Central Limit Theorem and the Use of Parametric and Nonparametric Statistics: Data Normality and Large Sample Requirements

1. Introduction

The Central Limit Theorem (CLT) is a fundamental principle of inferential statistics. It enables the use of parametric statistical tests even when the raw data are not perfectly normally distributed.

Since ideal normality is rare in real-world research, the CLT provides the mathematical basis for reliable inference, especially with large samples. When samples are small or distributions are highly skewed, nonparametric methods are recommended.

2. What Is the Central Limit Theorem?

The Central Limit Theorem states that as sample size increases, the **sampling distribution of the mean** becomes approximately normal — regardless of the distribution of the population (Field, 2013; Howell, 2012).
This means the raw data **do not have to be normal** for the means to behave as if they follow a normal distribution.

This is crucial because parametric tests (e.g., t-tests, ANOVA, Pearson correlation) assume normality of **sampling statistics**, not necessarily of the raw variables.

3. How large should a “sufficiently large” sample be?

Based on established literature:

- **n > 30** – The sampling distribution of the mean begins to approximate normality (Tabachnick & Fidell, 2013).
- **n ≥ 100** – Deviations from normality become less problematic (Howitt & Cramer, 2017).
- **n ≥ 200** – Parametric tests remain reliable even with considerable non-normality (Kim & Feldt, 2008).

Conclusion:
In applied research, **samples larger than n = 200** allow highly reliable use of parametric statistics due to the stabilizing effect of the CLT.

4. What types of data can be used in parametric statistics?

1) Interval and ratio scales — ideal

Examples: age, height, weight, time, total scores.

2) Likert scales (5+ categories)

Although ordinal, in large samples they behave like interval data.
Thanks to the CLT, their means are approximately normal → parametric tests are appropriate.

3) Dichotomous data (0/1, yes/no)

Can be used in regressions or as covariates.

4) Nominal categories

Usable via:

- dummy coding,
- ANOVA (with sufficient group size),
- regression models.

5) When to use nonparametric tests?

Nonparametric procedures (Mann–Whitney U, Kruskal–Wallis, Spearman) are suitable when:

- $n < 30$,
- heavy skewness is present,
- extreme outliers exist,
- the scale has very few categories (e.g., 2–3),
- variances are highly unequal.

5. Practical Example

Estimating average customer support response time

A company wants to estimate the average response time of its customer support team.
Response times are typically **right-skewed** — a few very long waiting times create a long tail.

The company receives around 5,000 support requests daily.

Procedure:

1. Draw a daily random sample of **n = 250 support tickets**.
2. Compute the average response time.

3. Repeat for 60 days.
4. Even though individual response times are highly skewed, the **sample means (n = 250)** are approximately normally distributed.
5. Parametric methods (t-tests, ANOVA, confidence intervals) can now be applied.

Conclusion:
Large samples and the CLT enable reliable parametric analysis despite non-normal raw data.

6. Conclusion

The Central Limit Theorem enables the widespread application of parametric tests even when data deviate substantially from normality. Samples above **n = 200** offer strong robustness, while samples above **n = 30** already provide acceptable approximations.

Parametric methods can be used with:

- interval and ratio data,
- Likert-type scales,
- dichotomous variables,
- nominal categories (through coding).

Nonparametric analyses are recommended for small samples or extreme violations of assumptions.

Link pdf: https://www.StatistischeBeratung.de/images/services/Central_Limit_Theorem-Parametric_Nonparametric_Statistics-Data_Normality-Large_Sample.pdf

https://www.StatistischeBeratung.de/images/services/www_Statistika_hr_Centralni_granicni_teorem_i_primjena_parametrijske_i_neparametrijske_statistike.pdf

Cite this article in your research paper:

APA

Statistische Beratung Leonardo Miljko (*01.12.2025.*) **Centralni granični teorem i primjena parametrijske i neparametrijske statistike: normalnost podataka i veliki uzorci**. Retrieved from: https://www.StatistischeBeratung.de/images/services/Central_Limit_Theorem-Parametric_Nonparametric_Statistics-Data_Normality-Large_Sample.pdf .

Literatura

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Sage.

Gravetter, F. J., & Wallnau, L. B. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Cengage Learning.

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Cengage Learning.

Howitt, D., & Cramer, D. (2017). *Introduction to statistics in psychology* (7th ed.). Pearson.

Khamis, H., & El-Sayed, A. (1999). On large sample tests for the normality of data. *Journal of Modern Applied Statistical Methods, 4*(1), 48–57.

Kim, J. O., & Feldt, L. S. (2008). Statistical procedures for classification and grading of items in educational measurement and evaluation. *Educational and Psychological Measurement, 68*(5), 829–844. https://doi.org/10.1177/0013164407313751

Opić, S. (2011). Centralni granični teorem i njegova uloga u inferencijalnoj statistici. *Školski vjesnik, 60*(2), 181–198.

Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Statistička konzultacija Leonardo Miljko. (2025). Određivanje potrebnog uzorka. Preuzeto s https://statistischeberatung.de/hr/04-forschungsmethoden-hr/02-11-01-12-hr-odredivanje-potrebne-velicine-uzorka-2

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic Press.