

7 Deadly Statistical Sins Even the Experts Make

Making statistical mistakes is all too easy. Statistical software helps eliminate mathematical errors—but correctly interpreting the results of an analysis can be even more challenging. Even statistical experts often fall prey to common errors. Minitab's technical trainers, all seasoned statisticians with extensive industry experience, compiled this list of common statistical mistakes to watch out for.

1: Not Distinguishing Between Statistical and Practical Significance

Very large samples let you detect very small differences. But just because a difference exists doesn't make it important. "Fixing" a statistically significant difference with no practical effect wastes time and money.

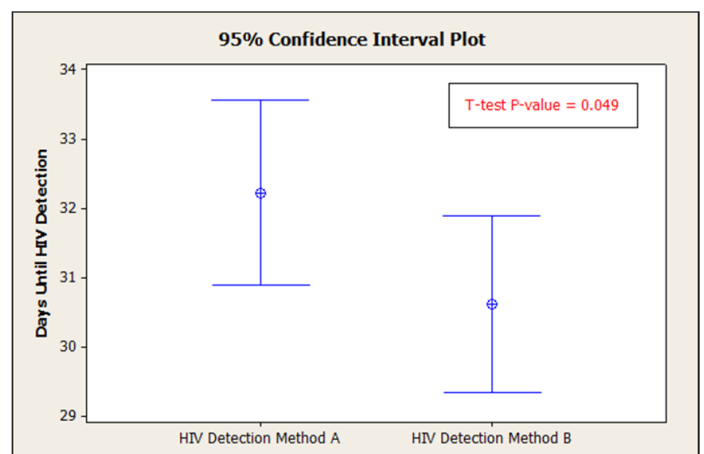
A food company fills 18,000 cereal boxes per shift, with a target weight of 360 grams and a standard deviation of 2.5 grams. An automated measuring system weighs every box at the end of the filling line. With that much data, the company can detect a difference of 0.06 grams in the mean fill weight 90% of the time. But that amounts to just one or two bits of cereal—not enough to notice or care about. The 0.06-gram shift is statistically significant but not practically significant.

Automated data collection and huge databases make this more common, so consider the practical impact of any statistically significant difference. Specifying the size of a meaningful difference when you calculate the sample size for a hypothesis test will help you avoid this mistake.

2: Misinterpreting Confidence Intervals that Overlap

When comparing means, we often look at confidence intervals to see whether they overlap. When the 95% confidence intervals for the means of two independent populations don't overlap, there will be a statistically significant difference between the means.

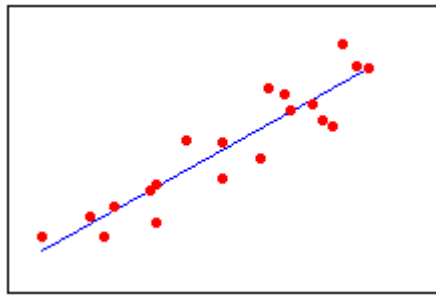
However, the opposite is not necessarily true. Two 95% confidence intervals can overlap even when the difference between the means is statistically significant. For the samples shown in the interval plot, the p-value of the 2-sample t-test is less than 0.05, which confirms a statistical difference between the means—yet the intervals overlap considerably. Check your graphs, but also the rest of your output before you conclude a difference does not exist!



Even though the confidence levels overlap, the 2-sample t-test found a statistically significant difference between the treatments.

3: Assuming Correlation = Causation

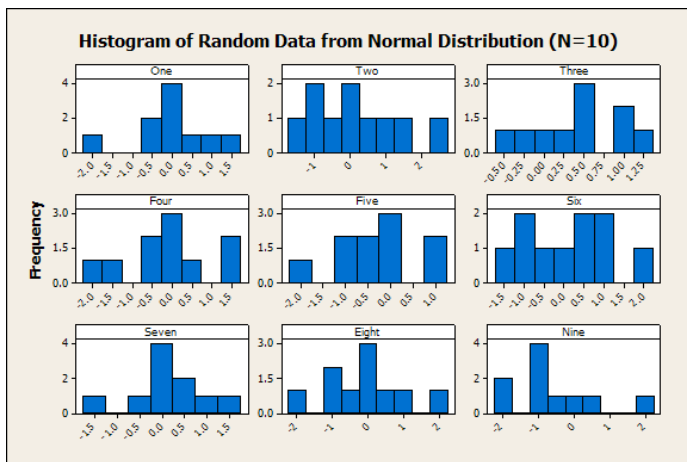
When you see correlation—a linear association between two variables—it’s tempting to conclude that a change in one causes a change in the other. But correlation doesn’t mean a cause-and-effect relationship exists.



Let’s say we analyze data that shows a strong correlation between ice cream sales and murder rates. When ice cream sales are low in winter, the murder rate is low. When ice cream sales are high in summer, the murder rate is high. That doesn’t mean ice cream sales cause murder. Given the fact that both ice cream sales and the murder rate are highest in the summer, the data suggest rather that both are affected by another factor: the weather.

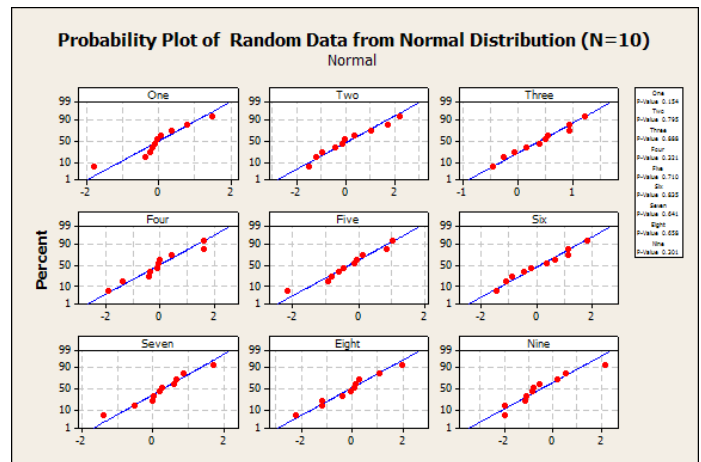
4: Rejecting Normality without Reason

Many statistical methods rely on the assumption that data come from a normal distribution—that they follow a bell-shaped curve. People often create histograms to confirm their data follow the bell curve shape, but histograms like those shown below can be misleading.



We drew these 9 samples from the normal distribution, but none of the histograms is bell-shaped.

Looking at the same data in a probability plot makes it easier to see that they fit the normal distribution. The closer the data points fall to the blue line, the better they fit the normal distribution.



If you have fewer than 50 data points, a probability plot is a better choice than a histogram for visually assessing normality.

5: Saying You’ve Proved the Null Hypothesis

In a hypothesis test, you pose a null hypothesis (H_0) and an alternative hypothesis (H_1). Typically, if the test has a p-value below 0.05, we say the data supports the alternative hypothesis at the 0.05 significance level.

But if the p-value is above 0.05, it merely indicates that “there is not enough evidence to reject the null hypothesis.” That lack of evidence does not prove that the null hypothesis is true.

Suppose we flip a fair coin 3 times to test these hypotheses:

$$H_0: \text{Proportion of heads} = 0.40$$

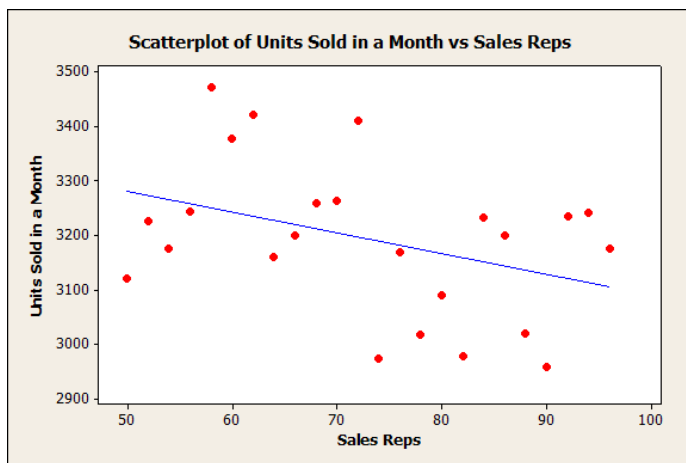
$$H_1: \text{Proportion of heads} \neq 0.40$$

The p-value in this test will be greater than 0.05, so there is not sufficient evidence to reject the hypothesis that the proportion of heads equals 0.40. But we know the proportion of heads for a fair coin is really 0.50, so the null hypothesis—that the proportion of heads *does* equal 0.40—clearly is not true.

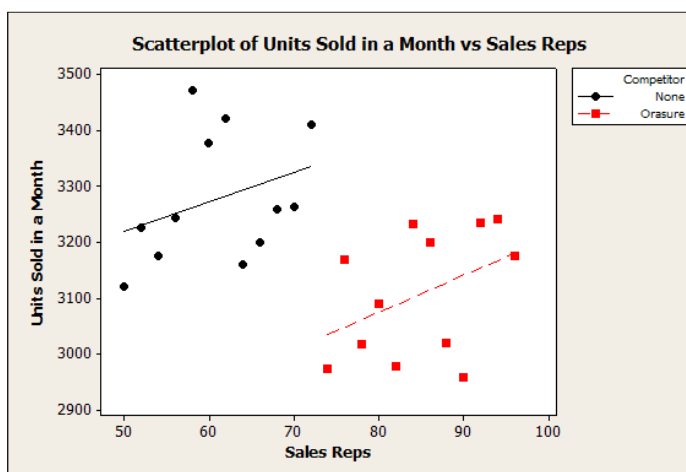
This is why we say such results “fail to reject” the null hypothesis. A good analogy exists in the U.S. criminal justice system, where the verdict when a prosecutor fails to prove the defendant committed a crime is “Not guilty.” The lack of evidence doesn’t prove the defendant is innocent—only that there was not sufficient evidence or data to prove otherwise.

6. Analyzing Variables One-at-a-Time

Using a single variable to analyze a complex situation is a good way to reach bad conclusions. Consider this scatterplot, which correlates a company’s sales with the number of sales reps it employs. Somehow, this company is selling fewer units with more sales reps!



But something critical is missing from this analysis: a competitor closed down while this data was being collected. When we account for that factor, the scatterplot makes sense.



Variables you don’t account for can bias your analysis, leading to erroneous results. What’s more, looking at single variables keeps you from seeing the interactions between them, which frequently are more important than any one variable by itself.

Avoid this mistake by considering all the important variables that affect your outcome.

7: Making Inferences about a Population Your Sample Does Not Represent

With statistics, small samples let us make inferences about whole populations. But avoid making inferences about a population that your sample does not represent. For example:

- In capability analysis, data from a single day is inappropriately used to estimate the capability of the entire manufacturing process.
- In acceptance sampling, samples from one section of the lot are selected for the entire analysis.
- In a reliability analysis, only units that failed are included, but the population is all units produced.

To avoid these situations, carefully define your population before sampling, and take a sample that truly represents it.

The Worst Sin? Not Seeking Advice!

Most skilled statisticians have 4-8 years of education in statistics and at least 10 years of real-world experience—yet employees in basic statistical training programs sometimes expect to emerge as experienced statisticians! No one knows everything about statistics, and even experts consult with others when they get stuck. If you encounter a statistical problem that seems too challenging, don’t be afraid to reach out to more experienced analysts!