Statistik Inferensial & Jenis Uji

Aryan Eka Prastya Nugraha

2018

Descriptive statistics

The major concern of descriptive statistics is to present information in a convenient, usable, and understandable form.

For example, once the data have been collected, some of the first things that a researcher would want to do is calculate their *frequency*, *graph* them, calculate the *measures of central tendency* (means, medians, modes), calculate the *dispersion* of the scores (variances, standard deviations), and identify *outliers* in the distribution of the scores.

These procedures are called descriptive statistics because they are aimed primarily at **describing the data**

Inferential statistic

on the other hand, is not concerned with just describing the obtained data.

Rather, it addresses the problem of making **broader generalizations** or inferences from the **sample data to the population**.

This is the more complicated part of statistical analysis, and this chapter will focus on the role that **inferential statistics** plays in statistical analysis.

Descriptive statistics will not tell the researcher, for example, whether the difference between a sample mean and a hypothetical population mean, or the difference between two obtained sample means, .

The basic aim of **inferential statistics** is to use the **sample scores** for **hypothesis testing.**

Hypothesis Testing

Fundamental to the strategy of science is the formulation and testing of hypotheses about populations or the effects of experimental conditions on criterion variables.

For example, in an experiment designed to investigate gender differences in IQ, the researcher hypothesizes that first grade girls have higher IQ scores than frst grade boys. She administers an IQ test to four boys and four girls in a frst grade class. The results showed that the mean IQ score for the girls (110) is higher than the mean IQ score for the boys (103).

Based on these findings, is the researcher justifed in concluding that her hypothesis is supported? The answer is obviously **“we don’t know.”**

That is, although the results clearly showed that there is a difference between the sample means, with the girls scoring higher on average than the boys, there is the possibility that the observed difference could have been due to the chance variability of **intelligence** among frst graders. In other words, given the variability of **intelligence** among first graders, and the **smallness of the sample size**, some difference in the means is inevitable (sarana) as a result of the selection procedures.

In order for the researcher to draw valid conclusions about her hypothesis, she needs to employ **inferential statistics to test the reliability** of the finding of apparent (tampak jelas) difference in intelligence among first graders. **A prime function of inferential statistics is to provide *rigorous (ketat)* and *logically (logis) sound* procedures** for answering these questions.

Types of Hypotheses

Research Hypothesis

Hypotheses derived (diturunkan) from the researcher’s theory about some social phenomenon are called research hypotheses. The researcher usually **believes that his research hypotheses are true**, or that they are accurate statements about the condition of things he is investigating.

However, theories are only suppositions (perkiraan) about the true nature of things, and thus hypotheses derived from theories must also be regarded (dianggap) as just tentative suppositions (anggapan) about things until they have been tested. Testing hypotheses means subjecting (menunjukkan) them to ***confrmation*** **or** ***disconfrmation*****.**

Null hypotheses are, in a sense, **the reverse of research hypotheses**. They are also statements about the reality of things, except that they serve to ***deny*** what is explicitly indicated in a given research hypothesis.

Testing hypotheses mean subjecting them to some sort of empirical scrutiny (pemeriksaan) to determine if they are **supported (didukung)** or **refuted (dibantah)** by what the researcher observes

Test Selection

When a researcher is ready to test a specifc hypothesis generated from a theory, or to answer a research question posed (diajukan), the researcher is faced with the task of choosing an appropriate **statistical procedure.**

While the researcher is faced with a multitude of statistical procedures to choose from, the choice of an appropriate statistical test is generally based on just two primary considerations (pertimbangan): (1**) the nature of the hypothesis** and (2) **the levels of measurement** of the variables to be tested.

Test of Difference versus Test of Relationship

Testing for differences means that the researcher is interested in determining whether differences in mean **scores between groups** are due (kebetulan) to chance factors or to real differences between the groups as a result of the study's experimental treatment.

Testing of relationships among two or more variables asks the question,
“Are variations in variable X associated (terkait) with variations in variable Y?”
For example,
do students who do well in high school also perform well in university?

Is there a relationship between socio-economic class and recidivism in crime?

All these questions concern the **relationships among variables**, and to answer these questions, researchers must choose statistical tests that will appropriately test for the ***relationships*** **among these variables**

Levels of Measurement

In addition to considering the nature of the hypothesis to be tested (**difference or relationship**), the researcher must also consider the **measurements of the variables to be tested**. This is because the levels at which the variables are measured **determine(menentukan)** which statistical test is used to analyze the data.

Most typically, variables in the behavioral sciences are measured on one of four scales: ***nominal***, ***ordinal***, ***interval***, **or** ***ratio*** **measurements**.

Nominal Scale

This is the **lowest level** of measurement and involves simply categorizing the variable to be measured into one of a number of discrete categories.

For instance, in measuring “ethnic origin,” people may be categorized as American, Chinese, Australian, African, or Indian.

Once people have been categorized into these **categories**, all people in the same category (e.g., those categorized as Americans) Numbers can be assigned to describe the categories, but the **numbers are only used to name/label** the categories. **They have no magnitude (besaran) in terms of quantitative value.**

Ordinal Scale

This level of measurement involves **ordering or ranking** the variable to be measured.

For example, people may be asked to **rank-order** four basketball teams according to their skills. Thus, a rank of 1 is assigned to the team that is the most skillful, a rank of 2 to the team that exhibits the next greatest amount of skill, and so forth (seterusnya). These numbers allow the researcher to quantify the magnitude (besaran) of the measured variable, by adding the arithmetic relationships “**greater than**” and “**less than**” to the measurement process.

Interval Scale

This level of measurement involves **being able to specify how far apart two stimuli are on a given dimension**.

For example, on a standardized intelligence measure, a 10-point difference in IQ scores has the same meaning anywhere along the scale. Thus, the difference in IQ test scores between 80 and 90 is the same as the difference between 110 and 120. However, it would not be correct to say that a person with an IQ score of 100 is *twice* as intelligent as a person with a score of 50. The reason for this is because intelligence test scales (and other similar interval scales) **do not have a true *zero*** that represents a complete absence (ketiadaan) of intelligence.

Ratio Scale

This level of measurement replaces the **arbitrary zero point** of the interval scale with a true **zero starting point** that corresponds to the absence of the variable being measured.

Thus, with a ratio scale, it is possible to state (menyatakan) that a variable has, for example, **twice, half, or three times as much of the variable measured than another.** Take weight as **an example**.

Weight has a **true zero point** (a weight of zero means that the object is **weightless**) and the **intervals between** the units of **measurement are equal** (sama).

Thus, the difference between 10 and 15 g is equal to the difference between 45 and 50 g, and 80 g is *twice* as heavy as 40 g.

Test Selection Grid

	Relationship	One Set of Scores	Related Two Sets	Independent Two Sets	More than Two Sets
Nominal	Point biserial (r_{pb}) (true dichotomy)	Single variable chi-square (χ^2) test	McNemar significance of change χ^2	Chi-square test of association	Chi-square goodness of fit
	Biserial (r_b) (artificial dichotomy)				
	Phi (r_Φ) (true/true)				
	Tetrachoric (r_t) (artificial dichotomy /artificial dichotomy)				
Ordinal	Spearman's rho	Kolmogorov-Smirnov test for ranked data	Wilcoxon matched-pairs signed-ranks test	Mann-Whitney *U* test	Kruskal-Wallis test
Interval/ Ratio	Pearson's product-moment correlation	One-sample *t*-test	Related samples *t*-test	Independent samples *t*-test	One-way ANOVA (independent)
	Linear regression				Factorial ANOVA (independent)
	Multiple regression				Multivariate ANOVA (related)

R E F L E K S I

1. **Informasi penting hari ini**

2. **Manfaat penting dari informasi penting hari ini**

3. **Tindak lanjut yang dapat saudara lakukan**

THANK YOU!

ANY QUESTIONS?