0 - INTRODUCTION

Statistika Terapan Semester Genap 2017/2018

COURSE ORGANIZATION

Kode Mata Kuliah	**CSIM603216**
Nama Mata Kuliah	Statistika Terapan
Prasyarat Mata Kuliah	Statistika dan Probabilitas
Total SKS	3 sks
Dosen Pengajar	**Prof. Drs. T. Basaruddin, M.Sc., Ph.D** chan@cs.ui.ac.id **Meganingrum Arista Jiwanggi, M.Kom, MCS** meganingrum@cs.ui.ac.id **Dipta Tanaya, M.Kom** diptatanaya@gmail.com
Asisten	TBA
Suggested Stats Tools	SPSS, AMOS, SmartPLS, atau tools opensource (e.g: R, Statistical Lab, dsb)

EXPECTED LEARNING OUTCOMES

Setelah mengikuti mata kuliah ini, mahasiswa diharapkan mampu untuk:

1. Menjelaskan berbagai teknik statistik untuk mengolah data
2. Memilih teknik statistik yang sesuai untuk memecahkan permasalahan tertentu
3. Menginterpretasikan dengan benar hasil pengolahan statistik sebagai bagian dari critical thinking
4. Mengkomunikasikan hasil pengolahan statistik baik secara lisan maupun tulisan
5. Menggunakan alat bantu statistik untuk mengolah data

COURSE OUTLINE

No	Deskripsi	Tanggal
0	Review	5 Feb 2018
1	Comparing Two Means (t-Test):	5, 7, dan 12 Feb 2018
2	Comparing Several Means: ANOVA (Analysis of Variance)	14, 19, dan 21 Feb 2018
3	Analysis of Covariance (ANCOVA)	26 dan 28 Feb, 5 Maret 2018
4	Factorial ANOVA	5, 7, dan 12 Maret 2018
5	Correlation	14, 19, dan 21 Maret 2018
6	**UTS**	**26 Maret - 3 April 2018**
7	Regression	4, 9 April 2018
8	Logistic Regression	11, 16 18, April 2018
9	Exploratory Factor Analysis and Principle Component Analysis	23, 25, 30 April 2018
10	Causal Modeling: Path Analysis and Structural Equation Modeling (SEM)	2, 7, 9 Mei 2018
11	Non-parametric stats	14, 16, 21 Mei 2018
12	**UAS**	**23 Mei - 1 Juni 2018**

GRADING

Diskusi Kelompok	: 40%
Kuis dan Partisipasi	: 5%
UTS	: 25%
UAS	: 30%

TOTAL **: 100%**

GENERAL NOTICE

1. Metode perkuliahan Statistika Terapan Semester Genap 2017/2018 merupakan kombinasi dari pertemuan tatap muka dan diskusi kelompok.
2. Mahasiswa perlu membentuk kelompok beranggotakan 4-5 orang yang akan menjadi rekan diskusi dalam satu semester ke depan. Nama-nama anggota kelompok dapat ditulis di SCeLE
3. Pada setiap akhir diskusi kelompok, akan diundi kelompok yang akan melakukan presentasi sehingga setiap kelompok sudah harus siap untuk mempresentasikan hasil diskusi kelompoknya
4. Setiap kelompok juga WAJIB mengumpulkan laporan hasil diskusi dengan mengikuti format yang telah disiapkan.

SUGGESTED TEXT

- A. Field. **Discovering statistics using SPSS (4th edition).** Los Angeles: Sage, 2013
- Sheldon Ross. **Introduction to Probability and Statistics for Engineers and Scientists 4th Edition**. Academic Press, 2009
- Dennis E. Hinkle, et al. **Applied Statistics for the Behavioral Sciences 5th Edition.** Houghton Mifflin, 2002

SAMPLING

Why sample?

Key Sampling Concepts

- The theoretical population: who do you want to generalize to?
- The study population: what population can you get access to?
- The sampling frame: how can you get access to them?
- The sample: who are on your study?

SAMPLING PROCESS

Defining the population

Specifying a sampling frame, a set of items or events possible to measure

Specifying a sampling method for selecting items or events from the frame

Determining the sample size

Implementing the sampling plan

Sampling and data collecting

Reviewing the sampling process

TYPE OF SAMPLING

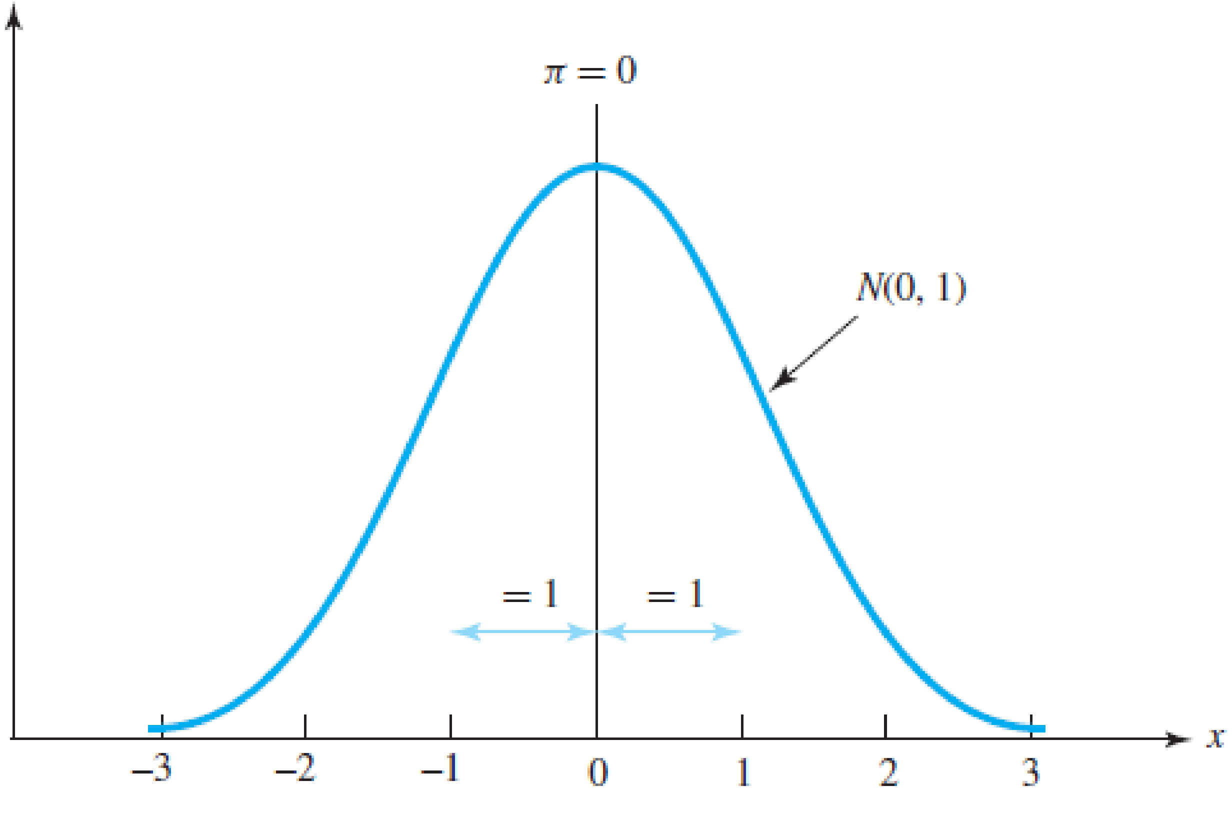
Probability

- Random Sampling : everyone in population has an equal chance of being selected
- Systematic / interval Sampling : randomly select number *y* between 1 and n, n=population size/sample size. Sample element y, $y+n$, $y+2n$, $y+3n$, etc.
  Example : population size = 64, sample size = 8, n = 8. Select number *y* between 1 and 8, ex:3. The sample are 3^{rd} , 11^{th}, 19^{th}, … .
- Stratified Sampling : divide the population into smaller groups (called strata), then take random sample from each strata.
- Cluster Sampling : divide the population into separate groups (called cluster), then take random sampling to choose the sampled cluster.

Non-probability

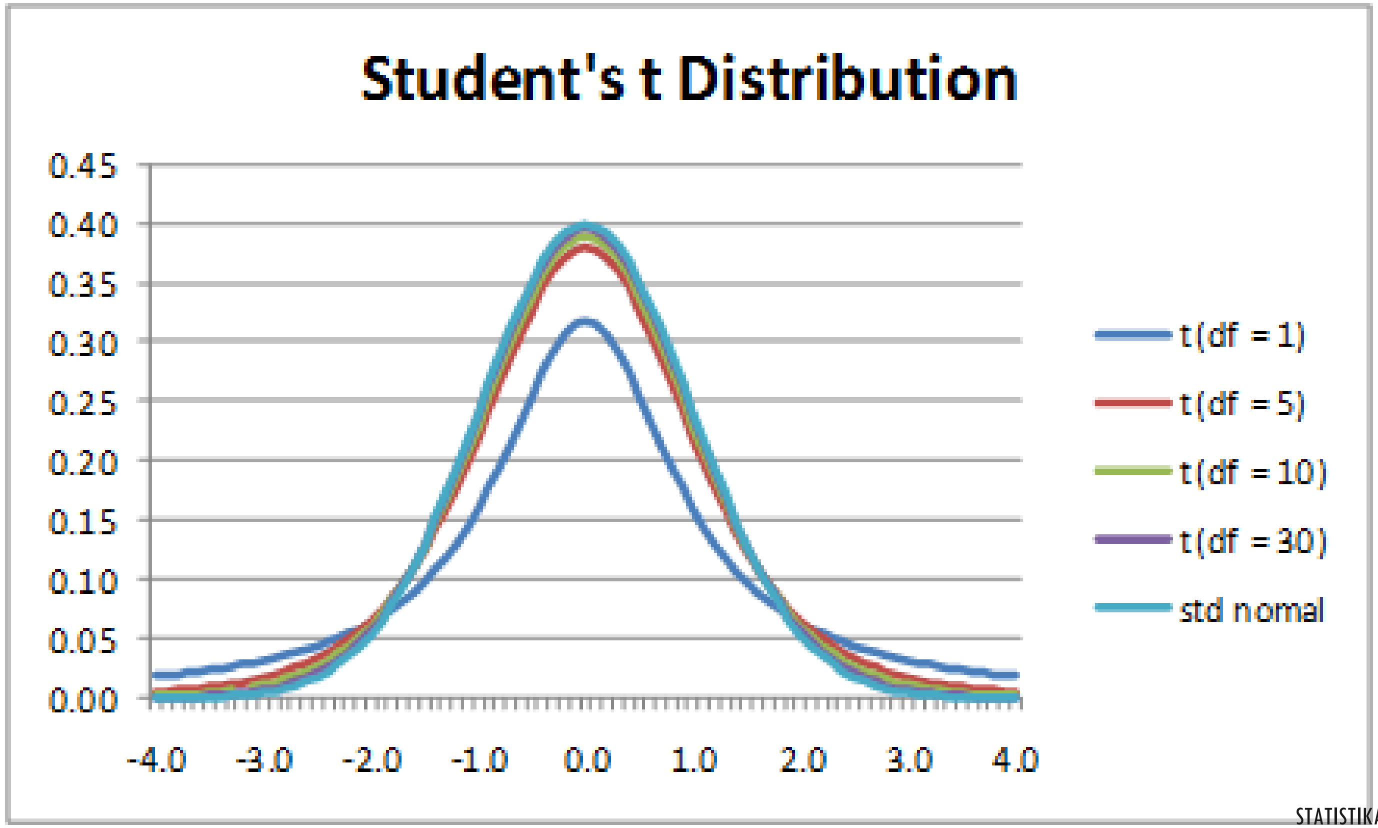
- Convenience : subjects are selected because of their convenient accessibility and proximity to the researcher.
- Quota : divide the population into smaller groups, then select the subjects or units from each segment based on a specified proportion
- Purposive : subjects are selected based on characteristics of a population and the objective of the study

NORMAL DISTRIBUTION

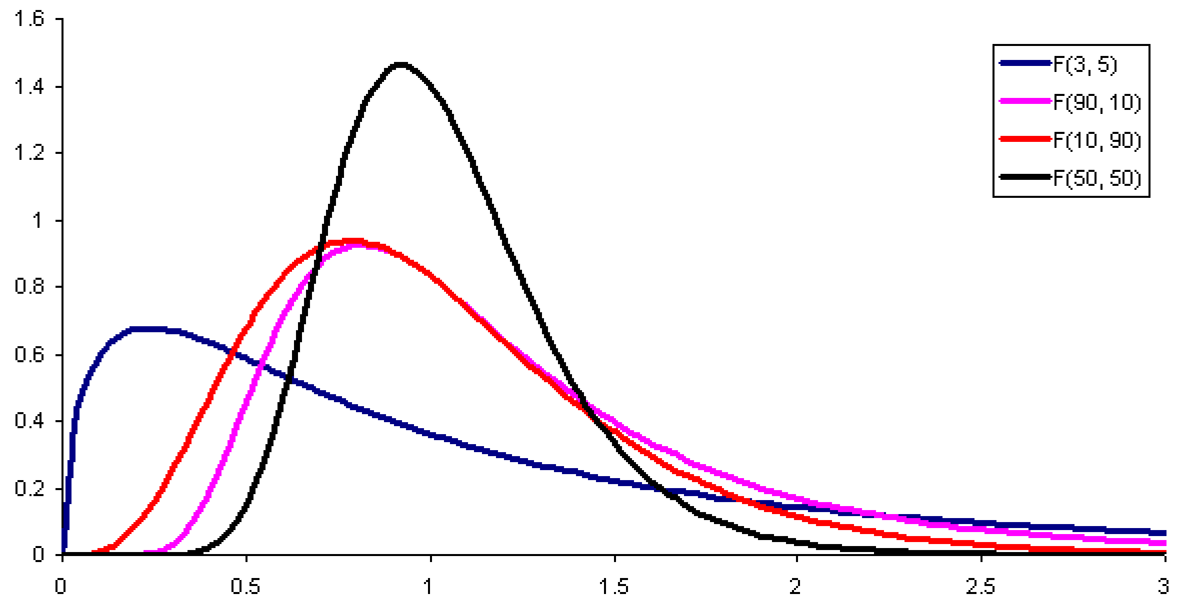


T DISTRIBUTION

Small sample size, unknown population standard deviation



F DISTRIBUTION



CONFIDENCE INTERVAL

A confidence interval for an unknown parameter θ is an interval that contains a set of plausible values of the parameter. It is associated with a confidence level $1-\alpha$, which measures the probability that the confidence interval actually contains the unknown parameter value.

CONFIDENCE INTERVAL FOR NORMAL MEAN WHEN THE VARIANCE IS KNOWN

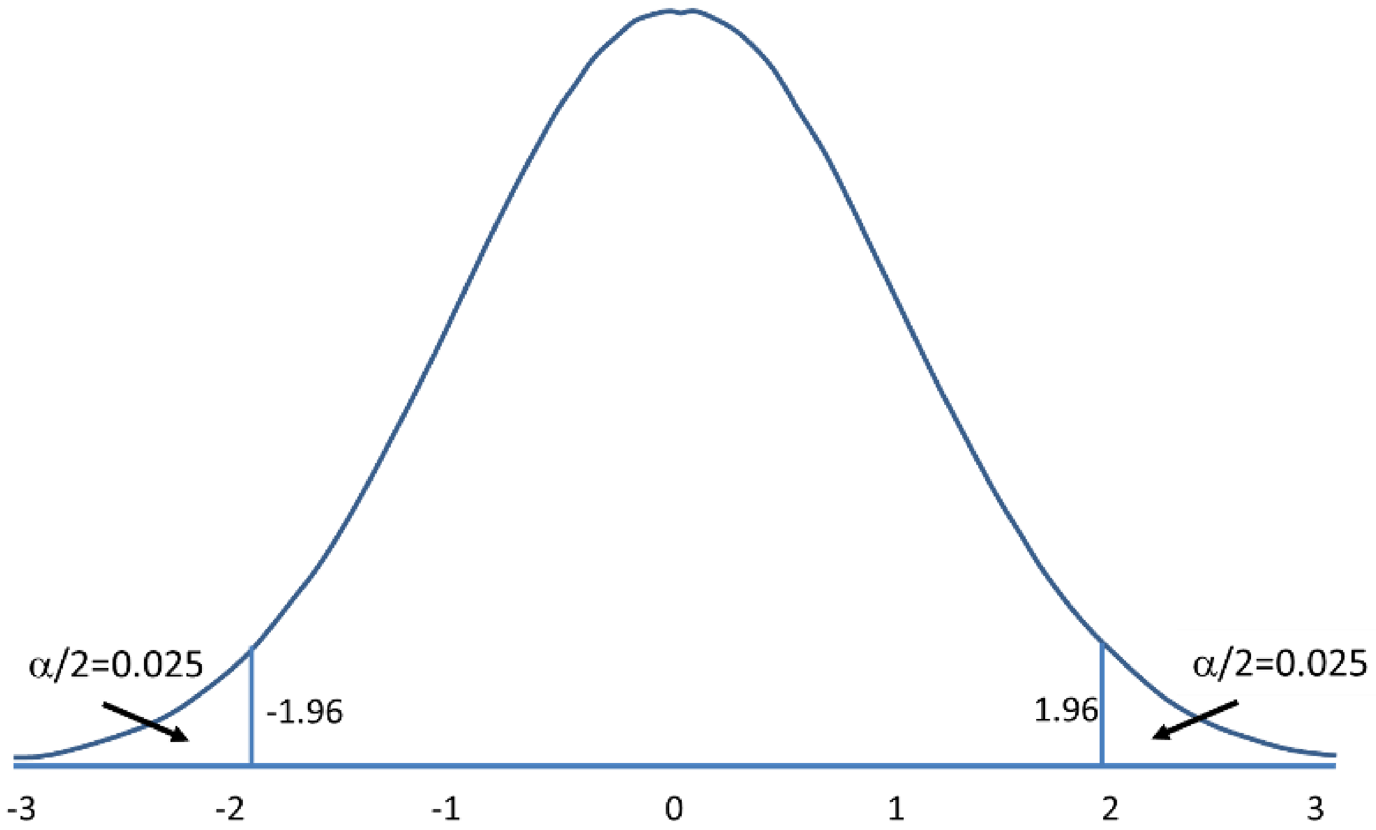
A $100(1-\alpha)\%$ two-sided confidence interval for μ is

$$\mu \in \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Example

- Suppose that data set of milk container weights is normally distributed with mean μ and variance 4. The sample mean $\bar{x}$ is 9. There were 16 samples in the experiment. Let us construct a 95% confidence interval for μ

CONFIDENCE INTERVAL



CONFIDENCE INTERVAL FOR NORMAL MEAN WHEN THE VARIANCE IS KNOWN

95% two-sided confidence interval for μ is

$$\mu \in \left(9 - 1.96\frac{2}{\sqrt{16}}, 9 + 1.96\frac{2}{\sqrt{16}}\right) = (8.02, 9.98)$$

How do we interpret that result?

- The observed interval (8.02, 9.98) brackets the true value of population mean with 95% confidence.

Which is better, wide or narrow interval?

CONFIDENCE INTERVAL

Two-sided CI

$$\mu \in \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

One-sided upper CI

$$\mu \in \left(\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

One-sided lower CI

$$\mu \in \left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

CONFIDENCE INTERVAL FOR NORMAL MEAN WHEN THE VARIANCE IS UNKNOWN

Two-sided CI

$$\mu \in \left(\overline{x} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}, \overline{x} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

One-sided upper CI

$$\mu \in \left(\overline{x} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \infty\right)$$

One-sided lower CI

$$\mu \in \left(-\infty, \overline{x} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

HYPOTHESIS

- Hypothesis : a statement about what findings are expected
- Null hypothesis : *"the two groups will not differ"*
- Alternative hypothesis :
  - *"group A will do better than group B"*
  - *"group A and B will not perform the same"*

POSSIBLE OUTCOMES IN HYPOTHESIS TESTING

	H_0 is actually	
	TRUE	FALSE
ACCEPT H_0	Correct desicion	Type II error
REJECT H_0	Type I error	Correct desicion

HYPOTHESIS TESTING

Steps :

- State the hypothesis : H_0 & H_a
- Set the criterion for rejecting H_0 : level of significance (α)
- Compute the Test-statistic : for example, compute z
- Decision about the Null Hypothesis
  - Accept the Null Hypothesis (H_0)
  - Reject the Null Hypothesis

HYPOTHESIS TESTING

- One sample case (test concerning the **Mean** of Normal Population)
  - Case 1 : known variance
  - Case 2 : unknown variance
- Two sample case (testing for **Equality of Variance** of Two Normal Population for independent samples)
- Two sample case (testing for **Equality of Means** of Two Normal Population for independent samples)
  - Case 1 : known variances
  - Case 2 : unknown variances, but $\sigma_1^2 = \sigma_2^2 = \sigma^2$
  - Case 3 : unknown variances, but $\sigma_1^2 \neq \sigma_2^2$
  - Case 4 : unknown variances, unknown relation between $\sigma_1^2\ and\ \sigma_2^2$
- Two sample case (testing for **Equality of Means** of Two Normal Population for dependent samples)

HYPOTHESIS TESTING (ONE SAMPLE CASE, WHEN σ^2 IS KNOWN)

	One-Tailed Test (i)	One-Tailed Test (ii)	Two-tailed Test
Hypothesis	$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic (z-test)	$z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$		
Reject H_0 if	$z > Z_\alpha$	$z < -Z_\alpha$	$\lvert z\rvert > Z_{\alpha/2}$
Accept H_0 if	$z \leq Z_\alpha$	$z \geq -Z_\alpha$	$\lvert z\rvert \leq Z_{\alpha/2}$

HYPOTHESIS TESTING (ONE SAMPLE CASE, WHEN σ^2 IS UNKNOWN)

	One-Tailed Test (i)	One-Tailed Test (ii)	Two-tailed Test
Hypothesis	$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic (t-test)	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ Where $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}, n = sample\ size$		
Reject H_0 if	$t > t_{\alpha,n-1}$	$t < -t_{\alpha,n-1}$	$\lvert t\rvert > t_{\alpha/2,n-1}$
Accept H_0 if	$t \leq t_{\alpha,n-1}$	$t \geq t_{\alpha,n-1}$	$\lvert t\rvert \leq t_{\alpha/2,n-1}$

TESTING FOR THE EQUALITY OF VARIANCE OF TWO NORMAL POPULATION FOR INDEPENDENT SAMPLES

Hypothesis	$\mathrm{H_0} : \sigma_1^2 = \sigma_1^2$ $\mathrm{H_0} : \sigma_1^2 \neq \sigma_1^2$
Test Statistic (F-test)	$F = \frac{S_1^2}{S_2^2},$ Where $S_1^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ and $S_2^2 = \frac{\sum_{i=1}^{m}(X_i - \bar{X})^2}{m-1}$
Reject $\mathrm{H_0}$ if	Case 1 $(S_1^2 > S_2^2)$ $F > F_{\frac{\alpha}{2},n-1,m-1}$ Case 1 $(S_1^2 < S_2^2)$ $F < F_{\frac{\alpha}{2},n-1,m-1}$