

Probability and Statistics Project Report

Kevin Widjaja 2019-05

Table of Contents

Pendahuluan	1
Desain Penelitian dan Hipotesa.....	2
Tipe data	2
Metode Pengumpulan Data/Metode Sampling.....	2
Hipotesa	2
Analisa Data.....	3
Dasar Teori	3
Statistika Deskriptif.....	3
Hasil Uji Hipotesa	7
Ringkasan dan Kesimpulan.....	7
Daftar Pustaka.....	8
Lampiran	9
Kuesioner	9
Data.....	12
Source Code	12

Pendahuluan

Pada penelitian ini, kami meneliti minat baca mahasiswa UMN (Universitas Multimedia Nusantara). Kami tertarik untuk meneliti masalah ini karena kami melihat minat baca mahasiswa UMN semakin menurun. Oleh karena itu, dengan penelitian ini kami berharap dapat lebih mengetahui tentang minat baca mahasiswa UMN. Selain itu, kami juga tertarik untuk mencari tahu hubungan antara hobi dan minat baca dari responden, tipe e-book yang biasa digunakan oleh mahasiswa UMN, dana yang dikeluarkan untuk membeli buku, dan hubungan antara minat baca dan semester.

Desain Penelitian dan Hipotesa

Tipe data

- Tipe data yang digunakan bertipe kategorikal, tetapi ada dua yang merupakan tipe data numerik yaitu dana baca per bulan dan durasi baca per hari.
- Tingkat pengukuran data yang digunakan yaitu tingkat nominal untuk variabel program studi, jenis kelamin, suka baca, genre buku, hobi, dan format e-book; tingkat ordinal untuk variabel Semester; tingkat interval untuk variabel durasi baca per hari; dan tingkat rasio untuk variabel dana baca per bulan.

Metode Pengumpulan Data/Metode Sampling

- Responden adalah mahasiswa dan mahasiswi semester genap UMN aktif (semester 2 sampai 8) pada tahun 2019.
- Alat yang digunakan adalah Google Form dan alat perekam video untuk merekam wawancara.
- Prosedur: Metode sampling adalah simple random sampling, kuesioner dijawab secara acak oleh sampel dari populasi Mahasiswa UMN. Variabel yang diteliti adalah variabel genre buku, hobi, semester, suka baca, durasi baca per hari, format e-book yang digunakan, dan dana yang dikeluarkan untuk membaca per bulan. Langkah-langkah pengumpulan data dimulai dengan perumusan hipotesa, dilanjutkan dengan perumusan variabel, membuat kuesioner untuk mencari data sesuai variabel, lalu data dirangkum dan terakhir data dibersihkan (cleaning) sebelum diolah.

Hipotesa

Hipotesa yang ingin diuji:

H_0 : Tidak terdapat hubungan antara hobi dan genre buku yang disukai seseorang.

H_1 : Terdapat hubungan antara hobi dan genre buku yang disukai seseorang.

H_0 : Tidak terdapat hubungan antara minat baca dan tingkat semester.

H_1 : Terdapat hubungan antara minat baca dan tingkat semester.

H_0 : Rata-rata Minat baca Mahasiswa UMN sama dengan rata-rata minat baca yang baik menurut UNESCO yaitu 5 jam/hari.

H_1 : Rata-rata Minat baca Mahasiswa UMN tidak sama dengan rata-rata minat baca yang baik menurut UNESCO yaitu 5 jam/hari.

H_0 : Tidak terdapat hubungan antara tipe e-book yang digunakan dengan apakah orang tersebut suka baca atau tidak.

H_1 : Terdapat hubungan antara tipe e-book yang digunakan dengan apakah orang tersebut suka baca atau tidak.

H_0 : Rata-rata biaya yang dikeluarkan untuk membaca buku per hobi sama.

H_1 : Terdapat setidaknya satu rata-rata yang berbeda.

Analisa Data

Dasar Teori

Rumus-rumus perhitungan yang digunakan dalam penelitian ini adalah *Chi Squared test of independence*., two-sided t-test, dan Kruskal-wallis test

Untuk Hipotesa 1, 2, dan 4, tes statistik yang digunakan adalah *Chi Squared test of independence*. Menurut buku *Business Statistics For Contemporary Decision Making* oleh Ken Black, *Chi Squared test of independence*. digunakan untuk menganalisa frekuensi 2 variabel dengan banyak kategori untuk menentukan apakah kedua variabel tersebut independen. Data frekuensi dari kedua variabel dijadikan satu menjadi sebuah *contingency table*. Hipotesa null dari Chi Squared test ini adalah kedua variabel independen. Kami menggunakan test ini karena kami hendak mencari tahu apakah terdapat hubungan antara 2 variabel nominal.

Untuk Hipotesa 3, tes statistik yang digunakan adalah two-sided t-test. Menurut buku *Probability and Statistics for Engineers and Scientists* edisi ke-empat oleh Anthony Hayter, two-sided t-test digunakan untuk menghitung p-value untuk hipotesa 2 sisi. Kami menggunakan test ini karena kami ingin melihat rata-rata durasi membaca mahasiswa UMN dan membandingkan dengan rata-rata menurut UNESCO, apakah ada perbedaan yang signifikan antara kedua rata-rata tersebut.

Untuk Hipotesa 5, tes statistik yang digunakan adalah Kruskal-Wallis test. Menurut buku *Business Statistics For Contemporary Decision Making* oleh Ken Black, Kruskal-Wallis test adalah alternatif non-parametrik dari ANOVA satu arah (*one-way analysis of variance*). Seperti ANOVA satu arah, Kruskal-Wallis test digunakan untuk menentukan apakah $c \geq 3$ sampel berasal dari populasi yang sama atau berbeda. Tidak seperti ANOVA satu arah yang berasumsi data berdistribusi normal, group independen, data minimal interval, dan varians populasi sama, Kruskal-Wallis test bisa digunakan untuk menganalisa data ordinal dan tidak didasari asumsi dari bentuk populasi. Kami menggunakan test Kruskal-Wallis karena data kami tidak berdistribusi normal sehingga tidak memenuhi asumsi normalitas ANOVA.

Statistika Deskriptif

```
## Parsed with column specification:
## cols(
##   Timestamp = col_character(),
##   Username = col_character(),
##   Nama = col_character(),
##   Prodi = col_character(),
##   Jenis_Kelamin = col_character(),
##   Semester = col_double(),
##   Suka_Baca = col_character(),
##   Alasan_Suka_Baca = col_character(),
##   Durasi_Baca = col_double(),
##   Alasan_Durasi = col_character(),
##   Genre = col_character(),
##   Hobi = col_character(),
```

```
## Format = col_character(),
## Dana_Baca_Sebulan = col_number()
## )
```

Graph 1

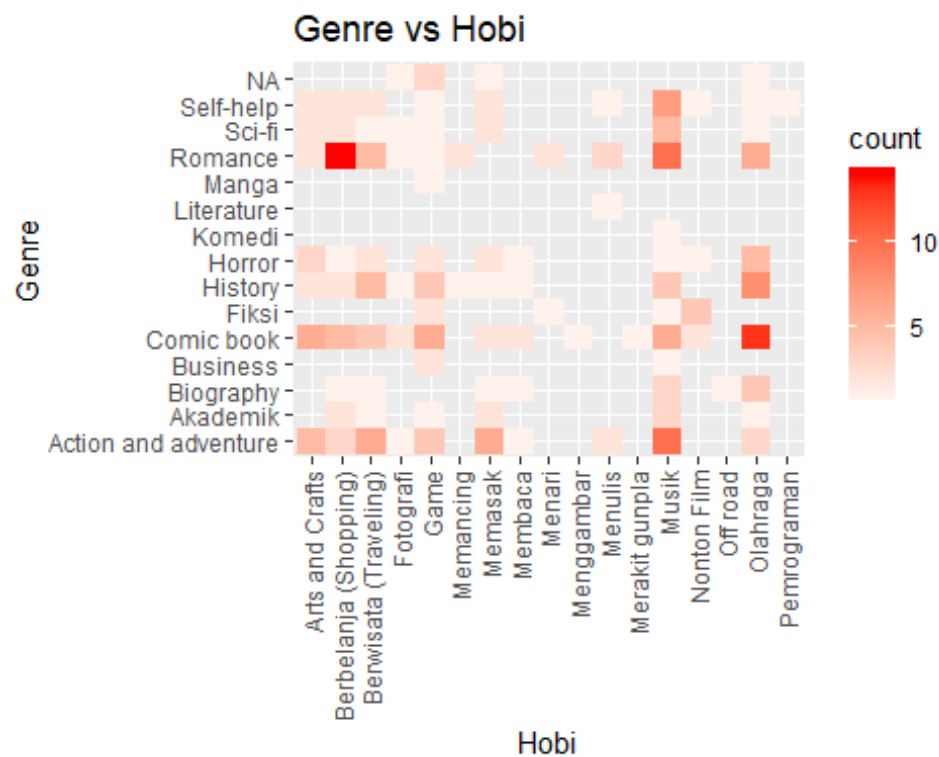
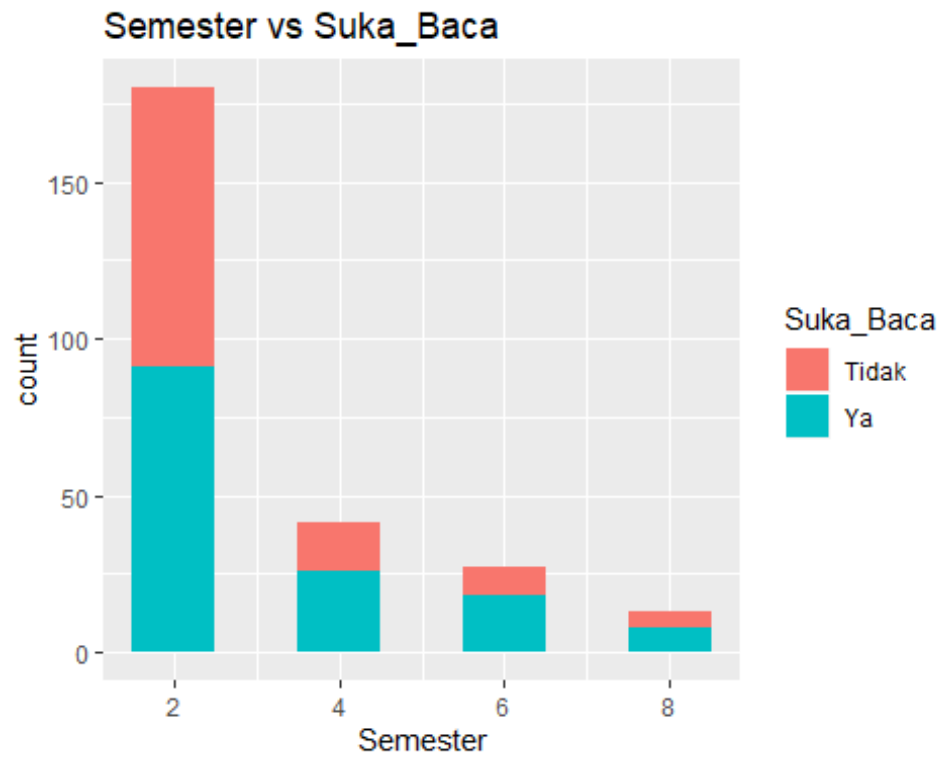


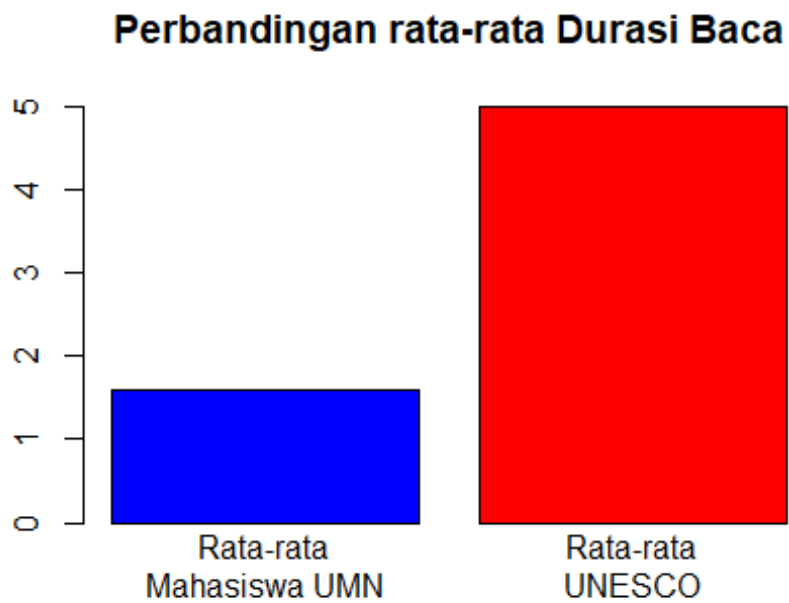
Table 1 Tabel Frekuensi Responden yang memilih Genre dan Hobi Tertentu

##	Genre	Hobi	Frekuensi
## 26	Romance	Berbelanja (Shopping)	14
## 215	Comic book	Olahraga	13
## 169	Action and adventure	Musik	10
## 180	Romance	Musik	10
## 217	History	Olahraga	8
## 182	Self-help	Musik	7

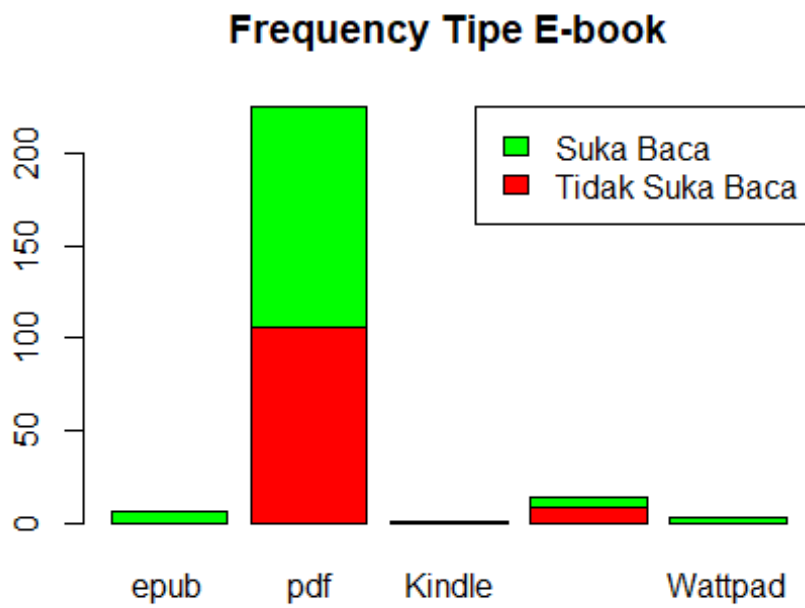
Graph 2



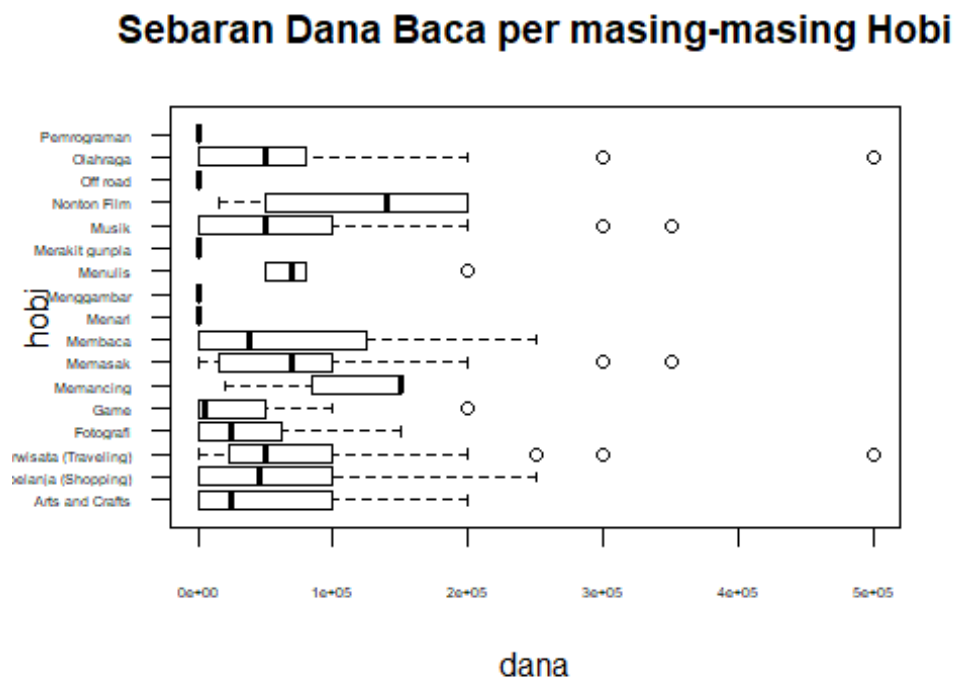
Graph 3



Graph 4



Graph 5



Hasil Uji Hipotesa

Pada Hipotesa 1, kami ingin menguji apakah terdapat hubungan antara variabel hobi dan genre buku yang disukai. Kami menguji hipotesa dengan uji independensi Chi-kuadrat. Hasil yang kami dapat adalah p-value senilai 0.002393 yang lebih kecil dari alpha 0.05, sehingga dapat disimpulkan terdapat hubungan antara genre buku yang disukai dengan hobi.

Pada Hipotesa 2, kami ingin menguji apakah terdapat hubungan antara variabel semester dan variabel minat baca. Seperti hipotesa 1, kami menguji hipotesa 2 dengan menggunakan uji independensi Chi-kuadrat. Hasil yang kami dapat adalah p-value senilai 0.2298 yang lebih besar dari alpha 0.05 sehingga tidak menolak H_0 . Dapat disimpulkan tidak terdapat hubungan antara semester dan minat baca.

Pada hipotesa 3, kami ingin menguji apakah rata-rata minat baca mahasiswa UMN sama dengan rata-rata minat baca yang baik menurut UNESCO yaitu 5 jam/hari. Kami menggunakan uji t-test untuk membandingkan kedua rata-rata. Hasil yang kami peroleh adalah p-value $< 2.2e-16$ yang lebih kecil dari alpha 0.05, sehingga H_0 ditolak. Kedua rata-rata tidak sama, maka dicari rata-rata minat baca mahasiswa UMN, didapat angka 1.59387 yang lebih kecil daripada 5. Dapat disimpulkan bahwa rata-rata minat baca Mahasiswa UMN lebih rendah dibanding rata-rata minat baca yang baik menurut UNESCO yaitu 5 jam/hari.

Pada hipotesa 4, kami ingin menguji apakah Terdapat Hubungan antara variabel format e-book dan Suka_baca. Kami menguji hipotesa kami dengan uji independensi Chi-kuadrat. Hasil yang kami dapat adalah p-value senilai 0.05331 yang lebih besar dari alpha 0.05, walaupun beda tipis. Maka kami tidak menolak H_0 dan didapat kesimpulan bahwa tidak terdapat Hubungan yang signifikan antara variabel format e-book dan Suka baca.

Pada hipotesa 5, kami ingin mencari tahu adakah kelompok hobi tertentu yang mengeluarkan dana untuk membaca lebih dari rata-rata kelompok hobi lainnya. Kami menggunakan uji Kruskal-Wallis dikarenakan data tidak memenuhi asumsi normalitas, yang dibutuhkan untuk melakukan uji ANOVA. Hasil uji Kruskal-Wallis adalah p-value senilai 0.01624 yang signifikan pada alpha 0.05. H_0 ditolak dan dapat disimpulkan terdapat setidaknya 1 rata-rata yang berbeda. Langkah selanjutnya adalah mencari rata-rata yang berbeda dengan melakukan Dunn Test, tetapi setelah diuji tidak ditemukan perbedaan yang signifikan.

Ringkasan dan Kesimpulan

Dari hasil Hipotesa 1, orang yang memiliki hobi tertentu cenderung suka dengan genre buku tertentu. Hasil ini sesuai dengan dugaan kami. Beberapa contoh adalah orang yang suka *Romance* cenderung suka berbelanja, yang suka komik cenderung suka olahraga, dan yang suka *Action and Adventure* cenderung suka musik. Hal ini dapat menjadi informasi yang berguna, misalnya lebih banyak memasang iklan buku komik di ajang pentas olahraga kemahasiswaan. Hal ini akan dapat membantu meningkatkan minat baca di kalangan mahasiswa sekaligus mengiklankan produk yang sesuai dengan target pasar.

Dari hasil Hipotesa 2, minat baca lintas semester cenderung sama, tidak seperti dugaan awal bahwa mahasiswa yang sudah berada pada semester lanjut cenderung memiliki minat baca yang lebih tinggi. Hal ini menunjukkan bahwa tidak terdapat peningkatan minat baca dari semester ke semester.

Dari hasil Hipotesa 3, Rata-rata Minat baca Mahasiswa UMN masih jauh di bawah rata-rata minat baca yang baik menurut UNESCO yaitu 5 jam/hari, sesuai dengan dugaan kami. Ini menunjukkan bahwa minat baca mahasiswa UMN harus ditingkatkan lagi. Universitas, orang tua, dan orang-orang di sekitar mahasiswa dapat turut mendukung dan membantu meningkatkan minat baca pribadi mahasiswa.

Dari hasil hipotesa 4, tidak terdapat hubungan antara format e-book yang digunakan dan apakah mahasiswa tersebut suka membaca atau tidak. Hal ini sesuai dengan dugaan kami. Format e-book yang paling banyak digunakan adalah .pdf, dan ada kemungkinan bahwa tipe pdf adalah yang paling sering digunakan karena kebanyakan e-book bajakan yang beredar di internet bertipe pdf. Hal ini juga dapat dilihat dari banyaknya mahasiswa yang tidak mengeluarkan dana untuk membaca (93 mahasiswa dari total 261 mahasiswa), dan mahasiswa yang suka membaca dan tidak mengeluarkan dana untuk membaca (33 dari 93 mahasiswa).

Dari hasil hipotesa 5, Terdapat setidaknya satu rata-rata biaya yang dihabiskan untuk membaca buku yang berbeda antar kelompok hobi. Dikarenakan keterbatasan kami, kami tidak menemukan kelompok hobi yang memiliki perbedaan signifikan, sehingga kesimpulan yang dapat kami tarik hanya sebatas mengetahui bahwa ada kelompok yang memiliki rata-rata yang berbeda.

Daftar Pustaka

- Black, K. (2010). *Business statistics: For contemporary decision making*. Hoboken, NJ: Wiley.
- Hayter, A. J. (2013). *Probability and statistics for engineers and scientists*.
- Kruskal-Wallis Test in R. (n.d.). Retrieved from <http://www.sthda.com/english/wiki/kruskal-wallis-test-in-r>
- Kruskal-Wallis Test. (n.d.). Retrieved from https://rcompanion.org/rcompanion/d_06.html
- Mendikbud: Sistem Perbukuan untuk Wujudkan Buku Bermutu, Terjangkau dan Merata. (2017, March 30). Retrieved from <https://www.kemdikbud.go.id/main/blog/2017/03/mendikbud-sistem-perbukuan-untuk-wujudkan-buku-bermutu-terjangkau-dan-merata>

Lampiran

Kuesioner

Minat Baca Mahasiswa UMN

Hey! Kuesioner ini dibuat untuk mengumpulkan data terkait penelitian kami tentang minat baca mahasiswa Universitas Multimedia Nusantara, Tangerang.

Your email address (kevin.widjaja@student.umn.ac.id) will be recorded when you submit this form. Not you? [Switch account](#)

* Required

Nama Anda? *

Your answer

Program Studi? *

Choose ▼

Jenis Kelamin? *

- ☐ Pria
- ☐ Wanita

Saat ini saya mahasiswa semester... *

Choose ▼

Apakah Anda suka membaca buku di waktu luang? *

- ☐ Ya
- ☐ Tidak

Mengapa Anda suka/tidak suka membaca buku? *

Your answer

Seberapa sering Anda membaca buku? (jam/hari) *

Choose ▼

Mengapa Anda sering/jarang membaca buku? *

Your answer

Genre buku apa yang paling Anda sukai? *



☐ Action and adventure



☐ Comic book



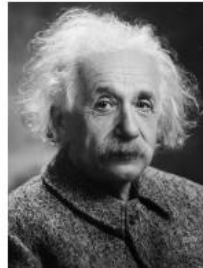
☐ Horror



☐ Romance



☐ Sci-fi



☐ Biography



☐ Self-help



☐ History

☐ Other:

Apa hobi Anda? *



☐ Olahraga



☐ Musik



☐ Memancing



☐ Berbelanja (Shopping)



☐ Berwisata (Traveling)



☐ Arts and Crafts



☐ Memasak



☐ Menulis

☐ Other:

Bila anda membaca e-book, format e-book apa yang paling sering anda baca? *

- ☐ .pdf
- ☐ .epub
- ☐ Amazon Kindle
- ☐ Google Play Books
- ☐ Other: _____

Berapa banyak uang yang Anda habiskan untuk membaca buku dalam sebulan? (dalam Rupiah) *

Your answer _____

☐ Send me a copy of my responses.

SUBMIT

Page 1 of 1

Never submit passwords through Google Forms.

Data

```
## # A tibble: 261 x 14
##   Timestamp Username Nama Prodi Jenis_Kelamin Semester Suka_Baca
##   <chr> <chr> <chr> <chr> <chr> <dbl> <chr>
## 1 2019/04/~ farras.~ Farr~ Desa~ Pria 2 Tidak
## 2 2019/04/~ william~ Will~ Desa~ Pria 2 Tidak
## 3 2019/04/~ metta.s~ metta Film Wanita 6 Ya
## 4 2019/04/~ bagas.d~ Bagas Desa~ Pria 2 Tidak
## 5 2019/04/~ justin.~ Just~ Film Pria 4 Tidak
## 6 2019/04/~ jennie.~ jenn~ Desa~ Wanita 2 Tidak
## 7 2019/04/~ christa~ Chri~ Desa~ Wanita 2 Ya
## 8 2019/04/~ anastha~ Anas~ Jurn~ Wanita 2 Ya
## 9 2019/04/~ rina.ku~ rina Desa~ Wanita 2 Tidak
## 10 2019/04/~ alvin.h~ Alvi~ Sist~ Pria 4 Ya
## # ... with 251 more rows, and 7 more variables: Alasan_Suka_Baca <chr>,
## # Durasi_Baca <dbl>, Alasan_Durasi <chr>, Genre <chr>, Hobi <chr>,
## # Format <chr>, Dana_Baca_Sebulan <dbl>
```

Source Code

```
# Hipotesa 1 H0 : Tidak Terdapat Hubungan antara variabel
# Hobi dan Genre Buku Ha : Terdapat Hubungan antara Variabel
```

```

# Hobi dan Genre Buku

# Langkah-Langkah pengerjaan import DATA_FINAL(CLEANED) ke
# variabel 'data' Lakukan Chi-Square test

head(data)
data1 <- table(data$Genre, data$Hobi)
chisq.test(data1)

# X-Squared = 270.19 df = 208 p-value = 0.002393

# p-value lebih kecil dari alpha 0.05 Tolak H0 dan terima Ha

# Kesimpulan : Ada hubungan antara genre buku yang disukai
# dengan hobi

# Statistika Deskriptif Graph 1.1
library(ggplot2)
ggplot(data) + geom_bin2d(aes(x = Hobi, y = Genre)) + scale_fill_gradient2(low = "white",
  high = "red") + ggtitle("Genre vs Hobi") + theme(axis.text.x = element_text(angle = 90,
  vjust = 0.5, hjust = 1))

# Graph 1.2
data1 <- data.frame(data1)
attach(data1)
data1new <- data1[order(-Freq), ]
names(data1new) <- c("Genre", "Hobi", "Frekuensi")
head(data1new)
detach(data1)

# Hipotesa 2 H0 : Tidak terdapat hubungan antara variabel
# semester dan variabel minat baca Ha : Terdapat hubungan
# antara variabel semester dan variabel minat baca

# Langkah-Langkah pengerjaan import DATA_FINAL(CLEANED) ke
# variabel 'data' Lakukan Chi-Square test

head(data)
data2 <- table(data$Semester, data$Suka_Baca)
chisq.test(data2)

# X-Squared = 4.3107 df = 3 p-value = 0.2298

# p-value lebih besar dari alpha 0.05 Tidak Tolak H0

# Kesimpulan : Tidak terdapat hubungan antara variabel

```

```

# semester dan variabel minat baca

# Statistika Deskriptif
library(ggplot2)
ggplot(data, aes(fill = Suka_Baca)) + geom_bar(width = 1, aes(x = Semester))
+
  ggtitle("Semester vs Suka_Baca")

# Hipotesa 3  $H_0$  : Rata-rata Minat baca Mahasiswa UMN sama
# dengan rata-rata minat baca yang baik menurut UNESCO yaitu
# 5 jam/hari  $H_a$  : Rata-rata Minat baca Mahasiswa UMN tidak
# sama dengan rata-rata minat baca yang baik menurut UNESCO
# yaitu 5 jam/hari  $H_0: \mu=5, H_1: \mu \neq 5$ 

# Langkah-langkah pengerjaan import DATA_FINAL(CLEANED) ke
# variabel 'data'

data3 <- data$Durasi_Baca
# cek normalitas data
shapiro.test(data3)
# data tidak normal, gunakan t test

t.test(data3, alternative = "two.sided", mu = 5)

# p-value < 2.2e-16

# p-value lebih kecil dari alpha 0.05 Tolak  $H_0$  dan terima  $H_a$ 

# Mencari rata-rata minat baca Mahasiswa UMN
mean3 <- mean(data3)
# Mean = 1.59387

# Kesimpulan : Rata-rata Minat baca Mahasiswa UMN tidak sama
# dengan rata-rata minat baca yang baik menurut UNESCO yaitu
# 5 jam/hari Setelah dicari mean minat baca Mahasiswa UMN,
# dapat disimpulkan Minat baca Mahasiswa UMN Lebih rendah
# dibandingkan dengan rata-rata minat baca yang baik menurut
# UNESCO yaitu 5 jam/hari

# Statistik Deskriptif
barplot(c(mean(data3), 5), main = "Perbandingan rata-rata Durasi Baca",
  names.arg = c("Rata-rata\nMahasiswa UMN", "Rata-rata\nUNESCO"),
  col = c("blue", "red"))

# Hipotesa 4  $H_0$  : Tidak Terdapat Hubungan antara variabel
# format e-book dan Suka_baca  $H_a$  : Terdapat Hubungan antara
# variabel format e-book dan Suka_baca

```

```

# Langkah-Langkah pengerjaan import DATA_FINAL(CLEANED) ke
# variabel 'data' Lakukan Chi-Square test

data4 <- table(data$Suka_Baca, data$Format)
chisq.test(data4)

# X-Squared = 9.3321 df = 4 p-value = 0.05331

# p-value lebih besar dari alpha 0.05 (beda tipis) Tidak
# Tolak H0

# Kesimpulan : Tidak Terdapat Hubungan yang signifikan antara
# variabel format e-book dan Suka baca

# Membuat barplot untuk mendeskripsikan hasil temuan
barplot(data4, main = "Frequency Tipe E-book", col = c("red",
  "green"), names.arg = c("epub", "pdf", "Kindle", "PlayBooks",
  "Wattpad"))
legend("topright", legend = c("Suka Baca", "Tidak Suka Baca"),
  fill = c("green", "red"))

# Setelah dibuat bar-plot, terlihat bahwa tipe ebook .pdf
# adalah yang paling sering digunakan oleh sampel sebanyak
# 269 mahasiswa UMN. Suka atau tidak suka baca buku, format
# yang paling sering digunakan adalah pdf.

# Hipotesa 5 H0 : Rata-rata biaya yang dikeluarkan untuk
# membaca buku per hobi sama. Ha : Terdapat setidaknya satu
# rata-rata yang berbeda.

# Langkah-Langkah pengerjaan import DATA_FINAL(CLEANED) ke
# variabel 'data'

hobi <- data$Hobi
dana <- data$Dana_Baca_Sebulan
data5 <- data.frame(hobi, dana)

# Cek Asumsi untuk ANOVA
shapiro.test(data5$dana)
# p-value < 2.2e-16 Data tidak memenuhi asumsi normalitas,
# sehingga digunakan tes Kruskal-Wallis

results <- kruskal.test(dana ~ hobi, data = data5)
results
# p-value = 0.01624

# p-value lebih kecil dari alpha 0.05 Tolak H0 dan terima Ha

```

```

# Kesimpulan: Terdapat setidaknya satu rata-rata biaya yang
# dihabiskan untuk membaca buku yang berbeda antar kelompok
# hobi.

# Mencari tahu kelompok hobi yang memiliki rata-rata dana
# baca yang berbeda Menggunakan Dunn test for multiple
# comparisons
library(FSA)
library(rcompanion)

dT <- dunnTest(dana ~ hobi, data = data5, method = "bh")
dT
dT <- dT$res
cldList(comparison = dT$Comparison, p.value = dT$P.adj, threshold = 0.05)
# Dari Hasil Dunn Test, tidak ditemukan p-value yang
# signifikan pada alpha 0.05.

# Statistika Deskriptif
boxplot(dana ~ hobi, data = data5, horizontal = TRUE, main = "Sebaran Dana Ba
ca per masing-masing Hobi",
        las = 1, cex.axis = 0.45)

# Terlihat kebanyakan mahasiswa tidak mengeluarkan dana yang
# besar untuk membaca.

```