dummies
A Wiley Brand

CHEAT SHEET

STATISTICS FOR DUMMIES CHEAT SHEET

From **Statistics For Dummies, 2nd Edition**

By **Deborah J. Rumsey**

Whether you're studying for an exam or just want to make sense of data around you every day, knowing how and when to use data analysis techniques and formulas of statistics will help. Being able to make the connections between those statistical techniques and formulas is perhaps even more important. It builds confidence when attacking statistical problems and solidifies your strategies for completing statistical projects.

UNDERSTANDING FORMULAS FOR COMMON STATISTICS

After data has been collected, the first step in analyzing it is to crunch out some descriptive statistics to get a feeling for the data. For example:

- Where is the center of the data located?
- How spread out is the data?
- How correlated are the data from two variables?

The most common descriptive statistics are in the following table, along with their formulas and a short description of what each one measures.

Statistic	Formula	Used For
Sample mean (average)	$\bar{x} = \frac{\sum x}{n}$	Measure of center; affected by outliers
Median	*n* odd: middle value of ordered data *n* even: average of the two middle values	Measure of center; not affected by outliers
Sample standard deviation	$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$	Measure of variation; "average" distance from the mean
Correlation coefficient	$r = \frac{1}{n-1}\sum\frac{(x-\bar{x})(y-\bar{y})}{s_x s_y}$	Strength and direction of linear relationship between *X* and *Y*

STATISTICALLY FIGURING SAMPLE SIZE

When designing a study, the sample size is an important consideration because the larger the sample size, the more data you have, and the more precise your results will be (assuming high-quality data). If you know the level of precision you want (that is, your desired margin of error), you can calculate the sample size needed to achieve it.

To find the sample size needed to estimate a population mean (*μ*), use the following formula:

$$n = \left(\frac{z^* \sigma}{MOE}\right)^2$$

In this formula, MOE represents the *desired margin of error* (which you set ahead of time), and *σ* represents the population standard deviation. If *σ* is unknown, you can estimate it with the sample standard deviation, *s,* from a pilot study; *z** is the critical value for the confidence level you need.

SURVEYING STATISTICAL CONFIDENCE INTERVALS

In statistics, a *confidence interval* is an educated guess about some characteristic of the population. A confidence interval contains an initial estimate plus or minus a *margin of error* (the amount by which you expect your results to vary, if a different sample were taken). The following table shows formulas for the components of the most common confidence intervals and keys for when to use them.

CI For	Sample Statistic	Margin of Error	Use When
Population mean (μ)	$\bar{x}$	$\pm z^* \frac{\sigma}{\sqrt{n}}$	X is normal, or $n \geq 30$; σ known
Population mean (μ)	$\bar{x}$	$\pm t^*_{n-1} \frac{s}{\sqrt{n}}$	$n < 30$, and/or σ unknown
Population proportion (p)	$\hat{p}$	$\pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$n\hat{p},\ n(1-\hat{p}) \geq 10$
Difference of two population means ($\mu_1 - \mu_2$)	$\bar{x}_1 - \bar{x}_2$	$\pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	Both normal distributions or $n_1, n_2 \geq 30$; σ_1, σ_2 known
Difference of two population means $\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\pm t^*_{n_1+n_2-2} \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$	$n_1, n_2 < 30$; and/or $\sigma_1 = \sigma_2$ unknown
Difference of two proportions ($p_1 - p_2$)	$\hat{p}_1 - \hat{p}_2$	$\pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$n\hat{p}, n(1-\hat{p}) \geq 10$ for each group

CHECKING OUT STATISTICAL CONFIDENCE INTERVAL CRITICAL VALUES

Critical values (z*-values) are an important component of confidence intervals (the statistical technique for estimating population parameters). The z*-value, which appears in the margin of error formula, measures the number of standard errors to be added and subtracted in order to achieve your desired confidence level (the percentage confidence you want). The following table shows common confidence levels and their corresponding z*-values.

Confidence Level	**z*– value**
80%	1.28
85%	1.44
90%	1.64
95%	1.96
98%	2.33
99%	2.58

HANDLING STATISTICAL HYPOTHESIS TESTS

You use hypothesis tests to challenge whether some claim about a population is true (for example, a claim that 40 percent of Americans own a cellphone). To test a statistical hypothesis, you take a sample, collect data, form a statistic, standardize it to form a test statistic (so it can be interpreted on a standard scale), and decide whether the test statistic refutes the claim. The following table lays out the important details for hypothesis tests.

Test For	***Null Hypothesis (H_0)***	***Test Statistic***	***Distribution***	***Use When***
Population mean (μ)	$\mu = \mu_0$	$\dfrac{(\bar{x} - \mu_o)}{\sigma / \sqrt{n}}$	Z	Normal distribution or $n > 30$; σ known
Population mean (μ)	$\mu = \mu_0$	$\dfrac{(\bar{x} - \mu_o)}{s / \sqrt{n}}$	t_{n-1}	$n < 30$, and/or σ unknown
Population proportion (p)	$p = p_0$	$\dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}}$	Z	$n\hat{p},\ n(1-\hat{p}) \geq 10$
Difference of two means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\dfrac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	Z	Both normal distributions, or $n_1, n_2 \geq 30$; σ_1, σ_2 known
Difference of two means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\dfrac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	t distribution with df = the smaller of $n_1 - 1$ and $n_2 - 1$	$n_1, n_2 < 30$; and/or σ_1, σ_2 unknown
Mean difference μ_d (paired data)	$\mu_d = 0$	$\dfrac{(\bar{d} - \mu_d)}{s_d / \sqrt{n}}$	t_{n-1}	$n < 30$ pairs of data and/or σ_d unknown
Difference of two proportions ($p_1 - p_2$)	$p_1 - p_2 = 0$	$\dfrac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	Z	$n\hat{p},\ n(1-\hat{p}) \geq 10$ for each group