatihan. Anda dapat melanjutkan menyelesaikannya nanti setelah pelatihan ini.

R adalah perangkat lunak statistika dan grafik. Perangkat lunak ini dikenal luas dan digunakan di seluruh dunia. R sangat luwes dan sangat baik untuk melakukan uji-uji statistika yang rumit maupun sederhana, juga dalam pembuatan grafik. Perangkat lunak ini dapat diperoleh secara gratis dan terus diperbarui, serta didukung oleh para ilmuwan yang selalu menyumbangkan paket-paket baru yang memperluas kemampuan R. R akan bertahan lama, sehingga usaha yang Anda lakukan untuk mempelajari perangkat lunak ini tidak akan sia-sia dan tentu akan terus bermanfaat sepanjang karier Anda. Terdapat pula pengguna setia R di seluruh dunia sehingga banyak laman batuan online dan kelompok diskusi dengan kode gratis yang dapat memberikan saran-saran.

Lihat <http://www.r-project.org/> untuk keterangan lebih lanjut.

Baca bab pertama dari dokumen R-intro.pdf untuk memperluas pengetahuan Anda (6 halaman).

Secara singkat, jika Anda mengerjakan statistika dengan lebih serius, maka Anda memerlukan paket statistika profesional, bukan Excel. Terdapat sejumlah paket perangkat lunak yang tersedia (seperti SAS, Stata, dan R), namun R termasuk yang terbaik dan gratis. R agak lebih sulit digunakan dibandingkan Excel karena Anda perlu mempelajari beberapa perintah penulisan kode komputer, namun usaha Anda tidak akan sia-sia. Kami akan mulai dengan memperkenalkan perangkat lunak apa yang Anda perlukan. Kami kemudian akan membimbing Anda melalui studi kasus 1. Anda dapat mulai menggunakan R.

### Perangkat lunak yang perlu diunduh

Hanya ada dua bagian perangkat lunak **gratis** yang harus diunduh. Kedua perangkat lunak ini akan menjadi perangkat lengkap bagi seorang statistikawan. Perangkat lunak tersebut adalah R dan RStudio. Beberapa dari Anda mungkin perlu minta bantuan staf TI untuk membantu karena mungkin Anda tidak memiliki akses yang diperlukan pada komputer Anda.

#### 1. R

Harap unduh R dari sebuah *CRAN mirror*, lebih baik melalui:

<http://cran.repo.bppt.go.id/>

Agency for The Application and Assessment  
of Technology

<http://cran.unej.ac.id>

The University of Jember

Sebagian besar dari Anda akan mengunduh versi R untuk Windows, yang memang lebih disarankan kecuali jika Anda pengguna Mac. Setelah diunduh, jalankan file dan install program.

## 2. RStudio

RStudio adalah lingkungan pengembangan untuk R. Salah satu fitur yang berguna dari RStudio adalah Anda dapat membuat file R code yang dapat Anda simpan, arsipkan, dan kirim dengan mudah kepada R untuk diimplementasikan. Hal tersebut membuat penulisan dan menjalankan kode R lebih mudah. Intinya RStudio memungkinkan Anda membuat sebuah file R code, yang bisa Anda kerjakan menggunakan kemampuan penyuntingan teks dan mengirimkannya ke R untuk dijalankan.

Lihat <http://www.rstudio.com/ide/> untuk keterangan lebih lanjut.

Harap unduh RStudio dari <http://www.rstudio.com/ide/download/>.

### Latihan

1. Buka R studio
2. Saksikan video untuk Lampiran 1
3. Ulangi Studi Kasus 1 lagi, namun dalam R (menggunakan R Studio) bukan Excel.

### Pengembangan lanjutan R dan keterampilan statistika.

Beli buku perkenalan statistik R, baca dan lakukan latihan. *Introductory statistics with R* oleh Peter Dalgaard (Dalgaard, 2008) merupakan buku yang sangat baik

(<http://www.springer.com/statistics/computational+statistics/book/978-0-387-79053-4>).

Selesaikan membaca 'An introduction to R' (R-intro.pdf).

Masuk ke laman Quick-R (<http://www.statmethods.net/>)

Berlatihlah dan jangan menyerah! Anda akan menjadi lebih baik dan usaha Anda akan terbayarkan.

## Lampiran 2: Kunci jawaban latihan dalam panduan pelatihan analisis data

### Studi kasus 1: Indikator Kinerja Veteriner

#### Latihan 1: Menelaah data

Bentuk kelompok beranggotakan 2-4 orang dan lakukan latihan berikut. Tulislah jawaban di buku kerja.

**Telaah dengan baik kolom worksheet dan data dalam setiap kolom.**

Apa arti tiap kolom?

ID= Sebuah nomor ID yang unik.

Tanggallaporan = tanggal laporan = tanggal laporan yang diterima oleh DINAS.

Jumlahkasus = jumlah kasus = Jumlah kasus sindrom.

Sindrom = satu dari sindrom berikut: Keguguran dan pembengkakan sendi; Menggigit dan perubahan perilaku; Demam pada babi; Pincang, berliur, dan berlesi; Unggas mati dan mati mendadak.

Sindrom ini dirancang untuk mendeteksi beberapa penyakit seperti HPAI, rabies, PMK, Brucellosis, dan Anthrax

Namapengirim = nama pengirim = Dokter hewan yang mengirimkan laporan.

Provinsi = provinsi tempat sindrom terjadi.

Kabupaten = kabupaten tempat sindrom terjadi.

Kecamatann = kecamatan tempat sindrom terjadi

Desa = desa tempat sindrom terjadi

Tanggaldiinvestigasi = tanggal diinvestigasi = tanggal sindrom diinvestigasi

Stafteknisdinas = staf teknis yang menginvestigasi sindrom

Tipeinvestigasi = jenis investigasi.

Diagnosasementara = diagnosa sementara = diagnosa sementara hingga diagnosa lanjutan tersedia.

Ada berapa laporan sindrom prioritas di dalam worksheet?

Terdapat 108 baris maka terdapat 108 laporan. Baris pertama adalah baris judul.

### **Apakah hipotesis mendukung tujuan analisis?**

Kolom data mana yang menurut Anda paling penting untuk memenuhi tujuan kita? Sebagai petunjuk, dua kolom penting disoroti dalam Gambar 3.

Dua kolom penting adalah Tanggal laporan dan Tanggal diinvestigasi. Kedua kolom ini dapat digunakan bersama untuk menentukan durasi waktu tunggu antara sebuah laporan sindrom dan investigasi.

**Laporkan temuan Anda ke seluruh kelompok.**

### **Latihan 2: Membuat tujuan secara terperinci**

**Tentukan suatu tujuan analisis yang lebih terperinci.** Petunjuk: apakah selisih yang ditunjukkan dua kolom yang disoroti? Gunakan data tersebut untuk menyusun tujuan analisis yang spesifik.

Kita perlu menganalisis selisih antara tanggal laporan dan tanggal investigasi sindrom prioritas. Kita kemudian mencatatnya sebagai jumlah perkiraan jam antara diterimanya laporan dan dilakukannya investigasi.

**Laporkan kesimpulan Anda kepada kelompok dengan tujuan yang telah disempurnakan.**

Fasilitator kemudian akan memandu peserta ke arah tujuan akhir yang sama. Kita perlu menganalisis perbedaan antara tanggal laporan dan tanggal investigasi sindrom prioritas yang dicatat dalam jam. Kita kemudian mendeskripsikan perbedaan ini dan tentukan apakah ada perbedaan antara waktu tunggu investigasi pada rabies dan pada keguguran.

Kita kemudian akan mengambil inferensia apakah secara keseluruhan layanan veteriner dilakukan cukup cepat dan apakah ada perbedaan tingkat urgensi untuk rabies dan keguguran.

### **Latihan 3: Hitung data yang hilang menggunakan Excel**

Hitung jumlah titik data yang hilang dalam kolom Tanggal diinvestigasi di isikhnas\_priority disease syndromes\_March\_2013.csv. Petunjuk: gunakan fungsi filter di Excel untuk memilih hanya sel kosong (data hilang), kemudian menghitungnya, atau menggunakan fungsi countif di Excel.

Jawaban: Terdapat dua metode yang bisa digunakan: metode filter atau countif (fungsi).

#### **Metode filter**

Bukalah data anda, klik filter, kemudian pilih kolom Tanggal diinvestigasi dan klik panah filter, kemudian pilih *deselect all*, gulir ke bawah dan pilih “blanks”. Langkah ini adalah untuk

memilih hanya baris yang tak memiliki data dalam kolom, sehingga bisa dihitung kemudian. Terdapat sekitar 40 baris yang kosong.

### **Metode Countif**

Pindahkah kursor ke sebuah sel kosong. Masukkan rumus berikut:

```
=COUNTIF(J2:J109,"")
```

Tekan enter dan jumlah sel kosong akan dihitung.

J2:J109 berarti periksa semua sel dalam kolom Tanggaldinvestigasi. Angka ini perlu diubah jika data yang dianalisis adalah data isikhnas\_priority disease syndromes\_March\_2013.csv yang memiliki 108 entry/baris/investigasi penyakit.

"" merujuk pada fakta kita sedang mencari data kosong/hilang. Dengan kata lain, tanda kutip ganda berarti data tidak berisi apa-apa.

Saksikan video pelatihan terkait untuk Studi Kasus 1, Latihan 3.

### **Latihan 4: Bias seleksi dan bias informasi.**

Lakukan diskusikan kelompok mengenai jenis bias seleksi dan bias informasi apa yang mungkin ada di data iSIKHNAS.

#### **Jawaban:**

Dapatkah Anda memikirkan data investigasi sindrom yang mungkin secara rutin hilang dari gugus data? Apa dampaknya terhadap kesimpulan yang ditarik berdasarkan data yang Anda miliki? Itulah bias seleksi.

Contoh: Anda berasal dari provinsi yang memiliki sejumlah besar industri susu komersial. Produsen ini memiliki dokter hewan sendiri dan tidak memanggil dokter hewan dinas ketika mereka memiliki masalah keguguran. Dengan demikian maka Anda memiliki bias seleksi sehingga sindrom aborsi hanya dicatat jika terjadi pada peternak kecil. Pendugaan Anda tentang berapa banyak keguguran akan ber bias karena Anda tidak memiliki informasi tentang jumlah aborsi pada sapi perah (dan prevalensi ini berbeda dengan aborsi pada sapi pada peternakan kecil). Mungkin terdapat lebih banyak atau sedikit keguguran sapi di provinsi Anda, tergantung pada apakah industri susu memiliki lebih banyak atau sedikit kasus keguguran dibandingkan peternakan kecil.

Dapatkah Anda memperkirakan kesalahan apa yang mungkin sering terjadi ketika sedang menyelidiki kejadian sindromik dan membuat diagnosa sementara? Inilah bias informasi.

Contoh: Misalkan rabies belum pernah didiagnosis di provinsi Anda sebelumnya. Jika terdapat beberapa anjing dengan sindrom Gilgalak, maka petugas mendiagnosa anjing-anjing tersebut sebagai keracunan. Inilah bias informasi karena anjing-anjing itu sebenarnya terinfeksi rabies.

### **Latihan 5: Membuat data baru (variabel waktutunggu investigasi)**

Petunjuk: Saksikan tayangan video untuk membantu Anda.

### **Latihan 6: Menghitung ukuran pemusatan dan ukuran penyebaran variabel waktutunggu investigasi**

Untuk variabel waktutunggu investigasi di file *isikhnas\_priority disease syndromes\_March\_2013.csv*, hitung nilai rata-rata, median, simpangan baku, selang kepercayaan, wilayah dan wilayah antarkuartil.

Petunjuk: Gunakan perangkat *data analysis tool* di Excel.

### **JAWABAN**

Untuk melakukan analisis data seperti ini dengan mudah di Excel, maka kita dapat menggunakan opsi *data analysis tool* dari *Analysis ToolPaks*. Untuk menggunakan *Analysis ToolPak* ini, pertama-tama Anda harus memastikan bahwa *Analysis ToolPak* tersedia di Excel Anda.

Untuk memeriksanya, lakukanlah langkah-langkah sebagai berikut: File>options>add ins>Analysis ToolPak>Go>Analysis ToolPak- check>OK

Anda kemudian akan melihat sebuah tab di pita data di Excel.

Perhatikanlah tayang video sebagai petunjuknya.

Anda dapat menginterpretasikan statistik deskriptif ini sebagai berikut:

Rataan adalah 750 jam dan nilai tengah adalah 0 jam. Terdapat perbedaan yang besar antara nilai rata-rata dan median, yang mengindikasikan bahwa data menjulur ke kanan atau tidak menyebar normal. Jika data tidak menyebar normal, maka median adalah merupakan ukuran pemusatan yang lebih baik daripada rata-rata. Dari nilai median tampak bahwa selisih antara waktu investigasi dan pelaporan yang paling umum adalah 0 jam. Tentu saja ini menunjukkan layanan veteriner yang baik.

Adapun nilai simpangan bakunya adalah 1234 jam dan selang kepercayaan untuk nilai rata-rata adalah ( $\pm 294$  jam). Simpangan baku dan selang kepercayaan ini tidak berguna digunakan sebagai ukuran penyebaran pada data ini, karena data tidak menyebar normal sehingga simpangan baku dan selang kepercayaan tidak akan akurat untuk menggambarkan ukuran penyebaran. Ukuran penyebaran yang sebaiknya digunakan adalah wilayah antarkuartil. Nilai wilayah antarkuartil tidak tersedia pada *data analysis tool* di Excel, sehingga kita harus menghitungnya secara terpisah, yaitu dengan menggunakan fungsi kuartil di Excel.

Cara perhitungannya adalah sebagai berikut:

Ketik rumus berikut pada sel kosong dalam worksheet Syndromasprioritas. Lakukan dua perhitungan, yaitu satu untuk memperoleh nilai kuartil pertama, dan kedua untuk memperoleh nilai kuartil ketiga. Kemudian Anda dapat mengurangkan kuartil ketiga ke kuartil pertama untuk mendapatkan wilayah antarkuartil.

Perhitungan untuk memperoleh nilai kuartil pertama adalah:

=QUARTILE(N2:N109,1)

Artinya kita ingin memperoleh nilai kuartil pertama dari data dalam sel N2 hingga N109.

Perhitungan untuk memperoleh nilai kuartil ketiga adalah:

=QUARTILE(N2:N109,3)

Artinya kita ingin memperoleh nilai kuartil ketiga dari data dalam sel N2 hingga N109.

Jawabannya adalah 0 dan 1218 sehingga wilayah antarkuartil adalah  $1218 - 0 = 1218$  jam.

Kita memperoleh hasil bahwa 50% waktu tunggu dari saat pelaporan hingga dilakukan investigasi adalah antara 0 dan 1218 jam di sekitar median 0 jam. Median waktu tunggu dari saat pelaporan hingga dilakukan investigasi adalah 0 jam dan wilayah antarkuartilnya 1218 jam. Hasil ini menunjukkan dilakukan waktu yang cepat untuk menginvestigasi laporan sindrom prioritas, tetapi waktunya sangat beragam.

Keragaman dalam waktu investigasi ini mengkhawatirkan karena jika terjadi suatu sindrom yang penting, maka penyakit tersebut dapat menyebar sebelum dilakukan investigasi. Hasil deskripsi data ini menunjukkan bahwa ada sebagian kecil investigasi penyakit hewan prioritas yang memerlukan waktu yang sangat lama untuk dilakukan investigasi. Namun dari data yang diperoleh, tampaknya waktu yang lama antara pelaporan dan investigasi tersebut umumnya terjadi pada periode awal entri data ke dalam gugus data yang mungkin masalahnya berkaitan dengan masalah perekaman data. Pada pengamatan selanjutnya yang lebih baru, tampaknya tidak terlihat lagi adanya waktu investigasi yang membutuhkan waktu yang lama. Pemeriksaan lebih lanjut mengenai keakuratan data diperlukan untuk memahami masalah yang ada di dalam gugus data. Hal ini dapat dilakukan misalnya dengan berbicara kepada para investigator di lapangan.

**Jika Anda tidak memiliki analisis ToolPak:**

Masukkan rumus ini ke Excel secara terpisah untuk menghitung tiap parameter.

Rataan: =AVERAGE(N2:N109)

Simpangan baku: =STDEV.S(N2:N109)

Selang kepercayaan: =CONFIDENCE(0.05,1233.55307,70)

Median: =MEDIAN(N2:N109)

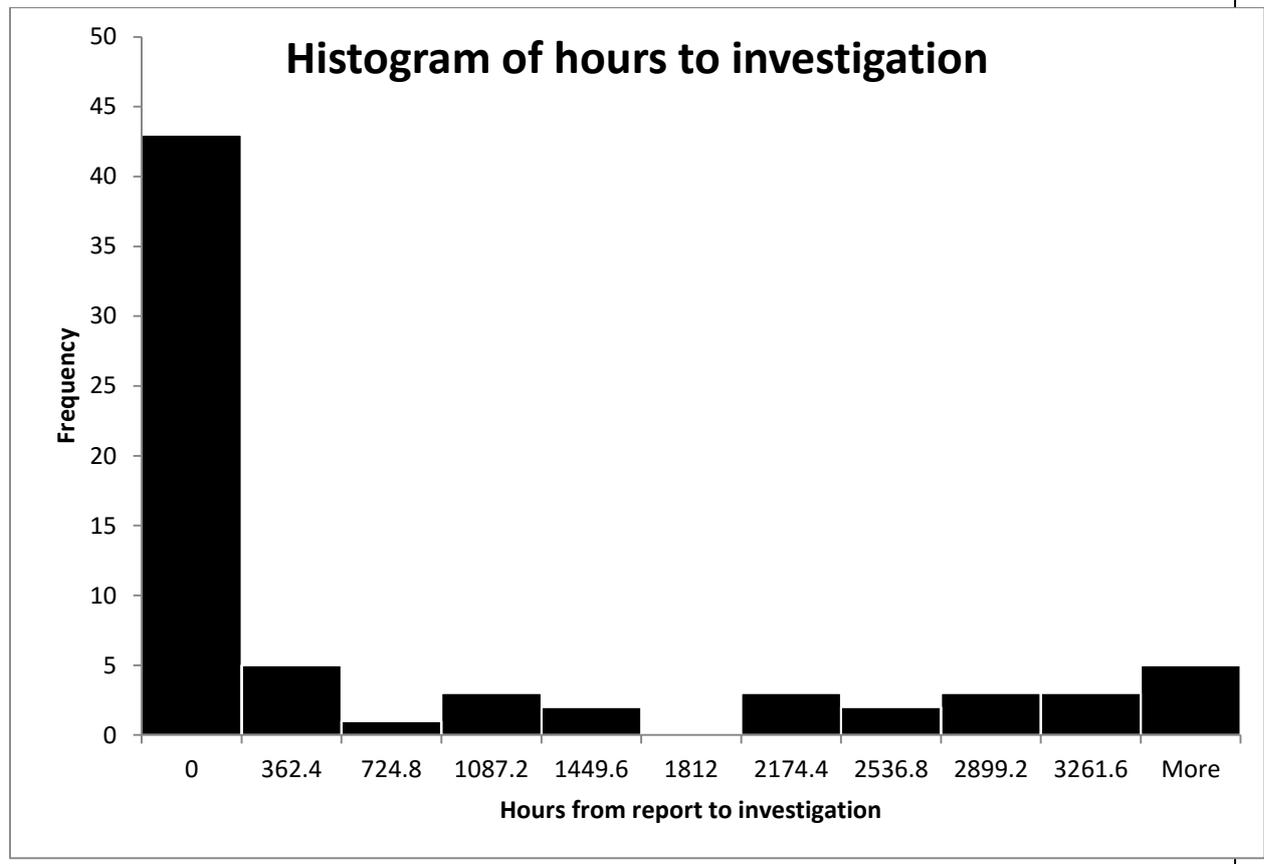
Q1: =QUARTILE(N2:N109, 1)

Q3: =QUARTILE(N2:N109, 3)

### Latihan 7: Histogram waktu tunggu investigasi dalam Excel

Buat sebuah histogram waktu tunggu investigasi.

Saksikan video untuk membantu Anda.



### Studi kasus 2: Sapi diare

Latihan 8: Tentukan variabel yang penting untuk memenuhi tujuan kita dalam file "Cattle clinical signs.csv".

Jawaban

Secara umum kita harus dapat menentukan jumlah kasus dan kapan ini terjadi. Selain itu untuk menghitung ukuran frekuensi penyakit kita juga perlu menentukan populasi berisiko. Kita juga perlu mempertimbangkan apa jenis bias yang mungkin ada dalam data.

### **Jumlah kasus:**

Terdapat satu kolom yang relevan yang bisa dipertimbangkan untuk menentukan kasus diare dalam populasi, yaitu kolom Tanda. Kolom diagnosis sementara terlalu terperinci dan tidak dapat mengenali kasus diare dengan mudah.

### **Populasi berisiko:**

Data ini dikumpulkan dari laporan penyakit dari peternak, sehingga merupakan data surveilans pasif. Kelemahan dari data surveilans pasif adalah datanya selalu berbias, sehingga sangat mungkin bahwa data yang kita kumpulkan dari iSIKHNAS akan memberikan perkiraan yang terlalu rendah mengenai jumlah hewan berisiko dan terinfeksi dalam populasi. Hal ini karena terdapat beberapa peternak yang tidak melaporkan kasus penyakit dan juga peternak tidak melaporkan hewan yang sehat. Namun demikian data ini masih dapat digunakan untuk menduga jumlah kejadian diare pada populasi target kita, karena meskipun datanya berbias dan perkiraan yang rendah terhadap jumlah sapi yang berisiko, kita dapat mengasumsikan bahwa pengurangannya bersifat konstan dan tidak berubah dari waktu ke waktu. Apabila perkiraan ini benar, maka kita masih dapat melihat perubahan jumlah diare pada sapi dengan akurat seiring waktu. Penduga yang pasti mengenai proporsi diare pada sapi mungkin tidak akurat, namun perubahan relatif dalam proporsi sapi diare tersebut akan akurat.

Jadi populasi berisiko dapat diasumsikan sebagai semua laporan sapi sakit dalam gugus data. Namun kita tahu bahwa pendugaan populasi ini berbias dan perkiraan yang rendah terhadap jumlah ternak dalam populasi tetapi masih memungkinkan untuk menentukan perubahan proporsi diare pada sapi dari waktu ke waktu.

### **Tanggal kejadian:**

Terdapat dua kolom tanggal, yaitu kolom Tanggal laporan dan Tanggal investigasi. Dari kedua kolom tersebut, kolom Tanggal laporan (tanggal pelaporan) mencerminkan waktu kejadian diare yang lebih akurat. Kita mengetahui sebelumnya dari Studi Kasus 1 bahwa terkadang ada penundaan laporan investigasi penyakit, meskipun laporan tersebut termasuk sindrom prioritas.

## **Latihan 9: Periksa kesalahan pada data dan pikirkan bias yang mungkin terjadi**

### **Jawaban**

Gunakan fungsi filter untuk mencari bagian kosong di tiap dua kolom data kita (Tanda dan Tanggal laporan). Tampaknya tidak terdapat bagian kosong.

Kemudian lihat kisaran tanggal dalam kolom Tanggallaporan. Periksa tanggal pertama dan terakhir dalam set data. Apakah tanggalnya masuk akal? Kisaran dari 1/5/2013 hingga 16/04/2014 tampaknya tidak masuk akal.

Periksa kategori gejala umum yang dilaporkan dalam kolom Tanda. Apakah masuk akal? Tampaknya ada beberapa keanehan. Misalnya ada banyak entri diare, misalnya diare terdaftar sebagai diare berdarah dan berbagai kategori diare lainnya.

Namun demikian, secara umum data tersebut baik dan tidak ada alasan yang jelas untuk tidak melanjutkan analisis data.

Kemungkinan bias yang utama adalah bias seleksi. Maksudnya adalah peternak dengan sapi sehat tidak akan melaporkannya kepada otoritas veteriner setempat sehingga menghasilkan perkiraan yang rendah terhadap denominator jika dilakukan pendugaan prevalensi diare.

### **Latihan 10: Data baru**

#### **Jawaban**

#### **Kolom bulan integer**

Agar dapat menganalisis data untuk melihat apakah terdapat perbedaan kejadian diare antara musim kemarau dan musim penghujan, maka kita perlu mengelompokkan tiap laporan gejala klinis ke dalam musim kemarau atau penghujan. Kita perlu melakukan beberapa langkah untuk melakukannya.

Data tanggal sulit digunakan dalam Excel. Maka kita harus membuat kolom baru yang berisi nilai integer (bilangan bulat) yang berhubungan dengan angka bulan (misalnya bulan 1-12, dengan 1 adalah Januari, 2 adalah Februari dan bulan 12 adalah Desember dan lain-lain.)

Langkah 1: format tanggal

Pertama periksa apakah Excel mengenali kolom Tanggallaporan sebagai sebuah kolom dengan format tanggal.

Langkah 2: Integer tanggal

Untuk membuat kolom bulan integer diperlukan rumus Bulan “=Month(serial\_number)”. Formula ini menghasilkan sebuah integer dari 1-12 tergantung pada data pada sel yang dirujuk. Namai kolom baru tersebut dengan 'integer Bulan', isi dengan rumus bulan dan tarik formula ke bawah di kolom baru tersebut sepanjang seluruh set data. Sebagai contoh, jika data tanggal ada di kolom B dan Anda ada di baris pertama di bawah judul maka formulanya: =Month(B2). Hal ini akan menghasilkan 1 jika data dalam sel B2 adalah Januari, 2 jika data dalam sel B2 adalah Februari, dan seterusnya.

Langkah 3: Pengelompokkan musim penghujan atau kemarau

Langkah selanjutnya adalah membuat kolom "musim" yang berisi data 'penghujan' atau 'kemarau' tergantung pada apakah bulan dari tanggal laporan jatuh pada musim penghujan atau kemarau. Untuk melakukannya kita menggunakan rumus 'If'. Anda mungkin telah mempelajari rumus seperti ini dalam pelatihan Excel sebelumnya. Namun kami telah menuliskan sebuah rumus untuk Anda di bawah ini. Pengkategorian ke dalam musim penghujan dan kemarau adalah, jika bulannya adalah 1,2,3,4, 11 atau 12 (Januari, Februari, Maret, April, November atau Desember) maka termasuk musim penghujan, dan jika selain bulan tersebut (Mei, Juni, Juli, Agustus, September, Oktober) maka termasuk musim kemarau. Pengelompokkan ini kemudian akan membuat kita dapat membandingkan kasus diare yang terjadi di musim penghujan dan kemarau.

```
=IF(C2=1,"Wet",IF(C2=2,"Wet",  
IF(C2=3,"Wet",IF(C2=4,"Wet",IF(C2=11,"Wet",IF(C2=12,"Wet","Dry"))))))
```

Tarik ini sepanjang kolom untuk mencakup seluruh gugus data sehingga tiap pengamatan akan diberikan klasifikasi musim penghujan atau kemarau.

*Diare atau gejala klinis lain:*

Kita akan membuat kolom baru dengan judul diare'. Kolom ini akan mencatat apakah kolom Tanda memiliki tanda diare atau gejala lainnya. Data ini digunakan untuk menghitung prevalensi. Kita akan kembali menggunakan rumus If.

```
=IF(ISNUMBER(SEARCH("diare",G2)),"Diare","Lainnya")
```

Rumus ini mengatakan bahwa, jika sel pada kolom Tanda (G2) mengandung kata "diare", maka dalam sel yang baru dicatat sebagai "Diare", dan jika G2 tidak mengandung kata "diare", maka di sel yang baru dicatat sebagai 'lainnya'.

Silakan saksikan video Case study 2\_Exercise 10 untuk membantu penyelesaian latihan ini. Anda harus mendapatkan hasil sebagai berikut:

ID	Tanggal laporan	Month Integer	Season	Jumlah ka Spesies	Tanda	Clinical sign
2	54817	16/04/2014	4 Wet	1 cattle	fever, Anorexia, abscess	Other
3	54809	16/04/2014	4 Wet	1 cattle	Anorexia	Other
4	54791	15/04/2014	4 Wet	1 cattle	lameness	Other
5	54783	15/04/2014	4 Wet	3 cattle	fever, other sign	Other
6	54775	15/04/2014	4 Wet	2 cattle	lameness	Other
7	54767	15/04/2014	4 Wet	2 cattle	diarrhoea	Diarrhoea
8	54742	15/04/2014	4 Wet	1 cattle	lameness	Other
9	54734	15/04/2014	4 Wet	1 cattle	diarrhoea, thin, Anorexia	Diarrhoea
10	54726	15/04/2014	4 Wet	1 cattle	hair loss, hyporesponsive, diarrhoea	Diarrhoea
11	54718	15/04/2014	4 Wet	1 cattle	abscess, sores	Other
12	54700	15/04/2014	4 Wet	1 cattle	Anorexia, fever, nasal discharge, lameness, drooling	Other
13	54684	15/04/2014	4 Wet	2 cattle	other sign	Other
14	54676	15/04/2014	4 Wet	2 cattle	other sign	Other
15	54650	14/04/2014	4 Wet	2 cattle	Anorexia, thin	Other
16	54643	14/04/2014	4 Wet	1 cattle	Anorexia, thin	Other
17	54635	14/04/2014	4 Wet	1 cattle	Anorexia, thin	Other
18	54627	14/04/2014	4 Wet	1 cattle	other sign	Other
19	54601	14/04/2014	4 Wet	1 cattle	thin, Anorexia	Other
20	54593	14/04/2014	4 Wet	1 cattle	thin, Anorexia	Other
21	54585	14/04/2014	4 Wet	1 cattle	Anorexia	Other
22	54577	14/04/2014	4 Wet	1 cattle	Anorexia	Other
23	54569	14/04/2014	4 Wet	1 cattle	Anorexia	Other
24	54551	14/04/2014	4 Wet	3 cattle	itching	Other

**Latihan 11: Mendeskripsikan kolom data yang baru dengan *pivot table*.**

**Jawaban**

Karena datanya adalah data kategorik (misalnya musim, atau bulan), maka tidak mungkin untuk menggunakan alat analisis data yang kita digunakan dalam studi kasus 1. Dalam latihan ini kita menggunakan *pivot table* untuk mendeskripsikan data yang kita miliki.

Video berjudul case study 2\_exercise 11 menggambarkan pendekatan tersebut. Namun, tabel berikut dapat dibuat dalam Excel menggunakan data yang tersedia dan pendekatan *pivot table*.

Row Labels	Count of Month integer
1	388
2	250
3	287
4	114
5	148
6	135
7	329
8	203
9	158
10	148
11	341
12	601
<b>Grand Total</b>	<b>3102</b>

Row Labels	Count of Season
Dry	1121
Wet	1981
<b>Grand Total</b>	<b>3102</b>

Row Labels	Count of Clinical sign
Diarrhoea	790
Other	2312
<b>Grand Total</b>	<b>3102</b>

**Latihan 12-14:**

Jawaban latihan 12,13 dan 14 disediakan dalam video dan dalam manual.

**Studi kasus 3: Statistika pemotongan dan swasembada daging sapi.**

**Latihan 15:** Apakah hipotesis mendukung tujuan analisis?

Jawaban latihan 15 bersifat subyektif sehingga tidak ada jawaban yang spesifik. Tujuan dari latihan ini adalah mendorong peserta berpikir mengenai hipotesis dan tujuan keseluruhan.

**Latihan 16:** Buat sebuah *pivot table* yang berisi statistika pemotongan berdasarkan bulan.

**Jawaban:**

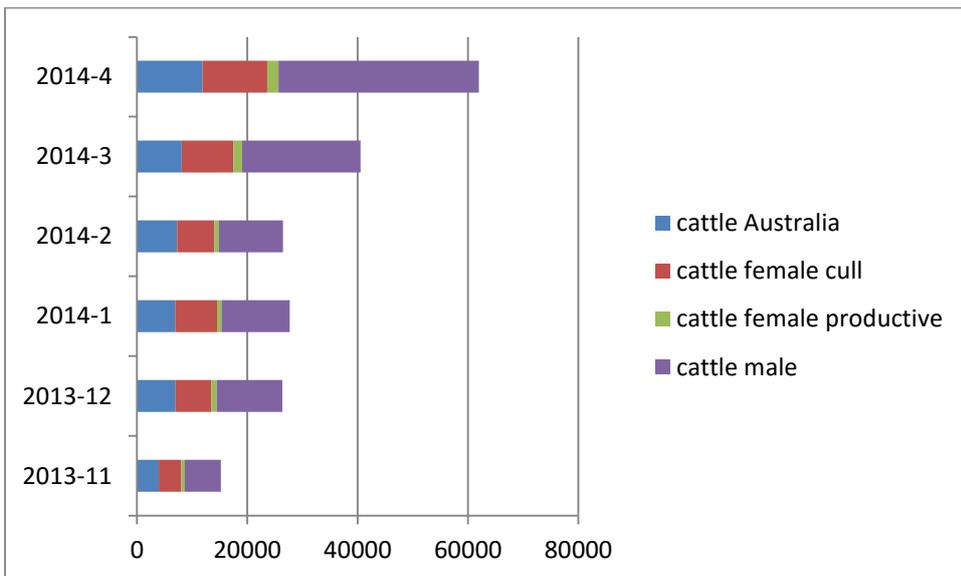
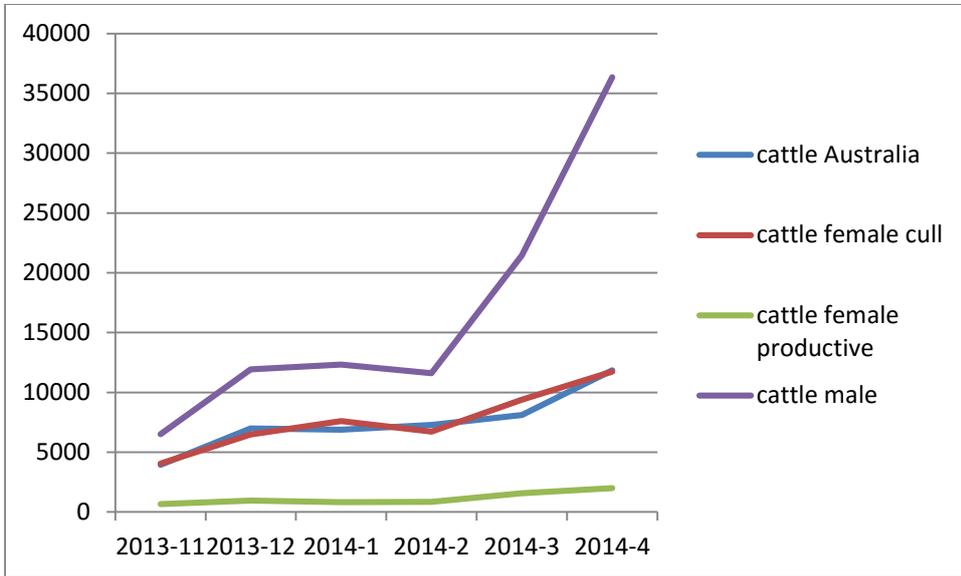
Tersedia dalam bentuk video. Namun disertakan juga di sini.

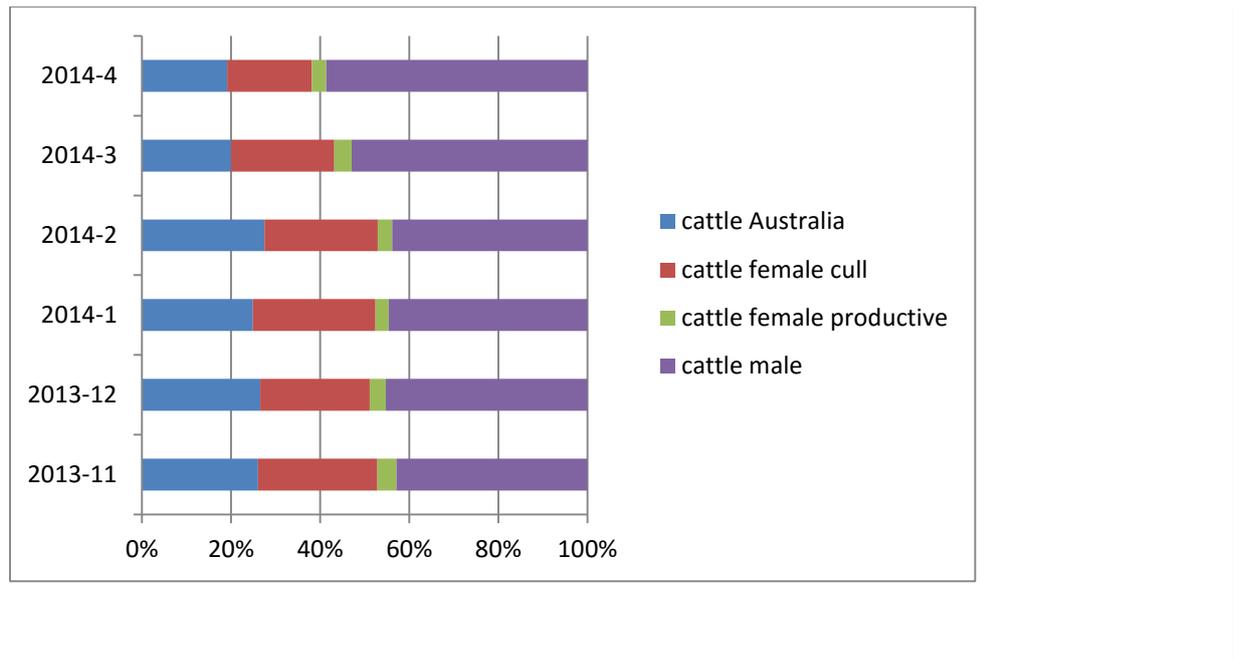
Row Labels	cattle Australia	cattle female cull	cattle female productive	cattle male	Grand Total
2013-11	3962	4066	654	6511	15193
2013-12	6988	6483	958	11921	26350
2014-1	6896	7611	829	12342	27678
2014-2	7277	6733	857	11607	26474
2014-3	8103	9390	1567	21455	40515
2014-4	11860	11751	2009	36357	61977
<b>Grand Total</b>	<b>45086</b>	<b>46034</b>	<b>6874</b>	<b>100193</b>	<b>198187</b>

**Latihan 17: Buat beberapa plot ringkasan dari tabel kontingensi.**

Jawaban latihan 17 tersedia dalam bentuk video

Namun, bagan/plot berikut dapat dibuat dalam latihan:





**Latihan 18:** Lakukan pengujian statistika khi-kuadrat pada *pivot table*.

**Jawaban:**

Statistik khi-kuadrat= 3599.6, derajat bebas = 15,  $p < 0.000000001$ .

Nilai kritik untuk menolak hipotesis nol adalah 25, maka 3599.6 sangat jelas melampaui nilai kritik ini!

Menolak hipotesis nol dan menerima hipotesis alternatif (terdapat hubungan antara jumlah sapi yang disembelih berdasarkan kategori sapi dan bulan disembelih).

**Latihan 19: Kesimpulan**

**Jawaban:**

Hasil yang paling penting dari pengujian statistika khi-kuadrat adalah bahwa nilai khi-kuadrat sangat besar. Nilai titik kritik untuk sebaran khi-kuadrat dengan 15 derajat bebas adalah 25, dan nilai 3600 yang kita miliki jelas melampaui ini! Maka kita yakin bahwa data tidak menyebar khi-kuadrat, dan tidak bebas.

Kita menolak hipotesis nol dan menerima hipotesis alternatif:

$H_A =$  Terdapat hubungan antara bulan dan jumlah total penyembelihan dalam kelompok sapi yang berbeda.

Kemudian kita memeriksa nilai khi-kuadrat untuk setiap sel dan nilai-nilai observasi dan nilai harapan, serta plot untuk menentukan di mana hubungan itu paling jelas terjadi dalam tabel kontingensi dan selanjutnya pengaruh apa yang terjadi.

Jelas terlihat bahwa nilai-nilai khi-kuadrat terbesar ditemukan di bulan penyembelihan akhir tahun, baik untuk sapi Australia maupun sapi jantan Indonesia. Informasi ini berguna untuk menunjukkan bahwa ketergantungan pada sapi impor menurun sampai batas tertentu sementara ketergantungan pada ternak potong Indonesia meningkat. Pengaruh ini menunjukkan bahwa terdapat kecenderungan adanya peningkatan swasembada produksi daging sapi.

Namun pengaruh ini patut diberi catatan. Data ini berasal dari kisaran waktu yang singkat (hanya 6 bulan). Untuk memastikan suatu kecenderungan, maka diperlukan pengamatan yang lebih lama dan penelaahan statistika yang lebih ketat. Demikian juga, beberapa pengaruh yang diamati adalah pasti disebabkan oleh peningkatan jumlah sapi dipotong yang dilaporkan seiring peningkatan laporan iSIKHNAS.

## Daftar Pustaka

- Dalgaard, P., 2008. *Introductory Statistics with R*. Springer.
- Dohoo, I., Martin, W., Stryhn, H., 2009. *Veterinary Epidemiologic Research*. VER Charlottetown.
- Petrie, A., Watson, P., 2006. *Statistics for Veterinary and Animal Science*. Wiley.